Український Технічно-Господарський Інститут позаочного навчання
в Подєбрадах (ЧСР).

На правах рукопису. Курс ч. 4.

У. 7964.

Левко Шрамченко
Професор Українського Університету в Празі
Доцент Української Господарської Академії

ОСНОВИ ТЕОРІЇ СТАТИСТИКИ

(Курс лекцій)

Видання друге

УТГІ

1936

Подєбради

Український Технічно-Господарський Інститут позаочного навчання
в Подєбрадах (ЧСР).

На правах рукопису. Курс ч. 4.

Левко Шрамченко
Професор Українського Університету в Празі
Доцент Української Господарської Академії

ОСНОВИ ТЕОРІЇ СТАТИСТИКИ

Курс лекцій

217743

Видання друге

УТГІ

1936

Подєбради

31

DAROVÁNO
Slovanským ústavem
s výhradou dohody čj. 690/32.

U 7964.



Copyright by Ukrajinska Hospodarska Akademija, Poděbrady.
Czechoslovakia.

SLOVANSKÁ KNIHOVNA
3186285149

Tiskl Josef Rašek - knihtiskárna litografie - Poděbrady Lázně.

П Е Р Е Д М О В А .

Курс "Основи теорії статистики" складений мною для задоволення потреби Українського Технічно-господарського Інституту позаочного навчання. В основу його покладені лекції, що викладались студентам Української Господарської Академії та Українського Вільного Університету в Празі.

До цього часу на українській мові ми маємо: курс проф.Остапенка, випущений друком ще в 1920 році за надзвичайно несприятливих умов для такого рода праці; "Нариси теоретичної статистики" проф.Дихляєва, подруге видані Статистичним Управлінням України в перекладі з російської мови на українську в 1927 році; "Варіяційна статистика" проф. Сапігина, вид. Книгоспілки в 1926 році та конспект лекцій математичної статистики (літогр.) Доцента УГА С.М.Романовського, виданий в Подєбрадах в 1928 році.

В своїх викладах основ науки про статистичну методу, автор звичайно не міг дати більш-менш повно все те, що потрібно б було дати в зв'язку з сучасним станом розвитку теоретичної статистики. Необхідність не виходити з рямців, поставлених Дирекцією Інституту, щодо обсягу підручника, а також відсутність в програмах навчання Економічно-кооперативного факультету обов'язкових викладів основ вищої математичної аналізи вимагали і відповідного пристосування викладів до поставлених завдань. Через те і беручи на увагу, що в програмі навчання для старших семестрів в Інституті мається ще спеціяльний мій курс: "Дослідження господарської кон'юнктури", автор вважав доцільнішим і з педагогічного боку перенести до цих викладів і ту частину статистичної методології, яка більш широко охоплює систему індексів, та аналізу часових статистичних рядів та способи відшукання нелінійних корелятивних зв язків. Спосіб позаочного навчання вимагає від студента дуже напруженої, вдумливої праці в умовах відсутности безпосереднього живого зв'язку з лектором, через те перенесення найбільш складних проблем статистичної методології до пізніших споріднених викладів, на мою думку, буде корисним. Для тих, хто хтів би тепер же більш широко ознайомитись з проблемами теоретичної статистики, автором і подана відповідна література. Тяжкі технічні умови праці і терминовість її не сприяли також й тому, щоб дати більш практичних прикладів та ілюстрацій для ліпшого засвоєння теоретичних основ науки та використання їх на практиці.

Будова системи курсу виходить з тої засади, що описова частина статистики, себто моменти одержання і підготовлення до наукового оброблення статистичних даних є не менш важливою і неподільною частиною загальної науки про статистичну методу, через що автор в своїх викладах, на розділ від багатьох інших авторів подібних викладів, вважав потрібним досить детально зупинитись на статистичному спостереженні і правилах його переведення. Цим самим курс може, на думку автора, бути корисним і для молодих статистиків - практиків.

Автор буде дуже вдячним українським спеціялістам за вказівки на помилки, неясності, бажані доповнення і т.п., за що наперед висловлює свою щиру подяку. Листи з зауваженнями автор просить надсилати на адресу Українського Вільного Університету в Празі (Praha II. ve Smečkach č.27/IV).

Нарешті висловлює свою щиру подяку лекторові Української Господарської Академії В. С а п і ц ь к о м у за допомогу при редагуванню цього курсу.

Л. Ш р а м ч е н к о .

Прага, грудень 1934.

Л е к ц і я п е р ш а.

ПРЕДМЕТ ТА ЗАВДАННЯ СТАТИСТИКИ.

I. Поняття статистики.

Загальною назвою статистика покриваються три зовсім різнорідні речі.

По-перше, під цією назвою розуміють ч и с т о о п и с о в у науку, що досліджує для певних умов місця й часу масові /колективні/ явища соціяльного життя. Систематичне дослідження масових явищ має великий інтерес, особливо в суспільствознавстві, де складниками масових явищ є люди й людська чинність. Статистика оперує тут з такими явищами, як із самостійним обєктом наукового досліду, що й визначає обсяг та зміст цієї науки.

По-друге, статистикою називають т е о р і ю с п о с о б і в д о с л і д ж е н н я особливих причинових взаємозалежностей, або співвідношень; т.т. так званих "свобідних причинових зв'язків". Під останнім поняттям відомий вчений статистик А. А. Ч у п р о в розуміє залежності і зв'язки,що зумовлюються багатьма причинами та характеризуються багатьма подіями, а також умови, в яких за тієї самої причини може наступити декілька різних наслідків та той же самий наслідок може викликатися декількома різними причинами. Характеристичною властивістю такого роду дослідження є вибір для спостереження не окремого якогось явища або предмету, а цілої групи явищ, подібних до себе в одних відношеннях і відмінних в других! Статистична метода /спосіб/ оперує тут з сукупностями. Вона розглядає, з яких елементів ці сукупності складаються, числить їх в цілому та по складових частинах, а потім, за допомогою властивих їй способів, встановлює закономірності в цих сукупностях. Остаточною метою статистичної методи, як це є і в індуктивній методі,- встановити причинові залежності між можливо простішими явищами. Таким чином, сукупності цікавлять статистика, як засіб до мети - встановити причинові залежності між явищами. Цілком зрозуміло, що число обєктів статистичного дослідження надзвичайно велике, поле статистики в цьому розумінні таке широке, як і світ органічної природи. І дійсно, статистичною методою досліджуються і біологічні явища і психологічні, нею користуються і антрополог, і язикознавець, і економіст, і соціолог. Цієї методи вживають навіть у фізиці та астрономії.

В цих рямцях статистика не має стисло окресленого об'єкта дослідження, а є необхідним знаряддям, як своєрідна метода дослідження для багатьох наук. В цьому відношенні статистика мислиться, як наука теоретична, методологічна, а обсяг способів і правил, що складають цю науку, найкраще можна б було включити до логіки, як окрему її частину, нарівні з теорією метод індукції /А.А.Чупров/.

Нарешті, статистикою називають кожне конкретне прикладання статистичних способів дослідження тих або інших проблем, до ділянки якої б науки вони не відносились.

В дальшому своєму викладі ми будемо розуміти під статистикою:

Спеціяльну методу наукового дослідження, яка систематичним числовим, масовим спостереженням колективних нетипових явищ описує ці явища та відшукує властиві їм закономірності, або постійні зв'язки, що керують цими явищами.

2. Природа колективних, атипових явищ. Вищенаведена дефініція в першу чергу вимагає від нас вияснення, що то є колективні /масові/, атипові/ нетипові/ явища, з якими оперують статистичною методою.

Формально, колективне явище протиставляється явищу індивідуальному. Бо коли б колективне явище ми спостерігали, не як масу, а по окремих його складових частинах, воно не дало б того образу, не виявилося б так, як виявляється у спостереженні цілого колективу, цілої сукупности. Напр.,спостерігаючи таке масове явище, як населення, ми можемо виявити його властивості, навіть правильності, що відбуваються в його стані,тільки в цілій його сукупності. Постійність числового відношення в населенні осіб мужеської й жіночої статей статистики встановили, спостерігаючи всю сукупність населення, його масу в різних країнах. Спостереження над народженнями в невеликому числі окремих родин, не може нам дати, чи виявити закономірності в народженні хлопчиків і дівчат майже в однаковому числі.

Але не всі колективні явища можуть бути об'єктом дослідження статистичною методою, а тільки ті з них, що складаються з одиниць, або явищ того ж самого роду. Однак ні в якому разі не тотожних, але щодо певних властивостей /ознак/ - відмінних між собою. Коли, напр., візьмемо населення, то окремі одинці цієї сукупности, окремі люди, будуть того ж самого роду, але щодо своїх ознак - віку, статі, родинного стану - відмінні між собою. Ці колективні явища і звуться атиповими /нетиповими/. Вони різняться від тих, теж колективних явищ, які складаються в сукупність з одиниць або явищ просто тотожних, або складені з одиниць, що виявляють певні властивості чи ознаки тотожні їм усім. Ці останні явища й звуться типовими. Напр., такі масові явища як вода, золото, щодо їх хемічного складу, уявляють з себе явище типове, бо складаються з одиниць тотожних. Теж типове явище буде й людина, коли розглядати її з погляду анатомічної будови тіла. Метода пізнання колективних атипових явищ є методою статистичною, методою масового спостереження ; метода пізнання другої групи явищ /типових/ є методою індивідуального спостереження - експериментальною методою.

Основні елементи води, напр.,пізнаються досвідом над якоюсь однією частиною її /шклянка, літр і т.п./. Постійне відношення між певною кількістю її складників - водня й кисня, установлене згаданим досвідом, лишається сталою величиною для всякої, довільно взятої, маси води. Навпаки, стан і структура населення пізнається лише на підставі статистичного спостереження всіх людей, що творять сукупність певного населення.

3. Поняття випадку і випадкової причини. Сказаного одначе не досить, щоб добре характеризувати природу зазначених вище двох категорій явищ та визнати необхідність статистичної методи для одних і непотрібність її для другіх. Щоб розібратись в цьому складному питанні, треба взяти за вихідну точку х а р а к т е р т и х п р и ч и н, які зумовлюють кожну категорію явищ, а з'окрема взаємовідношення постійних і випадкових причин у них.

Зясувати явище - це значить знайти найближчі причини, що викликали це явище до життя. Тому вимоги наукового знання задовольняються в тому разі, коли зможемо встановити послідовний ряд причинових відношень, що неминуче призводять до певних наслідків. Логіка формулює цю вимогу в категоричній формі так: с к р і з ь і з а в ж д и, де здибаємо явище А, за ним слідом іде В; с к р і з ь і завжди, де надибуємо явище В, іде перед ним А. Такої стисло зумовленої причинової залежности ми, одначе, не спостерігаємо в конкретній дійсності; лише в законах природи /закон тяжіння і под./ знаходимо найбільше наближення до формулювання причинових зв'язків логікою. Ми мусимо одначе пристосувати наше знання до певних частин простору й часу, а тоді із змісту нашого знання неминуче випадає поняття "скрізь і завжди" / П.Віхляєв/. Наше знання є обмежене, присто-

соване до певної конкретної дійсности. А ті поодинокі явища, з яких складається конкретна дійсність, вилучені з однорідних явищ, можна розглядати, як явища випадкові. Характер випадкового явища можна ясніш уявити в такого прикладу. Коли підкинемо вгору монету, то явище падіння монети додолу здається нам неминучим. Але чого ж саме ця монета падає якоюсь одною стороною, напр., решкою. Цей останній факт здається нам звичайним випадком. Одначе, коли простежимо рух монети в повітрі, то помітимо цілу низку послідовних умов, які призводять до того, що монета падає саме цією стороною - догори решкою, а не орлом. Коли, підкидаючи монету, кинемо її догори решкою, то обернувшись у повітрі парисге число разів, ця монета впаде додолу тією самою стороною, якою її підкинули, тобто решкою угору, а обернувшись непарисге число разів, навпаки-впаде на землю вже не решкою, а орлом. Якою стороною вгору лежить монета в руці того, хто її підкидає, з якою силою підкинено монету, вага цієї монети, опір повітря з моменту як ми кинули монету і аж поки впала вона на землю - це ті умови, що дають число обертань монети, та через що саме монета падає на землю тією або другою стороною догори. Ми неспроможні вилучити з цього складного згрудку умов ту найближчу причину, що викликає парисге чи непарисге число обертань, через те й саме явище здається нам випадковим. Основна властивість випадкового явища - це відсутність видимої потреби в його повстанні; одначе повстання явища може випливати із залежности явища від множности причин. Ці причини є занадто складні, щоб можна було вилучати з складного комплексу причинових залежностей ту, найближчу з них /умов/, що викликала це саме явище до життя.

Навколо випадкового явища в нашій уяві згромаджується багато причин і наслідків. Що складніше явище, то більш заплутаний клубок згаданих причинових залежностей, які його /явище/ викликають. Порівнюючи з явищами неорганічної природи, явища органічної і соціяльної природи визначаються особливо великою складністю і множністю причинових залежностей, через це й дають ширший простір для повстання явищ випадкових. Характеристичною рисою зазначених явищ є ще й те, що вони виникають інколи з неспільномірної ваги причин з наслідками, що з них випливають. Мала обставина, що непомітна для нашої уваги, може викликати великі наслідки /П.Віхляєв/. Напр., великі зміни в погоді відбуваються там, де атмосфера є в нетривалій рівновазі. На підставі спостережень на своїх станціях, метеорологи, напр., установлюють, що порушена рівновага атмосфери. Але де саме відбулось порушення рівноваги, зазначити точно не можуть. Протеріжниця температури на 1/10 ступня, не встановлена організованим спостереженням, направляє путь, якою йде циклон./

Відомий статистик-теоретик Чубер / Czuber / визначає поняття випадку так: „ випадок - це ті мінливі умови, які в безперервній своїй мінливості не піддаються нашому пізнанню. Випадковими подіями є такі, існування або наступлення яких не може бути з певністю встановлено на підставі відомих нам умов або передпосилок /логічних/". Таким чином, слово "випадок" не розуміється як негація причинового зв'язку. Вживаючи його, ми лише визнаємо, що "причиновий зв'язок в даному випадкові або взагалі не піддається пізнанню, або настільки складний, що дальше його констатування піти не можемо".

Одначе сказаного ще не досить, щоб освітити поняття "випадку" і "випадкової причини", особливо останнього, на якому базується теорія імовірности та наукові висновки статистичної методи. Повстає питання: невже все, що не піддається нашому пізнанню - все, чого ми з певністю не можемо вивести з відомих нам логічних посилок, буде "випадковою подією" або проявом випадку? Коли роздивлятись поступування людини на протязі цілого дня в конкретних проявах - як вона встала, що й як одягала, що та в якій кількості їла, для чого й якою дорогою пішла - очевидячки не зможемо віднести цієї сукупности вчинків людини до числа "випадкових" явищ; хоч ні одного з них, в їх конкретній дійсності, не зможемо вивести

з відомих нам посилок. Однаково не визнаєм "випадковостями" кожного випадку видужування або смерти хорого, кожного обвинувачення або виправдання підсудного, успіху або збанкротування промислового підприємства, хоч причини в кожного індивідуального з цих прикладів "не піддаються нашому пізнанню". Найбільш досвідчений лікар часто-густо не може сказати через що саме його пацієнт "А" видужав, а пацієнт "В" помер від тої самої хороби при однакових об'єктивних ознаках та однакових методах лікування. Про "випадок", "випадковість" ми будемо говорити лише при певній комбінації обставин. Коли людина спізнилася, через те що проспала, а проспала через те, що до пізньої ночі працювала, або прогуляла, ми не будемо тоді вбачати в цьому чогось "випадкового". Про "випадкове" запізнення ми будемо говорити, коли воно зумовлено зіпсуттям трамвая, або смерть хорого сталася через хибне заготовлення ліків і т.п. Себто, для того, щоб ми могли говорити про " випадок", мусить бути наявним не тільки наше незнання, але й ще щось інше - об'єктивне; мусить бути щось приєднане до нормальної течії події, мусить бути зустріч, збіг двох незалежних одно від одного явищ, подій, процесів.

Таким чином, істотна ознака поняття "випадку", "випадкової події" - це випадкове співіснування в часі, або двох чи декількох незалежно одно від одного існуючих, явищ, або також певної сукупности індивідуальних, випадкових причин поруч з деяким комплексом постійних причин. Філософ Віндельбанд пояснює це поняття як "співіснування, збіг фактів, які не є між собою у відношенні причини і наслідку та не залежать від якоїсь загальної причини - себто між якими не існує будь-якого необхідного зв'язку". Через те "випадок" не тільки є вираз нашого незнання, але й об'єктивна ідея комбинації декількох незалежних систем причинностей" /А. Cournot /. Ці поняття надзвичайно важливі для розуміння статистичної методи. Коли взяти таке явище як рух планетної системи, то для того, щоб воно одержало свій конкретний вигляд, потрібно було, по-перше, чинности певних механічних сил і законів, але, крім того, також певне первісне положення, чи розміщення частей матеріяльної системи, що склалась у форму нашого світу.

Збіг того й другого не був необхідним, себто він може розглядатись як елемент випадковости, через яку наш світ одержав свою конкретну форму. Випадок, як бачимо, відіграв свою роль, так мовити, в основі всіх основ. Його роля росте в міру ускладнення явищ. Уявіть собі перехід від логічного й математичного порядку до фізичного й механічного, далі до рослинного і тваринного царства, нарешті до людини і безмежного кола історичних подій. І скрізь, у всіх цих сферах і областях, істота справи лежить у зустрічі подій, що стосуються до незалежних серій причин, а не в нашому незнанні. Випадок не випливає з нашого незнання світових законів. Відношення випадковости існують між самими речами і роля випадку не зменшується від розширення нашого знання.

Коли це так, то для нас дуже важливо встановити відносно явищ чи то подій, що саме є в них постійно діючими причинами, а що причинами випадковими.

Візьмемо за приклад таке явище, як урожай збіжжя. Він залежить від дуже складної сукупности причин, бо тут діють: I/ природно-історичні умови - фізичні властивості і склад ґрунту, рел'єф місцевости, підсоння. 2/ аґрікультурні умови - система полеведення і сівозмін, обробка, угноєння ґрунту і т.п. в певному господарстві; 3/ від випадковостей хитання протягом певного року кількости опадів, змін температури, вітрів, від наявности чи відсутности шкідників і т.п. Умов цієї останньої категорії ми навіть не можемо передбачити; умови другої /аґрікультурні/ ми знаємо, але не можемо певним способом зв'язати з умовами першої категорії /природно-історичної/. Отже, коли студіюємо урожай в певній місцевості взагалі, до числа випадкових умов відносимо перш за все всі умови другої й третьої категорії, але далі, дивлячись на те, з яким саме міром підхо-

димо до студіювання урожаю, може бути до випадкових умов віднесена й частина умов першої категорії. Коли нас цікавить вплив на урожай ґрунту, можуть бути віднесені до випадкових - рел'єф і "мікрокліматичні" властивості. Коли студіюємо вплив мікрокліматичних властивостей, тоді до випадкових можемо віднести рел'єф і ґрунт. Коли ми студіюємо урожай певного року - причини третьої категорії увійдуть до групи постійних, а причини першої категорії опиняться частинно у постійних, а частинно у випадкових, дивлячись по тому, чи будемо студіювати урожай певного року взагалі, чи в зв'язку з іншими факторами. Коли нас цікавлять урожаї в певному господарстві, постійними робляться причини першої і другої категорії, випадковими залишаються лише умови третьої категорії.

Подібні комбінації постійних і випадкових причин лежать в основі кожного реально-існуючого об'єкту наукового дослідження. Випадковости нема лише в дедуктивних та в математичних виводах і побудованнях.

Виходить, що поділ умов на постійні й випадкові має дуже серйозне значіння, але разом з тим має релятивний /відносний/ характер в залежності від того, з якого погляду і з якою метою досліджуємо взаємозалежність умов певного явища, події, процесу.

В статистиці процес розподілу умов відограє величезне значіння, бо на ньому будується аналітичні таблиці і розподіл ознак об'єкту дослідження на факторіяльні ознаки і функціональні.

Через те, що одні явища складаються цілком, або, головним чином, під впливом постійних причин, і випадкові причини не мають на процес їхнього складання жадного впливу, або зовсім незначний, - такий, що його можна в розрахунок не брати, а другі явища складаються так, що роля випадкових причин надзвичайно велика і то настільки, що вони зовсім закривають діяння причин постійних - ділять явища, до якої б царини знання вони не належали, на дві категорії - на типові й атипові /індивідуальні/. В перших панують постійні причини, в других - рішучий вплив мають випадкові причини. Відомий статистик Р.Орженцький визначає цей розподіл явищ так: в одних випадках виділена ознака /або ознаки/ властива цілому ряду об'єктів, дає вистарчаюче, або майже вистарчаюче, для певних цілей пізнання групи об'єктів, об'єднаних певною ознакою або ознаками. Решта ознак кожного об'єкту групи, як подібних, так і різних, або зовсім не мають значіння для даних цілей пізнання, або мають другорядне значіння. В других випадках вибір однієї або декількох ознак хоч і характеризує групу об'єктів, але в малій степені, або так, що вона непридатна до цілей пізнання. В першому випадку виділені загальні ознаки звуться т и п о в и м и, а група обєднаних цим поняттям об'єктів - т и п о в о ю г р у п о ю. В другому - виділені від об'єктів загальні ознаки звемо - а т и п о в и м и, а об'єднану ними групу об'єктів - з б і р н о ю г р у п о ю.

До першої групи явищ відноситься велика частина явищ неорганічної природи, напр., властивості структури заліза, золота, хемічний склад води, рух землі навколо сонця і т.под. До другої, явища органічної природи і особливо явища біологічні, суспільні, напр., мінливість і дідичність рослин і тварин, довжина віку людини, добробут селянства, заробітна платня промислового робітника і інше.

Типовим явищам відповідають р о д о в і п о н я т т я, себто такі, де ознаки, що входять в дане поняття в повній мірі відносяться до кожного поодинокого явища даного роду і доступні нашому спостереженню. Родовим поняттям буде, напр., поняття води. Тут кожна частина чистої води має тотожний хемічний склад, замерзне й розстане при певній температурі, при нагріві або охолодженні підлягає певним законам розширення. Так само родовим є поняття пшениці - бо кожне стебло, кожний колос і кожне зерно характеризується певною сукупністю ознак, встановлених ботанічною систематикою. Атиповим явищам відповідають збірні поняття. Тут всі поодинокі явища, що підходять під збірне поняття, мають одну або небагато загаль-

них для всіх ознак, решта ж ознак /яких може бути дуже багато/ одним членам групи, що об'єднується збірним поняттям, належать, а другим ні, т.т. ці решта ознак мінливі, випадкові в тому розумінні, як це було зазначено вище. В кожному поодинокому явищі сукупність оцих мінливих або випадкових ознак поруч з родовими складається в своєрідний малюнок, в окрему індивідуальність. Напр., збірні поняття: студент, український студент, подільський селянин, малолітні злодії і т.п.

Родове поняття може перетворитись в збірне, коли в інтерес дослідника входять, крім загальних для всіх явищ, ще й індивідуальні, мінливі ознаки. Навпаки, збірне поняття перетвориться в родове, коли при дослідженні на мінливі ознаки уваги не звертається, коли вони ігноруються.

4. Визначення об'єкту статистики. Сказане нами вище, визначає об'єкт статистичної методи. Отже об'єктом статистичної методи є колективні, атипові явища, складені з маси явищ, що обєднуються в єдине збірне поняття одним або декількома загальними ознаками, а рештою ознак ріжняться між собою. Через те до царини вживання статистичної методи можна віднести всі ті галузі знання, які мають діло з обєктами такого роду, як тільки що було визначено, незалежно від того, чи вони належать до сфери природничих чи соціяльних наук. Метою дослідження зробленого за поміччю статистичної методи є характеристика відповідаючих збірним поняттям масових об'єктів певного конкретного обсягу та пояснення цих обраних сукупностей певним комплексом зведених ознак, а також встановлення збірних залежностей причинового характеру, що виявляються тільки в масі та виявлення в цих зведених ознаках і збірних залежностях діяння загальних причин, що керують певним колективним явищем.

Тільки тоді, коли статистичними способами вияснюється постійність певних впливів та основних причин явищ, статистика стає знаряддям наукового досліду.

5. Статистичне спостереження. Вище ми зупинялися на природі об'єктів, до дослідження яких прикладається статистична метода. Ми вказали на складність об'єктів, на множність причин і наслідків, які діють серед колективних атипових явищ. Із природи цих явищ випливають і способи їх спостереження. Само це спостереження зветься с т а т и с т и ч н и м. Воно є неподільною частиною науки про статистичну методу.

Спостереження взагалі може бути або п р и р о д н е, або н а у к о в е.

Перше базується на безпосередньому сприйманні людиною - субєктом спостереження - за допомогою своїх змислів тих вражінь, які робить на нього зовнішня природа, та в формулюванні цих вражінь в якусь загальну формулу або положення. Таке спостереження має в собі характер відірваности, несистематичности й суб'єктивности, а через те й висновки його не мають наукового значіння, не доведені. Це будуть лише особисті узагальнення побудовані на більшій або меншій здібності окремого індивіда правильно сприймати та передавати своїй свідомості зовнішні явища. Часто такого роду спостереження ставть підставою т.зв. "громадської опінії" та ріжних забобонів і повір'їв. Часто воно не є виразом справжньої об'єктивної дійсности, а перекрученням її в наслідок недосконалости наших змислів.

Спостереження наукове відмінне тим, що переводиться по певному, наперед обдуманому, пляну, по певній системі, та в таких, по змозі, умовах, щоб виключити із спостереження зовнішніх явищ вплив особистих властивостей самого дослідника. В цьому останньому розумінні прагнеться до того, щоб спостереження показували яким певне явище /об'єкт спостереження/ в дійсності є, а не яким воно "здається" дослідникові.

Таким чином, кожне наукове спостереження, а в тому числі й статистичне, мусить бути м е т о д и ч н и м, с и с т е м а т и ч н и м.

6. Суб'єкт статистичного спостереження. Між різними способами наукового спостереження, над будь-якими явищами, чи процесами, ми мусимо розріжнювати дві різні категорії. До першої відносяться ті способи, де цілий процес спостереження та висновки з нього є приступний для однієї особи, де дослідник складає плян сам і сам переводить дослід, студіює факти та відшукує причинові залежності. До другої категорії відносимо ті, де процес спостереження одній особі недоступний, де нормально спостереження переводиться цілою групою дослідників, яка одначе зведення спостережень в єдине ціле та аналізу здобутків спостереження переводити сама не може. Це є категорія с п і л ь н и х спостережень.

Отже спостереженнями **с п і л ь н и м и** ми звемо такі способи дослідження, коли по єдиному загальному пляну численні особи беруть участь у спостереженні явища не для перевірки один одного, як то роблять, напр., хеміки, фізики і т.под., але для уложення з суми своїх часткових спотережень єдиного суцільного спостереження. Коли хемік, напр., спостерігає зміну чи склад якогось фізичного тіла одним способом, а другий хемік рівнобіжно це ж робить другим способом, то тут обидва вони одержують два індивідуальні спостереження, що перевіряють одно одне. Але це не є спільне спостереження, як ми його розуміли вище.

7. Масовість спостереження. Зовсім інакше, ніж у попередньому прикладі, переводиться спостереження климатології. Тут спостереження, зроблені різними особами, в різних пунктах /місцях/, не контролюють одне одного, а служать складовими частинами загального дослідження климату певної місцевости або країни. Таке спостереження зветься масовим /Sistematische Massenbeobachtung / і дає в результаті спотереження над цілим складним /атиповим/явищем. Чім простіші явища, чім менше вони виявляють мінливость в просторі й часі, чім менше на них впливають причини випадкові, тим вони типовіші, тим більшою є можливість переводити одній особі спостереження.

Статистичне спостереження якраз позбавлене такої можливости, воно є м а с о в и м. Ціла царина соціяльного життя, соціяльних явищ визначається саме складністю, мінливістю, множністю дівчих причин. Через те статистичне масове спостереження є одним з найважливіших знарядь наукового дослідження соціяльних явищ, остільки важливим, що частина видатних представників наукової статистики, напр., А. Боулі /A.Bowley / вважає статистику наукою суспільною, що ставить собі метою "дослідження соціяльного організму у всіх його проявах".

Мінливість соціяльних та інших атипових явищ, як показало масове статистичне спостереження не уявляє з себе одначе якоїсь непереможної сили, що приводить ці явища до "анархічного" стану в процесі їх існування і динамічного розвитку. Навпаки, і в них статистичне спостереження знайшло правильності, певні постійності відношень, що є такі важливі для наукового пояснення цих явищ.

Для прикладу подамо смертність населення Спол.Ш.Півн.Америки на І.000 душ населення цілої країни. Вона дає таке число осіб:

Вікові групи населення	сільське	міське	столичне
Для всіх вікових груп	I5,34	22,I5	24,6I
Віком до I року	I2I,2I	243,32	264,35
" " 5 років	37,I2	80,40	89,25
" від 5-I5 "	4,03	6,2I	6,I6
" " I5-45 "	6,89	I0,80	I2,07
" 45-65 "	I5,I9	26,29	3I,52
" вище 65 років	67,83	88,60	96,62

В цих цифрах разом з мінливостю і ріжницями, ми бачимо певні правильності; розглядаючи числа прямовісно, бачимо, що й для сільського і для мійського населення смертність дітей до першого року найвища, можна сказати, навіть, виключна; далі, до 15 років життя вона зменшується і в селах і в містах, а з більшим віком зростає знов таки всюди. Розглядаючи числа поземому напрямкові, бачимо, що як для всіх вікових груп, взятих разом, так і для окремих вікових груп смертність сільського населення значно менша ніж мійського, а серед мійського нижча ніж серед столичного. Все це говорить, що в мінливості /ріжницях/ смертности по віку й характеру поселень є певний напрям, відбуваються правильності.

Спостерігаємо також числовим спостереженням, що у своїй ріжноманітності і мінливості явища виявляють також до певної міри стійкість, тобто хитання відбуваються лиш у певних не дуже широких межах.

Наприклад, число шлюбів у Швеції щорічно не дає вже таких великих ріжниць як можна було б думати відносно такого індивідуального явища: так в 1901-1905 р. найменше число шлюбів було - 30.088, а найбільше - 32.750. Обчислення за 1898-1906 р.р. показує, що число шлюбів по кожному року за цей період не перевищує 577 - 624 шлюбів на 100.000 населення Швеції. Для Англії за 1901-1906 р.р. це число дає від 737 - 760 шлюбів на 100.000 душ населення. Для України відповідні дані будуть такі: 1891-1895 р.р - 900; 1896-1900 р.р. - 870; 1901-1905 р.р. - 810; 1906-1910 р.р. - 880 шлюбів на 100.000 населення. Ще більше стійкости уявляють із себе числа, що торкаються заставлених річей в Королівському Льомбарді в Берліні, де заставлені кишенькові годинники складали серед решти заставленого майна такий відсоток:

1889 р.	1900 р.	1901 р.	1902 р.	1903 р.
16,4%	16,02%	16,71%	16,75%	17,12%

Пересічна ціна кожного екземпляра заставлених годинників складала для тих самих років: 21,40 мар.; 21,50 мар.; 21,15 мар.; 21,01 мар.; 21,26 мар. Виходить, що коли б заставодавці умовились зробити так, щоб щорічно повторювати свої застави так щодо їх числа, як і щодо ціни, то й тоді, навряд чи одержалася б більша одноманітність в цих числах з року на рік, аніж без такої змови. Ці правильності дають нам можливість передбачення та врахування для багатьох практичних цілей. Щоб встановити явище скільки людей щорічно хоріє і не тільки взагалі, а й від ріжних хороб, скільки дітей щорічно підростає і потребує школи, скільки людей можна взяти до війська і т.д. - треба й організувати спостереження цього, себто через реґістрацію кожного окремого випадку та через підрахунок їх всіх для кожної місцевости окремо. А щоб спостерегти й ті зміни, що відбуваються в часі, треба їх зафіксовувати щорічно, або й ще частіше. Ясно, що таке спостереження мусить бути організованим, с и с т е м а т и ч н и м.

Щоб находити як ці складні явища виявляються в своїй ріжноманітності, напр., смертність у містах, в селах, ріжних професіях і под., ми не можемо обмежитись лише декількома місцевостями, декількома містами, селами або фабриками, бо через спостереження одного або декількох явищ ще не можемо виводити те ж саме відносно других явищ того самого роду.

Через те мусимо розширити спостереження на цілу масу явищ того ж самого роду, при якому виявляться та будуть охоплені всі відмінности об'єкту.

Масовість спостереження й є важливою ознакою статистичного спостереження. Це останнє крім того, мусить бути організованим так, щоб переводилось о д н о ч а с н о для певного роду явищ, бо тільки тоді з послідовних аналогічних спостережень можна бачити як і де явище змінилося. Звідси випливає необхідність п о в т о р е н н я спостережень, щоб судити на підставі їх про зміни в явищах не тільки в просторі, але й в часі.

Щож до явищ, що змінюються постійно, то вони мусять і спостережуватись п о с т і й н о.

На підставі сказаного, також можна переконатись, що масове спостереження, охоплюючи цілу сукупність явища в просторі, мусить бути в и ч е р п у ю ч и м. Одначе вичерпливість ця є ідеалом мало здійсним. Статистична метода вчить і має способи переводити м а с о в е с п о с т е р е ж е н н я і н е в и ч е р п у ю ч е, але задовольняюче вимогам наукового дослідження, про це ми скажемо в іншому місці, коли будемо характеризувати т.зв. вибіркову методу спостереження, як окрему форму статистичного спостереження. Тут мусимо лише зазначити, що при статистичному спостереженні мусимо поневолі користуватися лише частиною тієї „маси", яка йому повинна підлягати. У всякому разі наші старання мусять бути направлені на те, щоб частина ця була можливо більшою, щоб для спостереження були вибрані місце й час найбільш характерні, щоб число об'єктів, пунктів місцевих було найбільше, а час спостереження можливо довший. Треба памятати, що вірність висновків при масовому спостереженні росте у значно слабшій степені, ніж саме число спостережень, себто точність висновку росте, при всіх однакових умовах, тільки як корінь квадратовий з числа спостережень / $\sqrt{N}$ /. Коли захочемо наблизитись до абсолютної точности висновку у прогресії I, 2, 3, 4 і т.д., то число спостережень мусить рости, напр., у прогресії I, 4, 9, I6 і т.д.

8. Числове спостереження. Статистика далі є м а т е р і я л ь н и м ч и с л е н н я м, а через те статистичне спостереження є ч и с л о в е, але воно ніколи не має до діла лише з рахунковими одиницями або рахунковими сукупностями /символічними/, але з одиницями й числами тісно зв'язаними з матеріяльною природою тих явищ, що їх числиться.

Специфічна властивість статистичної методи є в тому, що будучи наукою про дослідження колективних атипових явищ, вона допомагає переводити теоретичні концепції у вимірні, якісні характеристики у кількісні. Як тільки статистик забуває про якісну характеристику, про матеріяльну природу колективів, що студіює, він сходить з наукового шляху, підпадає впливу ідеальних числових математичних схем та ігнорує реальні сили розвитку, що керують певним колективом / М.Сміт/

Суть статистичного наукового спостереження /числового, масового і методичного/ можна коротко подати так:

I. Складний об'єкт /явище/ розкладається на неподільні, що його складають, а кожне неподільне - на свої характерні ознаки. 2/ Спостереження переводять виключно констатуванням наявности тієї або іншої ознаки для кожного неподільного. 3/ Здобуті одиничні свідчення сумуються і одержані величини /суми випадків, в яких ознака була зафіксова/ дають вислів тим властивостям або якостям, що відшукуються в явищах. 4/ Наслідки спостереження завше мають кількосний образ, а через те й наукове оброблення цих наслідків складається з правил і способів, що вимагаються математичною аналізою, поскільки аналіза ця допустима для величин не обстрактних, а конкретних, статистичних. 5/ Три стадії, що обовязково мають бути в кожному дослідчому процесі - спостереження, аналітичне зведення одержаних фактів і висновки у масовому спостереженні різко в і д д і л ю ю т ь с я: особи, що переводять реєстрацію, можуть не займатися аналітичним зведенням матеріялу і не оброблять його для одержання наукових висновків/Ю.Янсон /.

Вияснення понять колективного атипового явища, та статистичного спостереження приводить нас до важливих міркувань, а саме тих, що для явищ суспільного життя, явищ повних, так мовити, атиповости, статистичне спостереження є тим науковим спостереженням, що заміняє експериментальне дослідження. Поскільки нам, економістам, дослідникам господарського життя таке спостереження безумовно потрібне, необхідним є його вистудіювання. Відомий економіст Зомбарт наводить такі слова славетного філософа Канта: „Нема сумніву, що все наше пізнання починається з

досвіду, бо що ж інше могло розбудити чинність нашої пізнавальної здібности, як не предмети та явища, що будять наші почуття і почасти самі викликають уяви, а почасти призводять до чинности нашу здібність розсуду, спонукавчи її порівнювати предмети між собою, зєднувати або роз'єднувати їх між собою, переробляючи таким чином сировий матеріял змислових вражінь у пізнання предметів, що й зветься досвідом." Наводячи ці слова для підкреслення думки про важливість і необхідність статистичного досвіду в економіці, Зомбарт каже: "Часто-густо буває, що якийсь недисциплінований розум пропонує нам економічну або соціяльну систему, пробуючи вивести цілу сукупність соціяльних фактів з декількох недоказаних посилок, подібно до того, як шовковиця змотується з опряді"... Далі він продовжує: "слід рішуче стояти на тому, що с о ц і я л ь н а н а у к а є н а у к а е м п і р и ч н а, у високому розумінні цього слова, що вона повинна в основу кожного із своїх пізнань класти безпосереднє спостереження над подіями самого життя. Наша наука, каже Зомбарт, /себто політична економія/ ніколи не може хоріти від перебільшення емпіричного наукового матеріялу, і зараз нас скорше гнітить недостача в придатному для справи фактичному матеріялі. Позитивного знання, знання к о н к р е т н о ї б у д о в и г о с п о д а р с ь к о г о ж и т т я, знайомства з історичним минулим, а особливо спостережень над реальними подіями сучасности - ось чого перш за все не вистачає більшості із представників нашої науки, навіть так званим політико-економістам реалістичного історично-емпіричного напрямку. Для студіювання, дослідження економічних явищ треба спостерігати і спостерігати". Головним знаряддям такого спостереження є безперечно спостереження статистичне, - систематичне, числове, масове.

9. Органи спостереження. Ми вже зазначали вище, що статистичне спостереження не під силу одній особі, воно обов'язково мусить переводитись кооперацією багатьох людей. Такого роду спостереження вимагають великих матеріяльних витрат та торкаються в більшій частині важливих інтересів широких мас населення тієї або іншої країни. Через це такі спостереження переводяться головним чином органами публічної влади і управління, чи то центральними чи то місцевими /самоврядування/. Але не має більших невигід для статистики, як доручення адміністративним установам таких стасистичних реґістрацій, які не контролюються їх службовою відчитністю й відповідальністю, та у правильному веденню яких не зацікавлено само населення. Через це в багатьох випадках завдання пляну спостереження дуже ускладнюється, та виникає необхідність створення або часових спеціяльних органів статистичної реґістрації, або постійних статистичних установ на які покладається систематичне, масове, числове спостереження, що й бачимо зараз у всіх майже культурних країнах світу.

ВПРАВИ. Напишіть коротко і пришліть відповіді на наступні запитання:

1. Як визначена в І-й лекції статистика яко наука?
2. Наведіть один-два приклади колективних, але типових явищ?
3. Наведіть один-два приклади атипових /нетипових/ колективних явищ.
4. Яка ріжниця у формулюванні причинових відношень і зв'язків в явищах логикою і статистикою?
5. Які явища визначаються особливою с к л а д н і с т ю та м н о ж н і с т ю причинових зв'язків?
6. Через що поділ умов /причин/, що викликають явище, на п о с т і й н і й в и п а д к о в і має релятивний характер?
7. Яким явищам відповідають р о д о в і п о н я т т я?
8. Які завдання ставить собі статистика яко наука?
9. Які головні ознаки статистичного спостереження і що його відріжняє від експериментальної методи наукового дослідження?
10. Що то є за визначення: матеріяльне числення?

ЧАСТИНА ПЕРША.

ОПИСОВЕ ЗАВДАННЯ. СТАТИСТИКА / ІДЕОГРАФІЯ /.

Лекція друга

ГЛАВА І. Статистична сукупність.

§ 10. Властивості статистичної сукупности. Кожне колективне, атипове явище, природу якого ми висвітлювали вище, може бути через числове, масове і систематичне спостереження перетворене в статистичну сукупність. Себто без статистичного опису явищ, чи подій, нема й статистичної сукупности. Кожна статистична сукупність є сума предметів певного роду, точніше, сума одиниць якогось загального поняття. Але ця сукупність звичайно не охоплює цілого необмеженого обсягу неділимих /одиниць/, що підпадають під дане загальне поняття[x]/ а лише деяку, певну частину одиниць цього поняття, обмежену певними рямцями простору й часу.

Наприклад, при перепису населення ми цікавимось тільки одиницями що об'єднуються поняттям "людина", але з цілого необмежено широкого обсягу цього поняття вибираємо лише певну частину, певну сукупність одиниць, що перебуває в певний момент на певній території.

Коли ставимо собі завдання статистичної ідіографії / опису/, тоді така сукупність одиниць - в певних рямцях простору й часу - цікавить нас як така / сама по собі/. Коли ставимо собі задачу статистичної номографії /відшукання закономірностей, причинових зв'язків/, тоді нас ця індивідуальна сукупність одиниць не цікавить сама по собі, а є лише засобом до пізнання деяких загальних зв'язків чи постійних відношень, але й в цьому випадку статистичний опис явища мусить бути переведений.

Основою статистичної сукупности може бути або реальна /природна/ сукупність - між неділимими[xx]/одиницями/, з яких вона складається, існує певне реальне співвідношення, реальний зв'язок, - або штучно утворена статистиком-дослідником така сукупність, між неділимими якої реального зв'язку не має.

Можемо, напр., займатись студіюванням такої реальної сукупности як українська робітнича кляса, окремі одиниці якої зв'язані певною єдністю класових інтересів та політики. Але можемо скласти, на підставі даних перепису, сукупність дітей 5 років, що живуть на Україні І січня 1932 року, з метою пізнання цієї цілости. Хоч ці діти між собою реального зв'язку не мають, проте такого роду сукупности дуже часто студіюються статистичною методою і якраз до них найбільш стосується прикладання методи рахування імовірностей.

Важливість поділу цих двох видів статистичних сукупностей є безперечна, бо тільки за реальною сукупністю можна, здається, визнати значіння самостійної логічної категорії. "Штучна" сукупність, або ліпше кажучи, "абстрагована сукупність", що конструюється для цілей статистичного числення і аналізи, є просто маса /символ/ якоїсь множности, а не істотна властивість певного масового явища. Вона потрібна як знаряддя наукового опису і наукової аналізи, щоб виявити існування постійности тих або інших відношень через масу незалежних одиниць сукупностей, яко збірних понять

---

x/ Під обсягом загального поняття логіка розуміє суму всіх предметів з такими властивостями чи ознаками, що стосуються до обсягу даного поняття.

xx/ Під "неділимим" сукупности будемо завше розуміти: одиниці, індивідуми, предмети, екземпляри, з яких сукупність може складатись.

Крім того, треба мати на увазі, що статистичною методою можна студіювати так с у к у п н о с т і п р е д м е т і в, як і с у к у п н о с т і п о д і й. Під предметами тут розуміємо сукупності предметів неживих і живих, навіть і людей, аби вони були дійсними предметами з більш менш с т а л о ю е к з и с т е н ц і є ю, не будучи подією, яка є явищем, що відбувається в ч а с і. Наприклад, всі сукупності населення, сукупності промислових або сільсько-господарських підприємств, також товарів, що знаходяться у певній країні - суть сукупності предметів. Сукупностями подій будуть: сукпність народжень, смертей, шлюбів, нещасливих випадків з робітниками у виробництві і т.п.

§ II. Ознаки неділимих сукупности.

Одначе статистикові не ходить лише про обсяг сукупности, себто про числення тільки неділимих, з яких складається сукупність, але також і про певні властивості, о з н а к и тих неділимих, що складають сукупність. Ці ознаки є завжди м і н л и в и м и, вони ніколи не є властивостями родовими самого поняття, до якого належить статистична сукупність. Родові ознаки цікавлять нас доти, доки нам потрібно з'єднання одиниць /неділимих/ в певну сукупність. Як тільки сукупність складена, нас, статистиків, цікавлять ознаки мінливі, що від одиниці до одиниці змінюють свій кількісний і якісний вираз.

Ця мінливість ознак розуміється в двох напрямках: поперше, ознака в одних випадках виявляється, в других ні, себто м о ж е з н а х о д и т и с ь л и ш е у п е в н о ї ч а с т и н и о д и н и ц ь с у к у п н о с т и; подруге, що ознака набуває різних квантитативних /кількісних/ вартостей у різних одиниць сукупности - є з м і н н о ю в е л и ч и н о ю. Наприклад, при перепису населення нас можуть цікавить дефективності окремих одиниць - сліпота, глухота і т.п. При статистичній реґістрації лише в таких випадках констатуємо - має чи не має кожна одиниця населення вказаний вище тілесний дефект, чи п о с і д а є в о н а ц ю о з н а к у ч и н і.

Коли спостерігаємо таку ознаку як зріст або вагу кожної одиниці сукупности, то нас цікавлять лише кількісні вартості /стільки а стільки сантиметрів або кілограмів/, яких набуває ця ознака у різних одиниць. Ознаки, які цікавлять нас лише в розумінні наявности чи відсутности їх у одиниць сукупности, звуться к в а л і т а т и в н о /якісно/ м і н л и в и м и, або просто к в а л і т а т и в н и м и о з н а к а м и, а о з н а к и, що цікавлять нас, яко різні кількісні вартості, звуться к в а н т и т а т и в н о - м і н л и в и м и, або к в а н т и т а т и в н и м и. Цей поділ мінливости ознак не може, одначе, залежати завше від самої істоти ознаки, бо часто тут рішає те, з якого погляду цікавить дослідника певне явище і в залежності від цього мінливість ознаки квалітативна може стати квантитативною. Наприклад, в статистиці населення, як було вже вище зазначено, ознаки - сліпота, глухота є квалітативні, але ці ознаки припускають і квантитативний вимір, коли побажаємо міряти силу зору або слуху. Аналогічно багато ознак, що цікавлять нас, яко квантитативні, можна розглядати лише з погляду їх наявности або відсутности, себто як ознаки якісні. Наприклад, робоча худоба в сільському господарстві найчастіше нас цікавить яко певна кількість тварин в різних селянських господарствах, але худоба могла б нас цікавить також з погляду наявности або відсутности у кожному господарстві цієї ознаки. В такім випадку ми характеризували б сукупність господарств, наприклад, певним процентом господарств, що зовсім не мають робочої худоби, себто ознакою "робоча худоба" користувались як квалітативною, якісною. Є все ж таки й такі ознаки, які по своїй істоті можуть бути лише квалітативними, бо не можуть набути певних кількісних вартостей. Наприклад, ознака родинного стану - жонатий, нежонатий або статі - мужчина, жінка - в статистиці населення.

З певного погляду квалітативна мінливість ознак може вважатись за спеціяльний випадок квантитативної мінливости, а саме тоді, коли ми випадки наявности певної ознаки зазначаємо як випадки, що міряються вартістю I, а випадки відсутности ознаки, як випадки, що міряються вартістю 0; таке, до певної міри, штучне означення досить часто вживається в статистичній методі /S.Kohn /. Треба ще зазначити, що досить часто наявність однієї ознаки вказує на відсутність другої, протилежної / мужчина - жінка/. Таке положення може бути й при декількох відмінах ознаки. Наприклад, при поділі людей по ознаці барви волосся, останнє розподіляється на світле, русяве, чорне і т.п., або теж саме народність та релігія, яко ознаки,можуть мати десятки відмін.

Характеристика сукупностей з таков "многовідмінной" ознаков складається з цілого ряду характеристик сукупности щодо наявности окремих відмін ознаки.Таким чином статистик завше оперує з статистичними сукупностями і ознаками тих неділимих сукупности, з яких вона складається. З якою б метою статистик не підходив до завдання свого дослідження - з метою статистичного опису сукупностей,чи з метою відшукання емпіричних закономірностей в них і через них, - він мусить виходити з кількісного виміру одиниць сукупностей і тих ознак, які він знайшов доцільним числити в одиниці кожної сукупности для досягнення наукової мети, яку собі поставив.

Виходячи з основного поняття статистичної сукупности, частина теоретиків статистиків у своїх викладах цієї науки переходить безпосередньо до табличного зведення статистичних матеріялів, укладання статистичних рядів і т.п., поступово розвиваючи операції статистичної методи та оминає зовсім статистичне числення і правила його переведення. Це,мовляв, статистична техніка, а не теорія методи.

Наприклад, виходячи з цієї засади, відомий статистик Р.Орженцький і визначає зміст статистичної науки і статистичної методи. Він каже, що зміст цієї методи складають "способи студіювання зведених ознак і зведених закономірностей в тих випадках, коли ці ознаки й закономірності можуть бути визначені в числових величинах". "Сама метода, каже він, складається з характеристики предметів і явищ через зведені ознаки, або засада методи є встановлення зведених ознак та дослідження і перевірка їх реального, конститутивного і елементарного характеру". Цей характер, ці властивості можуть встановлюватись або індуктивно, через порівняння між собою цілих сукупностей, або статистично, через зіставлення сукупностей з ознаками індивідуальних об'єктів сукупностей".

З вищенаведеного бачимо, що Орженцький не відносить до істоти статистичної методи м о м е н т у о д е р ж а н н я т и х "і н д и в і д у а л ь н и х о з н а к", п і д с у м о к я к и х д а є з в е д е н і о з н а к и т а з в е д е н і х а р а к т е р и с т и к и. Себто, момент статистичної реєстрації явищ буцімто непотрібно вводити в науковий виклад статистичної методи. Так думає не один тільки названий вище вчений, а й інші; згадаємо хоч би видатного вченого англійського статистика Юла /G.Undy Yule /.

Думаємо, що таке звуження статистичної методи є некорисним. "Зведені ознаки" одержуються через статистичне спостереження, з додержанням необхідних правил статистичної реєстрації фактів, подій і явищ. Незнання цих правил і недодержання їх тягне за собою надзвичайно важливі наслідки, через які дальша наукова обробка статистичного матеріялу може бути непевною, а інколи й зовсім неможливою. Яскраві приклади можна подати з практики переведення переписів. Наприклад, під час промислового перепису в Німеччині у 1907 році,через неправильну дефініцію о д и н и ц і ч и с л е н н я - "промислового підприємства" та невідповідне інструктування осіб, що подавали відомості у формулярах цього перепису, довелося велику частину матеріялу перепису повертати з центра завідування переписом для виправлення на місцях, себто поновлювати в значній частині реєстрацію.

Через те саме дальше оброблення матеріялу перепису, його результати справедливо були піддані немилосердній критиці, особливо в талановитій праці німецького статистика Рудольфа Меерварта /R.Meerwarth / під назвою "Вступ до господарської статистики".

З неменшою силою переконання відомий англійський статистик Боулі піддав критиці видання ріжних матеріялів офіціяльної англійської статистики та зробив висновок, що "дійсне значіння друкованих цифр офіціяльної статистики не відповідає зовсім тому, яке можна їм надати на підставі зовнішнього вигляду і конструкції. Важливі дефекти загального перепису населення, переведеного у Польщі в I92I році, а також сільсько-господарського в I9I7 році в Росії й на Україні, пояснюються недодержанням або порушенням засад правильного статистичного числення. Можна було б подати досить прикладів невдалого переведення основної статистичної реґістрації в земських переписах в Росії та на Україні.

Все це свідчить про те, що необхідною частиною ознайомлення із статистичною методою є також і ознайомлення з тими формами статистичної реґістрації та важливішими правилами її переведення, які більш-менш забезпечують статистичними сукупностям добру якість дальшої їх наукової характеристики та наукового оброблення. Сучасне наукове оброблення статистичних сукупностей, що складається з цілого ряду чітких, математичних способів опису явищ та відшукання в них постійности відношень і зв'язків, на нашу думку, не може переводитись і є недоцільним тоді, коли статистичні сукупності цих явищ та многовідмінні ознаки неділимих цих сукупностей здобуті без належної свідомости, краще сказати, без статистичної свідомости, фахового спостереження.

В цій частині викладу вважаємо доцільним зупинитися на тій методології статистичного спостереження, в тісному розумінні цього слова /реґістрація явищ/, яке відноситься до явищ суспільного, соціяльно-економічного життя. Це через те, що статистикам-економістам найбільш доводиться мати діло з такого роду явищами, де статистична метода, метода кількісної оцінки подій і явищ, в цій царині знання має виключне значіння.

Прикладення статистичної методи, правда досить широке, в інших царинах знання, особливо в біології, зокрема в аґрономії / в селекції рослин/, вимагає і своєрідного освітлення правил цього прикладання, але таке освітлення занадто поширило б наш виклад лекцій, через те, не з принципових мотивів, а технічних, мусимо обмежитись лише характеристикою статистичного спостереження явищ соціяльної природи, як на важливішому для нас, економістів, серед досить широкого інтернаукового поля прикладання статистичної методи, дослідження[x]/

ГЛАВА II.

СТАТИСТИЧНЕ ЧИСЛЕННЯ, ЙОГО ОСНОВНІ ФОРМИ Й ПРАВИЛА ПЕРЕВЕДЕННЯ.

§I2. Одиниця числення. Сукупності, що досліджуються статистичною методою, створюються в наслідок числення. Розміри цих сукупностей виявляються числом більшої або меншої величини, що стосується як до обсягу відокремленої маси числення, так і до частости в складі цієї маси неділимих одиниць, позна-

---

x/ Для тих, хто цікавився б приміненням статистичної методи в інших царинах знання, можемо рекомендувати дуже цікаву колективну працю російських вчених під редакцією М. С м і т і А. Т і м і р я з є в а - "Статистический метод в научном исследовании" Москва. I925 г.

чених певною прикметою. Наукову уяву про сукупність /масу/ ми одержимо лише тоді, коли знаємо с к і л ь к и входить в певну сукупність одиниць, що мають кожна наперед означені певні прикмети. Таким чином, раніш, ніж переводити таке числення, треба обрати о д и н и ц ю ч и с л е н н я т а ї ї п р и к м е т и. Треба перевести попередню якісну аналізу, об'єкту, тієї маси, бо така аналіза необхідна для числового масового спостереження та є його змістом і метою. Відціль вимога доброго знання того об'єкту, з яким мусимо оперувати статистичною методою. Не знавчи біології вважалі, не можна ставити собі наукових проблем, навіть не можна дати доброї характеристики біологічних явищ статистичною методою. Не знавчи соціяльної природи явищ, або економічної, не можна цією методою досліджувати науково такі явища. Для цього треба наперед орієнтуватися в об'єкті, який досліджуємо. Від влучности такої орієнтації, її глибини та широти, залежить і більший, або менший успіх статистично-наукової праці. Тут треба добре пам'ятати, що спочатку мусимо створити в своїй голові якісну характеристику явища, бо потім обов'язково доведеться перетворювати її в кількісну.

Візьмемо, наприклад, таке питання /явище/, як безробіття. Щоб дати його статистичну кількісну характеристику, щоб зважити його значіння в соціяльному житті певної країни, треба наперед вирішити, що будемо розуміти під безробіттям, треба встановити що то є безробітний, а це не так легко, як здається на перший погляд. Коли не знаємо, що є види праці, де робітник нормально зайнятий лише на протязі певної доби року, в сезоні, то може статися, що при численні зарахуємо до безробітних і так званих сезонових робітників, змішаємо їх з дійсно безробітними. Робітник може бути зайнятий тільки певну кількість днів в тижні, а може бути й зовсім незайнятий цілий тиждень, себто безробіття може бути повне і часткове. У промисловому виробництві країни можуть бути такі роди промислу, де панує страшенне безробіття, в других його майже немає; треба знати клясифікацію виробництва. Навіть треба встановити якого віку людей мусимо вважати робітниками, бо працюють і діти, мужчини й жінки. Надзвичайно тяжко підрахувати й ті страти, що має виробництво від безробіття.

Увесь цей клубок понять, ознак треба розмотати, кожне та кожну з них можливо чітко встановити і лише тоді числити. Такий попередній логічний процес думки над об'єктом обов'язковий для кожного статистичного спостереження.

§I3. Моменти часу й простору в статистичному численні.

Виходячи з матеріяльної природи об'єкту дослідження, кожний статистик, як то раніше ми зазначили, повинен встановити одиницю числення та зробити вибір тих матеріяльних ознак одиниці, що потрібні для наукової характеристики явища, або сукупности. Тим самим дослідник розбиває масу на більше або менше число груп, при чому до складу кожної групи входять неділимі, що мають одну з прикмет /ознак/, обраних для характеристики цілої маси.

Наприклад, характеризувати населення можемо по статі й віку, тоді обидві статі, кожна зокрема, будуть розподілені на вікові групи, що мають вік до I-го року, I рік, 2 роки, 3 роки і т.д..

Групи, в які ми об'єднуємо таким чином неділимі одиниці по окремих ознаках мають назву г о м о л о г і ч н и х /однорідних/. Але щоб одержати гомологічні групи, наприклад, населення, або іншої сукупности, ми мусимо своє числення обов'язково о б м е ж и т и я к и м и с ь р я м ц я м и ч а с у й п р о с т о р у. Не можемо числити населення зразу на цілій земній кулі, але лише в державі, країні, повіті, окрузі і т.п. при тому охоплюємо його реєстрацією в певному і однаковому часі на цілій території, яку собі відмежували.

Відмежувати якесь явище в часі, через однорідність виміру часу-день, тиждень, місяць, рік, - можна досить точно, бо найбільше тут роз-

ходження маємо в неоднаковому числі днів місяця, себто від 28 днів до 3I дня. Таким чином, перемежки часу, в які укладаються поодинокі реґістровані факти, будуть приблизно рівні. Необхідність одночасового числення випливає також з того, що переведення його в ріжні, навіть не дуже довгі перемежки часу, може потягти за собою наслідки подвійного почислення значної кількости одиниць маси, яку числимо, особливо тоді, коли ця маса в певні моменти часу інтенсивно рухається.

Наприклад, при переведенню загальної реґістрації населення якоїсь країни, треба мати на увазі, що в окремих місцевостях цієї країни можуть бути періоди інтенсивного пересування населення з одного місця на друге, і тоді не виключена можливість подвійного урахування певної групи населення спочатку в одному, а потім тієї ж групи і в другому місці осідку.

Це й сталося при переведенні загального перепису населення у Росії в I897 році. Закавказзя має дуже нерівний рельєф території. Значна частина населення займається скотарством, при чому восени пересувається на зимові пасовиська в низини, а повесні, навпаки, до високих гірських пасовиськ. В другій половині квітня значні маси господарств, з жінками, дітьми та худобою, тягнуться по ріжних шляхах з низини до гір, часто з однієї губерні до другої, не то що з повіту в повіт. Центральний уряд статистичний, що призначив на січень ст.стилю по цілій Росії перепис населення, на прохання місцевої адміністрації, яка вказала на те, що в січні на Закавказзю є високі місцевості /Сванетія/ зовсім відрізані від решти поселень, дозволив перевести реґістрацію в квітні місяці. Тоді й сталося те, що в деяких місцях перевели реґістрацію з початком квітня, але в другій половині квітня ця вже переписана частина населення опинилась у другій зовсім губерні, на горах, де якраз переводили той перепис в другій половині квітня. Через те декілька тисяч населення було враховано подвійно.

Коли, таким чином, є необхідність у статистичному численні зробити якесь відступлення, порушення правил одночасовости реґістрації на цілому просторі спостереження, то треба передбачати шкідливі наслідки такого порушення та усунути їх відповідними заходами і розпорядженнями.

В статистичному численні населення дуже важливо не тільки встановити одночасовість періоду спису його, напр., місяць грудень такого то року, а й так званий критичний момент реґістрації / день і час/.

Напр., охоплене численням I926 року населення УССР є те населення, яке до I2 годин ночі I7 грудня I926 року, в тому числі й немовлята, що народилися в цей день, почислено на цілій території республіки. Дванадцять годин ночі I7 грудня й є к р и т и ч н и й м о м е н т ч и с л е н н я н а с е л е н н я для даного перепису.

Таким чином, правильний плян статистичного числення вимагає влучного п е р і о д у ч и с л е н н я, в залежності від того, в який час, коли саме, знаходиться об'єкт вашого числення в стані більш менш стабільному /тривалому/. Не кожний, звичайно, об'єкт числення допускає такий точний критичний момент часу, як це бачимо у численні населення, але тут діє загальна засада: щ о б з а б е з п е ч и т и т о ч н і с т ь м а с о в о г о ч и с л е н н я, т р е б а я к о м о г а н а б л и з и т и м о м е н т и н а с т а в а н н я в і д п о в і д н о ї п о д і ї т а ї ї р е ґ і с т р а ц і ї.

Незалежно від того, треба в кожному численні встановити також р е ч и н е ц ь т е х н і ч н о г о п е р е в е д е н н я р е ґ і с т р а ц і ї. Він визначається першою й останньою часовими датами того періоду, в який обов'язково потрібно статистичне числення перевести. Щодо відмежованности якоїсь маси в просторі, то цей момент в статистичному численні практично відограє може ще більшу роль, ніж момент одночасовости. В практиці статистичне числення у великому маштабі переводиться або на кошти держави для задоволення цілодержавних інтересів, або органами місцевого самоврядування в інтересах цього останнього, або, нарешті, тими й другими разом.

Наприклад, земські оціночно-статистичні праці в Росії і на Україні переводились за допомогою державних коштів та за зверхньою контролею державних органів, охоплюючи або цілу територію або великі її частини губерні. Доводиться поневолі дані статистичного числення стосувати до адміністративного їх поділу.

Величезна праця цифрової групіровки здобутого численням матеріялу, що вимагає значних коштів і фахового персоналу, не дає можливости пристосувати цей матеріял до дрібної територіяльної одиниці, а лише до таких одиниць, як губерні, повіти, округи і т.п. Але ще більш важливим є те, що адміністративний поділ території є поділ досить штучний, створений в інтересах адміністративного управління, а не з метою науковою, себто такою, де поділ території, її районування забезпечує яко мога більше для певних частин статистичної сукупности однозначність, рівномірність свого впливу.

Це особливо має значіння тоді, коли гомологічні групи сукупности щодо характеру своїх прикмет знаходяться в зв'язку з силами, що існують і діляють з приблизно рівною інтенсивністю на певній території. Напр., в сільському господарстві з великою силою діють на його структуру, його розвиток /динаміку/ природно-історичні фактори, що скупчені на певній території і стабільність цих сил та їх властивості определяються не згідно з адміністративним поділом території, а згідно з комплексом природно-історичних умов /рельєф, ґрунт, клімат, ступінь культури і т.д./, що скупчені в певному територіяльному районі.

Отже відповідно до матеріяльної природи явища, яке досліджуємо, мусимо часто відшукувати для статистичних гомологічних груп цього явища і відповідний районовий поділ території. Лише приблизно рівні умови, зв'язані з певним територіяльним районом дають дослідникові-статистикові можливість відшукувати та міряти /статистичною методою/ ті зв'язки, взаємовідносини, які існують між окремими прикметами - факторами певної сукупности.

Для зрозуміння важливости районового поділу території в статистичній методі дослідження візьмемо такий приклад. Поставлено завдання - визначити прибутковість в якомусь місті будинків, помешкання яких наймаються за плату, та з'ясувати, які головні прикмети - фактори впливають на величину цієї прибутковости і як вони на неї впливають.

Зробивши числення таких будинків в даному місті, переводимо його так, що для кожного будинку маємо зареєстрованих певну кількість важливих, на нашу думку, ознак, а саме, де будинок знаходиться /вулиця, дільниця/, з якого матеріялу він зроблений, який кубічний об'єм жилих помешкань, по кожній кватирі окремо, число кімнат кожної кватири і ще багато інших прикмет.

Зробимо далі відповідні обрахунки того, кільки кубічних метрів мають помешкання і цілий будинок. Тоді почнемо шукати, наприклад, як зв'язаний розмір платні за помешкання в будинках з їх кубічним об'ємом. Для цього звичайно треба взяти будинки однаковісенькі щодо матеріялу, поверху, числа квартир, кімнат і т.д., а лиш тільки різні щодо свого об'єму. Така група складає, припустимо, 200 будинків. Спочатку поділивши суму плат за помешкання всіх цих будинків на суму кубічного об'єму цих помешкань знайдемо, що середня гуртова прибутковість І-го кубічного метра для цілої групи будинків рівняється, допустимо, 6 карбованцям. Коли візьмемо гуртову прибутковість по окремих об'ємних групах, менших і більших, то виявляється, що, напр., будинки з об'ємом в межах 5-І0 куб.метрів дають прибуток І0 карб. на І куб.м, будинки від І0-І5 куб.м. - 8 карб.за куб.м., І5-25 куб.м. - 6 карб. і т.д.,-зі збільшенням об'ємної групи будинків, плата за І куб.м. поступово зменшується. При тому різниця в платні за куб. м. тих будинків, що входять в окрему об'ємну групу, дуже незначна, так що всі будинки, наприклад, першої групи розміром в 5,5 куб. м. в 6 куб.м., 6,50 куб.м. і т.д., аж до І0 куб.м. дають майже однакову, дуже близьку платню за куб одиницю, що дорівнюється І0 крб.за І куб. м.

21424 43

Теж саме спостерігаємо і в других групах. В результаті мали б право зробити висновок, що між об'ємом жилого будинку і гуртовим прибутком його існує зв'язок, при тому такий, що чим більше об'єм будинку, тим на кубічну одиницю цього об'єма припадає гуртовий дохід менший і що для даного міста, для певного періоду часу /коли взяти прибуток не за один, а декілька років/, цей зв'язок характеризується певними величинами.

Але повстає питання, чи не зробили ми помилки щодо величин цього зв'язку? Чи нема факторів, на які ми не звернули уваги, що змінюють цей зв'язок; приміром таких ознак, що, не порушуючи загальної тенденції зв'язку, зменшують або збільшують цифрові покажчики цього зв'язку.

При вдумливому відношенні до справи дослідника, дійсно знайдемо, що є фактор, який безперечно впливає на цей зв'язок. Ми раніш не звернули уваги на те, що будинки знаходяться в різних частинах міста, в різних його районах; раніш ми тільки дивились на місце будинку, як на адміністративний, поліційний його куток. Подивимось зараз чи не впливає також на гуртовий прибуток куба помешкання , район міста, де будинки знаходяться?

Беремо однакові цілком будинки, в тому числі однакові й щодо об'єму, та простежемо по вулицях, що пересікають ціле місто, чи не міняється розмір платні за куб.одиницю цих будинків? Дійсно, виявляється, що по одній і по другій вулиці, які тягнуться через ціле місто,- в центральній частині міста платня за куб. одиницю вища, а чим далі до окраїн менша, або що якась специфічна частина міста, напр., фабричний центр, показує значно дорожчу платню на куб.одиницю, чим решта окраїн. Виявляється далі, що ціна землі в цих частинах міста неоднакова, що є в ньому райони з більшою і меншою прибутковістю. Беремо ще ряд ознак та знаходимо, що місто можна розбити дійсно на декілька районів прибутковости. Вони показали, припустимо, що й в I-му районі і в 2-му і 3-му той зв'язок, що ми раніш встановили теж існує, але виявляється інакше. Наприклад, куб.метр будинків в першій групі від 5-I0 куб.м. оплачується в I-м районі не I0, а I2 карб., а в 2-му районі не I2 крб., а I0 крб., а в третьому не I2 і не I0 крб., а 8 крб. за I куб.м. Аналогічні зміни бачимо й для других об'ємних груп будинків. На підставі цього рішаємо, що зв'язок між гуртовим прибутком помешкання і районом міста, де воно знаходиться, є і при тому дуже інтенсивний. Щоб простежити вплив об'єму на прибутковість будинку, виходить, треба спочатку встановити райони, а в межах вже цих більш-менш однозначних територій простежити та встановити розмір зв'язку між об'ємом та прибутком за іншими ознаками.

З поданого тут робимо важливий висновок, що для в і д ш у к а н н я з в ' я з к і в в я в и щ а х, о с о б л и в о с о ц і я л ь н о - е к о н о м і ч н и х, в е л и ч е з н е з н а ч і н н я м а є в і д м е ж у в а н н я с у к у п н о с т е й в п р о с т о р і, що базується не на поділі адміністративному, більш-менш механічному, а на п о д і л і т е р и т о р і ї р а й о н о в о м у.[x]

Через те ми й бачимо, що статистичні публікації подаються, напр., для України, не тільки за адміністративним поділом її, а й за районовим - Полісся, Правобережжя, Лівобережжя і Степ. Подібний поділ безперечно дуже потрібний й для Західних Земель України. Нарешті, треба ще зазначити, що у випадках частих змін адміністративного поділу, як то ми спостерігали недавно на Великій Україні, ці часті зміни не дають потім можливости порівнювати результати двох або кількох статистичних чисельних у часі. А між тим порівняння результатів, наприклад, двох сільсько-господарських переписів, переведених в I920 та I930 р.р, має величезне значіння.

---

x/ Економічна географія, курс якої викладається в УТГІ, з особливою увагою ставиться до поняття і значіння району, через те ми й не зупиняємося більш над ним.

Л е к ц і я т р е т я.

§I5. Експедиційний і блянковий спосіб числення.

Статисчичне числення може провадитись двома різними способами: I/ Е к с п е д и ц і й н и м і 2/ Б л я н к о в и м.

Перший спосіб числення - це безпосереднє переведення опитування по певній програмі осіб на місцях спеціяльним кадром регістраторів для збирання потрібних відомостей числення. Другий - це одержання потрібних для числення відомостей через розсилку блянків /формулярів/ до тих осіб і установ, від яких ці відомості бажано здобути. Експедиційний спосіб вживається найчастіше в тих країнах, де населення станом своєї культури стоїть на невисокому рівні, має багато неписьменних, які не можуть добре зрозуміти тих питань, які поставлені у формулярі числення, не можуть дати на письмі відповідь на такі запитання, та взагалі не розуміють мети і значіння такого роду регістрацій, через що часто вороже ставляться до їх переведення. Через те, для успішного переведення числення експедиційним способом керуючому органу статистичного числення треба раніш підготовити потрібний кадр людей, інструктувати їх так, щоб вони одноманітно розуміли програму числення та зуміли здобути від населення відомості, що вимагаються формуляром. Опитування мусить переводитись т а к т о в н о без непотрібного дратування населення та з відповідним освітленням перед ними важливости мети числення.

Таке переведення числення та його добре підготовлення вимагає досить великих коштів, але проте забезпечує повноту статистичного матеріялу і його добру якість.

Блянковим способом статистичні органи безпосередньо звертаються до населення через органи адміністрації, місцевого самоврядування та сільського управління, через це цей спосіб значно дешевший. Цим способом числення взагалі частіш можна переводити та в досить короткий термін одержати потрібні відомості. Але в цьому випадку немає повної гарантії, що числення буде переведено без пропусків, що частина матеріялу буде задержана та несвоєчасно прислана центральному статистичному органу. Може бути також багато помилок і неправильностей, які вимагають інколи повторних запитів або й поновного заповнення формулярів.

Наприклад, при переведені в Чехословацькій республіці складного сільсько-господарського перепису в I930 році блянковим способом виявилось, що й процес збирання відомостей і вже присланий матеріял для Підкарпаття був значно тяжчим і значно гіршої якости, ніж то було зроблено в культурнішій частині республіки - Чехії. Навіть в таких культурних країнах як Англія, або Сполучені Штати Півн. Америки, доводиться порушувати принцип повної вичерпливости числення, так що маленькі, дрібніші сільсько-господарські підприємства, розміром I-3 акри, зовсім виключаються з перепису.

Звичайно, вибір того чи іншого способу числення залежить також і від складности самої програми числення. Лише в дуже культурних країнах, напр., масове складне числення сільсько-господарських підприємств, що, звичайно, ставить метою розкрити структуру сільського господарства, а почасти й його динамику / коли провадиться порівняння з попереднім численням/, може бути переведене блянковим способом.

Через те саме теорія статистичної методи вимагає, як каже відомий статистик А. К е т л е, щоб від осіб, що опитуються, одержувались лише відомості, що -"а/ безумовно потрібні для цілей даного статистичного числення, б/ які напевно можна одержати і в/ які не викликають у населення підозріння та побоювання"

Питання у формулярах треба ставити так ясно й просто, щоб усунути яко мога с у'б є к т и в н і с т ь спостережника. Найлегше це може статися тоді, коли населення само себе числить, себто в умовах самочислення. До певної міри це усунення с у б'є к т и в н о г о цінування явищ, що числяться, необхідно зробити і при експедиційному способі. Реґістратори мусять л и ш е р е ґ і с т р у в а т и зовнішні ознаки, а їх легко розпізнавати, бо стосуються вони до сучасного, а з тих, що стосуються до минулого, реґіструються такі, що заховують видимий слід у сучасному /Борткевич/.

Відсіль випливає, що й керівництво статистичним численням мусить бути у фахових руках, у осіб добре підготовлених і практично і теоретично. Вимоги, що стосуються до правильно поставленого статистичного числення, важливі ще й тому, що воно є н а й д о р о ж ч а с т а т и с т и ч н а о п е р а ц і я серед тих, які треба перевести аж до моменту оброблення і публікації здобутих матеріялів числення. Відомий статистик А.А.Чупров каже: "досягнути встановлення причинових співвідношень, коли маємо до діла з масовими явищами, можна тоді, коли будемо знати як для цього створювати сукупності, що студіюємо, себто, як п е р е в о д и т и ч и с л е н н я. Але мало знати, треба в м і т и п р и к л а д а т и с в о ї з н а н н я н а д і л і. З теорії статистики треба одержати знання, через практику опанувати техніком; лише після цього можна з надією на успіх приступити до наукової статистичної праці".

§16. Типи статистичного числення.

Характеризуючи природу явищ, які досліджуються статистичною методою, ми вже зазначили /перша лекція/, що важливою властивістю цих явищ є їх мінливість. Стану їх буття завше властива невпинна зміна. Але швидкість змін мінливих явищ не є однакова, себто різні явища мають неоднакову сталість в часі. Мінливість така може визначатись або через ті результати, які лише з бігом часу виявляються в явищах, або ці зміни можуть бути визначені тільки тоді, коли ознаки їх зафіксовуються /позначаються/ у момент наставання самого явища. Є явища, які мають такий короткий час існування, що статистична реґістрація їх може бути переведена тільки в мент їх наставання, інакше вони зовсім пропадають для статистичного числення, не можуть бути відновлені і не повторюються в тих самих умовах та з однаковими прикметами своєї властивости. Коли, наприклад, кожний випадок смерти не буде записаний в мент наставання, або в мент дуже близький до нього, то такий випадок не може бути потім записаний або зовсім, або буде записаний не повно, без тих обставин, в яких відбувався та без тих прикмет, якими певний померлий різнився від кожного з інших померлих. Коли товар не буде зареґістрований в момент перевозу його через кордон країни, пізніше не можна одержати даних про кількість довозу товарів та про складові частини цього довозу. Навпаки таке явище, як населення, хоч і мінливе щодо своєї кількости в складових своїх частинах, але разом з тим є більш-менш стале, постійне. Так само промисловість, сільське господарство і т.п. суть явища, хоч і мінливі, але більш-менш сталі, р е л я т и в н о п о с т і й н і.

Для того, щоб знати, як змінюється населення, його загальна кількість, склад і т.п. досить переводити спостереження над ними через певні перемежки часу. Тут нема необхідности постійно записувати кожну зміну, що відбувається в кожному неділимому /одиниці/, через те, що коли б така реґістрація навіть була можлива, вона не дасть нам більш ясного й точного поняття про явище, або подію, ніж те, яке будемо мати через періодичне повторення нашого спостереження.

З характеру мінливости таких явищ виникають і д в а т и п и с т а т и с т и ч н о г о ч и с л е н н я - 1/ п е р е п и с і 2/ п о т о ч н а р е ґ і с т р а ц і я.

Отже перепис є статистичне числення, що переводиться для явищ або об'єктів, що виявляють малу мінливість та застосовується до якого-

будь певного моменту часу в стані цих об'єктів. Поточна реєстрація є тип числення, що переводиться над явищами, яким властива велика мінливість та які є подіями, що їх мінливість і прикмети мінливости статистично можуть фіксуватись в короткі перемежки часу.

Кожний з цих типів має властиву йому т е х н і к у п е р е в е д е н н я, кожний потребує відповідно пристосованих органів числення та особливих умов, що забезпечують правильність первісної реєстрації.

Вище ми зазначили основні правила, яких вимагає статистичне числення, в тому числі й кожний перепис. Тут мусимо зупинитись лише на важливіших одмінах переписів в зв'язку з об'єктом статистичного числення та практикою їх переведення в різних країнах.

а/ Переписи населення. На першому місці треба поставити п е р е п и с и н а с е л е н н я, бо вони щодо свого теоретичного й практичного розроблення є найстаріші й найдосконаліші та торкаються дуже важливого елементу соціяльного життя - людности.

Переписи населення мають метою: I/ встановити кількість обивателів, індивідів, що перебувають у певний мент в межах певної території, і 2/ встановити так званий склад населення, себто розкласти на групи відповідно статі , професії і т.п. та встановити їх абсолютну кількість.

Починаючи від першого міжнароднього конґресу в Брюсселі / I853 р./ вчених статистиків і видатних практиків, а потім і на інших таких конґресах /зокрема на конґресі в Петрограді в I872 році/ було встановлено в загальних рисах головні засади доброго переведення переписів населення, які було прийнято практично майже у всіх культурних державах світу.

Першим важливим питанням було вирішити яке саме населення треба охопити переписом. Керуючись вказівками конґресів, адміністративна статистика розрізнює такі відміни населення: I/ Н а с е л е н н я н а я в н е /population de fait, ortsanwesende Bevölkerung/ - це всі ті обивателі, які знаходились у мент переведення перепису в певному місці. 2/ Н а с е л е н н я п о с т і й н е / population de sé jour habituel ou domiciliée dans l'aception ordinair du domicile, Wohnbevölkerung / - це всі ті обивателі, які стало перебувають в певному місці, при тому однаково чи знаходяться вони під час перепису в цьому місці, чи ні. Це населення буде рівнятись наявному населенню без тимчасо присутніх плюс тимчасово відсутні. 3/ Н а с е л е н н я ю р и д и ч н е, інакше переписне / population de droit ou légale, vohnsitz-oder heimatsberechtigte Bevölkerung / - це всі обивателі, що формально /правно/ належать до даного місця. Зокрема науковий інтерес має кількість людности н а я в н о ї, ф а к т и ч н о ї, бо людність ця є сукупність, що витворена комплексом /поєднанням/ умов, які зумовчих розміщенням людности на території держави.

Але для адміністративних цілей важливе значіння мають останні два види населення, бо з осілістю, приписком або постійним перебуванням зв'язуються звичайно в державі як певні права, так і обов'язки, розподіл повинностей і податків.

Брюссельський статистичний конґрес ухвалив, що перепис "мусить переводитись поіменно та базуватись на принципі наявного населення". Для виділення населення юридичного відомості лише можуть збиратись і не є обов'язковими. Між іншим, для конґресу було тяжко встановити, що треба розуміти взагалі під юридичним населенням, бо кожна держава розуміє цей термін різно, а через те й не можна встановити одноманітних ознак для його визначення. Спроби розв'язати це питання на Брюссельському і Флюрентинському конґресі не вдалися і через те ах у I872 році на Петроградському міжнародньому статистичному конґресі було ухвалено, що треба при переписах населення виходити із наявного населення та з нього виділяти один з двох інших видів населення, надаючи кожній державі право встановити відповідні видові ознаки згідно з потребами кожної держави. На цих конґресах була встановлена і п е р і о д и ч н і с т ь п е р е п и с і в ч е р е з

10 р о к і в, при чому бажано аби рік перепису завше кінчався нулем /1930, 1940, 1950 і т.д./. Окремим державам, одначе, на випадок необхідности, дається можливість переводити переписи в з а г а л і в п е р е м е ж к и д е с я т и р і ч н і. В останні часи, як відомо, в деяких культурних державах переписи людности переводяться і через 5 років, як от у Франції, Німеччині, Сполуч. Шт. Півн.Америки. Щодо змісту переписів населення, себто до тих ознак, які мусять реґіструватись для кожної одиниці числення /кожної особи/, то на міжнародньому статистичному конґресі в Петрограді /1872 р./ було встановлено 15 обов'язкових[x]/ ознак, а саме: 1/ Ім'я й прізвище, 2/ Стать, 3/ Вік, 4/ Відношення до голови родини - господарства, 5/ Родинний стан - жонатий, заміжня, нежонатий, дівчина, удівець, удова, розведений, розведена, 6/ Заняття, 7/ Визнання /віра/, 8/ Рідна мова, 9/ Письменність, 10/ Місце народження й національність, 11/ Постійна домівка та з яких мотивів перебуває на місце перепису, 12/ Сліпота, 13/ Глухонімота, 14/ Ідіотизм і кретинізм, 15/ Божевілля.

Інші ознаки не є обовязкові,а лише факультативні / можливі/ для кожної держави.

Щодо поданих вище 15 обов'язкових питань, які ставляться у формулярах для перепису населення треба зазначити, що на деякі з них, на практиці, не вдавалося одержати легко точні відповіді, через те з часом довелося спеціялістам дати до них пояснення, а саме: питання про вік /2/ не ставиться в загальній формі - скільки вам років або який маєте вік?,а запитується про рік народження, а ще краще додати місяць і день народження, щоб дійсно можна було встановити для кожної особи число прожитих років. Питання про відношення до голови господарства /4/ дуже важливе, бо дає нам потім матеріял для виявлення складу господарства. Через те, що тут можуть сплутуватись відношення кревні у родині - батько, син, дружина, а з другого боку економічні - хозяїн, робітник, наймит - краще питання розбити на дві частини: а/ відношення до голови родини і б/ відношення до голови господарства.

У питанні 5-му - "шлюб" і "розвід" треба розуміти лише у законоуправленій формі, а не в формі фактичного співжиття, або роздільного життя. Щодо визнання /віри/, то беручи на увагу, що воно є свобідним внутрішнім переконанням людини, та взагалі мати його не є обов'язком людини, - це питання в деяких культурних країнах, напр., в Англії, не вважається можливим і зовсім ставити.

Питання 8-ме - рідна мова - через свою неясність дуже часто служило не тому, щоб помогти об'єктивно визначити національне походження людини, але, навпаки, щоб затемнити його, або й зовсім перекрутити. Коли під р і д н о ю м о в о ю розуміти лише розмовну / Umgangssprache, langue parle /, як це було ухвалено на деяких міжнародніх статистичних конґресах, то, в умовах слабої національної свідомости, значна частина людности, особливо меншостей, буде почислена за національною ознакою неправильно. Наприклад, українці, що часто розмовляють російською або польською мовами /державними/, занесені будуть не до української,а чужої національности. Ще більше це може статися, коли з боку панівчої державної національности в ріжноманітних формах робиться "вплив" під час перепису використати це питання у власних інтересах, а не з метою об'єктивно встановити національне походження як своє, так і певної або певних меншостей.

Після світової війни в деяких країнах це виявилося досить яскраво і тяжко вражає кожного чесного статистика. Не може послужити справедливости також і розуміння під"рідною мовою" тієї, що вживається в сем'ї, родинної мови / Familiensprache/, бо, в умовах розповсюдження мішано-національних шлюбів, цілком випадково чоловік або дружина можуть "навмисне" занести і самих себе і дітей до тої національности, яка є кожному з них

---

x/ Обов'язковість тут треба розуміти лише як моральну, бо ці конґреси не мають права робити правнообов'язуючих ухвал для держав.

до вподоби, а не до справжньої. Визнається зараз найкращим за рідну мову вважати материнську мову / Muttersprache /. Але треба пам'ятати, що ця ознака, навіть і в умовах об'єктивного статистичного числення, не може вважатись за єдиновистарчавчу. для встановлення етнографічних ріжниць у людности країни. Для виявлення цих ріжниць треба, в залежності від конкретних обставин, завжди ставити, крім питання 8-го, ще декілька контрольних запитань в більш конкретній формі: "Підкресліть" яку мову - російську, польську, українську, жидівську, або іншу та яку саме вважаєте своєю материнською мовою"? При чому перелік на першому місці найважливіших мов робиться звичайно в залежності від того, в якій країні переводиться перепис. Не тільки з мотивів політичного характеру визначення національної ознаки через мову в багатьох переписах людности ріжних країн є не досконале , але через нерозробленість методи фіксації прикмет національности. Ще й доси бракує досконалої дефініції поняття "національности" та "нації" в науковій соціології.

Інструкція до вселюдного перепису 1926 р. в СССР досить детально зупиняється на виясненні значіння мовної ознаки для цього перепису, але й ця інструкція не може задовольнити наукових вимог, що могли б та й мусили б бути поставлені до виконання при переведенні переписів людности. Один з критиків цієї інстркції, А. Х о м е н к о, не безпідставно каже: "Визначення, подане в інструкції і теоретично невитримане і не передбачає і не розв'язує тих труднощів, що мали виникати в практиці перепису." x/

Нарешті, 12 - 15 питання, що торкаються тілесних дефектів, можуть викликати досить непорозумінь, бо, наприклад, ідіотизм, або божевілля є такі стани, що не визначаються коротко ясною ознакою, а через те в умовах масового числення /не спеціяльно медичного/ дуже тяжко піддаються реґістрації, через те часто й викидаються зовсім, хоч цю групу населення почислити є дуже важливим. Сліпота і глухота значно легше реґіструються і через те частіше вводяться до формулярів перепису. Теоретичною вимогою доброго переведення перепису населення є також не ускладняти його іншими завданнями, не використовувати його для вияснення якихось спеціяльних питань соціяльно-економічного життя. Між іншим, закон у Німеччині про переписи населення прямо забороняє це робити.

Нарешті треба зазначити, що ціла операція кожного числення населення складається з трьох частин. Поперше, треба обов'язково точно з'ясувати число заселених місць /пунктів/. В список заселених місць мусять входити не тільки міста, місточки, села і хуторі, але й залізничі або лісові сторожівні, навіть куріні лісових робітників та на пасовисках, бо інакше якась частина населення не буде занотована. В більших поселеннях, за декілька днів до к р и т и ч н о г о м о м е н т у п е р е п и с у треба зробити реґістрацію квартир та передати кожному господарю окремої квартири переписні блянки /формуляри/ в потрібній кількості. В останній день перепису реґістратор має знов обійти квартири, відібрати заповнені блянки та переглянувши їх зараз же виправити або доповнити, якщо це буде потрібно. Тільки після контролю матеріялу особою, що керує численням в певному переписному районі, цей матеріял може бути надісланий до центрального статистичного управління переписом.

Вселюдні переписи населення, які мають особливо важливе значіння для з'ясування властивостей української людности наступні: перепис у Росії 1897 р. /28.січня ст.ст./, а також за радянської влади 1920 р./28. серпня /, але цей останній переводився під час громадянської війни на Україні та не охопив цілої людности її; пропуски, що відносились до території не тільки цілих повітів, а й губерній, є занадто великі, хоч програма цього перепису й цікава щодо свого обсягу. Найбільш важливим і цікавим як щодо змісту, так і щодо фаховости переведення, останній перепис

---

x/ В своєму курсі Д е м о л о г і ї, що викладається на Економічно-Кооперативному відділі УТГІ ми детально й докладніше зупинимося на проблемі правильного числення національних груп людности.

людности в Радянському Союзі 1926 року /критичний момент 17 грудня/. В Польщі, після війни, переведено переписи людности 30 вересня 1921 р. і 9 грудня 1930 р., в Румунії в 1919 році /Тільки в Буковині та Бесарабії/ і 29 грудня 1930 р. та в Чехословаччині 15 лютого 1921 р. та 1 грудня 1930.

Для економічної історії українських земель важливе значіння має також, так званий, "Румянцевський опис", переведений на Україні 1767-1769 р.р., що охопив широко різні боки господарського життя української людности, а для українських земель, що були під Австро-Угорщиною, перепис 1910 року, переведений в цілій цій державі.

б/ Переписи професійні. Питання, що звичайно ставиться в переписах населення з метою вияснення професійного /заняття/ поділу людности, дає можливість лише стислого нарису головних занять населення, а проте велике значіння професійного поділу суспільства в цілому його обсязі і деталях, не тільки для науки, а й для практичних завдань сучасної економічної і соціяльної політики кожної держави, викликає потребу й окремих професійних переписів. Але для економії видатків в більшості країн збирають відомості про професійний склад людности під час перепису населення. Наприклад, досить великий матеріял щодо цього дає і останній перепис людности в Радянському Союзі у 1926 році. Проте, наприклад, в Німеччині окремі професійні переписи переводились декілька разів: 1882, 1895, 1907 і 1925 р.р. та відзначаються як і взагалі переписи цієї країни добрим методичним підготовленням та досконалістю способів числення. Тут переписи професійні зв'язуються з переписами промисловими. Найскладнішою частиною праці в цьому перепису є утворення клясифікації виробництв і самих професій у кожному виробництві. Система цієї клясифікації, крім того, збудована так, що дає можливість зв'язувати дані професійного перепису з переписом промисловим, вірніше з переписом закладів промислового виробництва, а це дає можливість освітлювати числово, статистично, не тільки технічну структуру виробництва, але й соціяльно-економічну його природу і значіння. Оскільки тяжко скласти клясифікацію професій в країні з розвиненим промисловим виробництвом, можна бачити з того, що статистикам та фаховим інженерам колективно доводиться складати величезний словник професій, в якому число визначених окремих професій доходить до двох і більше десятків тисяч. Кожну визначену в словнику професію треба відповідно зв'язати з тою чи іншою галуззю виробництва. Ця відповідальна праця потрібна як для самого переведення перепису, так і для оброблення матеріялу ним здобутого. Одиницею числення є кожна особа професійно чинна з урахованням всіх членів родини, коли вона є. Такий перепис дає можливість зважити, так мовити, соціяльну питому вагу окремих професійних груп людности в цілому суспільстві певної країни, а до певної міри і окремих галузів виробництва цієї країни. Відціль його велике практичне і наукове значіння.

в/ Переписи сільсько-господарські. Щодо переписів господарської чинности людности, то статистичне масове числення визначилось в двох важніших видах переписів: сільсько-господарському та промисловому. Збірання відомостей про сільсько-господарську чинність населення є одним з найтяжчих завдань статистичного числення. Складна система явищ, що охоплює цю чинність населення, та труднощі визначення складових елементів, на які сільсько-господарська чинність розкладається, довго не давали можливости виробити та переводити в життя сільсько-господарські переписи з тою систематичністю та одноманітністю програми, як то ми бачимо в переписах населення.

Промисел сільсько-господарський не концентрується, подібно фабричному або гірничому, в певних місцевостях у невеликій порівнюючи кількості підприємств, але розкиданий цілою своєю масою по всій території кожної держави. Самий склад підприємств цього роду уявляє з себе величезну різноманітність розмірів, форм і способів їх ведення, а результати господарювання тяжко надаються до визначення та виміру. Така різноманітність об'єкту спостереження ускладняє числення ще й тим, що не припускає єдиного

способу його переведення. Тоді як певна частина ознак складного об'єкту надається до періодичного числення, решта ознак є надзвичайно мінливою, іноді настільки, що не дає можливости використати спосіб поточної реєстрації. Одиниця числення - окреме підприємство, господарство - з якої мусить починатися числення, в більшості випадків не надається до розкладання на такі простіші ознаки, які легко могли б бути записані, занотовані. Господар часто сам для себе не може виділити складові елементи свого господарства та визначити їх кількісно. В менш культурних країнах господар часто не знає точного розміру свого господарства, не подасть витрат виробництва, не подасть навіть розміру врожаю. Це все добре відомо статистикам, що переводили переписи сільського господарства в Росії та на Україні /так звані подвірні земські переписи/. Через те і в працях перших статистичних конґресів майже не знаходимо тих загальних принципів та вимог, які з наукового погляду мусять бути виконані у сільсько-господарському перепису. На одному із статистичних конґресів /у Парижі/, правда, було висвітлено, що певні відомості для цієї галузі статистики мусять збиратися через більші перемежки часу, себто через 10 років, другі, навпаки, щорічно. Загальна програма подана цим конґресом для періодичного /що десять років/ збирання даних охоплює в собі такі питання: 1/ Вимір площі під різними культурами і кількости зібраного врожаю. 2/ Визначення способів угноєння. 3/ Встановлення кількости сільсько-господарських робітників - постійних і тимчасових - та розмірів їх заробітньої платні. 4/ Встановлення в господарстві кількости худоби з розподілом її на породи та розміру прибутку від неї. 5/ Встановлення числа підприємств /господарств/ з визначенням права користування землею - власники, фармери, орендарі, половинщики і т.п. 6/ Встановлення кількости і роду машин і взагалі мертвого інвентарю.

Щодо щорічних відомостей, то конґрес визначає їх так: площа культур, кількість врожаю в мірі і вазі. Це й є головні вказівки, що їх було зроблено для таких складних переписів, як сільсько-господарські.

Лише після світової війни, в період сучасного господарського лихоліття, міжнародні інституції, починають шукати способів, як поставити ліпше в кожній країні організацію господарської статистики взагалі і сільсько-господарської зокрема, як засіб раціонального розв'язання величезних соціяльно-економічних проблем, що повстали у світовому господарстві в наслідок світової війни. Останні міжнародні статистичні конґреси, Ліґа Націй з її міжнароднім Урядом Праці та конференціями при ньому статистиків праці та Міжнародній Сільсько-господарський Інститут / Institut international d'agriculture/ у Римі - ось ті важливіші інституції, які взяли на себе ініціятиву організувати та об'єднати господарську статистику в різних країнах. Мусимо тут зупинитися на визначній події в історії статистики, а саме на переведенні першого світового сільсько-господарського перепису в 1929-30 р.р. з ініціятиви Міжнароднього Сільсько-Господарського Інституту в Римі. З 1924 року розпочинає Інститут свою величезну відповідальну працю та звертається до держав членів цього інституту з проханням дружньої співпраці в переведенні світового сільсько-господарського перепису, в якому зацікавлені три міжнароднього значіння інституції: Міжнародній Інститут Статистичний, Господарський постійний Комітет при Союзі Народів та Рада Союзу Народів. Відгукнулося на це 55 держав, з них 51 в засаді прихильно /37 без застережень, 14 з деякими застереженнями/, три держави /Бельгія, Угорщина та Норвегія/ зробили завваження щодо року перепису і одна держава /Коста-Ріка/ заявила, що не може перевести перепису. Таким чином, із 71 країн, що належать до римського Інституту, не дало відповіді лише 16.

Спочатку Інститут з'ясував як організована сільсько-господарська статистика в тих державах, що повстали в наслідок світової війни, з метою доповнити свою раніш видану працю "Організація сільсько-господарської статистики в окремих країнах". Далі було організовано при Інституті окремий Уряд для світового сільсько-господарського перепису. Це зроблено було за матеріяльною допомогою американського "International Education Board"

/10.000 долярів щорічно на протязі 1925-1929 р.р./. Директором Уряду призначено комісара сільсько-господарського департаменту Спол.Шт.Півн.Америки - Л е о н а Е с т а б р у к а / Leon Estabrook /, який привіз з собою до праці в уряді цілий кадр співробітників.

Поперше, уряд почав збирати дані про виробництво, вивіз і довіз сільсько-господарських продуктів та про господарську худобу в окремих країнах, з тою метою, щоб з'ясувати де цих даних ще не збирається зовсім, яке значіння має хілборобство для кожної країни та яку вагу має сільсько - господарське виробництво в світовому обміні продуктами. Далі складено було словник сільсько-господарських плодів та тварин з перекладом кожної назви на латинську, французьку, англійську, італійську, німецьку, еспанську та російську мови.

Важнішою, одначе, справою було вироблення формуляру /блянку для опитування/ для цього перепису. Було вироблено власне три формуляри: вужчий - для всіх обов'язковий, ширший - рекомендований для використання в міру можливости та найширший - з факультативними додатками.

Визнавчи за формуляром, що мав бути основою перепису у всіх країнах, дуже важливе значіння, президія Інституту вирішила скликати тіснішу комісію статистиків та вчених спеціялістів сільського господарства для попереднього обговорення цього формуляру. Комісія опрацювала наступні засади організації перепису:

Замість трьох формулярів, виробити лише один, по можливості стислий та менший обсягом. Перепис треба переводити за участю спеціяльних комісарів /реґістраторів/ і то так, щоб або сам реґістратор заповнив із слів господаря формуляр або заповнений самим господарем формуляр реґістратор одібрав, а відомості, подані в ньому, за особистою відповідальністю, перевірив. Формуляр має бути складений так, щоб господар міг відповісти на кожне поставлене питання, а коли б господар не зміг подати відомостей, реґістратор неодержані дані мусить добути сам. Чим у таких випадках мусить керуватись реґістратор для забезпечення точности відомостей, виясняється в окремій для нього інструкції. Для цілої низки питань, на які господарі, як відомо наперед, часто не можуть дати певних відомостей, як наприклад, про вартість землі, про прибутки та видатки господарства, радиться організовувати спеціяльні анкети x/

Головною метою перепису є з'ясувати всі елементи важливі для продуцента у власному господарстві, для торгівця та посередників, що задовольняють національну і міжнародню потребу в продуктах сільсько-господарського промислу та для влади, якій ходить про розв'язання загальних питань господарської політики. В зв'язку з цим, при вирішенні питання, чи всі господарства мусять увійти до перепису, а також чи охопити переписом ліси і лісову продукцію - вирішено, що вибір розміру господарств предоставити окремим державам, з тою умовою, що коли яка з них захоче охопити переписом господарства менші як 40 арів, /норма запропонована Дирекцією перепису/ то мусить опрацювати матеріяли про ці господарства окремо від тих, що перевищують 40 арів. Площі лісів, що складають суцільну частину господарства, переписом визначаються, але прибуток від них ні. Днем перепису для північної півкулі світу слід визначити І січня 1930 року, щоб переписом охопити врожай зібраний в 1929 році, а для південної півкулі І липня 1930 р. з урожаєм 1929 плюс 1930 р.

Щодо вияснення при перепису розміру врожаю окремих рослин в кожному господарстві, то, знавчи, що господарі у великій своїй частині врожаю не вимірюють і не важать, радиться щоб реґістратори самі визначали розмір урожаю в тих господарствах, де господар не дасть його, бо приблизний вимір урожаю, визначений на місці перепису реґістраторами, буде ліпший, точніший, аніж той, який довелося б потім визначити спеціялістам у вигляді нормального або середнього врожаю для певної частини даної країни. Далі визнано потрібним віднести до компетенції широкої комісії статистиків

---

x/ Про цю форму статистичного числення скажемо далі.

вироблення клясіфікації сільсько-господарських тварин. Згідно з пропозицією Міжнароднього Уряду Праці в Женеві, введено в перепис питання про вік і стать робітників та їх мешкання в сільському господарстві.

Щодо словника з клясіфікацією рослин, то комісія, визнавчи його видатне наукове і практичне значіння, ухвалила, щоб кожна окрема країна вваела його як окрему частину до інструкції реґістраторам, розробивши відповідно до місцевих умов цю частину інтрукції. Наресті, відносно цікавого внесення проф. Беніні, щоб певні результати перепису, здобуті про нарочито визначені місцевості / напр., біля великих міст/, були внесені в спеціяльні картки з метою пізнішого пов'язання цих даних Міжнароднім Сільсько-господарським Інститутом з географічними, фізичними, метереологічними і іншими умовами сільсько-господарського виробництва, у формі спеціяльних природних господарських країн /районів/ або характерних "гнізд" / територій охоплювчих великі міста/ - комісія висловилась принципово позитивно, але через велику коштовність такого оброблення матеріялу, ухвалила цю пропозицію віддати на рішення Інституту.

Ширша Комісія статистиків - спеціялістів при Загальних Зборах Міжнароднього Сільсько-господарського Інститута розпочала свої праці в травні 1926 року на чолі з професором Беніні /R.Benini /, Т.Томсоном / /R Thompson/ за участю делеґатів від Англії, Бельгії, Бразілії, Голяндії, Данії, Египту, Італії, Іспанії, Мароко, Мексіки, Німеччини, Нової Зеляндії, Польщи, Спол.Шт.Півн.Америки,, Півн.Ірляндії, Чехословаччини, Швеції і Швейцарії. Від кожної країни брало участь декілько делеґатів спеціялістів наукової і урядової статистики. Завваження цієї комісії на проект формуляра перепису виробленого тіснішою комісією, можна звести до таких коротких положень: необхідним є о б с я г ф о р м у л я р а с к о р о т и т и та по змісту надати йому більшої єдности, щоб перепис не зустрів на місцях великих трудностів переведення та не викликав дуже великих видатків. Перепис мусить охопити всі господарства, починавчи розміром від І гектара /2 І/2 акра/, беручи на увагу трудноші перепису великого числа маленьких /до І гек/ підприємств та невелике значіння цих підприємств для обміну сільсько-господарськими продуктами.

Після довгих дебатів відносно реченця перепису ухвалено перевести його для північної півкулі світу з тим розрахунком, щоб охопити врожай у період від І-го липня 1929 р. до 30 червня 1930 року. Критичний момент /день/ числення не був вирішений, бо цьому стояла на перешкоді неможливість визначити день для з'ясування стану кількости господарських тварин, яка з бігом року значно міняється. Через те було ухвалено лише порадити охопити кількосний стан тварин у 1929 календарному році, а самий день числення віддати на рішення урядів кожної держави. Серьезний критиці була піддана також пропозіція проекта внести у формуляр питання про розмір урожаю. Комісія вважала, що определення врожаю в умовах індівідуального опросу господарів, які в більшості випадків або самі не зможуть встановити розміру врожаю або можуть подавати навмисно неправдиві відомості, не дасть позитивних наслідків. З істоти справи також випливає, що результат врожаю в кожному господарстві не береться звичайно яко ознака для групування підприємств у статистичному обробленні, через те краще залишити спосіб виділення врожаю за методою здогаду, себто через спеціялістів для більших груп господарств,- напр., для земельної дачи /катастральної дільниці/ цілого села, определити пересічний врожай на гектар, а цілий врожай одержати через множення цієї пересічної врожаю на площі засіву. По цим міркуванням було з формуляру викинено питання про дані індивідуальних врожаїв, а урядам країн залишено повну волю у виборі способа встановлення врожаю.

Теоретичні міркування, що були подані яко мотивування до загального формуляра перепису, остільки важливі, що мусимо коротко їх зазначити:

"Перепис є інвентар. Перепис сільсько-господарський є інвентар сільсько-господарського виробництва, себто вимір продукційних площів господарських тварин, сільсько-господарського робітництва, сил механічних,

праці робочої худоби, якож і решти складників, які мають значіння для сільсько-господарського підприємства та які мають вплив на виробництво предметів споживи і сировин та на ціни, що виробництво дістає, а споживачі платять. Багацтво, розквіт та добробут кожної країни залежить безпосередньо на розвитку та розумному використанню природніх джерел, з яких земля і користання з тварин є найважливішими. Добробут цей залежить далі від маси і якости споживи, яку забезпечує передовсім сільське господарство та його тварини, а також від маси і якости промислових сировин, що додає ця галузь господарської чинности та лісове виробництво. Він залежить також від стану торгівлі і транспорту предметом яких є в більшій мірі сировина та продукти сільсько-господарського виробництва, скотарства та лісівництва. Всі ці умови народнього добробуту, вимагають не тільки досконалого знання виробництва сільсько-господарських продуктів та їх споживи за цілий ряд років, але й досконалого й своєчасного знання ситуації в певні моменти їх стану та передбачення попиту й пропозиції.

Можливо, що деякий платник не має прямого інтересу на цих складниках народнього добробута, або на статистиці, що їх дотикається, але кожний громадянин має прямий і живий інтерес до цін, що одержує як виробник та які платить як споживач за їжу, одяг, паливо, меблі, транспорт та за різні послуги. Він не може не цікавитись обміном, купівною силою свого грошевого доходу, як також і добробутом та купівною силою осіб, з якими пов'язаний спільними інтересами, або з якими має торговельні стосунки. Ціни, в яких кожний громадянин безпосередньо зацікавлений, підлягають постійно великому впливу місячних і щорічних статистик, що вказують сучасну й будучу готовність господарства і розмір взаємовідношення між цією готовністю та правдоподібним попитом. Повнота, точність і вага місячних і щорічних сільсько-господарських статистик в значній мірі залежить на періодичних переписах.

x x x x

ВПРАВИ. /до 2-ої лекції/. Напишіть коротко і пришліть відповіді на наступні запитання:

1. Дайте приклад реальної сукупности, що є основою статистичної сукупности.
2. Подайте приклад штучно утвореної статистиком сукупности.
3. Дайте приклад кваліятивних ознак для одиниці сукупности.
4. Так само квантітативних ознак.
5. Дайте приклад "многовідмінної" ознаки.
6. Яку б "одиницю числення" Ви запропонували у численні кооперативів? Подайте стислу дефініцію /поняття/ цієї одиниці.
7. Який к р и т и ч н и й м о м е н т перепису кооперативів в Галичині Ви запропонуєте і через що ?
8. Який на Вашу думку р а й о н о в и й п о д і л Галичини мав би значіння для сільсько-господарського перепису?

ВПРАВИ /до 3-ої лекції/.

1. Які загальні статистичні переписи в Галичині Ви знаєте?
2. Які з цих переписів переводились експедиційним, а які блянковим способами?
3. Хто в Вас керує переписами людности?
4. Чи переводився в Вас загальний сільсько-господарський перепис у зв'язку з світовим сільсько-господарським переписом і коли?
5. Чи не брали Ви якої-будь участи в переведенню перепису? Коли брали, то коли, де, в якому саме перепису та які функції виконували?

Лекція четверта.

Продовження розд. I6 літ. в.

переписи сільсько-господарські. В своєму роз'ясненні до формуляру світового перепису Міжнародній Сільсько-Господарський Інститут зупиняється далі на наступному:

"В певній країні пропозиція предметів споживання і сировини у формі сільсько-господарських продуктів, продуктів лісівництва або від плекання тварин залежить не тільки від виробництва цієї країни, але також від її довозу і вивозу. Виробництво само залежить від доброго оброблення землі та доцільного господарення на ній, відповідно до умов клімату, складу ґрунтів, як також і від розміру купівельних і продажних цін у відношенні до інших галузів виробництва, нарешті, від конкуренційних цін продуктів, що походять з чужини.

Щоб можна було підняти сільсько-господарське виробництво, лісівництво та плекання тварин є необхідним, щоб Міністерство Хліборобства кожної країни, досвідні сільсько-господарські установи, фахові сільсько-господарські школи, а також всі приватні і громадські сільсько-господарські організації, наприклад, сільсько-господарські товариства, кооперативні і т.д. мали для своїх потреб точну і надійну статистику площ, призначених до сільсько-господарського виробництва.

Ці інституції мусять мати також статистичні відомості про плекання тварин, про наслідки лісового господарства, мусять знати стан навіть окремих видів худоби та щорічні результати врожаю різних плодів, за цілий ряд років, щоб з'ясувати чи зростає чи падає продукція виробництва в певних країнах та чим викликані в певних країнах ті або інші зміни.

На основі всіх цих данних державні чинники і всі ті, що стоять на чолі аґрарного руху, можуть будувати на ряд років програму поліпшень та вживати заходів щодо розвитку сільського господарства. Досконалість рішень і поступування тих, хто займається сільським господарством залежить від правильности та досконалости періодичних пепеписів та щорічних сільсько-господарських даних поточної статистичної реґістрації. Виробництво і ціни предметів споживання і сировини певної країни, які цікавлять цілий світ, залежить від точного знання і доброго розв'язання справ тими, хто повинен ними займатись.

Одержаний через періодичні переписи інвентар сільського господарства вказує не лише на багатство країни в певний момент, але довволяє міряти поступ або підупад його в межичасі від одного перепису до другого, себто стає сталою історією одного з найважніших складників господарського життя держави - його сільсько-господарського виробництва.

Не менше значіння періодичні переписи такого роду мають через те, що вони дають надійну основу доброї організації щорічної статистики та поточної щомісячної статистики сільсько-господарського виробництва в міжнародньому обсязі, що мають безпосередній вплив на розвиток цін цього виробництва. Природним є, що такі великі переписи переводити щороку неможливо, бо вони дуже дорого коштують та вимагають значного часу на переведення та оброблення статистичного матеріялу і його публікацію. Відсіль виникає необхідність для щорічної статистики рільничих плодів та продуктів від плекання сільсько-господарських тварин вживати інших способів добування потрібних даних з витратою також і менших коштів. Одначе, різні способи одержання такого роду даних, що основані на здогаді, хоч і широко практикуються в різних країнах, але дають недосконалі наслідки, підлягають сільніше хибам, ніж періодичні переписи, що провадяться раз за 5, I0 років, або I5 років.

Більшість статистичних даних, що базуються на здогаді, спираються на статистичні відомості попереднього року або на пересічних величинах за

декілька попередніх років, наприклад, за 5 або 10 років. Через те помилка, яка сталася, наприклад, при обчисленні площ засіву рослин, або кількости худоби в одному році, може повторюватись, або збільшуватись в роках дальших та повести, наретті, до значних ухилень від дійсного стану річей. Через те потрібно переводити своєчасно правильні переписи, щоб перевіряти попередні здогади, та одержати надійну основу для здогадів наступних. Є таким чином, взаємозалежність між періодичними переписами та щорічними статистиками, а через те обидва види її є необхідними. Великі переписи дають основу і контроль щорічної статистики, яка лише заповнює порожнечу між періодичними переписами та забезпечує правильність щорічних необхідних відомостей великого торговельного і господарського значіння. Щорічна статистика, що періодично перевіряється переписами, має велике значіння і ціну для різноманітних цілей. Щоб з'ясувати важливість правильної сільсько-господарської статистики, треба зазначити ті вигоди і можливості, які вони дають для різних цілей.

1. С т а т и с т и к а д а є в и р о б н и ц т в у н а п р я м. Статистика про площі засіву та наслідки врожаїв дозволяє сільскому господареві, щоб він на основі досконалого знання причин приправив наближчий час плян організації свого сільсько-господарського виробництва. Без статистики про попередні врожаї та сучасний врожай неможливо скласти доброго пляну засіву рослин, навпаки, є небезпека, що господарі будуть користатись із застарілого, здавна звиклого пляну господарювання, або вносити лише випадково щось нове в старі способи господарювання, а це може повести для певного періоду господарського або до перепродукції та катастрофально низьких цін для виробників, або до недостачі продукції та дуже високих цін для споживачів.

2. С т а т и с т и к а д а є н а п р я м к и щ о д о р о з п о д і л у т а п р о д а ж у п л о д і в. Правдиві дані про розмір виробництва перед врожаєм або підчас жнив, дозволяють як виробникам, так і продавцям та купцям скласти розумну програму дистрибуції /розподілу/ виробництва. На основі таких даних транспортові товариства і підприємства можуть своєчасно заготовити та подати на місця потрібну кількість вагонів для скорішого розподілу лишків врожаїв. Цілий великий механізм, що має метою розподіляти та переробляти лишки врожаїв - себто, млини, склади, цукровари і т.п.може оперти виконання своїх зобов'язань перед контрагентами на згаданих статистичних урядових даних сільсько-господарського виробництва і взагалі добре опрацювати господарський плян свого поступування. На основі цих даних про загальне рільниче виробництво в країні, можуть сільсько-господарські товариства та кооперативи чи їх союзи вигідніше опрацювати остаточну програму збуту плодів і продуктів, які вони постачають на ринок.

3. С т а т и с т и к а з а с і в і в т а в р о ж а ї д а є о с н о в у і н а п р я м д л я з а г о т о в л е н н я к о ш т і в, п о т р і б н и х д л я п е р е в е д е н н я з а с і в у, з б о р у в р о ж а ї т а п р о д а ж у л и ш к і в й о г о. Рослинне та тваринне сільсько-господарське виробництво у великих підприємствах вимагає великого капіталу для закуплення насіння, угноєння, господарського знаряддя, ріжних інших предметів та оплати робітників. При купівлі і організованому продажу лишків врожаїв потребують також значних коштів торговці сільсько-господарськими продуктами, через те мусять місцеві грошові інституції своєчасно заготовити оборотові кошти для кредитування як виробників, так і тих, що купують та продають продукти виробництва. Отже організація планового фінансування мусить спиратись на статистику та організовану публікацію статистичних відомостей очікуваного врожаю плодів.

4. С т а т и с т и к а в р о ж а ї в є п о к а ж ч и к о м к у п і в е л ь н о ї з д і б н о с т и. Купівельна сила сільсько-господарських країн залежить від лишків їх виробництва; через те природно , що дані про врожай та стан худоби певної країни є покажчиком її сучасної і будучої

купівельної сили.

5. Статистика врожаїв є важливою поміччю для фабрикантів і продавців промислових товарів. Статистика яко покажчик купівельної сили є надзвичайно корисна для промисловців і торговців, бо уможливлює їм розумно розраховувати розміри попиту на предмети їх виробництва, а відповідно до того будувати організаційний плян власного виробництва та його дистрибуції на ринках, що має і для цілого народнього господарства важливе значіння.

6. Статистика дає відсіч ріжним неправдивим та шкідливим чуткам і спекулятивній чинності. Старанно опрацьовані урядами періодичні дані сільсько-господарської статистики мають метою застерегти та усунути поширення фалшивих відомостей тими, хто зацікавлений, щоб внести заколот у суспільство та викликати у нього невірну уяву про сучасний та будучий стан сільсько-господарського виробництва, щоб з цієї фалши та заколоту зискати особисті користі за рахунок втрат суспільства. Коли урядова статистика збирається особами незацікавленими, вона заслуговує довір'я і викликає довір'я щодо своєї точности. Самий той факт, що така статистика систематично і правильно публікує для широких кол суспільства свої відомості, робить неможливим поширення шкідливих та фалщивих даних, подібно до того, як сама лише присутність органів публічного порядку в якому будь місті вже забезпечує охорону суспільства від злочинних елементів та дає необхідні умови спокійної чинности.

7. Статистика спиняє недобросовісну спекуляцію. Приватна спекуляція у великій мірі суперечить інтересам громадським, бо вона дуже часто викликає примусове та свавільне коливання цін та стоїть на перешкоді природному руху продуктів і товарів між продуцентом та споживачем. Шкідлива спекуляція просто живить себе непевність та браком інформацій у громадянства. Через те, оскільки офіційна статистика сільсько-господарського виробництва та скотарства усуває таку непевність, вона нищить спекуляцію.

8. Статистика визначає також розмір виробництва. Правдиві місячні й щорічні дані статистики, які контролюються періодичними переписами, тому, що є дороговказом для виробництва і споживання, що забезпечують виробникові та суспільству точну господарську орієнтацію, що знищують або обмежують спекуляцію та усталюють ціни, - дають добру заруку для удержання потрібної і певної кількости /маси/ виробництва. Способи господарювання, що залишаються лише завдяки старим звичкам або викликані нагодою, які не доведуть скористатись пляновістю, що дається статистикою - ведуть до двох наслідків: або до років недостатньої продукції, бідування та високих цін для споживача, або до років перепродукції з низькими цінами, що нищать саме виробництво.

9. Статистика є важливий чинник творення цін. Ціни, які виробники одержують, а споживачі платять, залежать перш за все від світового попиту і пропозиції. Коли допустимо, що світовий попит в цілому змінюється дуже мало, можемо твердити, що головним складником, який визначає ціни різних товарів, є їх готові запаси на ринкові так сучасні, як і будучі, або одні й другі разом. Найліпшою мірою чи покажчиком цього підготовлення запасів на ринкові є статистика виробництва, споживання та запасів, дані якої систематично публікуються окремими державами.

Відповідно до того, як статистичні дані оцінюють релятивне збільшення або зменшення цієї підготовлености запасів, створюється рівновага цін майже одночасово на всіх головних світових ринках. Ця рівновага цін стала дуже важливим чинником для народнього й світового господарства, бо як тільки статистика вказує, що заготовлення на ринок релятивно знижується, звищення цін веде автоматично до зниження споживи, до збереження на-

явних запасів, щоб витримати довший попит та тим вплинути на нову продукцію.

Навпаки, коли статистика вказує на релятивно велику пропозицію, викликане нею зниження цін веде автоматично до збільшення споживи - обмеження зростаючого виробництва. Ця важлива властивість цін, що приводить до регулювання торговлі, споживання, виробництва продуктів сільського господарства, а також значний вплив, який має міжнародня статистика виробництва та запасів різного роду товарів на творення цін, вказують нам оскільки є важливим, щоб міжнародня статистика виробництва була побудована для цілого світу на єдиних, одноманітних засадах.

IO. С т а т и с т и к а д а є н а п р я м к и д л я п л я н о в о ї ч и н н о с т и п у б л і ч н и х у р я д і в т а о с н о в и м у д р о г о з а к о н о д а в с т в а. Важливішою і ціннішою функцією сільсько-господарських та щорічних статистик є те, що вони дають основу для розумних заходів влади та законодавства, які разом обумовлюють розвиток сільського господарства та громадського добробуту взагалі.

Без цієї надійної основи громадські і державні інституції і уряди не можуть мати знання правдивої дійсности, що є необхідною як для кожного розумного рішення в сучасности, так і для справді ділової програми на будуче, і недостача таких важливих інформацій може спричинитись до великих господарських шкод не тільки для окремої держави а й для великого об'єднання народів. Проте, за інформаціями Міжнароднього Сільсько-Господарського Інституту, за останніх 25 років цього століття із загального числа 200 країн /включаючи сюди і колонії/ лише 37 перевело сільсько-господарські переписи. Територія цих країн складає лише біля половини світової території і біля 30% світової людности. Решта I60 країн, що складають біля половини світової поверхні і біля 70% населення світу, або зовсім не перевело переписів, або не переводило їх за останні 25 років. Але й ті переписи, що були переведені в різні періоди в окремих країнах, та з прикладенням різних метод їх переведення, через неможливість порівняння між собою, не дають підстав зробити наукові висновки міжнародньої ваги.

Спроба перевести світовий сільсько-господарський перепис, як ми вже зазначили, є першою в історії цих переписів, через те практика переведення та вислідки цього перепису мають величезну наукову і практичну цінність.

Треба одначе сказати, що цей, таким трудом підготовлений, перепис не охопив значної частини площі і населення земної кулі, як також в окремих державах або країнах не було можливости додержати строго єдиної програми, виробленої Міжнароднім Сільсько-Господарським Інститутом.

Щодо часу переведення перепису, то вимога, що її було поставлено програмою, виконана фактично наступними державами.

Країни:	Час перепису.
I. Англія з півн. Ірляндією, 2. Свобідна держава Ірляндія, 3. Норвегія і Естонія	в червні I929 року.
4. Данія і Латвія	в липні I929 року
5. Швайцарія	в серпні I929 року
6. острів Цейлон і Японія	у вересні I929 року
7. Австралія, Бельгія, Барбадос /острів/, Брит. Індія, Греція, острова Маврикія, острова Сешелі	кінець I929 року, або початок I930 р.
8. Алжір, Італія	в березні I930 р.
9. Аляска, Америк. Дівочі острови, острів Гуам, остров Гаваї, Гватемала, О. Ісляндія, острів Порторіко, Румунія, Острова Самоа, Спол. Шт. Півн. Америки.	р. квітні I930 року.
IO. Мехіко і Уругвай	в травні I930 року.

II. Австрія і Чілі — в червні 1930 року
I2. Кенія — в серпні 1930 року
I3. Південно-Африк.Унія і острови Сошані — у вересні 1930 року
I4. Мозамбік — в листопаді 1930 року

Держав, що перевели перепис в інший, ніж було вказано програмою, реченець є всього шість, а саме: Німеччина - в травні 1925 р.; Швеція - в вересні 1927 р.; Чехословаччина - в травні 1930 р.; Голяндія - в червні 1930 р.; Литва - в кінці 1930 р.; Канада - в червні 1931

В Єгипті перепис переводився в трьох різних добах року, відповідно до врожаю культур - навесні, влітку та восени. В Еспанії формуляри заповнювались також у 4-х різних періодах року. На Новій Зеляндії більша частина перепису була переведена в лютому 1930 р., але відомості по деяких питаннях збирались за декілька місяців наперед або після визначеного реченця перепису.

Цікаво також зазначити, що спірне питання програми, відносно того, чи слід опитувати безпосередньо господарів при перепису про розмір врожаю рослин, яке комісією спеціялістів було вирішено негативно, наприкінці, в більшості країн поза Европою розвязувалось позитивно. З числа 32-х країн, в яких безпосередньо запитувалось господарів про врожай, лише 8 країн відносяться до Европи.

Щодо обсягу програми, то коротко мусимо зазначити наступне: з числа країн, що перевели перепис, крім загально необхідних частей програми, - I7 з'ясували стан спеціяльного садівництва /8 в Европі і 9 поза Европою/; I8 країн - ліси /6 в Европі і I2 поза нею/; I5 країн - обводнення, I0 - осушення землі; 20 країн - моторизацію /I3 в Европі і 7 поза нею/; 34 країни /I8 в Европі і I6 поза нею/ - сільсько-господарське знаряддя; I3 країн /5 в Европі і 8 поза Европою/ - транспортові засоби; 49 країн /22 в Европі і 27 поза Европою/ - сільсько-господарських тварин; дрібні галузі - 42 країни, бджільництво - 30 країн, продукти скотарства - 26 країн /7 в Европі і I9 поза Европою/; сільсько-господарське робітництво - 32 країни /в Европі I8 і поза Европою I4/.

По даних, які одержав дотепер Міжнародній Сільсько-Господарський Інститут, взяло ту чи іншу участь у світовому перепису 63 держави, країни і колонії, з яких в 45 країнах було дійсно переведено правильний x/ перепис сільсько-господарських підприємств /господарств/, в 7 спеціяльні дослідження сільського господарства, а решта /II країн/ країн надіслала до Інституту на підставі своїх щорічних статистик детальні відомості про сільське господарство, пристосовуючи ці дані до встановленого програмою реченця перепису. З числа 45 вищезазначених країн - 24 відносяться до Европи, 4 до Північної Америки, 4 до Середьої Америки, I до Азії, 7 до Африки і 5 до Океанії.

В 4-х країнах одиницею перепису було не окреме господарство, а ціле село, або інша адміністративна одиниця, але програма відповідала обсягу світового формуляра перепису./Франція, Алжір, Перу, Сальвадор/. Три держави перевели перепис частково - Аргентіна перевела повний перепис худоби, Брит. Індія - перепис худоби, сільсько-господарського знаряддя та засобів транспорту і Японія - площу сільсько-господарських земель і розподіл згідно з пропорцією культур та правної форми землекористування /власник, орендар і т.п./.

Зовсім не могли взяти участь у перепису: Бразілія, Угорщина, Голяндська Індія і Польща /через фінансові утруднення/, а також Союз Совітських Республік, Сербія, Турція, Китай, Португалія, Персія, Авганістан, Венецуела, Еквадор, Колумбія і Парагвай; а на острові Цейлоні, в Кенії і Мозамбіку перепис обмежився лише господарствами нетубольців.

---

x/ В Болгарії перепис був переведений раніше /I926 р./ і матеріяли його розроблені по так званому вибірковому способу статистичного спостереження.

Оброблення матеріялів світового перепису є величезною статистичною працею, через що, хоч від першого моменту перепису /червень 1929 року/ минуло вже більш 3 років, а від пізнішого /Канада червень 1931 р./ більше року, проте результати перепису ще /за невеликим виключенням/ не опубліковані.

По мірі закінчення праці в окремих країнах Міжнародній Сільсько-Господарський Інститут згідно з принятим рішенням бере на себе:

а/ концентрувати в себе відомості про методи переведення перепису в кожній державі /формуляри, інструкції і т.п./

б/ студіювати всі ці доклади про перепис в окремих державах.

в/ одержані наслідки перепису частково публікувати в своєму щорічнику / Annuaire / і в місячнику / Bulletin mensuel /,

г/ скласти єдиний плян про публікацію наслідків перепису кожної країни в особливих вісниках.

д/ публікувати відповідно єдиного пляну /пакт г/ наслідки перепису в повному обсязі.

Ці останні /літ.д/ публікації світового перепису вже для кількох країн випущені з друку Міжнароднім Сільсько-Госпоpським Інститутом.

Щоб з'ясувати цілком значіння цього сільсько-господарського перепису для пізнання структури світового сільського господарства, подаємо тут також і збірний формуляр цього перепису, вироблений Міжнароднім Сільсько-Господарським Інститутом. /Див. окремий додаток /.

В світовому сільсько-господарському перепису, як вже було зазначено, не брали участь Совітський Союз і Польща, себто держави, для економіки українського населення яких світовий перепис мав би першорядне значіння. Мусимо через те зазначити, що в Польщі основним матеріялом про сільсько-господарське виробництво можуть служити поки що лише дані сільсько-господарського перепису 1921 року.

В Росії сільсько-господарський перепис було переведено в 1916 році на цілому просторі держави в час харчової скрути, як для цівільного населення, так і для війська. Характерною рисою його було те, що потрібні відомості збирались при перепису по двох формулярах. Перший формуляр з більш короткою програмою - обчислення загальної кількости населення, площу засіву під окремими культурами і кількости худоби - служив для господарств як селянського, так і приватновласницького типу, з додатком для останніх ще питань про збір врожаю та запасів збіжжя і кормів за останній рік. Другий більш детальний формуляр, в якому збирались також відомості про зміни, що відбулись у засівній площі культур і в зборі врожаю та в стані худоби за два роки служив для перепису тільки пяти відсотків. селянських господарств. Для одного відсотку господарств з'ясовувався також і річний оборот харчових продуктів.

Результати перепису опубліковано під назвою "Предварительные итоги Всероссійской сельско-хозяйственной переписи 1916 г.", де подано лише дані про засівні площі окремих культур та кількість худоби по окремих її видах.

В 1917 році було переведено новий перепис з поширеною програмою доповнено питаннями для характеристики землеволодіння і землекористування, розподілу земель на вжитки, реманенту та промислових занять кожної особи, що належала до окремого господарства.

Лише в 1923 році опубліковано матеріяли цього перепису, під назвою "Поуездные итоги Всероссійской сельско-хозяйственной поземельной переписи 1917 года" для 57 губерній і областей.

В 1920 році в зв'язку з загальним переписом населення переведено було третій сільсько-господарський перепис, матеріяли якого опубліковано під назвою "Итоги Всероссійской сельско-хозяйственной переписи 1920 года", причому підсумки цього перепису щодо УССР видано в спеціяльній публікації в числах 7,11,13,15,21,23-му "Статистика України".

Л е к ц і я п'я т а.

г/ Переписи подвірні земські. Окреме місце серед сільсько-господарських европейських переписів займають "земські подвірні переписи", що переводились в Росії і на Україні органами місцевого самоврядування /земствами/, починаючи з 70-х років минулого століття майже на протязі пятидесяти років. Ці переписи не були одночасними і загальними для цілої території держави, але переводились в різні моменти губерніяльними земствами, охоплюючи щонайбільше територію губерні, а інколи і частину її.

Хоч матеріяли здобуті цими переписами для характеристики тогочасної сільсько-господарської економіки зараз мають історичне значіння, але способи організації земських подвірних переписів, методи їх переведення і оброблення здобутих ними матеріялів уявляють із себе своєрідне, надзвичайно цінне, джерело для кожного, хто студіює методологію і техніку статистики, а особливо для тих, кого цікавить окрема галузь матеріяльної наукової статистики, що зветься аґрарно-економічною статистикою.

Відданість справі особистого персоналу земської статистичної організації, його ідейність і самоофіра задля загального добра трудових народніх мас, що жили в умовах російського царського режиму, не має прикладів в історії статистичних орагнізацій інших країн. Через те й спадщина, що залишена кращою частиною так званого "третього елементу" в земствах як у Росії, так і на Україні є і ще довго буде невичерпним, цінним джерелом для теоретика і практика статистики.

Першим на Україні почало /1869 р./ статистичне дослідження селянського господарства Херсонське земство, а потім Чернігівське /1876 р./.

В своєму поясненню завдань перепису один з керівників статистичного відділу Чернігівського земства П.Червінський підкреслює[х]/що метою перепису є економічне дослідження селянського побуту і до того не тільки сільсько-господарської, вузько хліборобської чинности селянства, але сукупної його економічної чинности.

Таким чином, з самого початку земські статистичні дослідження на Україні ставлять собі завданням широко і глибоко охопити економічне життя селянства.

Не в довгому часі після Чернігівського земства перепис переводиться у Воронізькому земстві, а потім у Полтавському і Харківському, при чому в цих губернях за одиницю числення приймається двір - господарство і переписом охоплюються всі двори-господарства цілої губерні, себто статистичне спостереження побудовано на принципі вичерпливого числення господарств.

В опублікованих матеріялах цих переписів, крім статистичних таблиць, знаходимо надзвичайно багату змістом текстову частину, в якій не тільки широко освітлюється стан селянського господарства цілої губерні, але й подаються детальні відомості і міркування про організацію подвірного перепису, техніку його переведення, програму числення і розроблення здобутого численням статистичного матеріялу.

Важливим для оцінки земських статистичних праць є й те, що вони в більшості випадків організовувались і переводились після спільного обговорення між собою спеціялістів-керівників земських переписів та представників статистичної науки на різних нарадах і з'їздах, а потім і в підсекціях статистики на періодичних з'їздах природознавців і лікарів цілої держави. Таким чином, земська статистична праця є не тільки продуктом індивідуальної творчости окремих, хочби й талановитих робітників, а продуктом колективної творчости кращих практиків і теоретиків статистики.

На одному з таких з'їздів в 1900 році два видатних російських земських статистика - Н. Анненський та Н. Черненков у своїх промовах висловили такі провідні думки дальших завдань земської статистики:

х/ С. Велецкій „ Земская статистика" т.2. стор. 722-725.

"Коли почалась статистична праця земств,- казав Н.Анненський,- то службову, спеціяльну працю встановлення цінности і прибутковости земель та іншого нерухомого маєтку, для виконання якої було створено в земствах статистичні бюра, дуже скоро покрило інше, більш широке завдання - студіювання економічних та інших загальних умов народнього побуту....

Тоді виникла величезна праця, створена умовами моменту, і був вироблений певний тип статистичних операцій загально-економічного характеру; тип дуже оригінальний, що досяг у нас великого ступня досконалости.

Це, власне, - був подвірний перепис. Що таке подвірний перепис? Це таке загальне спостереження, яке охоплює всі сторони селянського господарства, але яке, разом з тим, бере лише певний момент історичного народнього життя. У нас широко розвинувся і досяг високого ступня досконалости цей тип статистики.

Другий тип статистики - це спостереження поточного економічного життя, статистика процесу життя. Перше ніж перейти до поточної статистики, треба було твердо поставити на ноги статистику загальних економічних умов і це було зроблено з великим успіхом переведенням подвірних переписів. Перед земськими статистиками повстають зараз не менш великі завдання".....

Ми живемо зараз в такий час, коли йде ламання всіх господарських підвалин життя.....

Ми мусимо дати тверді і ясні віхи, по яких можна б було судити про економічний стан, про нормальні умови народнього побуту. Нам зараз важливо бачити загальне обличчя селянського господарства, ті умови, в яких складається сільсько-господарське виробництво ....

Зараз, коли все економічне життя і сільське господарство знаходяться в переходовому періоді, коли чистий тип господарства порушений, висуваються нові форми його; ми присутні при боротьбі нових форм із старими і зараз є важливим не тільки студіювання середніх умов даного моменту, але й перепис сам по собі, спостереження над поточним процесом. Цим умовам мусить відповідати наша поточна статистика. Із цього зерна нашими сукупними зусиллями мусить розвинутись така загальна праця, яка освітлить еволюцію сільського господарства"....

"В сучасний мент, продовжуючи Анненського думку, говорить Н.Черненков,[x] головним завданням досліджень земської статистики стає вже не статика, а динаміка народнього життя, народнього господарства".

На цьому з'їзді статистиків було ухвалено про необхідність переведення в кожній губерні в однаковий період часу нових подвірних переписів сільських господарств, коли й невичерпливих, то хочби в найбільш характерних районах або по групах сел, але щоб перепис обов'язково торкався тих господарств, які були охоплені попередніми переписами. Цим малось на увазі зафіксувати всі ті зміни в господарствах, які відбулись у них за час між двома переписами.

Як бачимо, необхідність переведення повторних, так званих динамічних переписів, для дослідження динаміки сільського господарства визнавалась земськими статистиками більш ніж 30 років тому /Напр. на Полтавщині було переведено аж три подвірних переписи в 1899 році, в 1900 році і в 1910 році і звичайно, що ці переписи дали такий прекрасний об'єктивний статистичний матеріял для глибокого дослідження процесів, що відбувались у господарському житті селянських мас Полтавщини, якого тяжко знайти навіть у культурніших країнах Европи.

Із поданого нами викладу про світовий сільсько-господарський перепис у 1929-1930 р.р. видно, що і в сучасний мент у багатьох країнах такого роду переписи не переводились.

Міжнародній Сільсько-Господарський Інститут ще й зараз закликає уряди різних країн до переведення сільсько-господарських переписів та детально і широко висвітлює їх надзвичайну важливість для організації господарського життя як окремих країн, так і цілого світового господарства. Через те власне ми мусимо так високо ставити земські статистичні подвірні

x/ У 1910 році, яко керівник Статистичного Бюро Чернігівського земства перевів подвірний перепис в цій губерні.

переписи і зосібна ті, що переводились на Наддніпрянській Україні.

Методологична цінність дослідження народнього життя земською статистикою в значній мірі залежала ще й від того, що в кадрах її працьовників було чимало освідчених і інтеліґентних представників суспільної агрономії. Мусимо також зазначити, що велику ролю в формуванню теоретичної думки земських статистиків, як і в виконанню високої практичної праці відограваліи товариські стосунки і співпраця між ними та ліпшими силами статистичної і економічної науки - А..І. Чупровим, А. Фортунатовим, Н.А. Каблуковим, А.А. Кауфманом, А.А. Чупровим, П. Постниковим та українським статистиком-економістом Миколою Зібером.[x] всі вони дуже спричинились до того, що земська статистика та її переписи досягли високого наукового рівня і високо оцінюється представниками західньо-европейської статистичної науки.

Значіння деяких подвірних земських переписів збільшується ще й тим, що здобуті ними дані економіки сільського господарства можна було зв'язувати з природно-історичними умовами, в яких відбувались процеси господарські.

Ще в 1894 році Підсекція статистики на Загальноросійському З'їзді природників і лікарів з ініціятиви і на підставі докладу видатнішого українського статистика О.О.Русова, розроблюючи детальну програму подвірних земських переписів, визнала важливість і необхідність студіювання і статистичного обчислення інвентарю селянських господарсьв та к о о р д и н а ц і ї с п о с о б і в н а у к о в и х д о с л і д ж е н ь ґ р у н т і в і природно-історичних умов даної території, яко загальних передумов ведення на них сільського господарства - і з с п о с о б а м и с т а т и с т и ч н о г о д о с л і д ж е н н я с і л ь с ь к о г о г о с п о д а р с т в а. За програмою і за керовництвом професорів Докучаєва і Сібірцева в ряді земських губерній, між іншим в Полтавській, а потім і в Чернігівській /під керовництвом Б.Полинова/ і були переведені наукові дослідження ґрунтів цих губерній та на підставі їх зроблені детальні мапи ґрунтів цих губерній. Це й давало можливість економічні дані пристосовувати і зв'язувати з природно-історичними районами там, де такі дослідження було переведено.

Земська статистика виробила також оригінальні формуляри для подвірних переписів. Спочатку числення переводились за допомогою селищних списків. На кожне селище складався окремий список; кожний поземий рядок цього списку призначався окремому господареві , про якого далі у прямовісних графах нотувалися відомості відповідно до питань програми.

З розвитком,одначе, в дальших переписах програми дослідження,списькова система дуже ускладнювала оброблення матеріялу, через що й була замінена картковою. Перший тип такої картки був простий, бо на ній графи розміщувались відповідно до логічної послідовности окремих питань та груп питань за програмою перепису.

Але Воронезькими статистиками в 1891 р. був створений новий вид подвірної картки, якій потім і був прийнятий іншими земськими статистиками. Важливою властивість цієї картки було те, що кругом всіх, або деяких країв картки розміщувались /виносились назовні/ окремі графи, при чому цифрові величини в цих графах писалися так, що після накладання заповнених відомостями карток одна на одну, окремі графи кожної картки утворювали поспільні стовпи однородних по змісту цифер, які дуже легко було сумувати. Ця картка уможливила скоротити майже на одну четвертину дуже важливу працю підсумування числових даних. Це мало особливе значіння ще й тому, що машинове підсумування може робитись лише великими центральними статистичними урядами, бо досить дорого коштує.

В той час як Західня Европа при переведенні переписів переважно користувалась і користується й зараз блянковим способом спостереження,земська статистика рішуче стала на бік безпосереднього опитування господарів через спеціяльно підготовлених осіб. Живий безпосередній зв'язок регістра-

---

x/ Дивись стор.153-155 цікавої розвідки В. С а п і ц ь к о г о про Зібера під назвою "До історії економічної думки на Україні", надрукованої в "Записки Української Господарської Академії" Вип.І т.ІІ. 1929 року.

тора з особою, що опитувалась, можливість контролю відповідів на сільських зборах господарів, перехресні допити їх щодо важливіших питань перепису - все це дуже сприяло одержанню таких відомостей і відповідів на такі питання, яких і які блянковим способом ні в якому разі одержати не можна б було навіть від грамотної частини населення.

Опубліковані матеріяли земських подвірних переписів не тільки служать цінним джерелом для вивчення методів організації і самого переведення статистичного спостереження взагалі і особливо сільсько-господарських переписів, але й дають важливі вказівки для належного оброблення даних цих переписів. Як треба переводити контролю здобутого статистичним переписом матеріялу, в які табличні форми треба укладати матеріял, як його групувати для того, щоб досягти через науково статистичну аналізу конечної мети дослідження - всі ці важливіші питання і завдання освітлюються в працях земських статистиків з належною повнотою, з добрим науковим знанням предмету, який трактують не рідко з видатною талановитістю, властивою окремим представникам земської статистики.

Праці Воронізького земства під керовництвом Ф.А. Щербини, Чернігівського під керовництвом спочатку П. Червінського, В. Варзара і О.Шликевича, а потім О.Русова, Є. Добровольського і Н.Черненкова, Полтавського під керовництвом О.О.Русова і Г.Г.Ротмістрова та Харьківського під керовництвом С.Жилкіна заслуговують особливої уваги українських економістів і статистиків.х/ На превеликий жаль всі ці праці публіковані російською мовою і не для всіх українських інтересантів достугні для студіювання, але звичайно не вина в тому земських статистиків.

д/ Переписи промислові. Як вище було зазначено, сільсько-господарські переписи цей досі не охопили в багатьох країнах цієї важливої галузі господарського життя. Ще в гіршому стані знаходиться справа статистичного дослідження промислового виробництва. Можна сказати, що до світової війни лише дві країни визначаються систематичним статистичним спостереженням промислового виробництва, студіюванням його структури і розміра продукції - це Німеччина і Сп.Шт.Пів. Америки. Вони й створили два типи статистичного числення промислових підприємств, з яких Німеччина висунула завдання дослідження організаційної структури промислового виробництва, а Америка - виявлення вартісного ефекту продукції виробництва по окремих його галузях. Промислові переписи в Німеччині / Betriebzählung / переводилися в 1875, 1882, 1895, 1907, 1917 і останній, після світової війни, в 1925 р. В Сп.Шт.Пів.Америки так звані індустріяльні цензи / Census of manufactures / можна зазначити початковою датою 1850 року, бо з цього року створена була нова центральна організація його переведення, а в програму його ввійшли майже всі головні елементи сучасного індустріяльного цензу - встановлення кількости промислових підприємств, капіталу, вкладеного в підприємство, службовців, робітників і платні, відсотку зиску; збиралися також подібні дані від закладів добувавчої промисловости, а зокрема і від домашнього промислу. До 1900 року ці цензи переводились в 10-річних інтервалах, з 1900 по 1920 рік у пятирічних, а з 1920 року і до цього часу в двохрічних. Скорочення часу між цензами, показує яку важливу ролю відограє промисловість в економічній і соціяльній структурі Сполучених Штатів Пів.Америки.

В Німеччині промислові переписи 1882 і 1895 і 1925 р.р. були переведені одночасно з сільсько-господарськими переписами.

В Англії промислові переписи були передені лише в 1907 і 1924 році, в Бельгії в 1896, в Швайцарії в 1905 р., в Австрії в 1902 і 1930 році

х/ Для тих, хто бажав би ближче познайомитися з земською статистикою, дуже корисними можуть бути видання: 1/ С.Велецький::Земская статистика" два томи. М. 1899 г. 2/ В.Н.Григорьев: "Предметный указатель матеріалов в земско-статистических трудах с 1960-1917 гг." Вид. Ц.С.У.С.Р.С.Р. 3/ З.М. і Н.А.Свавицкие: "Земские подворные переписи 1880-1913 гг." Вид. Ц.С.У.С.Р.С.Р. 4/"Труды подсекции статистики" съездов русских естествоиспытателей и врачей, особливо "Труды" IX,X,XI,XII з'їздів.

разом з сільсько-господарським переписом, в Чехословаччині в 1930 р. разом з сільсько-господарським переписом, в Росії й на Україні для частини промислових підприємств були переведені промислові переписи Міністерством Торгу і Промисловости в 1900, 1908 і 1912 рр. під керовництвом відомого Чернігівського статистика В.Варзара, а потім у 1918 році Центральним Статистичним Управлінням СРСР.

Слабий розвиток промислових переписів пояснюється не тільки великими труднощами матеріяльними і методологічними, що зв'язані з їх переведенням, але й силою капіталістичної кляси, що не усвідомлювала собі загального економічного значіння такого роду числень та боялася і не хотіла допускати освітлення та розкриття таємниць промислового виробництва замкнутої капіталістичної системи господарки. Англійському парламенту довелося в 1906 році, при ухвалі акту про перепис, спеціяльно гарантувати збереження таємниць окремого підприємства та не використання статистичних даних перепису для цілей фіскальних /податкових/. Значну боротьбу і цілу дипломатію довелося вести і дирекції американських цензів, щоб поставити справу їх здійснення в більш-менш нормальні умови.

"До 1920 року, каже один з керівників американського цензу, політикою бюро було давати перевагу виявленню релятивних покажчиків значіння індустріяльної чинности по окремих штатах, або географічних районах. Значні труднощі були зажиті при набуванню звітів від окремих промисловців, більша кількість яких думала, що дані збираються для податкових цілей. Звіти при попередніх цензах, до 1909 року, коли жадались від підприємців доповнюючі відомості, що охоплювали ріжні видатки по експлоатації підприємств, звичайно, давали приблизний образ прибутків кожного індивідуального підприємства. Через те, що завданням бюро цензу не було проводити дослідження з цією специфічною метою, в наступних цензах ці запити були упущені і, я думаю, що це мало добрий наслідок для цілої індустрії, яка зараз виявляє рішучу спробу співпрацювати з бюром цензу при виготовленню звітів для кожного цензу".

Одначе слабий розвиток промислових переписів в ріжних країнах до певної міри пояснюється і недосконалим науковим розробленням методології цих специфічних переписів.

Для кожного вичерпливого статистичного числення муситься строго і точно встановити одиницю числення і ті ознаки, якими треба наділити її та які треба через числення одержати для поставленої мети дослідження. Ця проблема в промисловому перепису, себто численню промислових закладів і підприємств, не така легка до розв'язання. Коли метою статистики промислового виробництва, яким ми цікавимося, є завдання емпірично / конкретно/ відшукати і дослідити складові чинники і самий організм промисловсго виробництва з виявленням тих процесів змін, які в ньому відбуваються, то само собою одиниця числення є тут лише засіб на шляху досягнення поставленого завдання, а не об'єкт /предмет/ цілої акції. Не може бути сумніву, що дослідженням, яке за допомогою критерія "заклад" або "підприємство" студіює свої феномени лише в межах чинности промислової або сільсько-господарської праці, ще не віришується головна проблема в цілій її теоретичній і практичній актуальності. Через те й тоді, коли розв'яжемо питання про "заклад" і "підприємство", означивши їх поняття та витворивши для них дефініцію, ми не сміємо забувати, що заклад і підприємство є лише формально усталені організаційні типи народньо-господарської техніки праці, які ані суб'єктивно, ані об'єктивно не вичерпують всі можливі способи технічного прикладення праці і капіталу в економічному процесі.

В сучасному житті зустрічаємо надзвичайну ріжноманітність закладів і підприємств, особливо, в галузі торгівлі або державної військової господарки, які дуже тяжко підвести під одне формальне поняття промислового закладу, або підприємства. Значна кількість господарської і виробничої праці в торговому обороті, або в державній військовій господарці виконує лише функцію трудової техніки, що ані проста, ані спеціяльна терміно-

логія не може втиснути в формальні категорії поняття "заклад" і "підприємство" / Meerwarth /.

Ці дефекти в досягненню субстанціонального /істотного/ критерія понять, ускладняються обмеженням понять "заклад" і "підприємство" царином промислу і рільництва. Але ми знаємо заклади і підприємства, що ані логічно, ані технічно не є промисловими або рільничими. Так само поняття "праця" заховує в собі при повному гносеологічному /пізнавальному/ трактуванню свого народньо-господарського змісту далеко більше, ніж праця в царині чинности промислової та сільсько-господарської. В сучасному промисловому перепису доводиться ураховувати такі роди праці й професії, які строго не відносяться до так званої "продукційної" праці в промисловому виробництві.

Отже, коли переписом треба охопити не тільки ту працю, яка створює мінові цінности, а працю яка існує і виконується взагалі для задоволення потреб населення, то з погляду економічного закону задоволення потреб, незмінно корисною працею буде й та праця, реальні плоди якої не виникають лише з вимог мінової цінности на ринку, але й з волі праці інтелектуальної, плоди якої дають спромогу для людини і людства етично вищого культурного життя.

Це все примушує науково працюючого статистика не тільки свідомо ставитись до оцінки і значіння технічної одиниці статистичного числення, а й ставити її в зв'язок з відповідно конструйованими покажчиками статистичного виміру тих явищ, які поставило собі метою дослідження.

В статистичному перепису є недопустимим, одначе, обмежуватися неясними або загальними фразами відносно основних понять досліду, від влучного і відповідного усталення яких залежить результат цілої статистичної акції. Через те інструкція до формулярів перепису всі такі поняття встановити мусить так, щоб кожний, хто подає відомості, розумів їх однаково.

Прикладом може бути промисловий перепис 1902 р. в Австро-Угорщині, коли, як каже видатний спеціяліст др.Вальтершіф / Walterschiff/, через неясну інструкцію "підлеглі часто ad hoc призвані для переведення перепису органи, мусіли ламати голову над інтерпретацією вказівок вищих спеціяльних органів. В такій ситуації одні ставали на шлях суб'єктивного розуміння втоптаних стежок статистики соціяльної, в тісному розумінні цього слова, другі розв'язували питання з погляду ділової доцільности, а треті плуталися в діялектиці правничих асоціяцій". Цілком зрозуміло, що наслідком цього велика частина матеріялу перепису 1902 року вимагала цілком нового опрацьовання.

Так само не легким є і встановлення поняття "промисловий", "промисел", які не стабілізовані в тій мірі, щоб їх можна було так би мовити, пустити до статистичного числення без відповідної іструкції.

Справа ускладнюється ще тим, що й позитивне право не дає точних однородних понять. Зовнішні і внутрішні моменти законодавчої практики в справі кодифікації законів сприяли тому, що законодавець не ставив своїм завданням досягнення в правних нормах, що торкались економіки і політики повноти і стислости понять народнього господарства, через що і в літературі і в практиці виникає чимало непорозумінь. Наприклад, в зв'язку з розвитком соціяльного права законодавець створюючи статути забезпечення робітництва від недуг і каліцтва в промислі, поневолі мусив був для своєї нормативної потреби вибрати інші й ширші рямці поняття "промисловий" ніж рямці "промислового статуту". Не можна також сказати, що і в економістів теоретиків все стоїть гаразд. Візьмемо двох видатних економістів: Зомбарта й Шмолера.

В своїй спробі підійти до визначення органічної суті закладів, які можуть ділати технічно і економічно, як самостійні одиниці, або входити в підприємства, як технічні заклади у виді окремих відділів та спеціяльних технічних об'єктів до цих підприємств, ці економісти кажуть наступне: Зомбарт[x] розуміє під закладом / Betrieb / урядження, яке забезпечує дальше

---

x/ В.Зомбарт. "Современный капитализм" т.I.

влаштування об'єкту і виконання його функції. Під урядженням він розуміє все те плянове й організоване, що є властиве кожному закладові. Ознаки організаційного порядку він вважає тим більше доцільними, що вони дають нам змогу відмежувати поодинокі заклади в окремі одиниці, їх індивідуалізувати, бо без такого критерія порядку це відмежування зробити дуже тяжко. Одноцільність порядку в закладі є для Зомбарта основними моментами індивідуалізації.

Що ж треба розуміти під одноцільним порядком? На його думку, урядження закладів складається з таких основних організаційних моментів:

а/ Ведення процесу праці; сюди належить право розпорядження приняттям, затрудненням і звільненням робітничого складу, так з боку кванитативного, як і квалітативного, а також право розпорядження необхідними для виробництва варстатами та засобами праці.

б/ Оформлення процесу праці - себто означення місця і часу праці.

в/ Провадження / Ausführung / процесу праці,- себто догляд над фактичним переведенням наміченого пляну і над приписаним розвитком процесу праці - іншими словами - провід мусить бути одноцільним, що назовні виявляється в існуванні провідного керувчого органу /наглядавчого заряду/."

Зомбарт строго відряжнає поняття закладу від господарства /ширше поняття, яке обіймає також і поняття підприємства/ і останнє визначає як сполуку консументів, а заклад як сполуку працівників. Він не погоджується з тим, буцім то єдність закладу вимагає єдности варстатів і як приклад подає фабрику ваґонів, льокалізовану на 30-40 гектарах, поодинокі варстати якої завше залишаються під строгою контролею і центральним наглядом, хоч самі вони й віддалені на півгодини ходу один від одного. Цілком зрозуміло, каже він, що окремо лежачі об'єкти і відливарні і гути можуть в певних обставинах творити один заклад. Наприклад, велика піч іоб'єкти до перетоплення металю, або об'єк сталевий і вальцівня.

З'ясовуючи теж саме поняття, Шмолер каже: x/

" Коли на самім фабричному подвірю знаходяться три між собою скомуніковані будівлі - прядільні, фарбарні і ткальні, коли велика піч лучиться безпосередньо із сталівнею і вальцівнею, то те все творить не лише одно підприємство, але й один або більше збірних закладів; льокальна єдність і ощадність на коштах є головною метою. Але коли якийсь великий об'єкт залізного промислу купує собі копальні вугілля і руди для забезпечення собі сирівцю, лежать вони звичайно в значнім віддаленні один від одного і мають власних директорів, то підприємства частей цього великого об'єкту в кожному разі будуть більше закладів - але їх злука має одноцільну економічно-технічну мету; краще в такім випадку казати не про мішаний заклад, а про збірне підприємство з кількома закладами. До цього роду належать рівнож такі сполучення, як залізниці з гутами і копальнями вугілля, з варстатами, доками, фабриками машин і т.ін."

Отже, думка Шмолера не зовсім ясна: з неї докладно не видно, який саме момент узасаднює одноцільність, єдність закладу. Вказівка на момент льокалізації закладів, що мають при тім власних директорів, свідчить, буцім то, про те, що місцеве розташування технічних закладів є одним з головних моментів їх відмежування. На думку Зомбарта на увагу слід брати тільки момент порядкування продукційними процесом, момент місцевого положення, територіяльного відокремлення не має значіння. Єдність управління може одночасно бути поділена /технічно/ по територіяльному принципу, але це не є обовязковим.

Отже обидва наведені поняття закладу не торкаються інших моментів детального технічного процесу виробництва, що мають значіння для відокремлення одного закладу від другого. Обидва вчені встановлюють різницю /поділ/ в технічному процесі, як головну ознаку відокремлення одного закладу від другого.

---

x/ Schmoller. Grundris der Allgemeinen Volkwirschaftslehre ст. 500.

Третій видатний економіст Кальмес намагається вияснити поняття закладу із становища чисто технічної структури продукційного процесу, але для определення його теж не дає потрібної стислости і ясности. По його думці заклад є "замкнена цілість, себто ряд технічно своєрідних декількох процесів виробництва, що уявляють з себе більше або менше самостійну фабрикацію, або краще, ступінь фабрикації".

Одначе, по технічних лише ознаках не можна з належною об'єктивністю встановити ріжницю одного закладу від другого. Тут повстає цілий ряд питань: що є процес виробництва? В яких випадках окремі процеси виробництва належать до близьких споріднених родів? Коли виявляється більш-менш самостійний ступінь фабрикації?

Думка відносно поділу закладів на підставі технічних ознак може допускати наприклад, підприємство, яке обіймає відливарню заліза і фабрику машин, яко один заклад, понеже це підприємство посідає, як каже Кальмес, два такі процеси виробництва, які технічно не належать до істотної відрубности, а через те і не уявляють самостійного ступеня фабрикації.

Одначе на підставі неясного определення Кальмеса можливо й навпаки - вважати відливарню заліза за один заклад, а фабрикацію машин за другий. Відливарня сама по собі посідає такі процеси виробництва, які є технічно зв'язані між собою, наприклад - столярню моделів, формовню, токарню, ліярню в тісному розумінню, шліфовню іт.п., з другого боку, фабрикація машин,знову,має такі процеси продукції, як кузня, токарня, монтаж, гоблярня, свердлівня, фризівня іт.д. Коли можливо б було ділити заклади якогось підприємства за допомогою технічної підстави поділу, то розділення це відноситься і до слушно порушеного Зомбартом принципу щодо одноцільности управи і порядкування. Оскільки відливарня заліза і фабрикація машин мають ширший обсяг, вони можуть знаходитися під окремими управами і доглядом,отже уявляти з себе два заклади.

Зазначене приводить до висновку, що для статистика не можна осягнути порівняння окремих одиниць числення, лише через поставлення у формуляр перепису питання про число закладів. Бо на таке питання підприємці,або директори закладів подали б ріжноманітні відомості /один під закладом розумів би одне,а другий зовсім інше/, які потім не можна б було систематично і доцільно обробити. Так, між іншим, і сталося з промисловим переписом 1907 р. в Німеччині, коли навіть після повернення величезної кількости формулярів для виправлення на місцях, все ж таки матеріял не став придатнішим для оброблення /одиниці числення були не до порівняння/ і публікацію його довелося зовсім взяти з обігу.

Отже, навіть у тих багатьох випадках, коли конкретно сучасне підприємство не розподіляється технічно на заклади, себто уявляє з себе єдину нероздільну цілість технічну і таким чином поняття підприємства і закладу здавалось би не можуть змішуватись на практиці,[x]/ треба мати на увазі, що сучасне промислове підприємство не є лише організацією продукції, а також і організацією торгу. Воно не виробляє лише самі продукти, але й купує сирівці та півфабрикати, а виготовлені продукти продає. Кожне підприємство розпоряджає торговельними відділами з меншим або більшим обсягом ділання.

Вищенаведене свідчить про те, які труднощі доводиться поборювати при встановленню основних понять /одиниць числення/ у промисловому переписі, для того щоб правильним численням технічних одиниць - закладів, можна б було розкласти при обробленню матеріялу окремі підприємства на заклади і по потребі окремі заклади об'єднувати в підприємства. Лише такий матеріял дає можливість статистично з'ясовувати організаційну структуру промислового виробництва. Гомогенність /однородність/ одиниць числення є одним з головних завдань промислового перепису.

---

x/ Підприємство є поняття економічне, а заклад технічне і через те змішання цих понять у промисловому переписі робить перепис не придатним для яких будь висновків.

Betrieb
Unternehmen

Лекція шоста.

Продовження розд.16 літ. д. Промислові переписи. Завданням промислових переписів є одержати статистичний матеріял, по можливості, гомогенних одиниць промислового виробництва. Цей матеріял повинен дати спроможність до оброблення його як у напрямку з'ясування природи технічної організації виробництва разом з виявленням динамічних процесів, що відбуваютьсяз бігом часу в його сукупності, так і в напрямку з'ясування економічної природи цього виробництва та тих змін, що характеризують еволюційний процес її розвитку.

"Заклад-підприємство" і "заклад у збірному підприємстві" /підприємство з декількома закладами/ є одиницями числення, що дають спроможність широко і детально освітлювати технічно-організаційну природу виробництва.

"Підприємство-заклад" і "підприємство з декількома закладами" є одиницями, що дають спроможність освітлювати економічно-організаційну природу виробництва.

Виходячи з цієї потреби розподілу підприємств за закладами і інтеграції /поєднання/ закладів у підприємствах, на нашу думку, можна б було зупинитись на такій дефініції закладу, як основної одиниці промислового числення.

Під промисловим закладом без ріжниці суб'єктивної кваліфікації підприємців /фізична особа, або юридична, приватно або суспільно-правна, індивідуум чи асоціяція в ріжних її правних формах/ вважається кожна виробнича єдність, яка або назовні правно самостійна, або в складі збірних підприємств технічно інтегральна та служить для промислового створення чи обороту господарських дібр споживчої або мінової цінности, або виконує трудові та службові функції мінової цінности.

Розуміється, інструкція мусить до цієї загальної дефініції подати відповідні приклади різного роду підприємств, які охоплює промисловий перепис, у цілях одноманітного розуміння дефіціції тими особами, що заповнюють формуляр перепису.

Помилки в нотованню одиниць числення звичайно можуть бути і при такій постановці справи переведення перепису, при блянковому способі збирання відомостей, але, на нашу думку, дефекти матеріялу не можуть бути дуже важливими і, безперечно, можуть бути виправлені після належної контролі.

Вироблення іструкції з прикладами для правильного розуміння дефініції одиниці числення безперечно мусить робитись з участю експертів-техніків, добре ознайомлених з технологією виробництва.

Методологічні труднощі переведення промислових переписів не обмежуються лише розв'язанням проблеми відшукання гомогенної одиниці числення. Не тільки в стадії оброблення добутого переписом матеріялу, але й для самого числення закладів промислового виробництва потрібно наперед звести в систему ріжні роди виробництва, бо і в стадії одержання відомостей дані підпадають під ріжноманітні якісно роди виробництва, відповідно до характеру промислової чинности закладів.

У всіх наступних операціях зведення і оброблення первісного матеріялу промислові підприємства обов'язково мусять систематизуватись, себто відповідно групуватися по родах виробництва в зв'язку з тими ознаками, які намічено зібрати переписом.

Така система для основного групування матеріялу і зветься клясифікацією виробництва.

"Клясифікація виробництва, каже один з ліпших теоретиків і практиків промислової сатистики В.Е.Варзар, є основою групування підприємств - вона є вихідною базою всіх дальших статистичних операцій, це - кардинальний пункт методології в науковій аналізі,це твердий кістяк цілої будови промислової статистики."

Але й до цього часу в теоретиків і практиків статистики нема потрібної єдности в поглядах на будову такого роду клясифікацій, через це ми не маємо й доси такої клясифікації промислового виробництва, яка б давала можність порівняння даних промислових переписів різних країн у цілях наукових студій цієї важливішої царини економічного життя.

В кожній країні, різні установи, що ведуть промислову статистику, групують підприємства, а через те і всі відомості зібрані від них, за різними схемами виробництва.В загальній своїй частині ці схеми здобиць подібні, але в детальній основі настільки,різні,що не дають ніякої можливости порівняння між собою, чим позбавляють статистичний матеріял наукової цінности.

Існуючі до недавнього часу класифікації, розподіляли промислове виробництво на три великі відділи: А - Сільське господарство, В - Промисловість і С - Торгівля. Пізніше до цих відділів стали додавати ще - Транспорт, а потім відділи таких промислововедених підприємств, як театральні, музичні, спортовні, або по охороні здоровля і забезпеченню гігієни.

Кожний з цих відділів поділяється на групи виробництва /визначаються рімськими цифрами/, групи на кляси /визначаються арабськими цифрами/ а кляси на підкляси /визначаються літерами алфавиту/. Ціла клясифікація містить в собі таким чином сотні окремих видів /підкляс/ виробництва. Внутрішнім своїм змістом суть і кількість родових і видових ознак клясифікації залежить в кожній країні як від характеру природи джерел, що дають засоби промислу, так і від характеру технічного і економічного розвитку даної країни,ступінь якого і визначає багатство і різноманітність виробничої техніки і самих продуктів індустрії.

Щоб мати конкретну уяву про форму і зміст сучасної клясифікації промислового виробництва,подаємо тут три схеми для новіших промислових переписів./Див.стор. 45-48 /.

Коли порівняти одну з старіших клясифікацій французького статистика Ж. Бертільона,прийняту в 1892 році Міжнароднім Статистичним конґресом у Чікаґо, з вищенаведеними, то побачимо, що з часом кількість кляс і підкляс /неподільних позицій/ росте все більше й більше. Два відділи клясифікації Бертільона, що охоплюють промисловість, мають у собі 113 кляс, з яких 72 розділені на 297 підкляс.

Останні американські цензи також довели вже свій поділ майже до 360 клясифікаційних видів виробництва.

Принципи, що покладені в основу сучасних клясифікацій, можуть у скорому часі привести поділ промислового виробництва до занадто великої диференціяції, на нашу думку, зайвої. Коли взяти на увагу, що при дуже великій економії в переведенню цензів, якої досяг сучасний директор Стюарт, втягнувши до ближчої співпраці з Бюром цензa 695 місцевих торговельних комор і 488 інших організацій індустрії, все ж таки доводиться витрачати біля міліона долярів тільки на переведення цензa - то цілий цензa з обробленням і публикацією матеріялів по такій клясифікації виробництва коштував би величезні гроші. x/ Не дивно, що Бюро дає одну лише просту, зведену таблицю всіх видів виробництва / у підметі / і всіх основних ознак спостережених цензом / у присуді /.

Цілком зрозуміло, що й інші статистичні уряди зводять свої публікації до значно меншого обсягу, не відповідаючого детальній клясифікації. Очевидно, треба шукати іншого шляху для компактної системи промислового виробництва.

---

x/ 400 видів виробництва в комбінації з територіяльним поділом країни наприклад по 50 округах давало б 20 тисяч рядків показових цифер для ознак виробництва, а в зв'язку з розміром підприємств по числу робітників, валовій продукції і тоді сотні тисяч рядків публікацій.

Порівняльна схема клясифікацій промислового виробництва новіших переписів у Німеччині в 1925 році, в Росії в 1918 році і в Чехословаччині в 1930 році.

1. Перепис 1925 р. у Німеччині		2. Перепис 1918 р. у Росії		3. Перепис 1930 р. у Чехословаччині	
Відділи і групи	Число неподільних видів виробниц. /підкляс/	Відділи і групи	Число неподільних видів виробниц. /підкляс/	Відділи і групи	Число неподільних видів виробниц. /підкляс/
A. Промислововедене городництво і плекання худоби та морське рибальство.					
I. Промислововедене /не сільсько-господарське/ городництво і плекання худоби	2	----------	------	XXII. Промислове плекання тварин і лісові маніпуляції /ця група є частиною відділу A - Промисл. виробництво/	2
II. Морське рибальство	I	----------	------	A. Промислове виробництво.	
B. Промисловість і ремесло		B. Промисловість			
III. Гірництво, добування солі і торфу.	I5	VIII. Гірнича і гірничозаводська промисловість	45	I. Гірнича промисловість, добування і виробництво торфу, коксу і брикетів	9
IV. Добування і оброблення каменю і землі	42	VII. Добування і оброблення каменю, землі і глин	47	II. Виробництво з каменю і землі	48
				III. Шкляна промисловість	I4
V. Добування залізної руди та руд інших металів	27	Дивись VIII графу клясифікації			
VI. Оброблення заліза, сталі та інших металів	30	IX. Оброблення металів	58	IV. Оброблення металів	67
VII. Індустрія машин, інструментів і засобів пересування	44				
VIII. Промисел електротехнічний, виробництво пристроїв точної механіки і оптики	28	X. Виробництво машин, інструментів і апаратів	52		
IX. Хемічна промисловість	23	XII. Хемічна промисловість	27	V. Хемічна промисловість	9

Відділи і групи	Число неподільних видів виробниц. /підкляс/	Відділи і групи	Число неподільних видів виробниц. /підкляс/	Відділи і групи	Число неподільних видів виробниц. /підкляс/
X. Текстильна промисловість	41	XVI. Виробництво з бавовни	20	VI. Текстильна промисловість	34
		XVII. " шерсти	21		
		XVIII. " шовку	11		
		XIX. " льону	13		
		XX. " коноплі та інших рослинних волокнистих речовин.	12		
		XXI. Виробництво з мішаних та неточно зазначених волокнистих речовин	9		
XI. Паперова промисловість і промисел друкарський	20	XXIII. Паперова промисловість	19	VII. Паперова промисловість	11
		XXIV. Поліграфична промислов.	13	VIII. Промисел графічний	6
XII. Промисловість шкіряна та лінолеума	6	XV. Шкіряна і хутрова промисловість	17	IX. Промисловість шкіряна	9
XIII. Промисловість кавчуку та азбесту	6	Входить у групу XII /хемічна промисловість/		X. Промисловість кавчуку і азбесту	2
XIV. Оброблення дерева і різьбярство	41	XI. Оброблення дерева	22	XI. Оброблення дерева	29
				XII. Промисел токарний /крім дерева/	3
				XIII. Виробництво щіток	1
XV. Промисел музичних інструментів і забавок /іграшок/	14	XXV. Промисел мистецтва і прикладної науки	46	XIV. Промисел музичних інструментів	4
				XV. Виробництво забавок /іграшок/	1
XVI. Харчова і смакова промисловість	53	XIII. Харчова і смакова промисловість та напоїв	63	XVI. Харчова і смакова промисловість	45
		XIV Оброблення твердих матеріялів тваринного походження	21		
XVII. Виробництво одягу	18	XXII. Виробництво одягу і туалетів	45	XVII. Виробництво одягу	26
XVIII. Будівельний промисел /включаючи і другорядні його види/	22	XVIII. Будівельний промисел	21	XVIII. Промисел будівельний та цивільної інженерії	27
		XXVII. Інженерно-будівельна справа	14		

Відділи і групи	Число неподільних видів виробниц./підкляс/
XIX. Добування і постачання води, газу і електрики	6
С. Торгівля і транспорт.	
XX. Промисел торгівлі	98
XXI. Асекурація	4
XXIII. Комерційне ведення ресторацій і шинкарства	6
XXII. Транспорт	22
Д. Комерційно ведені підприємства театральні, музичні, розважові та навчання.	
XXIУ. Промислові підприємства театральні, музичні спорту та розваги.	9.
XXУ. Промислово ведене навчання	3

Відділи і групи	Число неподільних видів виробниц./підкляс/
XXУI. Виробництво і передача фізичних сил і водопостачання.	7
A. Залізничні майстерні	
XXX. Б. Трамваї	
B. Елеватори	

Відділи і групи	Число неподільних видів виробниц./підкляс/
XIX. Водоводи	I
XX. Газовні	I
XXI. Електрарні	2
XXIII. Промислове ведення громадських інституцій, благодійних установ, в'язниць і карних колоній	I.
В. Торгівля, банківництво, асекурація.	
XXIУ. Мандрівна технічна праця та послуги і мандрівна торгівля	3
XXУ. Торгівля в тісному розумінню	60
XXУII. Банківництво	I2
XXУIII. Асекурація	6
XXIX. Решта спорідненого промислу	25
XXУI. Комерційне ведення готелів, ресторацій і шинкарство	8
С. Транспорт /крім залізниць і пошти/	
XXX. Транспорт /крім залізниць і пошти/	6
Д. Комерційно ведені підприємства театральні, музичні, спортовні, розваги та навчання.	
XXXI. Промислові підприємства театральні, музичні, спорту та розваг.	7
XXXII. Промислово ведене навчання	3

Відділи і групи	Число неподільних виробництв /підкляс/	Відділи і групи	Число неподільних виробництв /підклясу/	Відділи і групи	Число неподільних виробництв /підкласу/
Е. Комерційно ведені підприємства для охорони здоровля і забезпечення гігієни.		ХХІХ. Очищення будов, дворів і вулиць	9	Е. Комерційні підприємства для охорони здоровля.	492
ХХУІ. Комерційні підприємства для охорони здоровля і забезпечення гігієни	10			ХХХІІІ. Комерційні підприємства для охорони здоровля.	10
Разом	591		612		

"Наука, каже С. Джевонс,[x] є відкриття тотожности, а клясифікація є сукупне розміщення або в мислі або реально в просторовій близькості тих предметів, між якими була з'ясована тотожність. Через те вартість клясифікації рівнозначна з вартістю науки і взагалі з висновком. Завжди, коли ми складаємо клясу, ми зводимо множність до єдности або викриваємо, як каже Платон, єдине в множности і. Клясифікація має наслідком узагальнене знання, що відріжняється від безпосереднього і чуттєвого знання окремих фактів... Тільки прикладення на ділі клясифікаційних і узагальнених здібностей дає людському розуму можливість збагнути до певної міри безмежну кількість природних явищ".

Таким чином, клясифікація виробництва мусить дати нам певну інтеґрацію ріжноманітної природи промислового виробництва для збагнення основних моментів його технічної і економічної структури, а через повторні переписи і динаміки важливіших процесів, які в ньому відбуваються, а не для опису безмежних дрібнот цього виробництва. "За деревами, не побачиш і лісу", каже гарна приповідка. Відсіль випливає необхідність для сучасної промислової статистики знайти шлях для уніфікації різних клясифікацій промислового виробництва в напрямку інтеґрації, а не дальшої диференціяції груп і кляс виробництва.

Метою клясифікації виробництва мусить бути така система розподілу і групування закладів аналогичних по своєму технічному і економічному характеру /прикметам/, яка, усуваючи дрібні неважливі відмінности, дала б основу необхідного узагальнення і легкого огляду та відповідала б кращому студіюванню промисловости і виявлення результатів такого студіювання в належних формах. Кожне підприємство, як економічно-неподільна одиниця числення промислових переписів може ввійти до клясифікаційного групування матеріялу в наступний спосіб:

І. Яко нескладна, проста одиниця, коли клясифікаційна характеристика виробництва збігається та охоплюється єдністю економічної самостійної відрубности /не беручи на увагу помічних і доповнюючих виробництв/, себто, коли заклад збігається з підприємством.

2. Яко складна одиниця, коли в економічній єдності з'єднано декілька виробництв, але вони входять в ту ж саму групу виробництв та складають послідовний процес перероблення однорідних матеріялів і одержання подібних продуктів.

---

x/ "Основи науки" стор. 627 - 628.

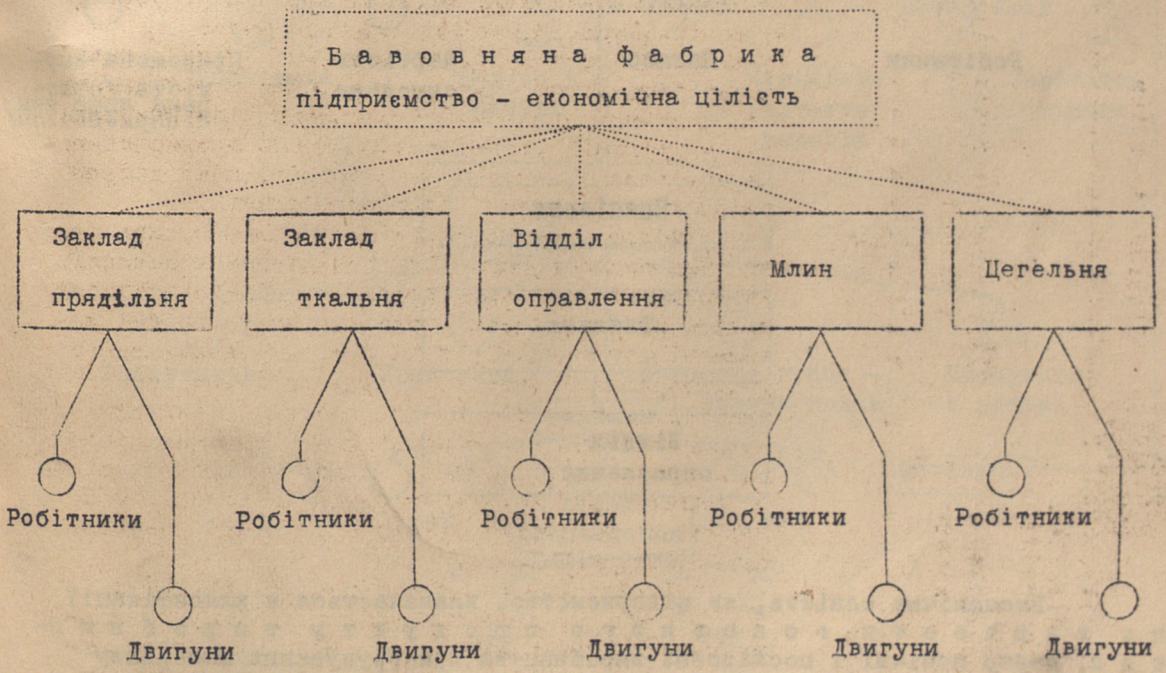
3. як комбінована одиниця, коли в склад підприємства входять різні щодо своєї природи виробництва, обробляють вони /або добувають/ різноманітні матеріяли або півфабрикати різними способами та в результаті дають різні вироби і продукти споживи.

В першому простому випадку всі системи клясифікацій - німецька, американська, російська легко гармонізуються. Але в випадках складних або комбінованих підприємств групування закладів і підприємств за ознаков "виробництва" у всіх зазначених системах дає глибоку різницю.

Користуючись цими міркуваннями, В. Варзара, подаємо далі і влучно ним поданий зразок різниць цих систем. Наприклад, візьмемо таке підприємство: бавовняна фабрика має прядільно-ткацьке виробництво і відділ оправлення /ушляхотнення/, а крім того два побічних виробництва: млин і цегельню. Для більшої ясности припустимо, що млин і цегельня за ознаками управління, встановленому порядку рахівництва у підприємстві і територіяльно злиті і неподільні з фабрикою бавовняною.

В нижченаведених схемах подаємо в вигляді чотирикутників окремі одиниці числення,/однаково як економічні, так і технічні/, з яких оточені суцільною лінією будуть ті, що при обробленню матеріялу можуть бути введені в рямці клясифікації, а точкованою лінією ті, що не можуть бути зв'язані з нею і при групуванню матеріялу не поєднуються з клясифікацією. Для повноти уяви про різниці - колом оточені основні ознаки, спостережені переписом в кожній одиниці числення. Схеми дадуть такий образ:

Німецька система.



Економічна цілість, як одиниця числення, розпалася на виробництва-заклади /Betrieb' /, при обробленню матеріялу вона не може бути поставлена в зв'язок з клясифікацією. Остання охоплює всі заклади, що пасують з її розподілом виробництва. Уяви про прямовісну і позему[x] структуру цілого

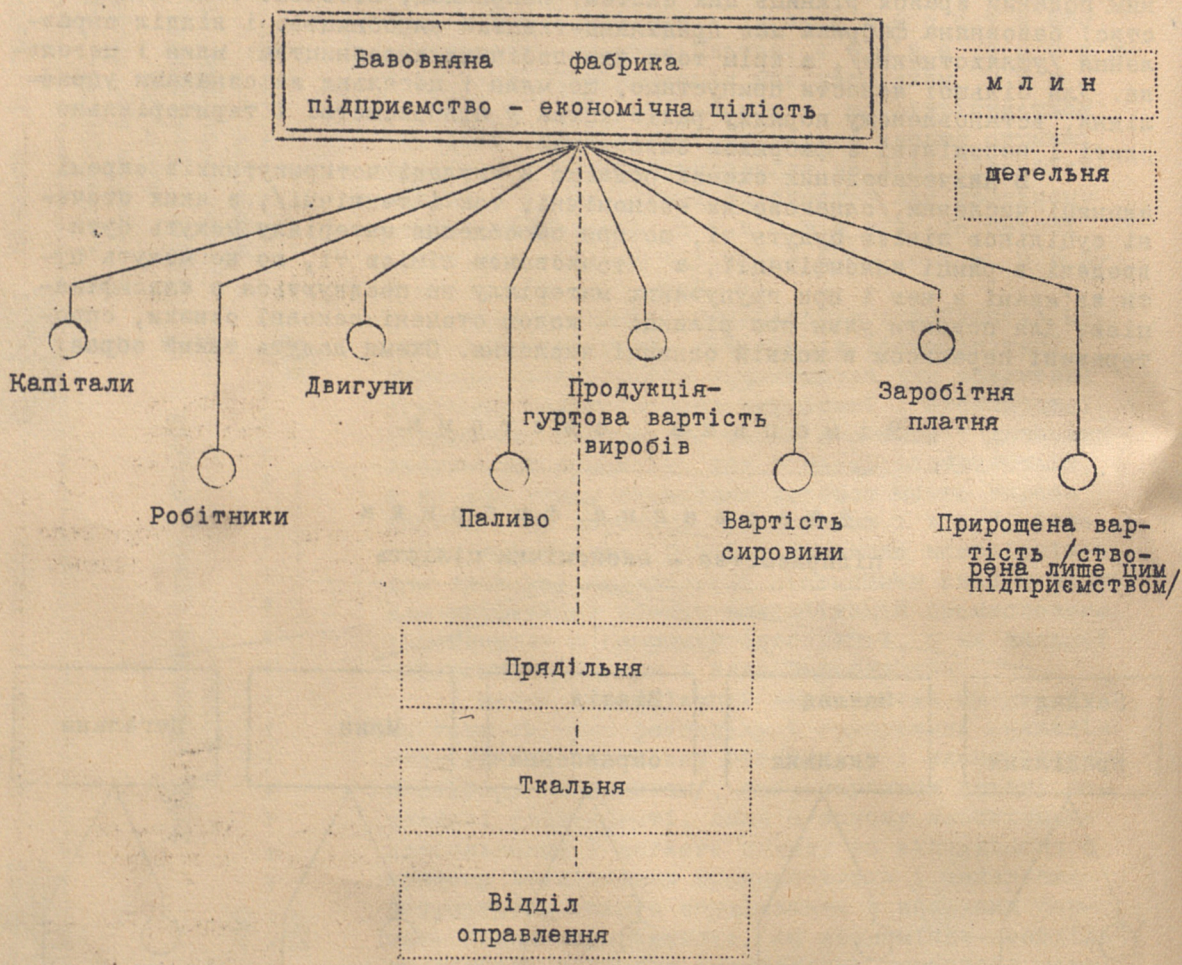
---

x/ Під прямовісною комбінацією розуміємо таку, коли підприємство одну галузь своєї продукції сполучує з попереднім або наступним ступенем фабрикації. Напр., виробництво заліза сполучено з копальнею залізної руди та вугілля, фабрика машин з відливарнею заліза і т.п. Під поземою комбінацією - таку, коли в одно підприємство входять різні технічно-виробничі заклади. Напр., тартак і олійниця. Копальні вугілля з хемічною фабрикою. Продовж.стор.50

підприємства мати не можна. Почислено для кожної одиниці дві ознаки - число робітників, що можуть поділятись за професіонально-демографічним переписом, і число механічних двигунів /рід, число і сила/. Переписний формуляр /Gewerbeboden / загальний для всіх закладів.

Американський перепис /ценз/, що охоплює майже всі промислові підприємства.

А м е р и к а н с ь к а с и с т е м а



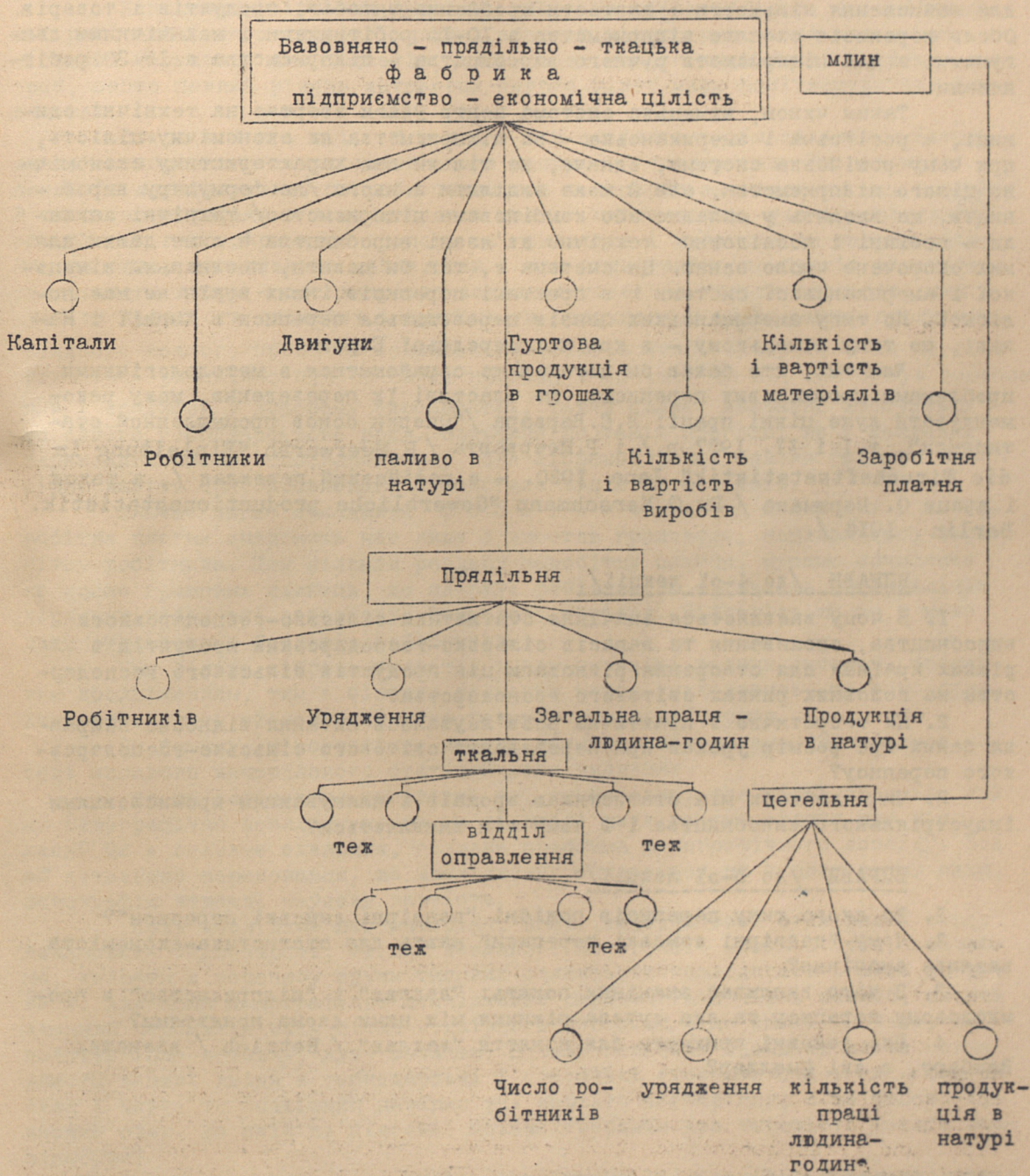
Економічна єдність, як підприємство, визначається в класифікації по вартости головного продукту виробництва, себто помічні і послідовні виробництва при групуванню матеріялу цезавть. Мати уяву про позему і прямовісну комбінації в підприємствах не

Завваження до німецької системи цілком відносяться до попередніх переписів. Щодо класифікації перепису 16 червня 1925 року, то до схеми її коротко додається, що автори її мають на увазі встановити для промислових закладів дві ознаки розподілу: територіяльну і економічну єдність /oertlich und Wirtschaftseinheit / і технічну єдність / Technische Einheit/; перші одиниці звуть вони підприємством /Unternehmung /, другі - закладом /Betriebe /. Це вже крок наперед, але сама внутрішня будова класифікації, що наперед фіксує форми комбінованих підприємств, може мати наслідком багато утруднень при обробленню матеріялів перепису і тим почасти вертає систему до старої бази.

можна. Кількість почислених ознак - велика. Числа подаються результативні за цілий рік. Для кожного підприємства одночасно, крім загального формуляру - скедули /General schedule /, вживається до 80 форм окремих доповнюючих формулярів для спеціяльних виробництв /ценз I925 року/.

В табличних зведеннях обчислення в натурі кількости головних, спожитих фабрикою, матеріялів і виготовлених виробів не подається. Переписом охоплюються всі рідприємства, річна продукція яких є вищою за 500 долярів.x/

Р о с і й с ь к а с и с т е м а.



x/ Для ясніщої уяви, що саме є американська скедула /формуляр перепису/, подаємо тут окремим додатком до курсу лекцій, як генеральну скедулу, так і одну з багатьох додаткових спеціяльних форм її за індустріяльним цензом I925 року.

Система в цілому зв'язана з класифікацією. Одиницею числення є підприємство, як економічна цілість, що визначається по складу своїх технічних виробництв /закладів/. Обсяг ознак-близький до американського цензу. Для побічних виробництв, що відносяться до інших груп, ніж ціле підприємство, а також для виробництв послідовних ступенів фабрикації, додаються окремі формуляри з скороченим числом ознак. Система дає можливість судити про прямовісну і позему комбінацію в підприємствах. Числа подаються результативні за річний період. Число формулярів - 37, рахуючи в тому один основний і по I2 форм / по числу груп виробництва/ - для технічного розподілу підприємства, для обчислення кількости і вартости спожитих матеріялів і для обчислення кількости і вартости зроблених виробів, продуктів і товарів. Обсяг переписів охоплює підприємства з I0-I5 робітниками і механічними двигунами, а для підприємств ручного виробництва - підприємства з I5-30 робітниками.

Таким чином, німецька система центр уваги звертає на технічні одиниці, а російська і американська - на підприємства як економічну цілість, при чому російська система, одначе, не тільки дає характеристику економічно цілого підприємства, але й може виділяти з нього /по формуляру виробництв, що входять у складне або комбіноване підприємство/ технічні заклади - побічні і послідовно технічно зв'язані виробництва й лише даючи для них скорочене число ознак. Ця система є, так би мовити, поєднанням німецької і американської системи і в практиці переписів інших країн не має подібної. По типу американських цензів переводяться переписи в Англії і Канаді, по типу німецькому - в країнах середньої Европи.

Для тих, хто бажав би детальніше ознайомитися з методологічними проблемами промислових переписів, що властиві їх переведенню, можу рекомендувати дуже цінні праці: В.Є.Варзара /"Очерки основ промышленной статистики". т.I і II. I927 р./,і Р.Меерварта / R.Meerwarth."Einleitung in die Wirschaftsstatistik" Jena I920. - є російський переклад /, а також і праці О. Нерштмана / Dr.O.Nerschmann "Gewerbliche productionsstatistik. Berlin. I9I6 /.

ВПРАВИ /до 4-ої лекції/.

I. В чому виявляється значіння статистики сільсько-господарського виробництва, споживання та запасів сільсько-господарських продуктів в різних країнах для створення рівноваги цін продуктів сільського господарства на головних ринках світового господарства?

2. Як теоретично і практично розв'язувалось питання відносно збирання даних про розмір урожаю при переведенню світового сільсько-господарського перепису?

3. Чим є зв'язок між статистиком врожаїв і плянуванням промисловцями індустріяльного виробництва і в чому він виявляється?

ВПРАВИ /до 5-ої лекції/

I. До якого типу переписів подібні "подвірні земські переписи"?

2. Чому "подвірні земські переписи" мають для статистика-економіста важливе значіння?

3. З чого випливає значіння понять: "заклад" і "підприємство" в промисловому перепису та яка суттєва різниця між цими двома поняттями?

4. Які головні прикмети для поняття "заклад" / Betrieb / визначає Зомбарт, а які Шмоллер?

Лекція сьома.

е/ Інші типи статистичного числення. а/ Монографічні дослідження. Під монографічним способом дослідження звичайно розуміють - спосіб детальнішого опису одного або групи явищ, вибраних з цілої маси однородних явищ, з метою прикласти результати цього опису і дослідження як типові, до всієї сукупности однородних явищ.

Наприклад, зібравши детальні дані, що характеризують бюджет типової робітничої чи селянської родини або декількох груп таких родин, на підставі характеру цих бюджетів, припускати аналогічний склад і висоту потреб, себто певний рівень життьових потреб цілої маси робітництва або селянства.

Селянські бюджети ставлять своїм завданням зображення економічного життя селянського двору - господарства. За допомогою бюджетів досліджується потреби селянства і способи задоволення цих потреб. В таких монографіях з'ясовується як наявне майно селян, так і подається обчислення господарського обороту та прибутків і видатків по господарству. Як при описі селянського майна, так і при студіюванню прибутків і видатків селянського двору - господарства встановлюється не тільки вартість, але й рід та кількість спожитих і зужитих та незужитих у виробництві предметів. Відносно кожного предмета селянського господарства показується чи пройшов він через ринок, або був об'єктом продажу й купівлі чи ні, себто був вироблений і спожитий у власному виробництві. Такий бюджет звичайно стосується до одного господарського року. Звичайно по такій програмі не є можливим переводити масове спостереження. Для зібрання даних одного бюджета потрібно навіть для вправленого статистика від 2 до 4-х днів.

Значно легше складання робітничих бюджетів. Заробітня платня знайомить нас лише з висотою грошового, номінального заробітку робітника. Щоб пізнати реальну заробітну платню, мусимо обчислити не число грошових одиниць, що одержує робітник, а кількість матеріяльних корисних дібр, які забезпечує йому його зарібок. Відповідь на це й може дати студіювання бюджетів робітників.

Чим більше зібрано бюджетів і чим типовіші родини охоплені бюджетним дослідженням, тим з більшою певністю можна сподіватись, що воно правдиво висвітлить економічний стан певної суспільної групи.

Як бачимо, монографічна метода дослідження є прямо протилежною методі масового вичерпливого статистичного числення.

Як же встановити типовість тої групи явищ, для цілої маси їх, які ми монографічно досліджуємо та на підставі яких хочемо судити про цілу масу? Це є головне завдання, головна проблема монографічного досліду. Без об'єктивного переконання, що дослідження спирається на типову групу явищ, монографія втрачає наукову вартість.

Коли дослідник має матеріял, напр., вичерпливого подвірного опису селянських господарств, то для детальної монографії можна б було, звичайно, вибрати з перепису групи типових селянських господарств і ними оперувати. Але коли монографічне дослідження переводиться через певний, досить значний перемежок часу після подвірного перепису, а це завше і буває, то повстає інше питання, дуже важливе. З часу перепису динаміка життя могла внести великі зміни в господарські і соціяльні стосунки цілої маси селянства і через те подвірний перепис дає занадто застарілий, а не типово-сучасний стан. На якій же підставі можемо твердити про типовість нашого дослідження?

Коли для бази монографії ми не маємо об'єктивного, незадовго перед тим переведеного вичерпливого спостереження, тоді тяжко відшукати типові елементи цілого явища, яке хочемо характеризувати. При таких умовах залишається досить широкий простір для суб'єктивних міркувань і припущень щодо типовости об'єкту монографічного дослідження. В цим і полягає головна не-

безпека такого способу дослідження. Таким чином, монографічне дослідження не можна назвати с т р о г о с т а т и с т и ч н и м в науковому розумінні.

Відомий дослідник селянських бюджетів на Вороніжчині Ф.А.Щ е р б и н а[x], який мав досить свіжий матеріял подвірного перепису селянського господарства, для перевірки типовости своїх селянських бюджетів, все ж таки мусів був зробити спробу перевірки цієї типовости, і для того за з значно зменшеною програмою зібрав 176 тисяч бюджетів для цілої губерні . Це дослідження його переконало, що типовість попередніх бюджетів ним осягнена була цілком задовольняюча. Але звичайно, така скількість скорочених бюджетів для перевірки типовости повних бюджетів виходить за рямці звичайних бюджетних досліджень і в практиці навряд чи може повторюватися. З вищенаведеного, звичайно, не треба робити висновку про повну непотрібність монографічного способу дослідження, зокрема для студіювання бюджетів і виявлення потреб споживання певних кляс суспільства. Монографічні дослідження не можуть заступати масового спостереження, але можуть бути, при обережному і розумному використанню цієї методи, бути корисним доповненням до загальних переписів; вони дозволяють малювати найдрібніші деталі явища, що досліджується та "розкривати інтимніші сторінки життя і через це надавати життєвість і забарвленність результатам більш екстензівних /менш глибоких/ масових досліджень" / Eulenburg /.

Через те монографічні бюджетні дослідження займають у статистичній практиці поважне місце. М а й о - С м і т зупиняється на трьох способах складання монографічних бюджетів: перший, що зветься екстензивним та визнається ним справедливо непридатним для такого роду досліджень - це збірання даних через розсилку бланків більшому числу родин. Другий - інтенсивний спосіб Л е п л е / Le Play / є детальне студіювання життя робочої родини кім небудь, хто буде з нею жити разом і особисто знайомитися із всіма обставинами життя такої родини. Третій - система родинних записів: намовити інтеліґентного робітника або його родину вести повний запис прибутків і видатків на протязі року. Майо-Сміт цей спосіб вважає за найдоцільніший. Нарешті, треба додати і четвертий спосіб, який полягає в складенні бюджета вправленим статистиком на підставі особистого опитування обраних для цієї мети родин на підставі раніш виготовленої програми. Цей спосіб поперше був досконало розроблений Ф.А.Щербиною і широко прикладався земськими статистиками в Росії і на Україні /Наддніпрянській/ для дослідження трудового селянського господарства. Щодо ведення прибутково-видаткових книг, то цей спосіб був використаний в Німеччині союзом робітників металістів /1909 р./, дослідженням якого зібрано було біля 320 робітничих бюджетів. За допомогою різних громадських організацій також Статистичний Уряд в Німеччині /1907 -1908 / доручив книжки 4136 господарствам, але дані річних бюджетів одержано лише для 960 родин, з них особливої уваги заслуговує оброблення 285 бюджетів Гамбурським міським статистичним бюром. Дуже цінні праці по такій системі переводяться /з 1901 р./ також швайцарським селянським секретаріятом на чолі з Laur'ом, при чому програма книжок пристосована для одержання даних, що характеризують рентабільність селянського господарства та для студіювання потреб споживання селян - господарів і найманих робітників.

На першому місті в Середній Европі треба поставити також бюджетні праці, що ведуться під керовництвом проф. В. Б р д л і к а Сільсько-Господарським Інститутом бюджетних досліджень у Чехословацькій республіці.

В Чехословаччині прибутково-видаткові книжки ведуться тут великою кількістю господарів і за надзвичайно детальною програмою, що дає можливість дуже глибоко охопити життя селянського господарства.[xx]

---

x/ Дивись його "Крестьянские бюджеты"

xx/ Результати цих досліджень освітлюються в таких працях: Dr.Prof.Brdlika: a/"Zemědělská spravověda" 1925.b/ "Rentabilitni soutěž vyrobnich směru skotu". Zemědělsky Archiv. 1925.

В значно меншому обсязі ведуть бюджетні дослідження статистичні урядн в окремих країнах, головним чином родин робітників і урядовців.

б/ А н к е т а. Під цим способом дослідження розуміємо або - I/ збірання особистих думок компетентних осіб про які будь факти і явища соціяльно-економічного життя, або 2/ одержання ясних відповідів від цілої маси осіб на такі питання програми, які освітлюють певне явище, що нас цікавить. До першого типу анкети можна віднести анкети англійських парламенських комісій, наслідки дослідження яких друкуються в так званих "синіх книгах". Це урядові анкети, але такого ж типу бувають і приватні анкети, як напр., дослідження Союзу Соціяльної Політики в Німеччині. В більшості випадків вони переслідують практичні завдання, але дають, звичайно, дуже багатий описовий матеріял для характеристики якогось важливого явища громадського життя. Вони не дають повного обчислення ознак, тісно зв'язаних з явищем, але за те мають досить цінних суб'єктивних думок і міркувань від таких осіб, які близько ознайомлені і можуть свідомо критично поставитись до оцінки певного явища, що досліджується. Одержуються такі свідчення як через особисте опитування певних осіб /Англія/, так по завчасно складеній програмі питань, що зобов'язують висловитись та подати відповідну оцінку.

Щодо другого типу анкети, то він характерний тим, що збираються не думки і міркування, а фактичні конкретні свідчення на питання поставлені програмою анкети. Цей тип анкети має характер статистичний, бо досить вичерпливо охоплює явище із значною кількістю більш-менш об'єктивних, фактичних відомостей. Часто буває, що з числа розісланих анкет 60% і більше одержується з відповідями, але все ж таки повного матеріялу анкета не дає. Напр., коли збираємо анкетно відомості про розмір заробітної платні і одержимо відповіді лише від 70% робітників, а серед цього числа виявимо за анкетою I5% робітників з вищею оплатою праці, то не можемо твердити, що ціла маса робітників теж має I5% з вищою платнею; цей відсоток може бути для всіх робітників і трохи більший і трохи менший. При статистичному вичерпливому дослідженні такої непевности бути не може.

Теж саме явище недосконалости /неточности/ спостерігаємо і при одержанню, напр., відомостей від добровільних кореспондентів про характер врожаю хліба, бо кадри їх складаються з більш ретельних, ліпших господарів, через те й відомості дають перевагу трохи вищої оцінки врожаїв, аніж мусіла б бути в дійсності.

Одначе для уяви змін урожаїв за цілий ряд років, або уяви про хитання заробітної платні, такого роду відомості дуже корисні - вони дають можливість без особливої небезпеки зробити відповідні важливі висновки про розмір мінливости певного явища.

Анкетний спосіб дослідження є, таким чином, сурогатом статистичного числення, потреба якого викликана самим життям для необхідного лише загального орієнтування серед явищ, що його вимагають.

В цілях практичних, анкетний спосіб збирання даних, особливо другий його тип, дуже часто вживається різними інституціями і організаціями і за відсутністю вичерпливого статистичного спостереження, безперечно є важливим і корисним засобом об'єктивного освітлення явищ і фактів, що з тих або інших мотивів мусять бути досліджені.

Прикладом анкети, яка має важливе практичне значіння, може служити і анкета, що переводиться Українським Технічно-Господарським Інститутом серед своїх студентів. Для кожного лектора УТГІ відомості подані анкетою мають важливе педагогічне значіння. Вік слухача, характер і обсяг його попередньої освіти, його професія, знання мов і т.п. дають не тільки загальну уяву про цілу масу студентів, але й ще більш цінні - для встановлення відповідного індивідуального зв'язку між студентом і педагогом, дуже важливою при методі позаочного навчання тої або іншої наукової дисципліни.

є/ Репрезентативна метода статистичного числення./Вибіркова метода/. Закінчуючи нариси різних видів статистичного числення мусимо зупинитись на так званій репрезентаитивній /вибірковій/ методі статистичного спостереження.

Суть цієї методи з'ясовується тим, що замість цілої маси одиниць об'єкта, охопленої вичерпливим численням /переписом/, свідомо і пляново числиться лише частина одиниць маси даного об'єкту.

Спосіб і правила переведення такого роду числення залишаються ті ж самі, як і в переведенню вичерпливого статистичного перепису, але обсяг маси одиниць числення значно менший. Напр., замість вичерпливого перепису 500 тисяч господарств певного краю, переписом охоплюються лише 5%, 10%, 20% цих господарств.

Для правильного переведення репрезентативного числення є необхідним такий вибір 5-ти, 10-ти або 20-ти відсотків із цілої маси господарств, який би усував яку будь суб'єктивність з боку дослідника та реєстраторів щодо вибору тих або інших господарств, які підпадають численню. Збереження об'єктивности в такому випадку можливо лише при м е х а н і ч н о м у в и б о р і потрібної суми господарств, себто для 5% кількости господарств - переписується кожне 20-е господарство, яке б воно не було, для десятивідсоткового вибору - кожне десяте господарство і т.д. Навіть переводити в таких випадках вибір по жеребковій системі не є добрим. Лише механічний спосіб вибору одиниць числення відповідає строго теоретичним вимогам. переведення репрезентативного перепису.

Із практики земсько-статистичних переписів можна мати приклад, коли частина прикмет одиниці числення нотувалась для всіх господарств певного краю, а частина прикмет нотувалась репрезентативно і для цих випадків зупинялись лише на таких господарях, які статистикові, що переводив перепис, здавались більш правдивими, поважними та охочими до правдивих відповідів. Коли одначе було перевірено який відсоток складають в цілій масі господарств господарства заможні у вичерпливому переписі і який у репрезентативному, то виявилось, що відсоток заможних, краще ведених господарств, у матеріялі вибіркового перепису був значно більший. Тут виразно виявився вплив суб'єктивности статистиків при виборі певних господарств. Коли для тієї ж місцевости приклали механічний вибір господарств, то виявилось, що відсоток певних господарських ознак серед господарств вибраних механічно, той же самий, як і серед загального числа господарств. Проф. А.А. Чупров зробив з матеріялу земського перепису Московської губерні /1898-1900/, за жеребковим способом вибору, вибір 500 селищ з цілої маси 5220 селищ для порівняння середнього рівня чоловічої грамотности. Виявилось, що для 500 селищ відсоток грамотних чоловіків складав 47,5%, а для всіх селищ - 47,6%. Далі, в числі 500 селищ - 55% сел мали грамотних чоловіків менше половини, а для цілої губерні таких сел було 53%. Частина сел, в яких грамотні складали менш четвертини у вибраних селах дорівнювалася лише 4,2%, а в цілій масі 4,1%. Майже тотожність результатів свідчить про те, яке серйозне значіння може мати практика прикладення репрезентативної методи числення для статистичних досліджень соціяльно-економічних явищ. Звичайне вичерпливе статистичне спостереження є операція дуже складна і коштовна, вимагає значної кількости підготовленого персонального складу та витрати часу. Через те, розуміється, репрезентативна метода, яка надзвичайно скорочує працю і зменшує кошти є цінним придбанням для статистичної практики. Метода ця з теоретичного погляду може вважатись цілком задовольняючою, особливо при додержанню певних правил спостереження, а саме: 1/ прикладати її для таких об'єктів, які своєю структурою є більш менш однородні, 2/ число вибраних одиниць об'єкту мусить бути настільки великим, щоб виявити як не тотожне, то дуже близьке співвідношення важливих прикмет об'єкту у вибраній масі одиниць і в цілій масі одиниць спостереження. 3/ в тих випадках, коли не є необхідним мати абсолютне число всіх одиниць об'єкту, що досліджується і 4/ вибір потрібної кількости одиниць з цілої маси об'єкту переводити механічно. Само собою розуміється, що принцип репрезентативної методи числення може бути прикладений не тільки в стадії переведення перепису, але і для моменту оброблення статистичного матеріялу, здобутого вичерпливим переписом, себто оброблюється не цілий матеріял перепису, а тільки певна механічно вибрана

частина.

Це особливо є важливим для таких наукових студій, метою яких є з'ясування загальних відносин і умов, що відбуваються або існують в явищах.

Вищенаведеного одначе не досить, щоб визнати репрезентативну /вибіркову/ методу статистичного числення, правдивим статистичним спостереженням з погляду строгої методології нашої науки.

Як доказати, що репрезентативна метода теоретично дійсно може заступати систематичне, вичерпливе статистичне спостереження? В яких випадках теоретично виправдується таке заступлення? Яка кількість одиниць об'єкту мусить бути вибрана з цілої маси одиниць об'єкту, щоб забезпечити виявлення постійности відношень і зв'язків через частину явища для цілої маси цього явища? Все це питання, на які статистична метода мусить дати відповідь.

Одначе проблеми, які мусимо розв'язати згідно з поставленими питаннями, тісно зв'язані з викладом теоретичних основ рахування імовірностей та викладом про т. зв. "закон великого числа", що виводиться на підставі теорії рахування імовірностей.

Цю частину /чисто теоретичну/ характеристики репрезентативної методи ми в цілях педагогичних викладемо в другій частині свого курсу лекцій, що буде трактувати про наукове оброблення статистичного матеріялу та номографічне завдання статистичної методи та яке базується на теорії рахування імовірностей і прикладення її до атипових колективних явищ.

§ 17. Поточна регістрація. В кожній державі по ріжних установах провадиться дуже багато поточних записів, що складають предмет статистичної поточної регістрації, вона є необхідна в зв'язку з широко розвиненим державним і суспільним життям та потребами державного і громадського управління. Ця регістрація дотикається різних сторінок життя людности - стану її фізичної культури, духовної культури і матеріяльної. В тих країнах, де збірання і оброблення даних, як основної / переписів/, так і поточної статистики концентрується в одному спеціяльно для того створеному органі - Статистичному Уряді, окремі відділи його займаються переведенням поточної регістрації, розробленням методології такої регістрації та публікацією важливіших даних, потрібних для внутрішнього управління країною та соціяльно-економічної політики.

Щоби мати уяву про обсяг поточної статистики сучасної держави наведемо тут програму діяльности Статистичного Уряду в Чехословацькій Республіці, організованого пізніше ніж у багатьох інших культурних країнах, та за короткий час свого існування зумівшого поставити статистичну працю на високий рівень як з технічного, так і наукового боку. Зазначаємо лише важливіші частини програму поточної регістрації, а саме: статистику п р и р о д н ь о г о р у х у л ю д н о с т и,- яка базується на записах народження, смерти, шлюбів, розводів, розлук і недійсних шлюбів.

Метрикація /записі народжень, смертей і т.д./ ведеться тут громадськими урядами, а обов'язковий огляд спеціяльним офіційним лікарем померщого і дозвіл його на поховання забезпечує можливість вести і статистику причин смертей по науковій клясіфікації хороб. Статистика здоровля подає відомості про захорування людности пошесними хоробами і іншими хоробами та лікування їх різними амбуляторіями, лікарнями й лічничими установами цінний матеріял ця статистика дає і для професійних хороб робітничої кляси.

Щодо статистики д у х о в н о ї к у л ь т у р и, то треба відмітити - с т а т и с т и к у ш к і л і о с в і т и. Сюди належать дані про всі типи шкіл в державі з обчисленням їх лекторського персоналу і учнів, національного характеру і п., а також яслів, притулків і подібних установ шкільного характеру. Нарешті статистики бібліотечна і періодичних друків, а також політичних виборів.

Статистика с у д о в а і к а р н а, інакше м о р а л ь н а статистика. Сюди входять статистика карних злочинів, що ведеться судови-

ми установами цівільними і військовими, як результат судового розгляду справ. На кожного злочинця складається особова картка, куди заносяться відомості про стать, вік, письменність, професію, національність і інше злочинця, а також про зроблений ним злочин та накладену на нього кару. В зв'язку з існуванням інституту умовного засудження подаються відомості і про випадки і характер таких присудів. Ведеться статистика також даних про чинність судових установ, цівільних і карних, а також комерційно-промислових. Періодично публікуються за ряд років великі статистичні збірники про цівільне і криминальне судівництво. Ведеться також статистика конкурсів. Яко спеціяльний відділ існує також с т а т и с т и к а с о ц і я л ь н а - тобто статистика праці. Тут опрацьовується і публікується відомості про працю переведену поза восьмигодиновою, законом установленою працею; про заробітну платню; професійні організації як робітничі так і пілприємців; про страйки і льокаути; про безробіття; статистика посередництва праці; про торг праці; соціяльне забезпечення бідних і недорослих. Ведеться також статистика колективних договорів, працьовної доби, а також чинности різних мирових камер і інституцій. Зокрема ведеться статистика бюджетів урядничих і робітничих родин.

Найбільше різноманітною є статистика матеріяльної культури. Вона складається з поточної регістрації даних сільського господарства, а саме: площ посівів і врожаю різних культур зі спеціяльним відокремленням хмелярства, виноградарства, садівництва, тютюнництва і бахчярства; статистика скотарства і дробу; статистика сільсько-господарського промислу: а/ виробництто, рух виробів та споживання сировин цукроварів, б/ виробництво і споживання пива, спирту, вина, мінеральних вод; в/ статистика виробництва і споживання штучних угноєнь; г/ солодовень і т.п.

Далі йде регістрація даних промислової статистики, а саме: гірництва і гут та виробництва сталі і інших металів, каоліну і магнезіту. Потім статистика водного транспорту, пошт і телеграфів і моторових возів /Статистика залізничого транспорту ведеться Міністерством Залізниць/.

Ведеться також статистика водяних сил і електрізації.

Статистика закордонної торгівлі подається як в місячних, так і в річних збірниках з детальним поділом товарів і країн ввозу і вивозу їх.

Статистика цін товарів ведеться як для гуртовоторговельних так і роздрібноторговельних з публікацією покажчиків цін /індексів/ найбільше чутливих до змін товарів; покажчиків продуктів споживи і всіх видатків на життя типових родин урядовців і робітників, покажчики валютарні і курсів девіз і багато інших важливих коефіцієнтів для студіювання господарської кон'юнктури.

Нарешті в широкому обсязі ведеться регістрація фінансової статистики, себто статистика державних і земських податків. Статистика кредитних установ, починаючи з банків і кінчаючи кооперативами. Статистика грошевого обігу; приватних асекураційних товариств. Зокрема треба зазначити спеціяльні видання по статистиці різних кооперативів, великих міст, дослідження житлових помешкань, чужинецького руху, платебного білянсу і народнього доходу республіки. Як бачимо з вищенаведеного джерела поточної статистики надзвичайно широкі в модерній державі.

Щодо ведення поточної статистики в сучасній Наддніпрянській Україні то обсяг її не менше широкий. Це й цілком зрозуміло, бо існуюча там система соціяльно-економічного будівництва, удержавлення промисловости і колективізація сільського господарства та взагалі плянова державна господарка без широко розвинутої статистики не могла б провадитись зовсім. Управління такою системою вимагає скорішого і безумовно точного та єдиного по змісту і меті обчислення явищ соціяльно-економічного життя. В радянському Союзі одначе не має цілком єдиної державної статистики і хоч Центральному Статистичному Управлінню і підлягають статистичні управління окремих республік та автономних країв, одначе ці останні не позбавлені можливости збирання і оброблення для краєвих потреб. З другого боку в Союзі досить розвинута і відомча/урядницька/ статистика, яку провадять напр., Наркомторг, Наркомфін, Наркомздрав, Наркомвост та інші уряди.

з іншого скрипта

Лекція восьма.

Б. Закон.

I. Значення слова „закон". Слова закон вживають в правознавстві однаково як в розумінні джерела позитивного права, так і в розумінні окремого виду права.

Коли кажемо: „закон встановлює правну норму", - цим підкреслюємо значення закону як джерела права; а це значить, що закон, як і правний звичай, має властивість підносити певне правило (норму) на ступінь позитивного права. В реченні ж: „закон є правна норма, встановлена державною владою" - ми підкреслюємо розуміння закону, як окремого виду права, який мусимо відрізняти од усіх інших видів його.

Щодо самого слова „закон", то дехто з правників (Сергєєвич) виводять це слово від слова „кін" або „кон". Часто прийменник „за", кажуть вони, приростаючи спереду до якогось слова, надає йому значення кордону, межі, краю; напр., загорода, загін, забережжя, задвірок та інш. В слові „закон" показується межу дозволеної людині свободи або чинности: людина не сміє перейти за „кон", себто переступити певну межу в своїх чинах. Отже, закон, мовляв, є той порядок, підлягати якому людина повинна та меж якого вона не може безкарно переходити. В кожному разі слово це є стародавнє загально-славянське слово, що ймовірно існувало вже раніш, аніж усталилося саме поняття закону в нинішньому звичайному його розумінні, себто як писана норма, встановлена державною владою. Означало це слово, очевидно, тогочасну правну норму взагалі, себто звичай (Сергєєвич); так само і слово „lex" за часів середнєвіччя часто вживалося для означення не тільки закону, але й правних звичаїв. Та власне і за наших ще часів у південних славян та на Україні звичайно вживається в народі слова „закон", коли йде мова про правні, і навіть обрядові звичаї (Проф.Ф.Леонтович).+) Розбіжність в змісті слів закон та звичай повстала у нас, очевидно, з того часу, коли наші предки зачали рецепувати візантійське право і коли грецьке слово „νόμος", що означало стверджену імператором, писану правну норму, перекладено було словом закон; за словом же звичай лишається попереднє його значення - неписаної народньої правної норми.

2. Повстання й розвиток закону. Вже раніше ми казали, що, на думку декого з авторів, закон є комбінуванням, історичним поєднанням правного звичаю з урядовим розпорядком (див. стор.66). Розвиваючи цю думку, Hauriou каже, що урядовий розпорядок або реґлямент (l'ordonnance ou le règlement) мав ту перевагу, що зміст його завсігди було вбрано в писаний, більш-менш точний текст, із зазначенням дати (часу) видання його, отже це полекшувало доказ існування та розшукання дійсного змісту реґляментарного права. Крім цього, розпорядок мав тенденцію до застосування

+) „Законне слово" або „законні речі" у нас ще й досі означають обрядові вирази, обов'язкові при народніх, наприклад весільних, звичаях: „Говорили старости законні речі про куницю", або „Поки староста се законне слово казав"... (Квітка-Основ'яненко).

його чинности на теренах цілої держави. Але натомість, розпорядок не мав згоди, чи, вірніше, прилучення до нього з боку людности, або принаймні ніщо не показувало, щоби людність на розпорядок погоджувалася. Він був виявом думки тільки самого монарха чи уряду і тому на народі не лежало, у нього не було внутрішнього, морального обов'язку підлягати нормам того розпорядку; це був основний дефект розпорядку. Розуміється, виконання розпорядку можна забезпечити принуком, силою, але послух силуваний не має тої вартости, що послух добровільний. Навпаки, звичай мав ту перевагу, що він був прийнятий народом, бож сам народ і створив його, і саме це робило звичай джерелом права, гідним переваги над розпорядком, бо право, сприйняте добровільно, є вищим за право, накинене згори. У свою чергу звичай мав свої невигоди, про які вже раніше було говорено: партикуляризм, неусталеність тексту й дати та інш. Отож, ніби фатально мусіло перейти взаємне наближення та злучення цих двох джерел права - звичаю та розпорядку. Злучення це сталося історично, а не логічно; спочатку воно було повільним, а відтак нову форму права, що вважалася за кращу, себто форму закону, продиктували вже самі історичні події.

В тих випадках, коли влада, видаючи розпорядки, що містили в собі правні норми, спиралася на правні звичаї, себто на норми, до яких людність була призвичаєна, справа була ясною, бо ж тоді малося вже прилучення людности до тих норм. Але в який спосіб можна було одержати від неї прилучення, і то в спосіб незабарний, до тих норм, що передбачали встановлення певного порядку на майбутнє, себто до цілком нових норм? Ці нові норми не мали опертя в минулому і відносно них невідомо було, чи пристане на них народ.

Проблему цю розв'язано було задовольняюче лише там, каже Hauriou, де політична організація, в достатній мірі демократична, привела до утворення апарату влади, одну частину якого становив народ або репрезентанти його, а частину другу - уряд, магістратура. Тоді редагування закону випадало як процедура, встановлена між однією частиною влади, що репрезентує магістратуру, та другою частиною тієї ж влади, що репрезентує народ. Між цими двома половинами однієї по своїй суті влади відбувався ніби діялог, підчас якого магістратура запитувала у народу або у його представників, чи приймає він даний закон, а народ відповідав сам безпосередньо (на вічах) або через своїх представників (послів), що він закон той приймає. Закон, у цей спосіб складений, лучив у собі важливіші риси розпорядку й звичаю: він мав писану форму, як розпорядок; він був виявом волі влади, як і розпорядок; але він був акцептований народом, як звичай.

В наведеній процедурі чи, вірніше мовити, в цій навмисно спрощеній схемі повстання закону є, як це визнає й сам автор її (Hauriou), щось від фікції, але, без сумніву, є також дещо й від реальности, від дійсного життя. „Діялог" між урядом та народом на всенародніх зібраннях (вічах), в парляменті чи через референдум (в новіші часи) міг відбуватися в країнах, що мали чи мають більш-менш демократичний устрій. Але, як знаємо, далеко не скрізь і не завсігди цей політичний режим існує. В деспотіях, тираніях, олігархіях, абсолютистичних монархіях і модерних диктатурах згоди у народу чи в його представників звичайно не питають. Йому впрост наказують підлягати велінням-нормам, які визнає за необхідне встановити особа фізична чи особа правна (владнучий колектив), що в даний мент зосереджують у своїх руках силу фізичну, або іноді й силу моральну.

Відомо, що найстародавнішим видом закону був <u>закон божий</u>, що містив у собі норми тогочасної моралі, видані в імени божества й підсилені звичайно авторитетом фізичної сили. А проте закон божий в стародавні часи не обмежувався лише моральною чи чисто релігійною стороною вчення, але часто обіймав різнобічно суспільне й родинне життя, містив кари за злочини, приписи гігієни та інш. Отже це був правний закон тих часів, заснований на авторитеті божества, як авторитет той народові тлумачили служителі культу.

Звичаєве право та „закон божий" становили основу старовинного права

навіть на перших етапах існування держави. Наявність права „законного" є вже ознакою порівнюючи високої культури. Навіть у державах ранньої доби осілої (сільсько-господарської) культури правотворчість не становила функції влади, а сама влада не намагалася зробити законодавство своєю функцією. Функції влади зводилися тоді до забезпечення зовнішнього й внутрішнього миру, а відповідно до того вони обіймали справи військові та судові, почасти ще й фінансові; останні - в міру необхідности роздобуття засобів на утримання війська. Так було в Західній Європі; так було і на Україні за часів Київської держави. Тодішня влада брала на себе роль законодавця лише виїмково, спорадично, під тиском якоїсь означеної необхідности, коли запровадження нових правних норм невідступно вимагало життя. „Предківщина, каже Тарановський, висовувала два питання, що вимагали законодатнього, так мовити, розв'язання, а саме - необхідність припинення самоправства, яке було судовою самопоміччю, аж доки держава не перебрала на себе правосуддя, та необхідність розрішення гострого соціяльного конфлікту, що його було зумовлено рідкістю капіталу та незвичайною дорожнечею кредиту" (себто лихварськими відсотками).

Надалі процес внутрішнього економічного об'єднання окремих країн, що розтягся в Західній Європі аж до XVI-XVII ст. та що припинився на Україні з XIII ст. під тиском несприятливих історичних подій, невідступно вимагав зосередження влади в одному центрі та організації складного державного управління, бо в цей час, під впливом змін в господарському й культурному житті, до сфери державних функцій повільно починає переходити керування задоволенням матеріяльних, культурних і духовних (в тім числі і релігійних) потреб людности. Держава нової історичної доби намагається встановити повну та всебічну опіку над індивідом і над цілим суспільством, не рахуючися з бажаннями цього суспільства та основуючи цілу справу управління на приписі, на наказі згори (так зв. „поліційна держава"). Цілком природно, що таке, майже безкрає розтягнення функцій управління вимагало відтоді з боку держави не тільки зміцнення, але й значного поширення правотворчої функції, бо ж треба було формулювати нові числені обов'язки підданців, визначити завдання й форми нового управління, регляментувати шляхи й засоби для осягнення тих завдань.

Але характерно, що держава тих часів, розвинувши вже широко свою правотворчу чинність, ще довший час наче не насмілювалася називати акти тої чинности законами (leges, lois), залишаючи цей термін для „закону божого", для римського права, для права природнього та взагалі для правних норм, так би мовити, вищої категорії. Законодатні по своїй істоті акти французьких королів з XVI аж до XVIII ст. називалися les ordonnances , себто королівські веління, розпорядки. На Україні це були „князівські устави", на Московщині - „государеви укази". Протягом XVII та XVIII ст. ст. остаточно та цілком свідомо визнається й зміцнюється законодатна функція державної влади. Закон набуває значення зверхньої форми позитивного права, а правотворчість стає монополем державної влади у вигляді законодавства (в формі законів).

Оскільки правні норми встановлював необмежений державець влади - самодержавний монарх, одночасний носій влади законодатної, виконавчої, а іноді й судової, розрізнити закон від розпорядку було дуже тяжко, власне - майже неможливо, бо в цих випадках закон і розпорядок - веління носія зверхньої влади (нім.- verordnung , франц.- ordonnance , російське - указ) мають однакову обов'язкову силу й однаково можуть творити право. Це дає підставу декому з авторів твердити, що правдиву правну конструкцію закону можна було заснувати лише на принципі поділу державних влад або, вірніше мовити, на поділі функцій державної влади на функції законодатну, виконавчу й судову. Тоді окремі державні акти можна розподілити по своєму змісту на закони, адміністративні розпорядки та судові вироки. Закон встановлює правні норми, творить право; судовий орік підводить конкретний випадок під абстрактну правну норму та розв'язує його на основі цієї норми;

адміністративний розпорядок здійснює управління країною. Одночасно з тим кожний з цих актів виконується окремим органом державної влади. Лише тоді дається цілком виразно прокласти межу між правом „законним" та правом „регляментарним",що випливає з розпорядку. В свою чергу право „законне" й право „регляментарне" відтоді одмежовується різко від права звичаєвого, і то не лише тим,що закон та розпорядок є правом писаним,в той час як право звичаєве є право неписане, але й тим,що перші два види права походять від певного одноособового авторитету (особи фізичної або правної),а звичаєве право - від авторитету анонімно-масового. Автори закону й розпорядку є нам,звичайно,відомі; автором правного звичаю є народ, може в кількох поколіннях.

Панування кожного з цих видів права,як це підкреслює Hauriou , відповідає певному,різко виявленому політичному типові суспільства: правний звичай є правом порівнюючи примітивних суспільних зв'язків; розпорядок або декрет - правом імперій (та диктатур - додамо ми); „законне право" - правом державного режиму,побудованого на демократичних підвалинах. А проте,це не означає,як уже раніше бачили, що „законне право" цілковито усунуло з життя право звичаєве й право регляментарне: закон лише взяв над ними гору,має зверхність. Не означає це також і того,що в попередні історичні епохи,принаймні епохи,коли людство винайшло вже писемність, закон був зовсім незнаним; він тільки не грав тої ролі,що почав відогравати в нову добу.

3. Французька доктрина закону. З кінцем XVIII ст. та в початку XIX ст. суспільна опінія сприйняла ідею панування „законного права",інакше кажучи,панування легалізму - з великим натхненням, бо на легалізм,під безпосереднім впливом ідей школи природного права,покладалося в цей час великих,перебільшених надій. Власне,добу легалізму започала і культ закону врочисто проголосила велика французька революція,хоч саме вона якраз і була великим руйником законів.

Популярними,незаперечними в ті часи тезами були такі: джерелом закону є розум або ж закон це є самий розум; закон впроваджує справедливість у взаємовідносини людей; він дає стабільність,стійкість суспільній організації та створює в суспільстві правний порядок. Закон - суверен, зверхній володар, бо ж він,мовляв,все визначає, йому все і всі підлягають. За революційною доктриною,закон має якість непомильности, бо випливає з непомильної загальної волі народу; тезу непомильности загальної волі доводив Жан Жак Русо („Contrat social" liv.IV, chap.II). Отже,законові приписувано було самі лише великі позитивні якості.

Декларація прав людини й громадянина 1789 р.,яку Georges Renard називає хартією публічних свобід, в цих словах визначила джерело походження та характер закону (артикул 6): „ З а к о н є в и я в о м з а г а л ь н о ї в о л і. В с і г р о м а д я н е м а ю т ь п р а в о б р а т и у ч а с т ь у т в о р е н н і з а к о н у о с о б и с т о а б о ч е р е з с в о ї х п р е д с т а в н и к і в. З а к о н м у с и т ь б у т и о д н а к о в и м д л я в с і х, ч и в і н о х о р о н я є, ч и в і н к а р а є ".

Теоретично звідси випливало передовсім те, що закон походить від волі народу; вияв цієї волі є зверхнім актом державної влади, є найвищим законом. Революційні чинники визнають творення закону за виключну прерогативу народнього представництва, як виразника загальної волі, а р.1793, під впливом вчення Руссо,серед них повстає навіть ідея необхідности народнього голосування, принаймні конституції та найважливіших законів (9/VIII 1793 було переведено перше голосування конституції). Разом із тим Конституанта (установчі збори) проголошує в 1789 та 1791 р.р., що „виконавча влада не сміє видавати жадного закону,навіть тимчасового;во-

на може видавати лише прокламації, які або наказують виконати закон, або закликують до його виконання". Отже, виявляється яскрава тенденція визнати за народнім представництвом, виразником народньої волі, монопольне право на видання правних норм, зробити його єдиним правотворчим органом. Одночасно робляться спроби перевести в державній системі повний розподіл державних функцій поміж владами законодатною, виконавчою й судовою.

Щоправда, революція кінця XVIII ст. недовго втрималася на цій принциповій позиції, бо Конвент (парлямент) вже з р.1793 зосереджує в своїх руках владу законодавчу й виконавчу та одночасно надає Комітетові Народнього Спасіння, а навіть окремим послам, що їх він відряджує на провінцію з певними дорученнями, право видавати постанови, які мали б силу „тимчасових законів". Таким чином різні функції влади знову зосереджуються в одному органі (Конвенті) та одночасно законодатна функція виконується цілою низкою державних органів: народом (при всенародньому голосуванні законів), Конвентом, Комітетом Народнього Спасіння та поодинокими депутатами-комісарами.

А проте принцип розподілу державних функцій поміж різними владами згодом і повільно вщіплюється в державне життя Франції, а законодатна функція стає виключною функцією народнього представництва. Протягом часу виявилися вади й незручності закону, але всеж залишається, як каже Hauriou „враження, що в основному знайдено було правну рівновагу".

Надалі традиційна французька доктрина, виходячи з конституцій Франції часів Великої революції та спираючись лише на формальний момент у визначенні закону, завсігди розуміла під поняттям закону правні норми, що походять від законодатних органів влади, себто визнала, що законом є норми, встановлені за безпосередньою участю народнього представництва, незалежно від змісту тих норм. Визначення того, які саме акти мають видаватися в законодатній формі, які питання повинно ставити на розв'язання законодавця, а значить, яку ухвалу зробити законом, тут цілком залежить від самого законодавця - парляменту.

Отже, за цією доктриною, законом є акт, що походить від органу, який визнається за виразника суверенної волі громадянства, колективу (Duguit). Разом з тим лише законові є властивим висловлювати суверенну волю держави (F.Moreau), а тому всі інші акти влади стоять супроти закону на нижчому щаблі та мусять підлягати йому; вони суть акти підзаконні.

Однак французька доктрина пореволюційних часів перестала вважати самий закон за єдине джерело права, а народне представництво за єдиного творця правних норм. За джерело права вона визнала також і розпорядок (reglement) виконавчої влади, але право видання регляментарних актів, як актів, що теж можуть творити право, французький законодавець зарезервував лише певній категорії органів влади - президентові республіки, префектам та мерам. Видавати подібні правотворчі норми дозволяє їм в певних наперед закреслених межах відповідний закон або рівнозначний правний звичай (G.Renard). Отже, правотворча чинність органів виконавчої влади тут пояснюється лише тим, що це право їм надано, як спеціяльна делегація, припорука з боку законодавця; самостійної правотворчої ролі державному управлінню французька доктрина не узнає. Згадуваний раніше Моро (F.Moreau) каже, що за традиційною французькою доктриною, істотність закону полягає в запровадженні принципів, а істотність розпорядку - в запровадженні деталів.

Таку загальну підставу для розуміння закону дала французька революція, поставивши на чолі цілого права закон, а на чолі державного управління - представників нової для того часу соціяльної сили - буржуазії.

4. Німецькі теорії закону. Теорія Шталя. Але докладніше розробити теорію закону довелося, однак, не французьким правникам, а державникам

і філософам німецьким. І це тому, що німецька буржуазія не спромоглася аж до революції 1918 р. остаточно скинути надвладу шляхти (юнкерства) і решток попереднього абсолютизму, як це зробила французька революція 1789 р. В конституційному німецькому праві, як і в німецькому правному житті, залишилося багато пережитків феодалізму й абсолютизму, а це примушувало німецьких теоретиків права будувати теорію підзаконної, правової держави, в якій цю надвладу повинен був би мати закон. Цілий час німецькій правничій думці доводилося мати діло з проблемою закону та надавати йому, принаймні теоретично, зверхньої сили, якій повинні були б коритися реальні державні сили та яка мала би корегувати дефекти існуючої в даний період державної системи. Німецька доктрина закону складалася на ґрунті німецької конституційної практики та еволюціонувала рівнобіжно зі зміною конституційного права. Під безпосереднім ідейно-науковим впливом німецької правничої школи перебували правники-теоретики двох інших колишніх європейських держав: Австро-Угорщини та Росії, державний устрій яких теж вимагав від представників науки права шукати в законі зброї проти надуживання владою з боку пануючих в державі елементів (ліберальна течія в науці права), або ж, навпаки, шукати в ньому підстав для зміцнення існуючого стану річей (консервативна течія). Взагалі ж треба сказати, що протягом XIX століття німецька правнича наука висунула цілий шерег талановитих вчених-правників (Савіньї, Пухта, Шталь, Л. Штайн, Гнайст, Момзен, Моль, Лабанд, Ієрінґ, Гачек, Єллінек та інш.), праці яких скрізь в Європі становили основу науки права. Впливів німецької науки права не уникла нарешті й французька правнича думка.

В першій половині XIX ст. великий вплив на побудову теорії закону мав Fr. I. Stahl („Die Philosophie des Rechts"). Шталь в своєму вченні про закон виходив з таких положень. Державне панування містить в собі два елементи - державну зверхність чи державну владу (imperium) та закон (lex). Державна влада є зверхньою, суверенною владою тому, що вона ні від кого незалежна. Одним із виявів цієї влади, і то виявом найважливішим, першим по силі, є законодавство, себто творення принципів права, загальних правних норм. Законодавство, як вияв суверенної волі, може здійснюватися тільки носієм цієї волі, а ним є в республіках народне представництво, а в монархіях - монарх. Але і в монархіях Шталь узнає необхідність участи в законодавстві різних суспільних елементів, бо від суверена закон набуває лише формальну важність, а матеріяльним джерелом і осідком закону є, на його думку, народня правосвідомість (про це раніше вчили Савіньї - Пухта). Тому законодавство треба узнати за справу не самого лише монарха, але й за справу цілонародню і на видання закону необхідно мати згоду від народу. Закони, як загальні правні норми, не можуть та й не мусять охоплювати цілу сферу державного управління. Мусить залишатися широке поле для регляментації чинности влади в царині управління урядовими розпорядками. За Шталем, правила, що призначені регулювати промисловість, сільське господарство, фінанси, поліцію та справи військові, мають творитися розпорядками влади, бо ці правила не встановлюють, мовляв, загальних правних норм, а вказують лише засоби, якими буде осягнено певних конкретних завдань. Отже, це не є норми об'єктивного права, а лише способи управління. До законодатньої компетенції народнього представництва слід віднести, на Шталеву думку, лише закони, що торкаються конституції та правосуддя. Видання розпорядків, якими вдійснюється управління, має становити самостійну чинність виконавчої влади, без втручання в цю справу народнього представництва.

Отже, теорія Шталя, що довший час помітно впливала на конституційне право в низці німецьких держав та одночасно й на конструкцію теорій закону у пізніших теоретиків права, визнавала за с д и н е д ж е р е л о права з а к о н, якому протиставився р о з п о р я д о к, як а к т н е п р а в о т в о р ч и й. Але тому, що розпорядкові було надано самостійного, незалежного від народнього представництва, значення, а головно тому, що до компетенції розпорядку чи, вірніше мовити, до самостійної компетенції виконавчої влади віднесено нормування важливіших галузів життя, са-

ме через це значення закону було обкраяно, принаймні що до кругу його чинности, звужено, так би мовити, в поземому перетині, бо ж очевидно, що в тих галузях, які Шталь назначив для самостійної чинности розпорядку, правні норми фактично творилися тим самим розпорядком. Припускають, що це з'явилося у Шталя в наслідок його бажання зміцнити позиції німецького абсолютизму, захитаного впливами французької революції, щоби бодай в цей спосіб зберегти для королівських і князівських влад широке поле для автократичного, незалежного від законодатних органів, державного управління. Але, здається, вірніше буде припустити, що, будуючи свою теорію закону й розпорядку, Шталь лише совісно зафіксував у ній тогочасну правну дійсність, констатував те, що існувало в тодішньому державному житті німецьких земель.

Треба сказати, що питання про взаємовідношення поміж законом та розпорядком, що його підніс Шталь в такий рішучий спосіб, відтоді не сходило та й досі не сходить з поля зору науки права. Довкола цього питання точаться постійні ідеологічні суперечки, що мають також велике практичне значення, бо від того чи іншого розв'язання питання в значній мірі залежить ціла система державної організації в окремих країнах або, точніше кажучи, від того залежить значення органів влади законодатної та виконавчої, питома вага кожної з них.

5. Теорія закону Лоренца Штайна. Дальшим етапом у розвитку німецької доктрини закону (та одночасно й розпорядку) треба визнати теорію найвидатнішого німецького правника Лоренца Штайна (Lorenz von Stein). Головну точку його вчення про закон становить теза, що законодатна влада є носієм з а г а л ь н о ї в о л і, яка складається з воль окремих індивідів. Істотність законодатньої влади полягає в тому, щоби „воліти", але зовсім не в тім, щоби „діяти". Виконувати в діях волю законодавця має державне управління. Отже закони - це воля законодавця, а управління - це чин, спрямований в першу чергу на виконання законів, себто волі законодавця. Але закон не може визначити всі функції й чини управління. Оскільки є прогалини, вони мають виповнюватись самостійно владою, що управляє державою. Тому ця влада мусить мати й свою волю, самостійну волю управління. Волю управління влада виявляє в розпорядках, що або приписують конкретні заходи для переведення вимог закону в життя (розпорядки виконавчі), або доповнюють існуючі неповні, неясні чи спірні закони, або, нарешті, творять цілком нові правні норми, не визначені в законах. Остання категорія розпорядків по істоті своїй є т и м ч а с о в и м и з а к о н а м и, які й мають бути на Штайнову думку, негайно по скликанні законодатних установ подані на їхню ухвалу. Взагалі ж Штайн узнає розпорядкам державної влади правотворчу силу, себто узнає, що розпорядки (за винятком розпорядків виконавчих) творять, як і закони, норми об'єктивного права. Це правотворчу якість розпорядків негували, як знаємо, французька доктрина закону часів революції та теорія Шталя. Супроти закону розпорядок, за Штайном, має бути на становищі підрядному, підлеглому, бо закон є дужчий за розпорядок; він, як загальний правний принцип, веде за собою розпорядок, що має конкретну, обмежену ціль.

6. Теорія закону Лабанда-Еллінека. Дальший помітний крок в розвитку теорії закону робить вчення Л а б а н д а, відтак розвинене й історично та філософічно обґрунтоване Е л л і н е к о м (кінець XIX стол.). Теорія Лабанда-Еллінека, з деякими змінами в ній, є пануючою в правознавстві й за наших часів. За цією теорією, слово „закон" має в правознавстві подвійне значення: м а т е р і я л ь н е та ф о р м а л ь н е (дуалістична теорія). Власне, різницю поміж законом у формальному розумінні та законом у розумінні матеріяльному вперше зформулю-

вав Ґнайст, але з того він не зробив відповідних висновків. У вченні Лабанда-Єллінека доводиться, що законом у розумінні матеріяльному треба вважати передусім загальний правний принцип, що встановлює право й обов'язок та набирає обов'язкової сили з наказу органів державної влади. Абстрактна правна норма чи інакше загальне правне правило реґулює звичайно велику кількість однакових випадків, але іноді правну норму встановлюється і для якоїсь індивідуальної події, на окремий випадок. Правне правило (норма) становить зміст закону, себто його матеріяльну сторону. Але, крім правного правила, закон невідмінно містить в собі також наказ, веління державної влади (про це говорено було раніше, див. стор. 44). Отже, в законі мусимо розрізняти два окремі, йому властиві елементи: закон-зміст і закон-наказ. Коли б закон не містив у собі наказу, закон-зміст не мав би обов'язкової сили. Цими двома елементами й визначається поняття закону в матеріяльному розумінні. Отже, за Лабандом, законом у матеріяльному розумінні треба визнати кождий вияв (наказ) державної волі, змістом якого є правна норма, незалежно від того, який орган держави та в якій формі його видав.

Єллінек формулює це саме коротше, кажучи, що закон матеріяльний це є веління влади, що містить у собі правну норму. Більш того - для Єллінека, як і для діячів великої французької революції, „закон є загальна воля".

Але сама теорія матеріяльної конструкції закону не може, на думку обох цих авторів, розв'язати питання про те, до якої категорії актів слід віднести той чи інший вияв (наказ) державної волі, і тому обов'язково потребує доповнення собі у вигляді формальної теорії закону. Лабанд вчить, що закон набирає своєї сили з ментом його санкції, себто затвердження законопроєкту відповідним державним органом, який то акт і перетворює проєкт закону в норму позитивного права. Тому що право санкції+) в конституційних державах належить монархові, Лабанд твердив, що й силу свою закон набирає від самого монарха, а парлямент лише визначає зміст закону, бо ж парлямент, мовляв, є покликаний лише для того, щоби виявляти народню волю і тому він не може діяти, а значить, не може й наказувати. Єллінек зм'якшує це категоричне твердження Лабанда, доводячи, що в конституційних державах закон твориться спільним актом волі монарха та парляменту, які, як це стверджує й офіційна англійська теорія, „в їхній єдності визнаються за джерело зверхньої волі держави". В демократичних республіках закон, схвалений законодатними зборами, визнається за відповідаючий змістові народньої, себто зверхньої державної волі, так ніби ухвала закону вийшла від самого народу. Отже законом у формальному розумінні буде те, що схвалить відповідний законодатний орган (парлямент) чи система цих органів (парлямент, сенат, монарх). Змістом закону у формальному розумінні цього слова, себто акту, прийнятого в описаний допіру спосіб, можуть бути не самі лише правні норми, як то встановлює поняття закону в розумінні матеріяльному, але й адміністративні, розпорядчі акти з приводу окремих випадків, та навіть судові вироки. В Англії, наприклад, парлямент є не лише законодавцем, але - шляхом так званих приватних біллів (розпоряджень) він бере безпосередню участь в державному управлінні; там же палата лордів одночасно є також зверхньою судовою установою Британської імперії. І приватні біллі нижньої палати, і судові вироки палати лордів з формального погляду суть закони, хоч ними й не встановлюється жадної нової правної норми.

+) Термін санкція раніше ми вживали в двох розуміннях: поперше для означення невигідних наслідків для порушника правної норми; подруге для означення (технічного) другої частини норми карного права. Тут слово санкція вживається в розумінні владного розказу, яким остаточно стверджується закон.

Лекція дев'ята.

в/ Факторіяльні і функціональні ознаки групування матеріялу./продовження./. Трудність вибору факторіяльних і функціональних ознак, які мусимо в статистичному досліді ставити у певний зв'язок, випливає з множности цих ознак в кожному атиповому колективному явищі, Напр., статистичні дані ясно малюють нам бідний стан якоїсь окремої групи селян - волости, гміни, району чи якоїсь національної групи, але ми не можемо легко вирішити чи цей стан є наслідком малоземелля, чи слабого робочого складу господарств, чи недостачою ремаенту, відсутности промислових заробітків, некультурности, або він є наслідком ще якихсь інших умов.

Виділити кожну умову з цілого довгого ряду умов не є можливим; статистична метода практично дозволяє ізолювати, з цілого великого ряду умов, лише небагато основних факторів або причин; відносно решти поневолі доводиться припускати, що вони знаходяться в стані взаємного урівноваження в масі.

Таким чином, суть статистичного "експерименту" зводиться до наступного: поперше, треба взяти ті групи, які, в межах переведеного в таблиці розмежування, різняться тільки однією ознакою - тою-самою, вплив якої хочемо дослідити, а по решті інших ознак, покладених в основу комбінаційного групування, ці групи однакові, а далі вже вияснити залежність ознак - ф у н к ц і й, що характеризують відібрані групи від змін основного фактору, вплив якого досліджується.

Щоб уявити ясніше процес статистичної аналізи щодо виявлення ролі факторів, наведемо тут працю пророблену студентами одного з семинарів під керовництвом проф.А.А.Кауфмана. Поставлено було завдання про залежність письменности й шкільного навчання від забезпечення землею, національної приналежности та віддалення сел від шкіл в Олександрівському повіті на Катеринославщині. Матеріялом були дані, взяті з поселкових таблиць, на підставі яких складена була комбінаційна таблиця за вищенаведеними трьома ознаками. З цієї комбінаційної таблиці зроблено групові таблиці, що дали наступні цифри:

	Письменних на 100 душ населення віком старшого 7 років.		Учнів на 100 дітей шкільного віку.	
	Чоловіків	Жінок	Хлопчиків	Дівчат
А. Групування по землеволодінню.				
До 5 десятин	23	4,0	31	7,0
5 - 10 дес.	29	1,5	28	4,6
10 - 25 дес.	33	2,9	33	6,0
25 - 50 дес.	40	20	61	58
більше 50 дес.	49	21	44	31
Б. Групування по національности.				
Українці	30	2,1	29	5,3
Німці-колоністи	45	27	70	74
Жиди-колоністи	49	10	66	12
В. Групування по віддаленню від школи.				
До 1-ої версти	31	2,9	32	7,0
1 - 3 "	21	2,1	13	2,3
Більше 3-х верст	15	1,5	13,4	2,5

Дані свідчать, що залежність взагалі до певної міри виявляється по відношенню до всіх трьох обраних факторіяльних ознак. Але німці майже

всі належать лише до великоземельних, а жиди до середньоземельних груп; і ті й другі при тому майже всі мають школу в себе в селі, тоді як українські села і в тому і в другому відношенні знаходяться в занадто різних умовах. Виникає питання: чи не через те вищі земельні групи і групи зі школою "до I версти" дають таку різку відміну із збільшенням догори, що це будуть переважно німці й жиди? Або, навпаки, - чи не через те зосібна високий відсоток письменних і учнів у німців і жидів, що вони добре забезпечені землею і при тому мають школи в себе в селах? Щоб виказати вплив земельного забезпечення незалежно від національного фактора, досить навести результати групування по землі в українців:

Земельні групи	Відсоток письменних чоловіків	жінок	Відсоток учнів хлопців	дівчат
До 5 десятин	22	3	29	4,4
5 - I0 дес.	28	2	28	4,6
I0 - 25 дес.	32	2	30	6
25 - 50 дес.	36	9	44	2I
вище 50 дес.	45	II	27	I3

Залежність від земельного забезпечення і тут виявляється досить чітко. Щоб виявити в чистому виді вплив національної ознаки, порівняємо українців і німців у вищих земельних групах

Земельні і національні групи:		Відсоток письменних Чоловіків	Жінок	відсоток учнів Хлопців	Дівчат
мають від 25 - 50 дес.	українці	36	I0	47	2I
	німці	43	28	72	94
мають більше 50 дес.	українці	45	II	27	I4
	німці	52	32	75	5I

Ці дані поперше підтверджують, що земельне забезпечення має свій вплив серед кожної з національних груп, а потім показують, що вплив національної ознаки виявляється і незалежно від різниці в земельному забезпеченні, при тому різниця по відношенню до письменности серед однакових земельних груп виявляється значно слабше. Щоб ізолювати вплив національної ознаки від ознаки забезпечення школою, беремо для всіх трьох національних груп лише одні села зі школою, "до I-ої версти":

національні групи	відсоток письменних чоловіків	жінок	Відсоток учнів хлопців	дівчат
Українці	30	2I	47	2I
Німці	45	27	70	75
Жиди	49	I0	66	I2

Бачимо, що при однаковім віддаленні школи різниця між національними групами залишається щодо письменности такою самою різкою, але помітно меншою щодо шкільного навчання.

Статистична практика одначе вчить, що комбінаційна таблиця, коли вона досить складна /напр., 4 ознаки/, часто не дає очікуваних наслідків - не виявляє тих складних залежностей, для розкриття яких вона складена. Себто, вона не дозволяє ізолювати вплив кожної з покладених в основу таблиці факторіяльних ознак. Це тому, що в комбінаційній таблиці матеріял роздрібнюється між сотнями груп і ці останні кількісно настільки малі, що випадкові причини, що в них діють, перестають взаємно урівноважуватись, а через те закривають чинність постійних причин явища. Тому й слід в таких випадках переходити від комбінаційної таблиці з більшим числом факторів, до помічних таблиць групових /по одній ознаці/, або до комбінаційних з меншим числом ознак.

Щодо оперування з функціональними ознаками, що відповідають змісту присудку статистичного речення та характеризують спостережене явище, як в цілому, так і по окремих природних або штучно-створених групах - то операція ця складається з двох основних форм. П е р ш а - це підсумковий облік квантитативних ознак, що характеризують окремі випадки або індивіди /екземпляри/ масового явища. Д р у г а - це груповий облік, себто облік числа випадків або одиниць класифікованих за тими або іншими статистично почисленими ознаками.

До цих основних форм треба додати ще т р е т ю - просте сумування екземплярів або випадків, що складають цілу статистичну сукупність та на якому ми вже вище зупинялись.

Напр., треба знати скільки всього людей, голів кожного виду худоби, землі по кожному виду дібр, засівів по кожному виду хлібів і т.п.мають селянські двори - господарства і відсіль вивести пересічний склад сем'ї, забезпеченність худобою, землею, засівами і т.д. у підсумках по селах, волостях і по цілому повіту. Це й буде підсумковий облік. Але треба також знати, як розподіляються двори господарства в повіті, волостях і селах щодо етнографічного складу, характеру промислової чинности і т.д.; ще більше того, як розподіляється населення щодо забезпечення робочою силою /двори без робітників, з одним робітником, з двома і т.д./,робочим скотом /без коней, з одним конем, двома і т.д./, землею, розміром засівної площі і т.п.; для цього всі двори - господарства в селі, волости, повіті - класифікуються по кожній знаведених вище ознак на групи і підраховується число дворів, віднесених до кожної такої групи. Це й буде груповий облік.

Розуміється, що підсумковий облік є можливий тільки для таких ознак, які мають квантитативний /кількісний/ вираз. Групове оброблення і обчислення даних можливе як для якісних ознак, так і для ознак кількісних. Можемо напр., групувати і числити населення по ознаках статі, письменности, віровизнанню, національности, або промислові підприємства по кількости робітників, по силі механічних двигунів; робітників - по розміру заробітної платні; кватири по числу кімнат і т.п.

г/ Плян зведення матеріялу. Вищенаведене свідчить про те, що зведення статистичного матеріялу вимагає певного пляну. Треба памятати тут, що основний первісний матеріял статистичного числення не є приступний для зацікавлених осіб громадянства, вони можуть користуватися тільки тим, що зведено й опубліковано в таблицях. Плян зведення через те мусить передбачати різноманітні потреби осіб, що зацікавлені в використанні статистичних даних. Кожне основне статистичне числення є важливим і для державного діяча - фінансиста, законодавця, і для вченого - соціолога, економіста, біолога і т.д. і через те розроблення пляну зведення вимагає знайомства з тими науками, або царинами практичної політики, для яких можуть бути потрібні результати певного статистичного дослідження. Необхідно також добре знайомство з об'єктом дослідження, із зібраними матеріялами, поскільки лише таке знайомство поможе вирішити, які саме дані заслуговують підрахунку та який напрям такого обчислення може дати найбільше повні і цінні результати.

ґ/ Централізоване і децентралізоване зведення матеріялу. Щодо техніки переведення табличного зведення матеріялу, то треба розрізнювати в цій операції два способи: I/ Децентралізоване зведення, коли воно виконується тими особами і установами, які переводять статистичне числення /реґістрацію/ і 2/ Централізоване зведення - коли реґіструючі об'єкт аґенти або місцеві установи передають сировий матеріял для оброблення у відповідну центральну установу.

В першому випадку нижчий реґіструючий орган сам підраховує зібраний ним матеріял по формах надісланих із центру, при чому підрахунки нижчого аґента йдуть далі до одної або двох інстанцій, кожна з яких робить підсумки для свого більшого району, обличення ж загальних підсумків виконується вже центральною установою. Напр., в Чехословаччині при пере-

писах худоби першою інстанцією, що зводить матеріял є старости сел; відсіль матеріял поступає до "окресного"/округи/ уряду і підраховується в межах території цього уряду, а потім вже надсилається до статистичного уряду в Празі. У Франції спочатку матеріял перепису населення підраховується мерами, а потім поступає до префектур і лише звідсіль йде вже до Центрального Статистичного Уряду.

В другому випадку нижчі органи надсилають матеріял без підсумків просто центральній установі. Напр., митниці можуть надсилати Центр.Стат. Уряду індивідуальні нотації про пройдені через кордон товари без якого будь зведення, або судові установи про дані своєї чинности та дані про переступи і злочини - безпосередньо в Статист.Уряд або М-во Справедливости.

Яка ж із цих систем зведення матеріялу є більш доцільною? Довший досвід і теоретичні міркування в сучасний мент дають статистичній науці підстави висловитись за централізоване зведення матеріялу. По словах німецького статистика Ґеорґа М а й р а ця перевага випливає з таких міркувань: "І. Спостерігаючі органи звільняються від обтяжної малоцікавої для них і через те не доброї щодо виконання рахункової праці. 2/ Ця праця переноситься на орган зацікавлений в справі. 3/ Матеріял обробляється одноманітно, зокрема в тих випадках, коли вимагається підсумування окремих випадків в узагальнюючі розділи. 4. Можлива організація досконалого, безперерви дієвого контролю. 5. Можливе використання ліпших способів зведення, зокрема зведення електричними і іншими машинами, що вимагає лише централізованої праці. 6. Матеріял може бути якнайповніше використаний шляхом складення банатих по змісту комбінаційних таблиць, складення яких при децентралізації зустрічає непереможні труднощі.

Взагалі треба сказати, що при централізованому зведенню матеріялу головний орган не є в'язаний початковими підрахунками, він оперує із сировим матеріялом, який групує й клясифікує в міру потреби і як схоче. Треба мати також на увазі, що цей спосіб зведення матеріялу також досягає більшої економії часу і грошових витрат.Мусимо одначе вказати і на деякі негативні риси централізованого зведення матеріялу. Загальні підсумки важливого статистичного числення чи то буде перепис населення, чи сільсько-господарський, або промисловий дуже важливо одержати яко мога швидче, в цілях практичного використання таких підсумків або органами місцевого самоврядування, або іншими важливими інституціями громадського значіння. При децентралізованому способі зведення, такі підсумки і можуть бути скоріше одержані і використані на місцях. Центральна установа може приступити до операції зведення лише після досконалого контролю засланого матеріялу, а інколи і переведення необхідних доповнень або виправлень даних на місцях, а ці вимоги до рахункової операції потрібують чимало часу. Цим затримується і публікація центральною установою даних спостереження. Щоб запобігти цьому сучасні статистичні уряди в багатьох країнах опубліковують часто лише "попередні підсумки" важливих даних переписів,з допущенням необхідних в такому випадку помилок і тим скорше задовольняють потребу в одержанню важливих корисних даних. Лише через деякий довший час громадянство може мати вже публікацію виправлених і повних даних нормального зведення матеріялу певного статистичного числення.

Другою негативною рисою централізованого зведення матеріялу може бути /але не завжди буває/ не добре знання установою, власне її персоналом місцевих особливостей, місцевих умов окремих частин держави або країни, особливо тоді, коли така країна щодо свого національного складу та соціяльно-економічного побуту є дуже різноманітною. Це стосується, напр., до такої країни, як колишня Росія або сучасний Радянський Союз, де не тільки багато територій з окремими національностями, але й територій з таким різким устроєм господарського життя, як Україна і Киргизький степ з кочовою господаркою. Отже виправлення і контроля матеріялу для центральних статистичних установ у таких країнах є особливо тяжкою, без участи осіб, що знають добре і місцеву мову і соціяльно-економічні відмінності та взагалі своєрідність побуту місцевого населення. Досвід розроблення

центральними установами матеріялу переписів в подібних країнах дає багато прикладів важливих утруднень та дефектів, що безперечно зв'язані з централізованою системою зведення первісних даних статистичного числення.

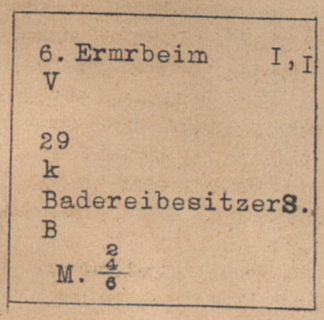
д/ Техніка рахування в статистичному зведенню матеріялу. Для операції рахування відповідних даних статистичного спостереження є необхідним первісний матеріял привести в такий вигляд, щоб числове зведення його переводити механічно, скоро і одноманітно по різних групованнях. В зведенні матеріялу бере участь велика кількість осіб. Праця їх мусить бути простим механічним рахуванням зовсім ясних по своєму значінню поодиноких свідчень, які б не вимагали від робітника яких будь міркувань, сумнівів та давали можність групувати їх однаково в процесі рахування. Проте досягти такого механічного процесу рахування можна лише тоді, коли основний матеріял записів буде відповідно підготовлений. Напр., для обліку населення по віку, що в основному формулярі нотується роками і днем народження треба перевести ці дані на певну кількість цілих "прожитих років". Так само основні записи про письменність або фізичні дефекти, а тим більше професій, причини смертности і т.п. треба підвести під точно встановлені категорії, або клясифікації на групи або замінити їх відповідними умовними знаками і тільки тоді рахувати. Напр., $\frac{133}{2}\frac{в}{-}$ може означати особу, професія якої відповідає числу 133 по словнику професій, літера "в" - положення в підприємстві - робітник, цифра в знаменнику - відношення до голови сем'ї - член сем'ї і т.п. Таке р о з м і ч е н н я основного матеріялу і треба проробити.

Дуже влучно була напр., побудована картка /фішка/ Георгом Майром при розробленні баварського перепису. Поперше, він для різного родинного стану населення вживав картки різної фарби: для нежонатих - білої фарби, дівчат - блакітної, для жонатих - жовтої, заміжніх - червоної, вдівців - сірої, вдів - гнідої, розведених чоловіків - зеленої і розведених жінок - жовтогарячої. Ознаки на картці визначались символами, умовними літерами, а саме:

Літери і знаки:		Їх значіння
V	-	голова господарства
f	-	дружина
s	-	син
t	-	дочка
vw	-	родич
d	-	прислуга
pf	-	на утриманню
n	-	ночувальник
b	-	гість
a	-	інше положення в господарстві
√	-	родився в томуж селі.

Літери і знаки:		Їх значіння
√ √	-	родився в другому селі тої самої округи
k	-	католик
rk	-	римо-католик
ak	-	старо-католик
p	-	протестант
ev	-	евангелицького
ref	-	реформатського
is	-	ізраїльського
і т.д.		
S	-	самостійний
B	-	власник
G	-	підмайстер
D	-	робітник
і т.д.		

Сама фішка мала такий вигляд:



Картка читається так. Жонатий, бо вона жовтої фарби, голова господарства, родився в своєму селі у 1829 році, віровизнання католицького, занятя його - господар купелів, самостійний, власник, родина складається з 2-х чоловіків і 4-х жінок, разом 6 душ.

В такий системі дуже важливо не сплутати значіння фарби картки і через те треба звернути особливу увагу

на індивідуальні властивості осіб, що опрацьовують матеріяли. Сама переписка з формулярів на картки має свою вигоду в тому, що кожний формуляр з поодинокими свідченнями уважно перечитується та кожний сумнів може бути розв'язаний керовником бюра раніш ніж свідчення буде занесене на картку. Її доручається ліпшим, відповідальним робітникам статистичної установи, а після такого підготовлення рахування стає вже механічним. Обличення переводиться або вживанням карбів /рис/ чи точок, або рахункових карток чи фішок; нарешті може переводиться і електричними машинами.

Для карбування /чи точкування/ заготовляються окремі рахункові таблиці, при чому графи цих таблиць компонуються не так, як в остаточній таблиці, де графи йдуть в порядку їх внутрішнього змісту і логічного зв'язку, а в порядку послідовности підрахунку, який дає також і найбільшу економію місця. Кожне одиноке свідчення зазначається в певній клітині таблиці точком або карбом, а щоб улегшити кінцевий підрахунок, точкування і карбування ведуть групами, напр., по пять точок або карбів, як це наведено в наступній таблиці./див. на стор. 73/.

Розуміється, що така праця вимагає великої уваги та акуратности. Контроля розмічення під час праці неможлива, перевірка може бути зроблена лише повторенням цілої операції. Самий спосіб вживається лише для групового підрахування, а не до переведення загальних числових підсумків, яке робиться вже рахівницею або на аритмометрі.

Щодо способу "фішок" і карток, то це є вживання спеціяльних листків, з яких кожний відповідає поодинокому екземпляру об'єкту дослідження. - особі в перепису населення, селянському двору в сільсько-господарському перепису, кватирі /перепис помешкань/, злочинцю і т.п. Операція обличення переводиться тут через розкладення карток /фішок/ по групах або купках, так, що до кожної з них кладуться картки з певною ознакою, або комбінацією ознак /чоловіки; глухонімі; або ковалі-господарі; чоловіки-глухонімі; безземельні-без коней/, а потім кожна група підраховується. Вживання карток підрахункових є більш модерний спосіб, бо при ньому менше робиться помилок, ніж при карбуванню. Цей спосіб також забезпечує більшу економію часу на операцію рахування, та видатків на її переведення.

Саме розкладення карток на групи треба тримати /не розкидувати/ доти, доки ознака, по якій ця група зроблена, буде вичерпана за всіма комбінаціями з іншими ознаками. Нарешті, зведення матеріялу електричними машинами робиться для загальних переписів, де рахуються сотні тисяч одиниць з комбінаціями ознак, або взагалі для матеріялу, що вимагає великих рахункових операцій.

Для такої машини треба всі необхідні ознаки перенести на невелику картку, подібну до нижче наведеної, що служить для підрахунку -"Вивозу" в Статистичному Уряді в Чехословаччині, для статистики закордонної торгівлі /Див.стор.74/.

Подібно до того, як у картковому /фішковому/ способі обличення, індивідуальні дані мусять бути перенесені з формуляра на картку /фішку/, так і при рахуванню машиною, напр., Powers'a[x], треба індивідуальні дані необхідні для підрахунку, зазначити на спеціяльній нищенаведеній картці, на якій визначена окремими цифрами ціла система потрібних ознак. Для того, щоб зробити таке перенесення, цілий кадр осіб на окремих діркувальних машинах, пробиває для певних ознак круглі дірочки в відповідних місцях системи на картці і лише після цієї операції картка поступає на рахункову машину Powers'a , яка автоматично групує матеріял відповідно до системи спеціяльного діркування, визначеного на картці.
Дальше підсумування даних вже робиться на інших машинах, табуляторах, після чого таблична зведення легко закінчується.

Операція зведення матеріялу електричними машинами вимагає багатьох осіб, певної спеціялізації . їх по здібностях виконання тої або ін-

---

x/ Крім цієї машини вживаються також і інші системи, напр., Hollerith'a, або Classificompteur March'a ,що приводиться в рух не електрикою, а руками.

Сіль-Госп.Робітники. 1930. Постійні. Село: Мрин. 14. Повіт. Ніжинський ч.23.

Число всіх робітників	Місце народження і приписки та національність		До 18 років		від 18-- 24 років		Від 24 - 40 років		від 40 - 50 років		і т.далі
	Місце народ. і приписка	Національність	чол.	жін.	чол.	жін.	чол.	жін.	чол.	жін.	
1	2		3	4	5	6	7	8	9	10	
127	Народжені і приписані в селі	Українці	I	III	⊠ ⊠ ⊠ II	⊠ ⊠ ⊠ ⊠ I	⊠ ⊠ ⊠ ⊠ III	⊠ ⊠ ⊠ ⊠ IИ	⊠ II	⊠ ⊠ III	
					I7	2I	23	24	7	I3	
		Росіяне			III	⊠ II	I	III	II	II	
					3	7		3	2	2	
		Інші									
59	Народжені в селі, але приписані в цьому повіті	Українці			⊠ I	⊠ ⊠ II	⊠ ⊠ ⊠ I	⊠ II	IИ	⊠	
					6	I2	I6	7	4	5	
		Росіяне	I			3	2		I	2	
		Інші									
48	Народжені в селі, а приписані деінде на Україні	Українці.	II	III	⊠ III	⊠	⊠ ⊠ I	⊠ I	III	III	
			2	3	8	5	II	6	3	3	
		Росіяне		II				III			
				2				3			
		Інші			полякI			поляк I			
2	Народжені в селі, а приписані в чужині	Українці					I				
		Росіяне									
		Інші					поляк I				
і т.д. разом 367	і т.д.										

Після опрацьовання цілого села мусить число всіх осіб у всіх графах, починаючи з третьої дати ціле число робітників у всіх господарствах села згідно з контрольною графою число I.

шої праці, а також великої дисципліни. В Статистичному Уряді в Празі при обробленню перепису населення /I92I р./ працювало 26 діркувальних машин і I2 клясифікаторів. На одній машині діркувальній за 5 годин зпрацьовано пересічно II00 карток, на клясифікатору - 40.000, одначе досягалось до I750 карток на першій і 70.000 на другій машині за одну пятигодинову зміну. З персоналу Уряду - 784 особи, на технічному опрацьованню було чинним I00 осіб у діркувальних машин і 32 у клясифікаторів. Працювалось по дві зміни денно, а біля клясифікаторів і по три зміни денно. Діркування почалось

26/IУ I92I р., а скінчено 5/У I922 р. Більшість всіх карток, числом I3,5 мільйонів, мусила пройти клясифікаторами 35 разів.

Рік.	Місяць	Рід перевязки	Порядкове число	Митний район	Статист. число	Вага Kg - g	Вартість в кор.ческ.	Країна походження	Країна призначення	Штуки
30	I0	0	● ●● 000	00 00	●000●	000 ●00 000	000 ●0 ● 0● 0	00	00	00 000
2I	I	●	I II III	II I I	III II I	I II I II I II I	I II II I III I	II	II	II III
●	2	2	2 22 222	22 2 2	2 2●2 2	2 22 2 22 2 2	2 22 22 2 2 2	22	22	22 22 2
23	3	3	3 3 3 ●33	33 3 3	3 ● 333	3 33 3 33 33 ●	333 33 3 3 3	●3	33	33 33 3
24	4	4	4 4 4 4●4	44 4 4	4 4 44 4	4 44 4 ●● 4 4 4	444 4●4 44 4	4●	44	44 44 4
25	●	5	5 5 5 555	55 5 5	5 5 55 5	5 55 5 55 5● 5	555 5 55 5● 5	55	55	55 55 5
26	6	6	6 6 6 666	66 6 6	6 6 6●6	6 66 6 66 6 6 6	666 666 66 ●	66	66	66 66 6
27	7	7	77 7 777	77 7 7	7 7 7 7 7	7 77 7 77 ●7 7	777 777 77 7	77	77	77 777
28	8	8	88 8 88●	88 8 8	8 8 8 8 8	8 88 8 88 8 8 8	888 888 88 8	88	88	88 88 8
29	9	9	99 9 999	99 9 9	9 99 9 9	9 99 9 99 9 9 9	999 999 99 9	99	●●	99 99 9

е/ Публікація зведеного матеріялу. Добра публікація зведеного статистичного матеріялу мусить задовольняти певні вимоги, характер яких випливає з того великого, значіня, яке мають для людей практики і для робітників науки статистичні збірники взагалі й основних переписів особливо. Ці вимоги можна коротко з'ясувати так:

I. Видання офіційної статистики мусять бути зосереджені в одному центральному органі, найліпше статистичному уряді, при тому кожна окрема галузь матеріяльної статистики мала б своє число й літеру. Видання тоді не розкидаються по великому числу різних інституцій і урядових установ, знаходити їх легко і вони маються в продажу, через що доступні кожному.

2. Заголовки таблиць збірників мусять друкуватись не тільки на своїй мові, але й в одній з більш поширених европейських мов /напр.,французькій, англійській або німецькій/. Дуже багато пропадає цінних статистичних даних лише через те, що вони друкувалися лише в мові своєї країни і не могли бути використані науковими робітниками для своїх праць.

3. Статистичні видання важливих матеріялів мусять завжди мати текстове пояснення до табличних даних, щодо методів самого спостереження, так і процесу оброблення матеріялу. Таке ознайомлення з досвідом статистичних праць в різних країнах сприяє поліпшенню об'єднання методів та уникненню помилок у практичній роботі.

4. Попередні сучасні статистичні конгреси нераз вказували і вказують на необхідність також одноманітних клясифікацій та термінології в статистичних працях окремих країн. Клясифікація причин смертей в демографії, а також промислового виробництва, професій, товарів у статистиці закордонної торгівлі і багато інших важливих поділів об'єкту дослідження вимагають порівняння даних у світовому маштабі. Через те сучасні статистичні центри мусять у своїх публікаціях дбати про можливе задоволення цієї настирливої потреби.

5. Вживання різними країнами різних мір - ваги, вартости, простору і т.п. дуже утруднює використання й порівнання даних для наукових дослідів. Праця переведення одних мір на другі є часто непосильною для окремих робітників, центральні установи завжди мають технічну можливість задовольняти і цю важливу потребу.

6. Дуже важливим також і корисним є подавати матеріял великих статистичних праць не тільки в таблицях, але й в загальних оглядах, що знайомлять читача з загальними висновками цілої праці, а також додавати до них і відповідну статистичну графіку /діаграми, картограми і т.п./

Л е к ц і я д е с я т а.

ГЛАВА IУ. ПОХІДНІ СТАТИСТИЧНІ ВЕЛИЧИНИ І СТАТИСТИЧНІ РЯДИ.

§ 23. Похідні величини. В результаті переведення операції зведення статистичних даних, здобутих спостереженням /численням/, ми одержимо в різних таблицях формах певні абсолютні величини. Але за помічч лише абсолютних величин, тільки""сирового"числового матеріялу, ми не можемо скласти собі ясну уяву про дійсну міру змін, про зростання або убування явищ, а тим більше не можемо зробити порівняння і зіставлення різних незалежних між собоч соціяльних мас. Абсолютні числа раніш мусять бути зведені до якоїсь однієї міри, виміру, щоб можна було одержати дійсно придатні до порівняння пропорційні величини. Сказане одначе не треба брати в такому розумінні, що абсолютні числа не мають в собі ніякої вартости. Навпаки, багато з них має велике практичне значіння. Нерідко студіювання змін ряду абсолютних цифр або зіставлення змін /хитань/ двох чи декількох таких рядів можуть привести до істотних висновків наукового значіння. Це особливо тоді, коли йде мова про хронологічні зміни явищ, що відбуваються в одній країні, або взагалі в якомусь єдиному осередку, а разом з тим, коли маємо підставу припускати незмінність чи малу мінливість цього осередку.

Одначе такі припущення робити доводиться дуже рідко, як правило вони не мають місця в реальній дійсності і через те для того, щоб робити висновки із статистичних чисел, ми мусимо попереду перетворити абсолютні числа в похідні величини, інакше кажучи - в статистичні коефіцієнти, що складають групи похідних величин - величини релятивні /відносні/, величини середні та індекси.

§ 24. Величини релятивної /відносної/ частоти. Як раніше ми зазначали, кожна статистична сукупність чи то реальна, чи штучно створена, складається з почислених одиниць певного об'єкту дослідження, а кожна одиниця має цілий ряд ознак кваліятативних, або квантитативних.

Коли нас цікавить факт лише наявности якоїсь ознаки в одній частині сукупности і відсутности цієї ознаки в другій частині її, тоді характеристика, яку надаємо сукупності з погляду цієї ознаки набуває форми релятивної, відносної частоти. Отже релятивна частота є поділ між числом одиниць, що мають спостережену ознаку /або певні її відміни/ та числом всіх одиниць сукупности. Цей поділ звичайно визначається процентовим відношенням /відсотком/.

Таким відношенням буде, напр., відношення числа жінок до цілого числа людности, число кваліфікованих робітників до цілого числа робітників даного міста або району.

В статистиці, одначе, маємо поруч з вищенаведеними відносними величинами і інші відносні величини, що мають велике значіння та які подібні лише своєю зовнішньою формою, але істотою, своїм внутрішнім змістом, більш-менш різняться.

Для величин відносної частоти в стислому розумінню слова є характерним те, що в поділі, який вони визначають, одиниці чисельника є також частиною одиниць знаменника, себто частиною вибраною з цілої сукупности одиниць за певною ознакою, відносну частоту якої хочемо знати. Себто одиниці чисельника і одиниці знаменника є одиницями єдиної сукупности з тією різницею, що перші мають крім того одну, тільки їм властиву, окрему ознаку. Таким чином, формально, з погляду логічної структури, ця властивість наближує величини відносної частоти до імовірностей, що як побачимо пізніше, є також поділами, відношеннями, в яких чисельник обіймає частину одиниць знаменника, обрану нами з погляду певної, спеціяльної ознаки. Себто в знаменнику маємо цілу суму однаково можливих випадків

явища, а в чисельнику кількість тих із них, що ведуть до наслідку, імовірність якого визначаємо. Ця спорідненість релятивної частоти, в тісному її розумінні, з імовірностями є дуже важливою через можливість прикладання основних правил теорії імовірностей до студіювання статистичних релятивних частот. Інші релятивні статистичні числа, які лише назовні подібні до відносних частот, в тісному розумінні цього поняття, цієї властивости не мають. Візьмімо тепер іншого характеру релятивні частоти - такі, як коефіцієнти шлюбности чи народжень. Вони можуть бути названі відношеннями інтенсивности, але в той же час не дають, подібно вищенаведеним, відношення части до цілого, не дають через те, що кожне з тих явищ, до яких вони відносяться /кожний даний шлюб, кожне дане народження/ не стосується до кожної одиниці, з яких складається знаменник поділу, але до двох одиниць одночасно, а іноді і до більшої частини одиниць. Наприклад, колиб хотіли знати відношення нещасних сутичок автомобілів або потягів до цілої людности в державі, кожний з цих випадків, що міститься в чисельнику, відноситься одночасно до цілої групи одиниць знаменнику. Але й такого роду відносні частоти, які не є релятивними частотами в дійсному розумінні цих слів, всеж означають якусь інтенсивність явища в певному оточенні, а через те можуть бути названі відношеннями інтенсивности. Ці відношення не мають відношення части до цілого, а через те своєю структурою менше подібні до відношень імовірностей ніж відносні частоти в тісному розумінні. Певним одначе способом, що веде до зміни їх числової вартости, відношення інтенсивности можуть бути змінені на форму аналогічну до відношень частот в тісному розумінні. Наприклад, вистарчить помножити коефіцієнт шлюбности на два, тоді в чисельнику одержимо число осіб /але вже не пар/, що вступили до шлюбу і тоді можемо мати відношення числа осіб, що вступили до шлюбу, до цілого населення, себто поділ аналогічний дійсним релятивним частотам. Правда, не завжди можливо так легко, як у даному випадку, перевести операцію перетворення відношення одного виду в другий, але важливим є, що таке перетворення нерідко можливе.

Треба мати на увазі ще одну різницю між указаними видами відношень. Статистичні числа, наведені яко приклад відношень інтенсивности, мають у собі ту відміну, що знаменник їх не подає цілком точної кількости одиниць, а лише приблизно ту цілість, з якої походять одиниці чисельника. Коли візьмемо в знаменнику загального коефіцієнту народжень цілу кількість населення за початком або половиною дотичного періоду, він не буде представляти точно число осіб, з яких складаються батьки дітей, що народились в певному періоді часу. Такими батьками могли бути й особи, які імігрували /вселилися/ до даної країни бігом того ж періоду часу, або бігом другої половини цього періоду. Навпаки, щодо числа народжень, то воно обіймає й число дітей тих батьків, що померли в першій половині періоду часу і т.п. Подібні неточності зустрічаємо й при обраховуванню коефіцієнта шлюбности.

Окремий вид релятивних величин уявляють з себе числа, що показують відношення частоти однієї частини сукупности до другої. Прикладом для того може бути відоме з демографії відношення числа народжених хлопчиків до числа народжених дівчат / число хлопчиків на 100 або 1000 дівчат/. Тут маємо в чисельнику число одиниць сукупности з певною ознакою /чоловіча стать/, а в знаменнику вже не цілу суму одиниць сукупности, але число одиниць з протилежною ознакою. Важливо зазначити, що й такі релятивні числа можливо відповідними способами змінити в дійсні відносні частоти, а саме: хай буде число одиниць, що мають певну ознаку, A, зазначено символом (A), а число одиниць, що не мають цієї ознаки або з ознаком протилежною, символом (a), тоді наведений вище приклад відношення можна висловити як $\frac{(A)}{(a)}$, тоді як дійсна відносна частота, себто відношення одиниць сукупности A до цілої суми одиниць сукупности рівнялося б: $\frac{(A)}{(A)+(a)}$. Таким чином, від першого виразу можемо перейти до другого тим, що помножимо перший вираз на величину $\frac{(a)}{(A)+(a)}$, яка є відносною частотою протилежної ознаки в цілій сукупності. Колиб захотіли від відношення числа хлопчиків до числа дівчат перейти до відношення частоти хлопчиків в цілій кількості народжень, до-

сить перше відношення помножити на відносну частоту дівчат в цілій кількості народжень. Існуючу в таких випадках залежність можна формулювати так:[x] зазначимо, що $\frac{(A)}{(a)} = u$, а $\frac{(A)}{(A)+(a)} = v$, тоді $\frac{(a)}{(A)+(a)} = \frac{I}{(A)+(a)} = \frac{I}{u+I}$

Зв'язок між "u" і "v" можемо з'ясувати ще так: $v = \frac{u}{u+I}$ відповідно цьому $v = \frac{v}{I-v}$. Себто величина "u" є функцією величини "v", що є дійсною відносною частотою. Виходить, що й цей вид відносних чисел може бути легко зведений до форми, яка є аналогічною імовірностям.

Є ще вид відносних чисел, який не має по суті нічого спільного з наведеними вище відношеннями частоти та імовірностями, - це відносні числа, що служать лише для наочности порівняння обсягу різних явищ, але також визначаються у відсотках. Напр., порівняння певної вартости ввозу до певної вартости вивозу /береться на 100 карб. або 1000 карб. вивозу стільки то карбованців ввозу/, для з'ясування активности чи пасивности торговельного балянсу, або відношення числа шлюбів до числа розводів і т.п.

В таких випадках не порівняємо частей з цілою сумою, або явищ, що відбуваються в певному оточенні з цим оточенням, або одну частину сукупности з другою, а лише спостерігаємо відношення величин відмінних по собі, з метою лише ясніщого порівняння.

Дуже близькими ф о р м а л ь н о до тільки що наведених відношень /наочности явищ/ є також відносні числа, що з'ясовують відношення обсягу явища за певний період часу /рік, місяць і т.п./ до обсягу цього явища, за який будь попередній період часу. Ці числа, таким чином, обчислюємо з метою наочної уяви змін якогось явища бігом часу. Такі відносні числа найчастіше звуться індексами. Розуміється, що подібні числа можемо обчислювати і для з'ясування змін явища в просторі, себто одержати відношення обсягу явища в одному місці до його обсягу в другому місці. В дальшому над індексними числами зупинимось докладніше.

Мусимо ще вернутись до характеристики величин відносної частоти, щоб звернути увагу на різницю їх, важливу з погляду аплікації понять теорії імовірностей на статистичні числа /Чубер/. З наведених вище прикладів бачимо, що ознакою, за якою вибираємо одиниці чисельника певної відносної частоти з одиниць знаменника, себто з цілої сукупности, може бути певний факт, певна подія, яка сталася з певними одиницями сукупности. В коефіцієнтах смертности вибираємо з цілої сукупности людности осіб , яких настигла смерть бігом певного часу; або з числа осіб даного покоління, що дожили віку "n" років, вибираємо осіб померлих в n + I році свого життя. Подібно робимо і з обчисленням коефіцієнту злочинности або захорування. Не можна не побачити, що між цими випадками і такими відношеннями частоти, як відношення чоловіків до цілого числа людности, відношення народжених дітей чоловічої статі до цілого числа народжених, є велика різниця. В наведених тільки що прикладах прикметою, за якою певні одиниці сукупности відносимо до чисельника поділу, буде щось абсолютно спільне, зв'язане з певними одиницями сукупности; зв'язане цілком незалежно від часу спостереження: кожна людина як така буде чи чоловіком чи жінкою, незалежно від протягу часу нашого спостереження. Кожний випадок народження буде або народженням хлопчика, або дівчини також незалежно від поділу доби спостереження. Одначе при вибору осіб померлих або вчинивших злочин в добі одного року, не буде основою, за якою ми ці одиниці вибираємо, щось абсолютне від первопочатку зв'язане з даною одиницею - бо є це п о д і я, що утворюється бігом часу. Через те питання чи припаде дана одиниця знаменника до числа одиниць чисельника залежить на тому як довгий час ми одиниць спостерігали. Ця властивість робить відносні частоти такого роду нестійкими в тому розумінні, що такі відносні частоти такої цілої сукупности змінюють свою числову величину /вартість/ в зв'язку з поділом часу спостереження. Ця непевність буває часто гострішою через те, що в статистиці найчастіше береться на увагу, яко доба спостереження явищ, один рік.

x/ St.Kohn. Zaklady teorie statisticke metody.

Отже, коли студійною прикметою одиниць сукупности є подія, що здійснюється бігом часу, слід такі відношення частоти вважати часовими відношеннями частоти на розділ від тих відношень частоти, які слід назвати відношеннями частоти ознак. Само собою розуміється, до часових відношень частот /або до відношень частот подій/ не можна відносити таких випадків, в яких кожна одиниця сукупности уявляє з себе подію і ознака, за якою вибираємо частину цих подій до чисельника, буде вже абсолютно зв'язана з подіями знаменника і не залежить від довжини часу спостереження./Відносна частота народжень хлопчика в цілій сукупності народжень бігом певної доби, частота крадіжок у цілій сукупності переступів за певну добу і т.п./. Поскільки використання правил теорії імовірностей, як побачимо далі, у статистичній аналізі атипових колективних явищ має важливе значіння, постільки і усвідомлення наведеного поділу відносних частот на часові /події/ і прикметні є необхідним.

\- - - - - -

Мусили б далі перейти до з'ясування другого типу похідних величин, що звуться середніми, але ці величини стосуються до такої маси одиниць сукупности, в яких спостережена ознака набуває різних кількосних вартостей. Середня величина є завжди характеристиком сукупностей з погляду квантитативно змінних ознак, себто таких ознак, відносно яких нас цікавить не факт тільки її наявности або відсутности, але числові вартості, яких ці ознаки набувають. Через те й мусимо спочатку з'ясувати, на підставі чого й в які форми мусимо вкладати сукупності з такого роду ознаками /мінливими/, щоб досягти дальшої, досконалішої характеристики цих сукупностей за допомогою середніх величин.

§25. Таблиці розподілу частоти /статистичні ряди/. Коли маємо діло з таков сукупністю, яка складається з великого числа одиниць, при чому спостережена ознака цих одиниць набуває різних числових вартостей /значінь/, тоді скласти собі уяву з погляду даної ознаки про цілу сукупність не мається можливости. Така сукупність уявляє з себе непроглядний ряд цифр, схопити значіння яких можемо лише перетворивши в спеціяльні табличні форми, в певні статистичні ряди, розподілу вартостей змінної кількісної ознаки даної сукупности. Такого роду таблиці і звуться таблицями розподілу частоти або таблицями фреквенції. Для складання таких таблиць треба цілу сукупність розподілити на ряд груп, клясів, відповідно до вартостей ознаки і тоді підсумувати кількість одиниць сукупности, що належать до кожної групи або кляси. Найліпше вияснити собі природу таких клясифікованих рядів на прикладах, з яких також побачимо наскільки велика користь досягається скороченням довгого ряду поодиноких спостережень.

В нижченаведеній таблиці У. Юла подані пересічні річні смертності, обчислені в ‰ числа людности /річне число померлих, що припадає на 1000 людности/ по 632 округах Англії і Валесу за десятиліття 1881-90 р.р.

Таблиця I.

Пересічна річна смертність в ‰	Число округів, що виказують смертність в межах поруч означених	Пересічна річна смертність в ‰	Число округів, що виказують смертність в межах поруч означених.
12,5 - 13,5	5	23,5 - 24,5	5
13,5 - 14,5	16	24,5 - 25,5	3
14,5 - 15,5	61	25,5 - 26,5	1
15,5 - 16,5	112	26,5 - 27,5	1
16,5 - 17,5	159	27,5 - 28,5	2
17,5 - 18,5	104	28;5 - 29,5	-
18,5 - 19,5	67	29,5 - 30,5	-
19,5 - 20,5	42	30,5 - 31,5	2
20,5 - 21,5	25	31,5 - 32,5	-
21,5 - 22,5	18	32,5 - 33,5	1
22,5 - 23,5	8	Разом	632

Колиб ми розглядали без такої таблиці індивідуальні дані по 632 округах, навряд чи змогли б мати якусь уяву про характер спостереженого явища, бо в смертності окремих округів мається великі неправильні розділи. В той час, короткий навіть погляд на таблицю зразу розкриває важливі речі. Бачимо, що річна смертність хитається лише в межах I3 до 33‰ /кругло/, але в більшій частині округ є ближче до нищої своєї межи ніж до вищої; в 60% округ смертність хитається в тісних межах від I5 до I8,5‰ найбільше число округ виказує смертність біля I7 ‰

Таблиця 2. Смертність осіб тої або іншої статі фіксувалась на Радянській Україні в I926 і I927 роках також за віковим їх поділом. Розподіл частоти цього явища є таким:

Вікові групи осіб	Чоловіків померло на I000 люду відповідного віку	Вікові групи осіб	Жінок померло на I000 люду відповідного віку
0 років	I59,5	0 років	I32,6
I - 4 "	3I,9	I - 4 "	29,8
0 - 4	70,9	0 - 4	59,6
5 - 9	8,3	5 - 9	7,8
I0 - I4	8,5	I0 - I4	8,4
I5 - I9	4,5	I5 - I9	4,5
20 - 24	6,6	20 - 24	6,0
25 - 34	6,I	25 - 34	6,I
35 - 44	7,8	35 - 44	7,2
45 - 54	I3,2	45 - 54	9,6
55 - 64	24,3	55 - 64	I7,9
65 - 74	49,8	65 - 74	42,3
75 - 84	I06,5	75 - 84	94,2
85 і більше	I95,6	85 і більше	I78,0
Пересічно	I9,5	Пересічно	I6,4

Розподіл частоти дає можливість легко замітити, що найменша смертність і чоловіків, і жінок припадає на вік від I0 до 20 років, що від цих вікових границь угору й вниз смертність поступово зростає, що найбільші коефіцієнти смертности припадають на групи немовлят і найстарших віком, що серед жінок смертність є слабшою ніж серед чоловіків і т.п.

Таблиця 3. Біометриками[x] було спостережено число лучів у макових головках певного сорту маку / Mak Shirle'я / для встановлення характерної риси цього маку. Через те, що між найменшим числом лучів і найбільшим не виявилось у цілій масі дуже великого розходження, різниця між клясами /групами/ встановлена всього одиниця /один луч/.

Число лучів	Число маковиць з відповідним числом лучів	Число лучів	Число маковиць з відповідним числом лучів
6	3	I4	302
7	II	I5	234
8	38	I6	I28
9	I06	I7	50
I0	I52	I8	I9
II	238	I9	3
I2	305	20	I
I3	3I5	Разом	I905

x/ Примітка: Школа англійських статистиків-біологів, яка з успіхом приклала статистичну методу до соматології рослин і взагалі багато сприяла розвитку статистичної методології з початком 20 століття.

Як видно з розподілу частоти, число лучів хитається від 6 до 20, найчастіше число лучів у маковиці I2, I3, і I4.

На яких же засадах треба будувати такого типа таблиці або статистичні ряди, як вищенаведені, або подібні їм таблиці розподілу частоти квантитативних ознак для одиниць певної сукупности ? Праця будови такого роду таблиць складається з наступних моментів:

I. Поперше треба встановити розмір клясового /групового/ розхилу - інтервалу, себто кілька ступневих клясових одиниць мусить мати кожний інтервал. В таблиці I-й було встановлено для кожного інтервалу однаковий розхил - одиниця, в таблиці 2-й розхил цей неоднаковий, а саме, виділена група немовлят, потім розхил в клясах до 24-х років має 4 одиниці, від 25 до 84 років - I0 одиниць і виділена окрема вікова група 85 і більше років.

2. Мусить бути вирішено, який певний розмір подати для нижчої і вищої границі інтервалу. Напр., в таблиці I-й необхідно було встановити чи границі інтервалу взяти I2-I3, I3-I4, I4-I5 і т.д. чи I2,5 - I3,5 , I3,5 - I4,5 і т.д.

3. Коли це вирішено, всі спостережені випадки розподіляються по певних клясах цілої таблиці.

4. Після віднесення різних випадків до певних встановлених кляс підсумовуються в кожному клясовому інтервалі всі віднесені до нього випадки, а потім і складається відповідна таблиця розподілу частоти.

Будова таблиць розподілу частоти вимагає одначе деяких методологічно-технічних пояснень.

а/ Кількість клясових інтервалів і їх границі. З наведених таблиць розподілу частоти певної квантитативної ознаки, ми бачимо, що деякі з цих змінних ознак, по розміру своєї величини змінюються не в широких межах. Напр., кількість лучів у маковиці хитається в межах від I-го луча до 20-ти, себто ціла варіяція ряду не має широких меж, при чому сама варіяція йде від серії до серії ряду уривцем так, що кожна з серій є цілим числом та різниться одна від другої одиницею - I,2,3,4,5 і т.д. Така варіяція ознаки зветься звичайно - змінною розривною. Навпаки, в інших випадках варіяція не переривається, а межі її дуже широкі, тоді ознака з такою мінливістю зветься змінною тяглою. Напр., змінна в таблиці першій.

Отже в випадках змінної розривної не мається підстав якогось спеціяльного вибору клясового інтервалу, бо звичайно таким інтервалом буде одиниця.

У випадках змінної тяглої такого природнього інтервалу не має і його вибір і будова мусять братись на увагу дослідником.

Поперше треба дбати про те, щоб кількість клясів /окремих серій з інтервалами/ змінної тяглої не була дуже великою, бо тоді таблиця не буде наочною. Не мусить їх бути і дуже мало, бо тоді перемежки клясові /інтервали клясові/ довелося б робити занадто великими і таблиця описувала б сукупність у дуже грубих рисах /загальних/.

Статистична практика учить, що найліпше в таких випадках ділити варіяцію /ряд/ на I5-20 і не більше як 25 клясів. Одначе, коли бачимо, що сукупність мала, а в перебігу частоти від інтервалу до інтервалу зустрічаємо значні неправильності, або деякі інтервали варіяції будуть зовсім без частоти, тоді необхідно обмежитись меншим числом клясів, взявши їх I0 - I2.

Далі виникає питання: як треба визначати границі перемежків кляс? При встановленню границь інтервалів інколи виникають утруднення з тим, як відносити окремі випадки змінної до того, або іншого інтервалу клясового. Можемо наприклад, встановити клясові інтервали для розподілу округів по площі в Km^2, від 0 - I00, від I00 - 200, від 200 - 300 Km^2 і т.д. Куди маємо віднести округ з площею розміром рівно в I00 Km^2 - до першої чи до другої кляси? Коли площа кожного округа мірялась не тільки в цілих а й з десятими кілометра, тоді можемо віднести всі округи, що мають I00, I00,I і більше Km^2ах до I99,9 Km^2 до другої кляси, а всі ах від 0-99,9 Km^2

до першої. Коли встановити більшу точність від нас не залежить, себто мусимо оперувати вже з готовими величинами, тоді треба всі випадки з граничним розміром ознаки /змінної/, напр., всі округи з площею I00 Km^2, розділити на дві однакові групи, рівні частини, і одну половину віднести до відповідної кляси /нижчої/, а другу до сусідньої вищої.

Ліпше таблицю укладати так, щоб границі інтервалу клясового визначались так: "О і вище, але менше ніж I00 Km^2, "I00 і вище, але менше ніж 200 Km^2 і т.д. Так визначені границі інтервалів можемо ще зазначити: "О -", "I00 -", "200 - " і т.д.

Для прикладу можемо подати розподіл сукупности 7749 дорослих чоловіків за ознакою ваги в Англії, поданий у У.Юла.

Таблиця 4.

Вага в фунтах англійських	Число осіб з відповідною вагою	Вага в фунтах англійських	Число осіб з відповідною вагою.
90 -	2	I90	263
I00 -	34	200 -	I07
II0 -	I52	2I0 -	85
I20 -	390	220 -	4I
I30 -	867	230 -	I6
I40 -	I623	240 -	II
I50 -	I559	250 -	8
I60 -	I326	260 -	I
I70 -	787	270 --	-
I80 -	476	280 -	I
		Разом	7.749

Коли важення таке зроблено з точністю лише до одного фунта, то до інтервалу клясового "90 і вище, але менше I00 " припадуть особи з вагою 90,9I,92... 98 і 99 ф., до інтервалу "I00 - " припадуть особи з вагою I00, I0I, I02 ... I09 ф. Границя цих інтервалів знаходиться десь між 99 ф. і I00 ф. Дуже легко рішити, де вона знаходиться. При важенню мусили вагу осіб заокруглювати по правилах аритметики, себто десяті долі фунта до 0,5 заокруглювати до нижчих цілих чисел, а десяті 0,5 і більше - до вищих. Себто вартости змінної 99,I; 99,2; 99,3; 99,4 відносили до кляси "90 -", а вартості 99,5 - 99,9 до кляси "I00 -". Отже границею цих двох інтервалів і є число 99,5. З цього випливає, що від точности, з якою міряння ознаки /змінної/ було переведено, залежать границі сусідніх інтервалів різних вартостей змінної. При важенню з точністю до 0,I фунта, припадали б до інтервалу "90 - " особи з вагою аж 99,9 ф. Одначе, при такій точности важення мусило б робитись заокруглення сотих і величини 99,90; 99,9I; 99,92; 99,93; 99,94 ф. відносились би до 99,9 ф., а величини 99,95; 99,96; 99,97; 99,98; 99,99 ф. до I00 ф. Наслідком цього границя між зазначеними двома інтервалами клясовими була б 99,95 ф., але ніколи 99,5 ф., як то було при важенню з точністю до одного фунта. Через те треба завжди при встановленню інтервалів подібного типу подавати точність, з якою переводилось міряння, інакше особа, що користується таблицями не може знати межі між сусідніми інтервалами. При великих інтервалах таке детальне встановлення їх границь, розуміється, не має значіння, але при малих інтервалах, - в 5 - I0 одиниць, в зв'язку з деякими умовами наукового оброблення, має своє значіння.

Можемо також визначати інтервали клясів як у таблиці 4-й, подавши нижчу і вищу границі кожного, але ці границі визначати з більшою точністю, ніж робилось міряння ознаки. Напр., в нашому прикладі, коли ознака мірялася з точністю до I-го фунта, було б потрібно визначати границі:

89,5 - 99,5
99,5 - I09,5 і т.д.

Але таке визначення є незручним, через те можемо визначати мінімальну і максімальну вартість ознаки, що належить до даного інтервалу. Себто, коли вона мірялась з точністю цілої одиниці, інтервал "90 і більше, але менше ніж 100" можна визначати:

90 - 99

а наступний 100 - 109 і т.д.

При ознаці, що мірялась з точністю до 0,1 ф. було б визначення таким:

90,0 - 99,9

100,0 - 109,9 і т.д.

При цьому способі визначення мусять бути числа границь інтервалу, розуміється, подані з тією точністю, з якою ознака мірялася. Цей спосіб визначення вживається переважно в російській статистиці, визначення 90 -; 100 -; 110 -; переважно в англійських статистиків.

Треба також пам'ятати, що будова клясів, якоїсь змінної може бути відповідно виконана тільки тоді, коли дослідник досконало знає той об'єкт з яким оперує. В кожній будові таблиці розподілу частоти квантитативної ознаки, до певної міри мусить бути і передбачення наукової мети, яку поставив собі дослідник. Особливо в тих випадках, коли маємо справу з тяглими змінними треба завжди виходити з природи явища, яке досліджуємо, в якому хочемо встановити якісь закономірності або причинові зв'язки.

Розподіляючи наприклад, на кляси змінну, яка мусить дати нам образ вікової структури суспільства, важливо знати також, що система клясів чи вікових груп змінної, яку ми вкладаємо в таблицю розподілу частоти може нам служити для різних цілей. Вона не тільки цікава нам для виявлення як кількісно і в якому відношенні розподіляється людність по вікуідля практичних яких будь завдань, але й може бути цікава для наукового виявлення цілого ряду взаємозв'язків, але різної інтенсивности чи сили між певним явищем і тими віковими групами, на які ми розподіляємо цілу сукупність. Лише знаючи наперед,що, наприк., смертність немовлят, а також дітей до 5 років взагалі, значно вища, ніж смертність інших вікових груп, або спостерігаючи це із зведеного статистичного матеріялу - треба й систему клясів змінної будувати так, щоб можна було ліпше виявити і аналізувати зв'язок між смертністю й певним віком осіб.

ВПРАВИ / до 8-ої лекції/

1. Яка сутєва ріжниця між простою й комбінаційною таблицею?

2. Як обчислюється число граф для простої таблиці і для комбінаційної ?

3. Дайте приклад статистичного речення та зазначте, що в ньому буде статистичний підмет, а що присудок, що факторіяльною ознакою, а що функціональною ?

ВПРАВИ / до 9-ої лекції/

1. Через що в комбінаційній таблиці не можна ставити в зв'язок дуже багато ознак ?

2. Які негативні риси способу централізованого зведення статистичного матеріялу ?

3. Як сконтролювати вірно чи невірно почислені по групах особи в рахунковій таблиці наведеній на стор. 73 /9-та лекція/.

ВПРАВИ / до 10-ої лекції/.

1. Наведіть приклад відносної частоти певної ознаки в якій будь цілій сукупності.

2. Наведіть приклад відносних чисел, що служать лише наочности порівняння обсягу різних явищ.

3. Подайте невелику таблицю розподілу частоти якої будь ознаки.

Лекція одинадцята.

в/ Рівність перемежків клясових. При складанню таблиць розподілу частоти якої будь сукупности треба мати на увазі, що такий розподіл може дати вірну уяву про певну сукупність тоді, коли в ньому інтервали /перемежки/ кожної кляси, на які розділена змінна встановлені однаковими. Нерівність перемежків утруднює або й зовсім унеможливлює прикладення до такого розподілу змінних деяких метод статистичних, при тому часто самі вартості частоти в таблиці подають на певний погляд фалшивий образ, себто таблиця втрачає на ясності. Візьмемо, наприклад, таблицю, що дає розподіл частоти жилих опадаткованих будинків в Англії, в якій подано клясифіковання річного доходу будинків за нерівними інтервалами, а також і за рівними інтервалами, розміром у 10 фунт.стерлінґів.

Таблиця 5.

річний прибуток в фунтах стерлінґів	частота /число/ будинків відповідно клясам	частота цих будинків відповідно клясам з рівними перемежками в 10 ф.с.
1	2	3
20 і менше 30	306.408	306.408
30 - 40	182.972	182.972
40 - 50	105.407	105.407
50 - 60	63.096	63.096
60 - 80	71.436	35.718
80 - 100	32.365	16.182
100 - 150	41.336	8.267
150 - 300	26.732	1.782
300 - 500	6.198	310
500 - 1000	2.098	42
1000 і вище	644	?
Сума будинків	838.692	-

Звернемо увагу на 1 і 2 графи цієї таблиці і бачимо, що частоти будинків кляс з прибутком від 60 до 80 і від 100 до 150 мають значне відхилення. В дійсності причиною цих ухилень є зміна розміру інтервалу з початку з 10 ф.ст. на 20 ф.ст., а потім з 20 на 50 ф.ст. Вище 150 ф.ст. кляси мають ще більші інтервали і частоти їх,так би мовити, не пасують до порівняння. Щоб зробити частоти дійсно здатними до порівняння між собою,треба привести їх /зредукувати/ до рівних перемежків клясових, напр., розміром в 10 ф.ст., ділячи частоти пятої і шостої кляси на два, сьомої на пять, восьмої на 15 і т.д. тоді і одержимо варіяцію подану в графі третій таблиці. Редукцію таку не можна одначе проробити для останньої кляси, розмір якої означений 1000 і більше, себто не точно. Такого роду неозначених стисло кляс треба з а в ш е у н и к а т и, бо для наукового оброблення матеріялу вони шкодять. В урядових публікаціях, в цілях економії праці та зменшення таблиць дуже часто спостерігаємо такі випадки. Викликані вони інколи й тим, що в цих публікаціях не допускається показувати, напр., прибутки двох або трьох найбагатших осіб в окрузі, як це могло б статися при точному розподілу конечних вищих кляс прибутковости, бо тоді кожній інтеліґентній особі, що знає місцеві умови такі особи стали б відомі.

Із загального правила, що всі інтервали клясові мусять бути рівними, треба одначе зробити виключення в тих випадках, коли ходить про розгляд на підставі менших рівних інтервалів тої частини варіяційного ряду, в якій частоті змінюється інтенсивно /прискорено/. Це можна бачити з таблиці 2-ої

на сторінці 79-й, де при розподілі частоти смертности дітей інтервали встановлені значно менші ніж для інших вікових груп; це цілком доцільно через те, що смертність дитяча дуже велика та крімтого ще значно більша в дітей віком до I року, ніж у старших дітей, а через те частоту смертности треба для дитячих вікових груп з'ясовувати по менших інтервалах клясових. В подібних одначе випадках бажано все ж таки давати поділ на кляси з дрібними інтервалами так, щоб на бажання можна було б ці менші кляси з'єднувати в більші з інтервалами однаковими як і в другій частині цілої варіяції певної ознаки.

23. Техніка обліку частот по клясах. Для складення цілої таблиці розподілу частоти треба також кожну одиницю сукупности віднести до відповідної кляси. Коли сукупність є велика /декілька сот одиниць чи екземплярів або й більше/, найліпше вживати рахункових листків /карток/, себто написати вартість ознаки кожної одиниці на окрему картку, а картки розкидати відповідно до вартостей ознаки на окремі групи /купки/, з яких кожна буде мати картки з однаковою вартістю ознаки / коли змінна є переривна/, або вартостями ознаки, що відносяться до відповідного інтервалу клясового /коли змінна є тягла/. Підрахувати потім картки в кожній групі дуже легко, а одержані суми частот занести в таблицю на відповідне місце.

Коли сукупність невелика, не має рації вживати карток. Облік переводиться так, що для кожної вартости ознаки або для кожного інтервалу помітимо крапкою /точкою/, або рискою кожну одиницю сукупности, а тоді точки або риси підрахуємо, щоб рахування облегчити, робимо облік по групах в пять точок або рисок /так: ⁘ або 𝍸/ як то я вже раніше показував. При великій сукупності спосіб точок або рисок /метода карбування/ може легко повести до помилок, що є не добре, бо контролювати такий облік не можна інакше, як тільки повторенням цілого підрахунку, в той час як при способі картков му можемо легко зконтролювати підрахунок, проглянувши кожну групу карток.

24. Графічне зображення розподілу частоти. Статистична таблиця розподілу частоти є перетворення статистичного матеріялу в ясний числовий образ, що дає нам змогу розглядати ці числа в їх взаємному зв'язку, вона облегчує нам наукову аналізу даних, розкриваючи в ладних рядах певні правильності. Але й найліпша таблиця все ж таки має незручність. Полягає вона в тому, що таблична маса цифер, у збудованих рядах, через хитання цифер, відтягає часто нашу увагу від загального характеру явища до його осібних властивостей та перешкоджає нам легко відшукати те загальне основне, що є важливим для наукового висновку. Щоб усунути до певної міри кількісні ріжниці в явищах та досягти більшого виявлення властивих їм загальних рис, статистика й вживає графічних зображень. Графічні зображення "кажуть оку прямо". Через них ми відразу охоплюємо сукупність явищ та відразу охоплюємо те, що в них є загальним або протилежним. Через те саме графічна метода прикладається не тільки в статистиці, а й часто в інших, так чистих, як і прикладних науках.

Графічне зображення розподілу частоти певної змінної можна бачити з наступних прикладів.

Нище наведена таблиця /див. на стор.85 / уявляє з себе розподіл сукупности I495 учнів за висотою їх зросту / Матеріял Ф. Штампаха. " Антропологія ". III. I925. стор. I85 опрацьований С. Коном/.

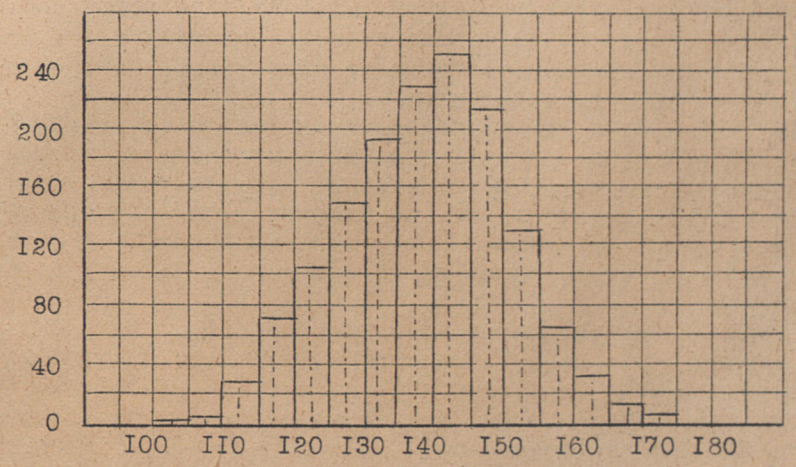
Цю таблицю можна графічно зобразити в наступний спосіб. На сантиметровому папері накреслимо систему ординат. На позему вісь /абсцісу/ наносимо поділи, що відповідають інтервалам клясовим змінної висоти тіла при тому так, що кожний інтервал у 5 см. буде, припустім, рівнятись I/2 сантиметру /на папері/. На прямовійсній ординаті зліва наносимо поділи частоти так, щоб кожні I/2 сантиметра відповідали, припустимо 20 учням /висота зросту яких зміряна/. Тоді з середини інтервалів /на обсцісі/, або з їх граничних точок проведемо прямовісні, рівнобіжні одна до другої, до пев-

ної висоти, що відповідає числу учнів, зазначених для кожного інтервалу клясового в таблиці, а також і на прямовісній ординаті зліва. Завершимо далі над інтервалом прямокутник так, щоб його висота відповідала клясовій частоті цього інтервалу. Отже, коли на інтервал "II0"-" припадає 30 учнів, висота відповідного цьому інтервалу прямокутника рівняється I I/2 поділу на прямовісній ординаті, себто на прямовісній частоти /або 7,5 м.м./. Це графічне зображення наведено в малюнку I.

Таблиця 6.

Зріст в см.	Число учнів	Зріст в см.	Число учнів
I00 -	2	I40 -	250
I05 -	7	I45 -	2I5
II0 -	30	I50 -	I30
II5 -	69	I55 -	64
I20 -	I03	I60 -	33
I25 -	I49	I65 -	I4
I30 -	I94	I70 "	6
I35 -	229		

Малюнок I. Розподіл I595 учнів за висотою тіла.

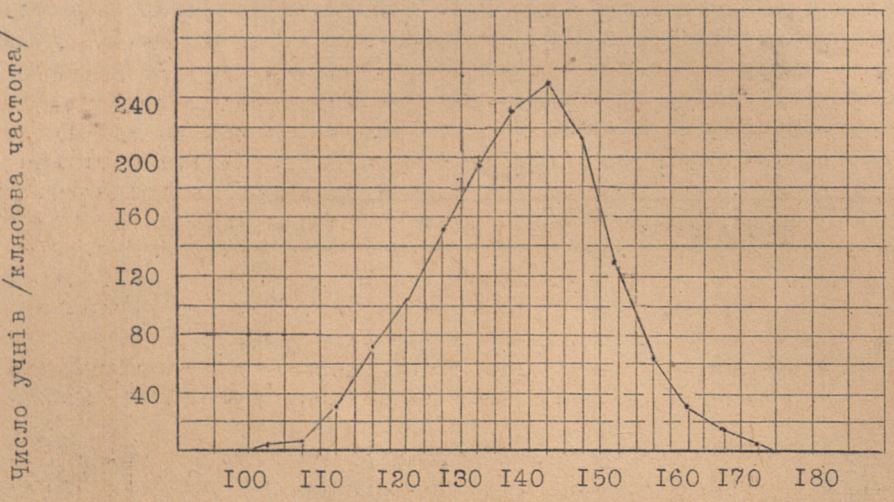


Як бачимо, графічне зображення розподілу даної частоти має вигляд ступневого малюнку, складеного з прямокутників різної висоти. Понеже основи цих прямокутників є взаємно рівні, є їх площі співмірні висотам, а через те є співмірні відповідним клясовим частотам. Площа цілого малюнку зображує цілу сукупність учнів при чому кожна частка площі, що рівняється I/4 см.^{2} /I/2 см.x I/2 см./ відповідає 20 спостереженням /20 учням/. Вище наведене графічне зображення розподілу частоти зветься г і с т о г р а м.

Можемо також цей розподіл частоти зобразити і іншим способом, а саме, так, що на сторчових рівнобіжних, проведених з середини кожного інтервалу зазначимо відповідну клясову частоту згідно з поділом на прямовісній /лівій/ ординаті, а тоді з'єднаємо вершини рівнобіжних простою лінією /Дивись малюнок 2 на стор. 86/. Таким чином одержимо многокутник або п о л і г о н ч а с т о т и. Ціла кількість спостережень не зображена вже на полігоні ступенчатою фігурою, але многокутниками, створеними з одної сторони абсцисою, згідно з обсягом змінної /від інтервалу "I00 - " аж до "I70-" а з другої сторони ламаною лінією, що з'єднує вершини рівнобіжних.

Наступний малюнок і подає нам полігон частоти розподілу учнів у Кралупах у Чехії за їх зростом.

Малюнок 2. Розподіл 1495 учнів за висотою тіла.

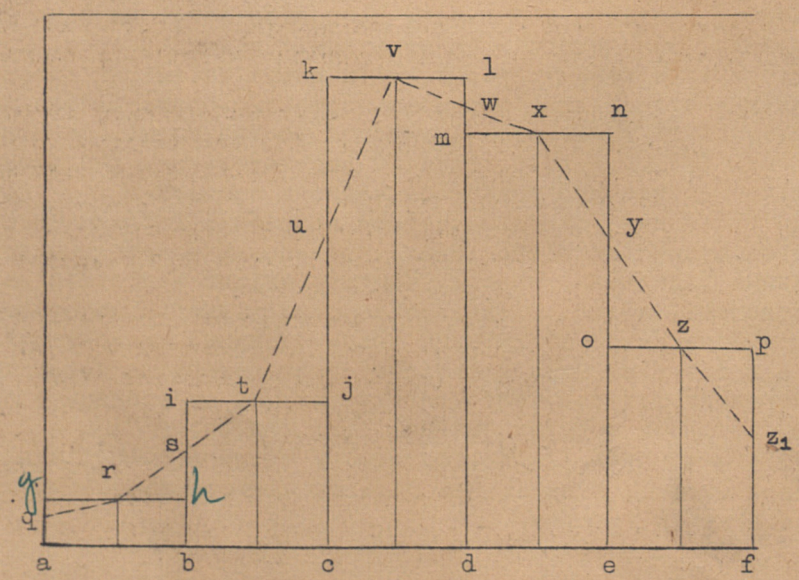


Щодо полігону, то треба мати на увазі, що в той час, як площа кожного прямокутника гістограму є точно співмірна відповідній клясовій частоті, частина площі полігону, яка припадає на кожний інтервал не є точно співмірна відповідній клясовій частоті, бо не завжди точно рівняється площі прямокутника, накресленого над тим же самим інтервалом гістограму. Це ясно буде з малюнку 3-го та з нищенаведеного порівняння.

На малюнку 3-му повною лінією означена частина гістограму agihijk lmnopf , а точкованою лінією відповідна частина полігону aqrtvxzz$_1$f . Коли порівняємо, наприклад, площу bijc гістограму з площею bstuc полігону, які накреслені над тим же інтервалом bc, то побачимо, що другу площу bstuc одержимо з першої, віднявши трикутника ist , та придавши tuj , який є більший ніж трикутник ist ; через що площа bstuc є в даному випадку більша ніж площа bijc.

Навпаки, коли порівняємо площу прямокутника dmne та площу многокутника dwxye , бачимо, що в цьому випадку трикутник, що віднімається, xny буде більший, ніж трикутник mwx , який додається. Через те частина площі полігону накреслена над інтервалом de буде менша, ніж частина площі гістограму, накресленого над тим же інтервалом.

Малюнок 3.



25. Ідеальна крива частоти. З вищенаведеного ми бачили, що полігон частоти якоїсь змінної геометрично є многокутником, себто гранична лінія його є ламана лінія. Уявім на хвилину, що маємо якусь сукупність з великим числом спостережених одиниць, а інтервали кляс, на які поділено цілий обсяг певної тяглої змінної будемо поступово все зменшувати та зменшувати. Тоді площі рівнобіжників /паралелограмів/ накреслені над кожним інтервалом все будуть зменшуватись та зменшуватись, ламана лінія буде мати все більше й більше кутів, аж перетвориться в лінію криву. Теж саме сталося б і з граничною лінією гістограму. В дійсності, звичайно, не будемо мати такої будови клясів і інтервалів конкретної змінної, яка б дала геометрично криву лінію. Але цілком допустимо дивитись на ламані граничні лінії гістограмів чи полігонів, як на лінії, що наближуються до кривих, а через те можемо емпирично спостережені розподіли частоти зображувати у формі площ, окреслених кривими, вирівнюючи ступневість або ламаність ліній, що окреслюють гістограм або полігон. Для чого нам потрібний цей перехід до кривої частоти /фрек - венційної кривої/ ?

З основ аналітичної геометрії знаємо, що різні типи кривих можуть розв'язуватись рівняннями кривих. Через те маємо можливість різні емпиричні спостережені типи розподілу частоти розв'язувати аналогічно відповідним ідеальним кривим, а через те можемо й клясифікувати розподіли частоти змінних, з якими зустрічаємось в статистичних дослідженнях. Отже емпиричні криві можемо аналогувати математичним кривим та користуватись прикладенням до них метод математичної аналізи.

В тих випадках, коли спостережена статистична сукупність уявляє з себе лише частковий вибір з більшої маси предметів, які нас конкретно цікавлять, крива частоти виявить нам ліпше характер розподілу частоти цілої маси ніж гістограм або полігон, бо вирівнює випадкові неправильності розподілу, що викликані малою кількістю частоти сукупности, в той час як полігон чи гістограм відбиває на собі всі сліди випадковости вибору даних певної сукупности.

В короткому курсі основ статистичної методи не можемо детально зупинятись на характеристиці типів кривих, наприклад в тому розмірі, як це подає відомий біометрик Пірсон. Вважаємо одначе необхідним подати лише важливіші типи розподілу частоти, з якими статистикам на практиці часто доводиться зустрічатись, не зупиняючись на їх математичній аналізі.

26. Типи розподілу частоти. а/ Симетричний /співмірний/ розподіл частоти. Під цим типом будемо розуміти такий розподіл частоти, при якому частоти кляс змінної співмірно спадають по обох сторонах найбільшої частоти кляси. Прикладом цього типу можуть служити наступні розподіли частоти: /дивись табл. 7 на стор. 88/.

Графічне зображення таблиці 7-А симетричного типу показує нам розподіл частоти приблизно співмірної форми. Полігон частоти складений з трохи ширшими інтервалами, не беручи на увагу його дрібних границь: 56 I5/I6 - 57 I5/I6; 57 I5/I6 - 58 I5/I6 і т.д., але 57 - 58, 58 - 59 і т.д. Друге графічне зображення /малюнок 5 / подає теж приблизно співмірний розподіл частоти, але менш правильний; все ж таки і цей розподіл можемо цілком визнати приблизно співмірним - симетричним.

Вищенаведені приклади взяті з статистичного дослідження явищ біологичної природи; в явищах соціяльно-економічної природи з співмірною формою розподілу частоти статистикам зустрічатись доводиться рідко. Це цілком зрозуміло, бо ще в I-й лекції, характеризуючи природу атипових колективних явищ, ми вказували на те, що соціяльно-економічні явища, досліджувати які статистикові доводиться найчастіше, є разом з тим і найбільше складні явища, в яких діють випадкові умови і причини, що затемнюють основні причини чи умови, керуючі цими явищами. Статистичні ряди і зображення таких змінних через те й дають найчастіше ту або іншу форму асиметрії.

Таблиця 7.

А

Розподіл частоти висоти тіла дорослих чоловіків, народжених в Англії і в Ірляндії.

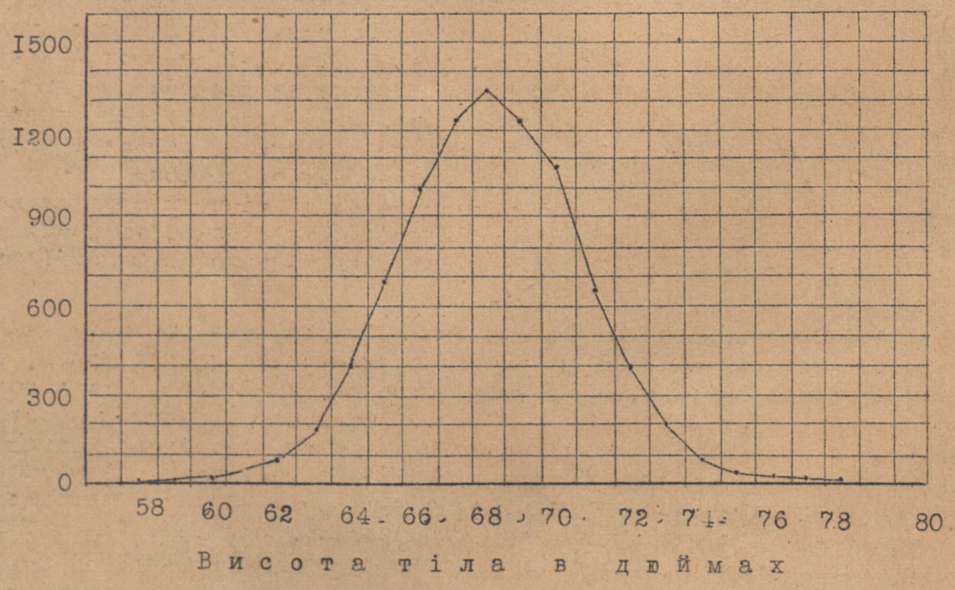
Висота тіла в дюймах /босих/.	Число осіб з відповідновисотою
57 -	2
58 -	4
59 -	I4
60 -	4I
6I -	83
62 -	I69
63 -	394
64 -	669
65 -	990
66 -	I223
67 -	I329
68 -	I230
69 -	I063
70 -	646
7I -	392
72 -	202
73 -	79
74 -	32
75 -	I6
76 -	5
77 -	2
Разом	8.585

В.

Розподіл частоти висоти тіла Кембріджських студентів.

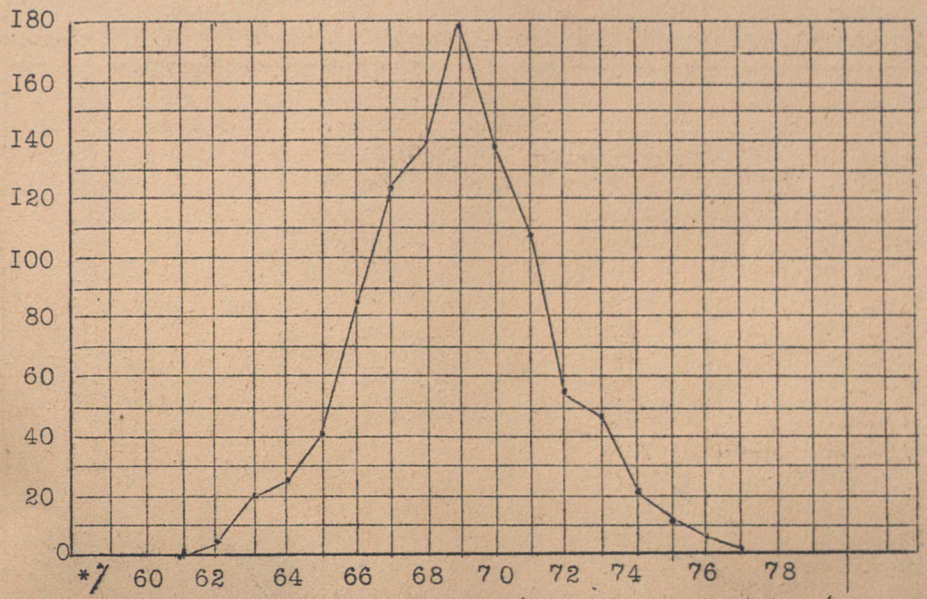
Висота тіла в дюймах.	Число студентів відповідної висоти.
59,5 - 60,5	-
60,5 - 6I,5	-
6I,5 - 62,5	4,0
62,5 - 63,5	I9,0
63,5 - 64,5	24,5
64,5 - 65,5	40,5
65,5 - 66,5	84,5
66,5 - 67,5	I23,5
67,5 - 68,5	I39,0
68,5 - 69,5	I79,0
69,5 - 70,5	I38,5
70,5 - 7I,5	I08,0
7I,5 - 72,5	53,5
72,5 - 73,5	47,5
73,5 - 74,5	2I,0
74,5 - 75,5	I2,0
75,5 - 76,5	5,0
76,5 - 77,5	0,5
77,5 - 78,5	-
78,5 - 79,5	-
Разом	I.000

Малюнок 4. / Для таблиці 7-А /..



х/ Матеріял поданий у Юла: дійсні інтервали: 56 I5/I6 - 57 I5/I6; 57 I5/I6-58 I5/I6 і т.д.

М а л ю н о к 5. / Для таблиці 7-В./.



б/ Мало - асиметричний тип /мало - неспівмірний/ розподіл частоти.

Другим основним типом розподілу частоти треба вважати такий, в якому частоти убувають /спадають/ по обох сторонах від максимума з нерівномірною швидкістю. При тому асиметрія може бути більшою або меншою. Такий розподіл частоти є найчастіший в практиці статистичних досліджень, особливо соціяльно-економічних.

Прикладом цього типу розподілу частоти може служити розподіл висоти тіла I495 учнів, поданий на малюнку I-му /гістограм/ стор.85 /.

в/ Дуже асиметричний розподіл частоти. Нарешті, третім типом розподілу частоти якоїсь змінної є дуже асиметрична, дуже неспівмірна форма розподілу, при якій максимальна частота лежить на одному кінці розхилу змінної і частоти в напрямку до другого кінця розхилу постійно убувають. В економічній статистиці часто такий розподіл частоти зустрічаємо в розподілах частоти осіб по ознаці величини їх прибутків або маєтку. В становити точну межу між мало-асиметричною і дуже-асиметричною формами розподілу досить тяжко. Найчастіш у випадку більшої деталізації першого інтервалу клясового з'ясовується, що максимальна частота знаходиться не на кінці розхилу змінної, а на певній віддалі від цього кінця. Рішення питання про те, до якої форми асиметрії треба віднести розподіл, залежить часто на тому, як детально проведено розділ на кляси цілого розхилу.

Прикладом дуже-асиметричної форми розподілу частоти може служити розподіл річної прибуткової вартости та числа маєтків осіб, що брали участь в повстанню Яковитів в I7II році /матеріял поданий за Юлом. Таблицю 8. та малюнок 6 див. на стор. 90 /.

Приклад показує, що при даному способі розділення на інтервали розподіл частоти цілком можливо віднести до дуже-асиметричної форми, бо число маєтків між 0 та I00 ф.ст. річного прибутку є більше в шість разів ніж число маєтків у клясі від I00 - 200 ф.ст. при тому разом із зростом прибуткової вартости, число маєтків постійно убуває. Детальний розгляд даних одначе показує, що в першій клясі найбільша частота не відноситься до розміру прибутковости, що лежить близько нулі, але максимум частоти в дійсності припадає на маєтки з річною прибутковістю розміром I ф.ст. I5 шілінгів, але все ж таки цю форму слід віднести до дуже-асиметричної.

Такої ж приблизно форми розподілу частоти набувають таблиці даних розподілу осіб за розміром їх загального маєтку, що публікуються в збірниках різних країн, але в них, на превеликій жаль, майже без виключення зістрічаємо надзвичайно нерівномірний розмір інтервалів окремих кляс змінної.

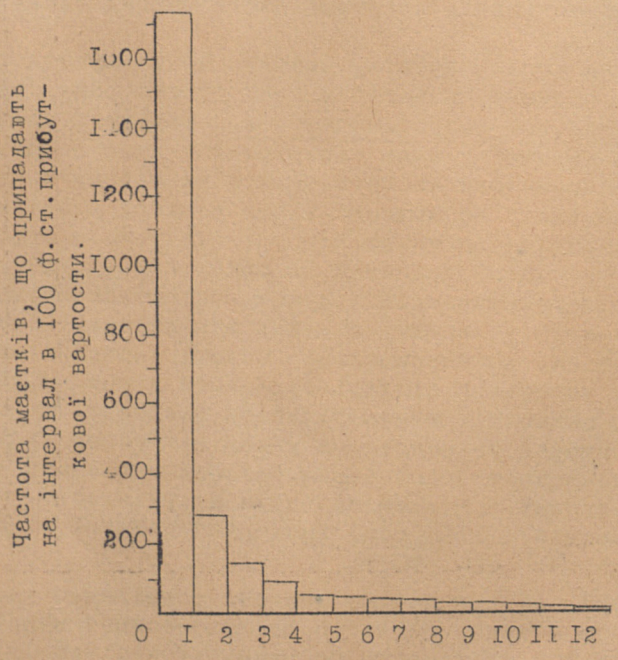
---

*/ На абсцисі подано висоту тіла в дюймах, а на прямовісній ординаті число студентів.

Таблиця 8.

Річна прибуткова вартість у ф.ст.	Число маєтків	Річна прибуткова вартість у ф.ст.	Число маєтків.
0 - I	I726,5	I7 - I8	I
I - 2	280	-	-
2 - 3	I40,5	20 - 2I	4
3 - 4	87	2I - 22	I
4 - 5	46,5	22 - 23	I
5 - 6	42,5	23 - 24	I
6 - 7	29,5	-	-
7 - 8	25,5	27 - 28	2
8 - 9	I8,5	-	-
9 - I0	2I	3I - 32	I
I0 - II	II,5	-	-
II - I2	9,5	39 - 40	I
I2 - I3	4	-	-
I3 - I4	3,5	45 - 46	I
I4 - I5	8	-	-
I5 - I6	3	48 - 49	I
I6 - I7	5	Разом	2.476

Малюнок 6.



г/ Тип кривої зворотній асиметричній або мало-асиметричній формі розподілу частоти. Поза основними трьома вказаними вище типами розподілу частоти зустрічаємо, правда дуже рідко, і окремий тип розподілу в вигляді /формі/ кривої, подібної до латинської літери V , при якому на низькі й високі вартості змінної припадає найбільше число одиниць сукупности, а на середні вартості менше. Цей тип і можна назвати зворотньо-співмірним або зворотньо мало-асиметричним.

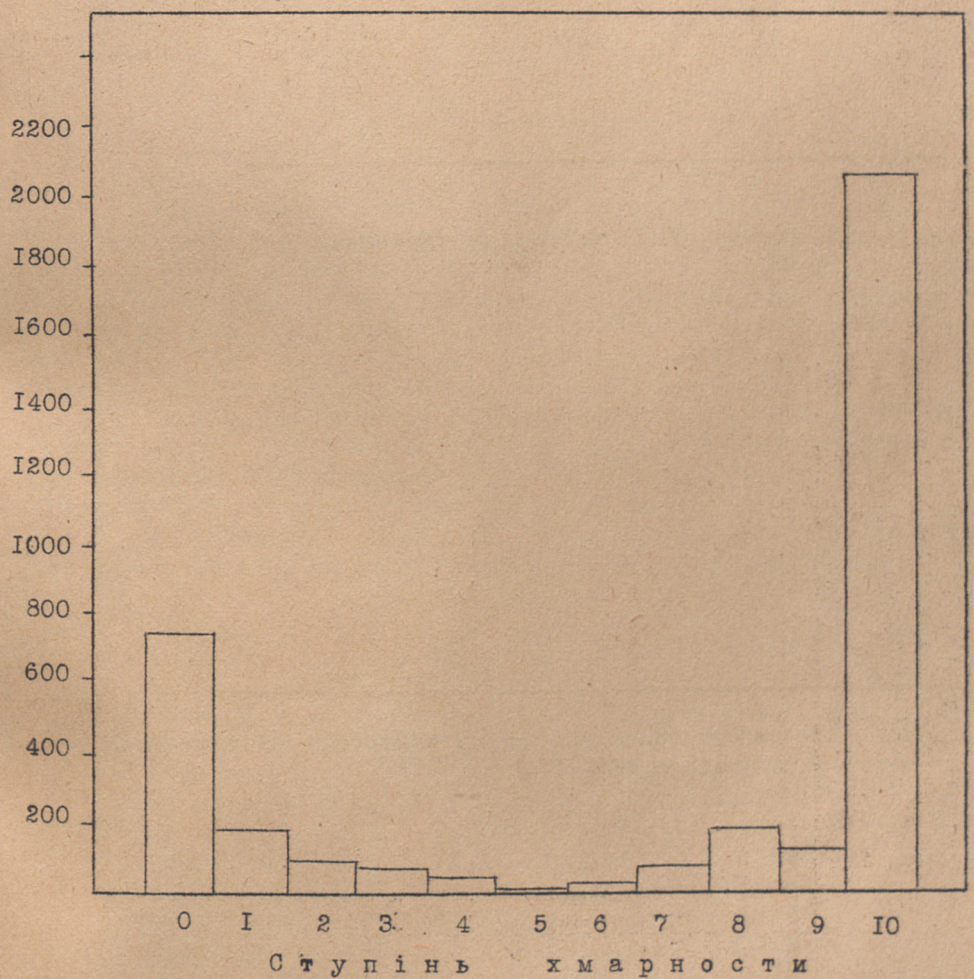
Прикладом може служити нам оброблені К. Пірсоном наступні метеорологічні дані /наводимо за Юлом/.

Таблиця 9.

Ступінь хмарности	Число днів	Ступінь хмарности	Число днів
0	75I	6	2I
I	I79	7	7I
2	I07	8	I94
3	69	9	II7
4	46	I0	2.089
5	9	Разом	3.653

Малюнок 7.

Розподіл 3.653 днів за хмарністю.

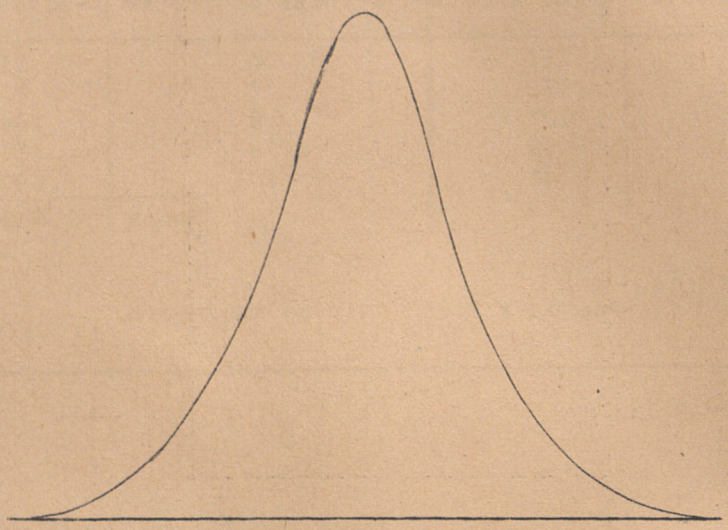


Отже велике число днів припадає на мінімальну хмарність, в той же час до середини розподілу змінної число днів значно зменшується /падає/, але в кінці розподілу знову росте, досягаючи в останньому інтервалі максимальної вартости, а крива набуває форми латинської літери U

Зображуючи всі чотири основні типи розподілу частоти в ідеальних геометричних формах будемо мати наступні фігури:

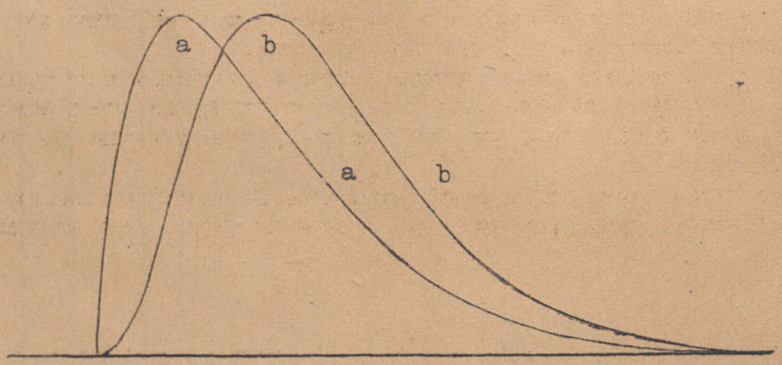
I. Сіметрична /співмірна/ крива розподілу частоти.

М а л ю н о к 8.



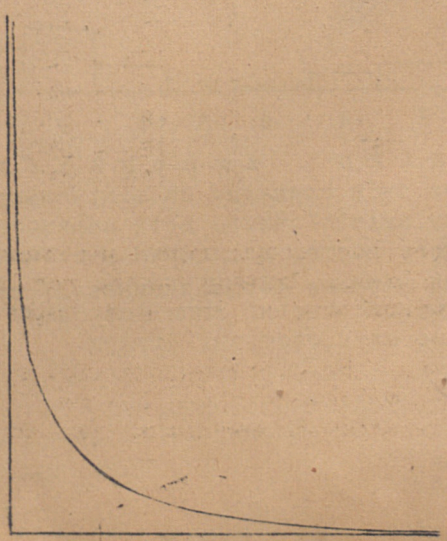
II. Мало-асиметрична крива розподілу частоти.

М а л ю н о к 9.



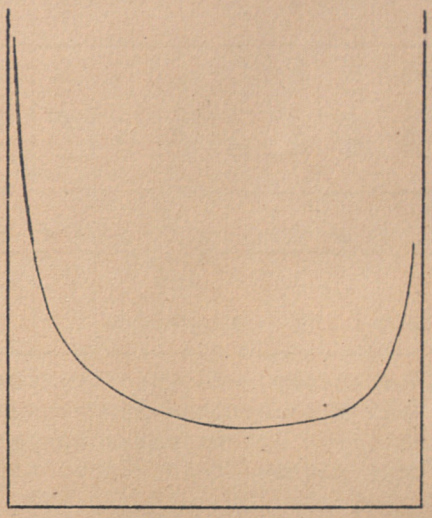
III. Дуже асіметрична крива розподілу частоти, яка інколи визначається в вигляді літери " J "

М а л ю н о к IO.



IУ. Розподіл частоти форми " U "

Малюнок II.



Крім наведених типів розподілу частоти зустрічаємо інколи і таку форму розподілу, в якій крива має не один максімум /вершину/, а два і більше. Неправильності в руху частоти від інтервала до інтерва у можуть бути цілком випадкового характеру, не виключена можливість, що ці неправильності счезнуть, коли збільшимо число спостережень..Можуть, одначе, ці неправильності бути сутєвими, дійсними, себто випливати з самої природи явища, бути характерною властивістю явища, яке досліджуємо.

Такий двовершинний або многовершинний розподіл частоти найчастіше свідчить про - неоднородність матеріялу з погляду спостереженої ознаки та про наявність двох й більше груп, що зосереджуються кожна на різних утинках розхилу ознаки.

Візьмемо як приклад розподіл частоти I65 старопразьких черепів, без розділу статі, за шириною тильної части голови /матеріял наведено за С.Коном./

Таблиця IO.

Ширина тилу в мм.	Частота.	Ширина тилу в мм.	Частота.
96 - 97	I	II2 - II3	26
98 - 99	-	II4 - II5	27
I00 - I0I	-	II6 - II7	I4
I02 - I03	4	II8 - II9	6
I04 - I05	II	I20 - I2I	7
I06 - I07	I4	I22 - I23	5
I08 - I09	3I	I24 - I25	3
II0 - III	I5	I26 - I27	I

Тут спостерігаємо два центри скупчення частоти: в інтервалі I08 - I09 мм. та інтервалі II4 - II5 мм. Причину такого розподілу не тяжко є висвітлити, коли дану сукупність розділимо за статтю на черепи чоловічі і жіночі./Дивись таблицю II на стор.94/.

В цій таблиці розподілу частоти максімум в інтервалі I08 - I09 дають жіночі черепи, а чоловічі в інтервалах II2 - II3 та II4 - II5. Змішання обох сукупностей в одну і мало наслідком двовершинний розподіл частоти.

Таблиця II.

Стать	Ширина тилу в м. м.															
	96- 97	98- 99	I00-I0I	I02-I03	I04-I05	I06-I07	I08-I09	II0-III	II2-II3	II4-II5	II6-II7	II8-II9	I20-I2I	I22-I23	I24-I25	I26-I27
	Число черепів.															
Чоловіки	-	-	-	2	5	8	I4	I0	2I	2I	I0	5	7	5	3	I
Жінки	I	-	-	2	6	6	I7	5	5	6	4	I	-	-	-	-

Неоднорідність матеріялу може мати наслідком не тільки згаданого типу розподіл частоти, але й форму мало-асиметричну. Допустимо, наприклад, що наша сукупність складається з двох однакових кількісно частей, характеризованих співмірним /симетричним/ розподілом, але з різними центрами згромадження одиниць.

I частина: 2. 5. I3. I8. I3. 5. 2. - - -

II частина: - - - 2 5 I3. I8. I3. 5. 2

В сукупності маємо двовершинний розподіл.

Припустимо одначе, що кожна з обох частин сукупности характеризується тою ж самою формою розподілу як раніше, але одна частина сукупности є вдвічі частіша ніж друга:

I частина: 2. 5. I3. I8. I3. 5. 2. - - -

II частина: - - - 4. I0. 26. 36. 26. I0. 4.

Ціла сукупність має ряд: 2. 5. I3. 22. 23. 3I. 38. 26. I0. 4.

Отже в такій умові сукупність має форму одновершинну мало-асиметрічну з максімумом, що міститься в правій стороні розподілу - асиметрія правостороння.

З антропологічного матеріялу в праці проф. Матейки / Антрополо - гія №№3 стор. I89, I924 р./ можна бачити приклад розподілу чоловічих і жіночих черепів, який відрізняється тим, що черепи чоловіків заступлені в сукупності більшою кількістю ніж жіночі. Спостереженою ознакою є поземий обвід черепів.

Таблиця I2.

Стать	Поземий обвід черепу в мм.									
	470- -479	480- -489	490- -499	500- -509	5I0- -5I9	520- -529	530- -539	540- -549	550- -559	560- -569
	Число черепів									
Чоловіки	----	2	II	20	28	26	I2	I2	3	I
Жінки	2	3	I7	20	I0	3	-	-	-	-
Разом	2	5	28	40	38	29	I2	I2	3	I

Ціла сукупність дає ясно мало-асиметричний розподіл частоти така форма розподілу може виникнути розуміється і через інші причини, ніж неоднорідність матеріялу; складна проблема з'ясування причин, що викликають ту чи іншу форму розподілу частоти змінної ще досить мало опрацьована в науковій статистиці, через це й не можемо тут більше зупинятись над нею.

Лекція дванадцята.

27. Середні /пересічні/ величини.

Подані вище таблиці розподілу частоти / таблиці фреквенції/, а також різні форми їх графічного зображення /лекція IO-II/, в значній мірі полегшують нам скласти уяву про статистичну сукупність, що складається з великого числа одиниць, спостережена квантитативна ознака яких набуває різних числових вартостей. Це полегшення для одержання зручної загальної уяви про такого роду сукупності, в зв'язку з спостереженими ознаками для кожної одиниці в сукупності, одначе не може нас цілком задовольнити. Мусимо відшукати таку характеристику сукупности й для цілого роду сукупностей, яка б давала можливість єдиною величиною, одним числом, дати нам образ розподілу вартостей ознаки в сукупності, а крім того дати й можливість надійного, легкого порівняння різних сукупностей між собою щодо спостереженої в них певної ознаки. Таку характеристику сукупности можна дати перш за все з погляду центральної висоти, рівня, який мають чи набули вартості квантитативної ознаки в сукупності.

Справа в тім, що є можливий однаковий тип розподілу частоти ознаки в сукупності при різній центральній висоті вартостей ознаки, себто за однакового типу кривої розподілу, центри розподілу, що міряються віддаленням по вісі абсцис, значно різняться в порівнюваних сукупностях. Коли візьмемо, наприклад, сукупність дорослих людей в порівнянні з сукупністю дітей щодо їх висоти тіла, то збудовані криві фреквенції обидві будуть мати однаковий симетричний тип, але в першій з них максимум частоти ознаки буде купчитись біля I60- I70 см./висота тіла/, а другій біля I20-I30 см. Себто при однаковому типу розподілу частоти будуть криві взаємно різнитись своїм положенням на вісі абсциси, а то так, що перша./дорослих/ буде на вісі від початку її уміщена далі.

Для наочного уявлення такого розподілу подамо тут дві симетричні криві однакового типу, проте розміщених навколо різних центрів, на підставі даних спостереження над зростом японських і американських салдатів./подаємо за проф. Віхляєвим/.

Таблиця I3.

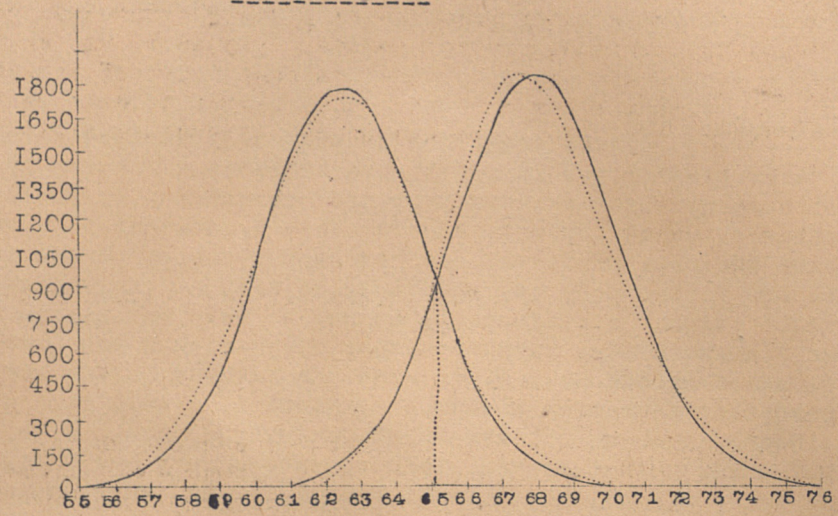
розмір росту в дюймах	число салдатів японських	число салдатів американ.	розмір росту в дюймах	число салдатів японськ.	число салдатів американ.
55,5 - 56,5	47	-	66,5 - 67,5	208	I822
56,5 - 57,5	I25	-	67,5 - 68,5	64	I695
57,5 - 58,5	3I6	-	68,5 - 69,5	I2	I294
58,5 - 59,5	640	-	69,5 - 70,5	-	868
59,5 - 60,5	I065	-	70,5 - 7I,5	-	5I0
60,5 - 6I,5	I486	-	7I,5 - 72,5	-	263
6I,5 - 62,5	I730	38	72,5 - 73,5	-	II4
62,5 - 63,5	I698	I92	73,5 - 74,5	-	42
63,5 - 64,5	I328	538	74,5 - 75,5	-	I2
64,5 - 65,5	839	I055			
65,5 - 66,5	442	I557		I0.000	I0.000

/ Малюнок I2. див. на слідуючій сторінці/

Отже треба мати таку характеристику, яка б давала нам ясне положення кривої на вісі абсциси, себто центральну висоту вартостей ознаки в певній сукупності. Такою характеристикою й є с е р е д н і в е л и ч и н и.

Раніш ніж перейдемо до опису різних видів середніх і способів їх обчислення /вираховування/ мусимо зупинитись на деяких загальних їх властивостях.

Малюнок 12.



Дійсний та нормальний розподіл за ростом японських і американських салдатів.

I/ Об'єктивна і суб'єктивна /ідеальна/ середня величина. Треба розрізняти ті середні, що одержані від повторного, декількаразового виміру якого-будь предмету. Наприклад, можна 10,20,30 разів зважити якийсь предмет, або певну особу і при кожному зважуванні одержимо ряд неоднакових величин ваги, хоч і близьких між собою. Різниця ваги в окремих випадках зважування пояснюється тим, що абсолютної точности й рівности кожне зважування дати не може. Склавши всі такі величини ваги і, поділивши їх на число разів зважування, одержимо дійсну, пересічну/середню/ вагу певного предмету чи особи. Така середня й зветься об'єктивною.

Але можемо міряти, наприклад, ріст 1000 новобранців; кожний з новобранців буде мати різну кількість сантиметрів росту, але склавши всі величини росту новобранців і поділивши їх на 1000, одержимо також пересічну величину росту новобранця, припустім, 1,5 метра, яка буде величиною абстрактною, лише близькою до конкретного росту великої кількости цих новобранців, що дає нам цілком добру уяву про пересічний ріст для цілої маси цілої сукупности. Така середня й зветься середньою суб'єктивною. Коли ми кажемо, що середня величина родини становить 5,8 душ, або що пересічно на одне господарство припадає 2,5 робітника, 150,5 пудів пшениці і т.п.,ми завжди маємо діло з суб'єктивними середніми. В статистиці найчастіше зустрічаємось з такого роду середніми.

Отже середню величину визначається, так би мовити, фактом її топографичного положення в середині, в центрі варіяційного ряду, а відповідна її ордината перетинає криву, що зображає геометрично цей ряд, в якій будь точці. Одначе з багатьох можливих центрів, для кожного з яких можна було б подати властиве йому матиматичне визначення, статистична метода цікавиться лише тими, що відповідають певним вимогам середніх величин, а саме: середня величина повинна спиратись у своєму виводі на всю кількість зроблених спостережень; 2/ вона повинна викликати певне конкретне уявлення щодо спостереженої сукупности і 3/ вивід її мусить бути математично якнайпростішим.

2/ Аритметична середня. Таким вимогам у найбільший мірі відповідає а р и т м е т и ч н а с е р е д н я, як один із видів родового поняття середніх величин. Аритметична середня є сума числових вартостей /частот/ ознаки у всіх одиниць статистичної сукупности, поділена на число цих одиниць. Маємо, наприклад, десять господарств, а ознакою, про яку нам ходить буде кількість коней. Ця кількість в кожному господарстві, наприклад є та-

ков: 4,I,3,?,0,4,6,5,8,7, що в сумі дає сорок коней. Поділивши цю суму на десять, одержимо аритметичну середню 4,0, яка й характеризує нам пересічну висоту вартостей ознаки /кількість коней/ у даній сукупності господарств.

Коли зазначимо числову вартість ознаки сукупности, що складається з N одиниць, літером "а", притому вартість ознаки першої одиниці означимо a_1, другої a_2 і т.д., а останьої a_n, саму ж середню величину через літеру M /media /, то математичне означення аритметичної середньої дасть наступну формулу:

$$M = \frac{a_1 + a_2 + a_3 \dots\dots + a_n}{N}$$

Користуючись грецьком літером $\sum$ для означення суми, можемо вищенаведену формулу подати в більш простій формі:

$$M = \frac{\sum a}{N}$$

З вищенаведеного випливає, що аритметична середня не дає нам уяви про окремі одиниці сукупности, а лише про цілу сукупність і в цьому розумінні аритметична середня є типовою характеристикою сукупности. За влучним виразом проф. О р ж е н ц ь к о г о її можна також назвати і "зведеною ознакою".

З поданого до цього в наших лекціях з'ясування статистики, власне статистичної методи випливає, що студіює вона завжди сукупності об'єктів у цілому, а не одиниці, що належать до цих сукупностей, хоч початкове спостереження і спирається на дані, що стосуються до окремих одиниць об'єктів.

Таким чином аритметична середня є п е в н е в і д н о ш е н н я, в якому чисельником буде завжди сума числових вартостей ознаки у всіх одиниць сукупности, а знаменником число одиниць даної сукупности. Той факт, що аритметична середня є певним відношенням, ставить перед нами питання: чім вона відрізняється від інших статистичних величин, що визначають також відношення, себто від релятивних /відносних/ чисел, а з окрема від чисел відносної частоти, з якими ми вже ознайомились вище?/ лекція IO /. Не рідко дослідники не усвідомлюють собі межі, що відокремлює середню від інших релятивних величин. З погляду цієї відрубности за критеріум можна взяти те, що аритметична середня є характеристикою сукупности, що узгляднює властивість квантитативно мінливих ознак, себто таких, при яких нас цікавить не факт їх наявности або відсутности, але числові вартості, яких ці ознаки набувають. Цікавить дослідника лише факт наявности або відсутности ознаки в одиницях сукупности, характеризує він цей факт відносною частотою, себто відношенням числа одиниць сукупности, що мають певну ознаку, до цілої маси одиниць даної сукупности. Коли цікавлять нас різні числові вартості ознаки в сукупності, характеризуємо цю ознаку аритметичною середньою, себто відношенням суми тих числових вартостей, яких набуває ознака у всіх одиниць, до цілої маси одиниць сукупности. При цьому для нас не грає ролі чи дана квантитативна ознака є наявною у всіх одиниць сукупности, чи в деяких з них вона відсутня, себто приймає вартість "О". На вартості "О" треба дивитися однаково, як і на решту числових вартостей квантитативної ознаки і наявність таких вартостей не позбавляє даної статистичної величини характеру аритметичної пересічної.

Декотрі статистики-теоретики бачать критеріум відрубности між аритметичною пересічною й відносною частотою в тому, що в першій мусить ознака бути наявною у кожної одиниці /може тільки міняти числову вартість/, а в другої є ознака наявною лише в певної частини одиниць, що не відповідає дійсности. Щоб відмежувати аритметичну пересічну від релятивної частоти, досить мати на увазі, що перша величина є характеристикою сукупности з погляду її квантитативних ознак, тоді як друга величина є характеристикою з погляду її квалітативних ознак.

Щоб відмежувати аритметичну пересічну /середню/ від інших релятив-

них чисел, треба усвідомити собі, що чисельник її є сумою числових вартостей ознаки всіх одиниць сукупности, число яких є в знаменнику, себто чисельник охоплює суму певних числових характеристик окремих одиниць знаменника. Через те чисельник аритметичної пересічної може бути розкладений на ряд складників, з яких кожний може бути застосований до певної одиниці знаменника, яко її характеристика. Ця властивість аритметичної пересічної дозволяє нам відмежувати її від тих відносних чисел, які служать для наочности порівняння обсягу різних явищ та визначаються у відсотках /див.лекцію IO стор. 77/ та в яких немає зв'язку між одиницями чисельника й знаменника. Тою самою властивістю відрізняється аритметична персічна й від відношень інтенсивности, бо й в цих останніх величинах одиниці, що складають чисельник, не можуть бути застосованими до окремих одиниць знаменника, але лише до цілих груп одиниць знаменника /див.лекцію IO стор.76 /.

а/ Властивості аритметичної середньої. Першою властивістю аритметичної середньої є та, що сума відхилень вартостей ознаки у всіх одиниць сукупности від аритметичної середньої рівняється нулю. Суму тут беремо в розумінні альґебричному,себто з огляду на знак / -, + /, при цьому відхилення нагору мають знак плюс, а відхилення донизу знак мінус. Дійсно, означимо літерою X спостережену квантитативну ознаку. X є величиною змінною й може набувати вартості X_1, X_2, X_3- X_n.

Означимо далі аритметичну середню літерою M, а літерами x_1, x_2, x_3- x_n розділи /відхилення/ між окремими вартостями ознаки й аритметичною середньою, тоді будемо мати такі рівняння:

$$\begin{aligned} X_1 - M &= x_1 \\ X_2 - M &= x_2 \\ X_3 - M &= x_3 \\ &\cdots \\ &\cdots \\ X_n - M &= x_n \end{aligned}$$

Підсумуємо праву й ліву сторону всіх рівнянь і одержимо:

$$\sum X - nM = \sum x$$

Але $\sum X$, себто сума вартостей ознаки у всіх одиниць сукупности, як випливає з дефініції аритметичної середньої,рівняється:

$$M = \sum X/n; \quad , \quad \sum X = n.M.$$

Підставимо це значіння $\sum X$ в рівняння й одержимо тоді, що

$$nM - nM = \sum x$$

себто $\sum x = 0$. / I /

З цією важливою властивістю аритметичної середньої будемо зустрічатися не раз.

Другою властивістю аритметичної середньої є, що за поміччю аритметичних середніх вартостей ознаки в окремих частинах чи групах даної сукупности, можемо одержати аритметичну середню цілої сукупности.

Коли поділимо певну сукупність на декілька частин /груп сукупностей/, то, щоб одержати на підставі аритметичних середніх цих частин аритметичну середню цілої сукупности, треба аритметичні середні окремих груп помножити на частоти цих груп та суму цих добутків поділити на частоту цілої сукупности.

Означивши аритметичну середню цілої сукупности літерою M, а частоту цієї сукупности через N , аритметичну середню першої частини цієї сукупности через M_1, а її частоту через N_1, аритметичну середню другої частини цієї сукупности через M_2, а її частоту через N_2- будемо мати

$$M = \frac{M_1N_1 + M_2N_2}{N}$$

Дійсно, коли назвемо вартості, що може приймати ознака в цілій сукупності літерою X, а ті вартості, що може приймати ознака в першій частині сукупности через X_1, а в другій частині через X_2 розуміється:

$$\sum X = \sum X_1 + \sum X_2$$

З дефініції аритметичної середньої випливає, що $\sum X$ є аритметична середня помножена на число одиниць цілої сукупности - себто:

$$\sum X = M.N$$

так само:

$$\sum X_1 = M_1.N_1$$

$$\sum X_2 = M_2.N_2$$

відціль: $M.N = M_1.N_1 + M_2.N_2$

Поділивши ціле рівняння на N одержимо:

$$M = \frac{M_1.N_1 + M_2.N_2}{N}$$

Звичайно це буде правдивим не тільки у випадку поділу на дві части певної сукупности, а й для поділу на яку хочемо кількість груп цієї сукупности, через те для довільного числа груп сукупности будемо мати:

$$M = \frac{M_1.N_1 + M_2.N_2 + M_3.N_3 \ldots\ldots + M_8.N_8}{N} \quad / 2 /$$

Це й є формулювання другої властивости аритметичної середньої, що зазначена вище.

Наприкінці, мусимо ще зазначити, що аритметична середня суми або різниці вартостей двох ознак сукупности є рівною сумі або різниці аритметичних середніх цих ознак. Коли одна ознака є X, друга Y, а їх сума або різниця Z то маємо для кожного члена сукупности

$$Z = X \pm Y$$

а для всіх членів: $\sum Z = \sum X \mp \sum Y$

коли поділимо на частоти сукупности, будемо мати: $M_Z = M_X \mp M_Y$

Це формулювання можна, розуміється, поширити й на більшу кількість ознак сукупности.

в/ Способи вирахування аритметичної середньої. З поданої нами формули аритметичної середньої ясно, що одержати її можемо, підсумувавши вартості ознаки у всіх одиниць сукупности й поділивши цю суму на число одиниць сукупности. Одначе така операція є дуже легкою в тих випадках, коли частота сукупности не дуже велика, в противному випадку обчислення займає багато часу. Правда, що в деяких випадках підрахування частоти сукупности є потрібними незалежно від необхідности виведення аритметичної середньої, або це підрахування вже зроблено раніше, себто маємо його готовим. Але, коли цього бракує, а є необхідним аритметичну середню обчислити, можливо виведення її зробити простішим способом, а почасти приблизно.

Маємо випадок, коли ознака змінюється переривно й набуває невеликої кількости вартостей. Наприклад, в 152 господарствах, в яких спостерігаємо стан молошних корів, хочемо обрахувати пересічне число молошних корів у господарстві. Складемо наступну таблицю розподілу частоти для нашої ознаки /кількість молошних корів у господарстві. Таблицю 14 див. на слід.стор/.

До наведеної таблиці розподілу частоти додано окремо графу 3-ю, в якій обраховано кількість вартостей ознаки для кожної кляси /графа перша/ через помноження вартости ознаки кожної кляси на її частоту /xf /. Сума добутків дасть нам цілу суму вартостей ознаки в сукупності. Поділивши цю суму на число господарств і одержимо середню аритметичну / 3,7 /. Спосіб цей безперечно є зручний, притому дає в результаті не приблизну, а точну величину середньої аритметичної.

Таблиця I4.

Число молошних корів у господарстві x	число господарств f	добуток xf
I	2	3
I	2	2
2	27	54
3	52	I56
4	36	I44
5	2I	I05
6	7	42
7	3	2I
8	2	I6
9	I	9
I0	-	-
II	-	-
I2	-	-
I3	I	I3
	I52	562:I52 3,7

В тих випадках, коли ознака міняється неперервно - є змінною тягло, або коли розхил її перервних вартостей є завеликий, мусимо, як це вже нам відомо /лекції I0-а й II-а стор.80-83 / в таблиці розподілу частоти подати змінну в клясових інтервалах /від - до / і до кожної кляси обчислити відповідаючі ним частоти. В таких випадках ми маємо діло лише з більш-менш вузькими границями, в яких вартості ознаки певної кляси будуть знаходитись у кожному інтервалі клясовому. Коли б ми знали аритметичні середні вартостей ознаки, що знаходяться в кожному інтервалі клясовому, могли б безпосередньо використати другу властивість середньої аритметичної, що формулювали ми вище, себто помножити середні кожної кляси на частоти цих кляс, підсумувати одержані добутки, а їх суму поділити на частоту цілої сукупности. Але нам дано лише границі інтервалів, через те можемо лише допустити що аритметична середня вартостей ознаки в кожному інтервалі клясовому є рівною середній вартості відповідного інтервалу клясового, одержаній яко півсума з нижчої й вищої границь інтервалу, або з мінімальної й максимальної вартости.

Візьмемо для прикладу змінну тяглу, що подали ми на стор.85 в 6-й таблиці розподілу частоти. В ній кожний інтервал клясовий є поданий в границях, розхил яких складає 5 см. Перша кляса I00-I04 см., друга I05-I09 см. і т.д., а ознака міряється з точністю до одного см.,отже, ясно, що середня вартість першого інтервалу буде рівною

$$\frac{I00 - I04}{2} = I02, \quad \text{другого} \quad \frac{I05 - I09}{2} = I07 \quad \text{і т.д.}$$

Ці середні вартості інтервалів клясових, що подаємо нижче в графі першій таблиці I5-ої, і будемо вважати за аритметичні середні вартостей ознаки, що знаходяться в границях кожного інтервалу, хоч таке припущення розуміється, зв'язане з певною неточністю; тоді таблиця розподілу частоти прийме такий вигляд: /див. таблицю I5 на слід.сторінці/.

Спираючись на д р у г у в л а с т и в і с т ь аритметичної середньої, можемо суму добутків 206485 поділити на I495 /суму частоти сукупности/ і одержимо I38,I2 см.; цю величину можемо вважати за майже рівну середній аритметичній. Наше припущення спиралось на те, що середня величина вартостей ознаки в кожному інтервалі є рівною, або знаходиться в середині інтервалів клясових. Аналогічно й в тому випадку, коли припустимо, що всі

вартості ознаки, що знаходяться в інтервалі клясовому будуть однакові й рівні середній величині відповідного інтервалу, також прийдемо до однакового результату, себто в обох припущеннях середня аритметична буде однаковою. Але треба здати собі справу в тому, за яких умов наші припущення будуть відповідати конкретній дійсності, конкретному розподілу вартостей в кожнім інтервалі клясовому.

Таблиця I5.

ріст в см. x	частота f	добуток xf	
I02	2	204	
I07	7	749	
II2	30	3360	
II7	69	8073	
I22	I03	I2566	
I27	I49	I8923	
I32	I94	25608	
I37	229	3I373	
I42	250	35500	
I47	2I5	3I605	
I52	I30	I9760	
I57	64	I0048	
I62	33	5346	
I67	I4	2338	
I72	6	I032	206.485:I495 = I38,I2 см.
	I.495	206.485	

Розуміється, що припущення наше буде тим правдивішим, чим вужчі будуть гряниці інтервалів клясових. Цим між іншим пояснюється й та вимога, яку ми ставили вище /лекція I0 стор.80 літ.а/ при викладі міркувань про будову таблиць частоти для змінної тяглої /себто з інтервалами клясовими/. Безперечно також, що аритметична середня тих вартостей змінної /ознаки/, що припадають на кожний даний інтервал, буде рівнятись середній цього інтервалу в тих випадках, коли вартості ознаки таких окремих екземплярів, розкладаються по цілій ділянці інтервалу / в його границях/ рівномірно. Наприклад, коли на інтервал клясовий від I25 - I29 якоїсь змінної припадає припустимо, 50 екземплярів /одиниць змінної/ рівномірність їх в інтервалі цілком вважалася досягнутою, коли б на частину інтервалу від I25-I27 припадало 25 екземплярів /випадків/, а на другу частину від I27-I29 теж 25 випадків.

Коли ж зустрічаємось з таким розподілом частоти змінної, в якому число випадків /екземплярів/ росте поступово від нижчих інтервалів до вищих, цілком справедливо буде припускати, що це зростання відбувається й в границях інтервалу. Тоді ясно, що умови рівномірности розподілу вартостей в інтервалі не має і коли все ж таки ми припустимо, що цей розподіл рівномірний, а середні вартості інтервалу приймемо за аритметичні середні, ми зменшимо в цьому випадку систематично оті середні аритметичні проти їх дійсної вартости. Аналогічно цьому, але зворотнім буде й наслідок тоді, коли число екземплярів змінної спадає від початку розподілу частоти в таблиці до його кінця; цьому спаданню буде відповідати також спадання кількости екземплярів змінної в границях кожного інтервалу, а аритметичні середні будуть збільшені проти їх дійсної вартости.

Відціль робимо висновок, що поки таблиця співмірна, себто частоти її в першій половині розхилу змінної ростуть, а в другій половині з тією ж скорістю /в тій же мірі/ спадають, випадки зменшення клясових середніх в одній частині таблиці компенсуються збільшенням клясових середніх в другій, а через те й хиба в обчисленні аритметичної середньої буде малою. Навпаки, при неспівмірности такого росту й спаду хиба буде великою і тоді вказаний нами спосіб виведення аритметичної середньої буде не точний і він буде тим більше неточним, чим більше розподіл частоти змінної, або, як кажуть, крива змінної буде асиметричної форми.

Наведений вище спосіб приблизного обчислення аритметичної середньої до певної міри є затяжким через складність підрахунків треба одержати добутки xf, потім їх сумувати, потім ділити на досить великі числа. Через те можемо використати ще зазначену вище властивість аритметичної середньої, після якої алгебрична сума відхилень окремих вартостей ознаки від аритметичної середньої є рівною нулю. Цей спосіб виведення аритметичної середньої звичайно звуть способом виведення від довільно взятої основи. Формулювання його буде таким:

Назвемо вартості ознаки, яких набуває кожна одиниця сукупности, яких одиниць є N, через X_1, X_2, X_3 ..... X_n. Довільну величину /основу/, від якої будемо обраховувати відхилення вартостей ознаки змінної, так само, як це робилось при обраховуванні відхилень від середньої аритметичної, і означимо цю величину літерою A, а відхилення вартостей ознаки від цієї основи літерами: ξ_1, ξ_2, ξ_3 - ξ_n. Тоді можемо скласти такі рівняння:

$$\begin{aligned} X_1 - A &= \xi_1 \\ X_2 - A &= \xi_2 \\ X_3 - A &= \xi_3 \\ &\cdots\cdots\cdots \\ X_n - A &= \xi_n \end{aligned}$$

Сумування обох сторін рівняння дає: $\sum X - NA = \sum \xi$

Поділимо рівняння на N, одержимо:

$$\frac{\sum X}{n} - A = \frac{\sum \xi}{N}$$

Але $\frac{\sum X}{n}$ ніщо інше як аритметична середня M. Через це:

$$M - A = \frac{\sum \xi}{n} \qquad M = A + \frac{\sum \xi}{n} \quad /3/$$

Таким чином аритметична середня від довільно взятої основи є рівною довільній основі збільшеній на алгебричну суму відхилень вартостей ознаки від величини довільно взятої за основу, поділену на число одиниць сукупности.

Практичне значіння цього способу є в тому, що при обчисленні аритметичної середньої можна взяти яку хочемо /довільну/ величину за основу A, обрахувати відхилення вартости ознаки від цієї основи й суму їх поділити на цілу частоту сукупности. При цьому можемо величину A - себто, початок нашого обрахування відхилень вибрати так, щоб відхилення було легко обчислити, наприклад, щоб вони були числами цілими і малими.

Коли маємо тяглу змінну, а інтервали в розподілу частоти не є рівні одиниці приймаємо, що в кожному інтервалі вартості є однакові, себто що в інтервалі, напр., I00-I04 см. росту осіб всі вони мають ріст I02 см. Тоді обрахування аритметичної середньої буде видно з наступного прикладу: /Див. стор. I03 /.

Раніше ми робили обчислення так, що середню вартість кожного інтервалу множили на частоти цих інтервалів. Виберімо зараз певну величину за довільну основу.: від якої будемо обраховувати відхилення. Вибір цей зробимо так, щоб довільна основа знаходилась в середині того інтервалу, який знаходиться в центрі її розхилу, що має найбільшу частоту /229/ - це буде I37 см. себто A = I37 см. Оскільки границі інтервалів клясових в таблиці розподілу частоти встановлені в 5 см.. можемо відхилення вгору й вдолину від довільної основи A брати не фактичні - 5; - I0; - I5, але - I, - 2 і т.д.

Таблиця I6.

	Зріст в см.	частоти	Відхилення від А в інтервалах	Добуток відхилення на частоту
	I00 - I04	2	- 7 /-35 /	- I4
	I05 - I09	7	- 6 /-30 /	- 42
	II0 - II4	30	- 5 /-25/	- I50
	II5 - II9	69	- 4 /-20/	- 276
	I20 - I24	I03	- 3 /-I5/	- 309
	I25 - I29	I49	- 2 /-I0/	- 298
	I30 - I34	I94	- I /-5 /	- I94
A = I37	I35 - I39	229	0	-I283
	I40 - I44	250	I /+ 5/	+ 250
	I45 - I49	2I5	2 /+I0/	+ 430
	I50 - I54	I30	3 /+I5/	+ 390
	I55 - I59	64	4 /+20/	+ 256
	I60 - I64	33	5 /+25/	+ I65
	I65 - I69	I4	6 /+30/	+ 84
	I70 - I74	6	7 /+35/	+ 42
		I495		+ I6I7
				+ 334

Помножимо далі відхилення на відповідні частоти, одержимо добуток їх в останній графі таблиці. Алгебричне сумування цих добутків дасть суму - I238 та +I6I7, себто +334. Це число й є $\sum \xi$ в рівнянні / 3 / на стор. I02 Щоб одержати $\sum \xi / N$ поділимо 334 на цілу частину сукупности I495, тоді $\frac{334}{1495}$ = =0,223 інтервалу, Щоб перейти зараз од інтервалу до одиниць, в яких означка є міряна, себто до сантиметрів, треба одержану величину помножити на розмір інтервала, себто на 5 см. Одержане з цього добутку число - / 0,223x5 см. I,I2 см. додамо до числа А /довільної основи/ і одержимо I37 см.+I,I2 см.= I38,I2 см.

Цим способом обраховується середня аритметична для тяглих значно скоріше.

При використанню цього способу обрахування аритметичної середньої треба завжди звертати увагу на знак / +або - /, з яким одержуємо число $\sum \xi / N$ і згідно з цим додати або відняти абсолютну величину від довільної основи А. Недогляд щодо цього дуже часто трапляється й веде до хибних результатів.

Слід також зазначити, що наведений спосіб виведення аритметичної середньої дозволяє також і легко контролювати її обрахування, бо коли візьмемо інший початок, іншу довільну основу для обрахування, то в результаті мусимо одержати величину аритметичної середньої.

Вкажемо ще на один спосіб виведення аритметичної середньої. Він хоч і не дає більшого збереження часу в праці обрахування, але дає полегшення в тому, що обрахування переводиться простим сумуванням величин, без помноження їх, що має значіння, коли не маємо по руці арифмометра або іншого технічного знаряддя для обрахування.

Візьмемо ту ж саму таблицю розподілу частоти, що наведена вище /див. табл. на слід. сторінці/.

Як і раніше за довільну основу обрахування буде середня інтервалу "I35--I39" =I37 см. Але при обчисленні величини $\sum \xi / N$ будемо чинити інакше. У верхній частині таблиці проти кожного інтервалу клясового, від першого аж до інтервалу I30-I34, в графі третій таблиці, напишемо суму дотичної клясової частоти і частот усіх попередніх кляс, починаючи згори / 2; 2+7=9; 9+30 = 39 і т.д./, подібний ряд сум обчислимо й в долішній половині таблиці, але йдучи зі споду догори , себто від кляси I70-I74 аж до I40-I44. Не беремо на

Таблиця I7.

ріст у см.	частота /f/	суми /s/	
I	2	3	4
I00 - I04	2	2	
I05 - I09	7	9	
II0 - II4	30	39	
II5 - II9	69	I08	
I20 - I24	I03	2II	
I25 - I29	I49	360	
I30 - I34	I94	S_0^- 554	S_1^- 729
I35 - I39	229		
I40 - I44	250	S_c^+ 7I2	S_1^+ 905
I45 - I49	2I5	462	
I50 - I54	I30	247	
I55 - I59	64	II7	
I60 - I64	33	53	
I65 - I69	I4	20	
I70 - I74	6	6	

увагу лише центральну клясу I35-I39 см. Останнє суму в горішній частині третьої графи - 554 назвемо S_0^- , останнє суму долішньої частини - 7I2 назвемо S_0^+ Далі підсумуємо верхню частину графи аж до кляси I25-I29 і цю суму - 729 назвемо S_1^- Відповідно зробимо це й в долішній частині графи 3-ої, йдучи аж до кляси I45-I49 і цю суму - 905 назвемо S_1^+ . Не досить тяжким альґебричним формульованням може бути доказано,[x] що величина $\sum \xi$ з нашого попереднього рівняння / 3 /, себто всіх відхилень від A /довільної основи/ обчислена в інтервалах буде рівною:

$$S_0^+ - S_0^- + S_1^+ - S_1^-$$

Через те для обчислення величини $\sum \xi$ досить провести такі операції

$$\begin{array}{rrr} - & 7I2 & - & 905 & + & I58 \\ & 554 & & 729 & & I76 \\ \hline & I58 & & I76 & & 334 \end{array}$$

$\sum \xi$ таким чином рівняється 334. Ділимо цю величину на число N = =I495, тоді $\sum \xi / N = 0{,}223$ інтервалу і коли цю величину помножимо на 5, а результат додамо до основи A, одержимо як раніш: M =I37 + I,I2 = I38,I2

---

x/ Доказ цей наведено в Czuber'a:" Die Statistischen forsehungsmethoden" стор. 62 - 63.

ВПРАВИ /До лекції II-ої/.

I. Подайте таблицю розподілу частоти й графічне її зображення у вигляді гістограму.
2. Ту ж саму таблицю розподілу частоти подайте у формі полігона частоти.
3. Поясніть яка з форм симетрії відповідає Вашій таблиці розподілу частоти.

Лекція тринадцята.

28. Середня аритметична зважена - media ponderata.

Характеризуючи властивості середньої аритметичної, ми вже раніше /Лекція I2. стор. 98/ зазначили, що при поділі якоїсь сукупности на довільну кількість груп, середня аритметична для цілої сукупности вираховується по формулі:

$$M = \frac{M_1N_1 + M_2N_2 + M_3N_3 \ldots\ldots + M_sN_s}{N}$$

Себто, мусимо кожну аритметичну середню кожної окремої групи /части/ в цілій сукупності помножити на частоти, що стосуються до кожної групи, а одержані добутки підсумувати й поділити на загальну частоту сукупности. Відціль зрозуміло, що було б невірним обчислювати звичайну аритметичну середню для цілої сукупности просто як аритметичну середню з окремих аритметичних середніх її частей/груп/, тобто підсумувати аритметичні середні першої, другої, третьої і т.д. груп. сукупности та поділити суму на число груп. Пересічну для цілої сукупности не можна обрахувати за формулою

$$M = \frac{M_1 + M_2 + M_3 \ldots\ldots + M_s}{S}$$

через те, що аритметичні пересічні окремих частей цілої сукупности не є для нас рівноцінними, з однаковою вагою, бо кожна з цих частин стосується до різної кількости одиниць /частот/. Це випливає також з самої дефініції аритметичної пересічної, яко суми вартостей ознаки у всіх одиниць сукупности, поділеної на число одиниць цілої сукупности. Отже суму вартостей ознаки у всіх одиниць сукупности одержимо з аритметичних середніх її частей тільки тоді, коли помножимо ці часткові аритметичні середні на число одиниць /на частоту/ в кожній групі, в кожній частині сукупности.

Маємо, наприклад, на меті встановити пересічний врожай в даному році для пшениці на одній десятині або гектарі в господаря, який на своїх трьох полях: I-му, II-му, III-му, площею 2,3,5 гект. нажав з одного гектара на I-му полі - 40 q , на другому - 30q і на третьому 30 q . Схематично ці дані можна подати так:

Поле	I	площею 2 ha	нажато	- 40q	себто на I ha	- 20 q
"	II	" 3 "	"	- 30 "	" " "	- I0
"	III	" 5 "	"	- 30 "	" " "	- 6
		I0 ha		I00q		36 q

Тут не можемо суму різних врожаїв на Iha = 36 q ділити на 3 /поля/ та вважати, що пересічний врожай на I ha є рівним I2 q , бо тоді господар мав би нажати з I0ha x I2 q = I20 q, в той час, як в дійсности, він нажав I00 q. Так обчислити пересічний врожай ми мали би право лише тоді, коли б кожне поле мало рівну кількість гектарів, себто вага, врожаю на I ha в кожному полі була однакова. Мусимо, таким чином, значіння врожаю на кожному полі важить розмірами кожного поля, себто відповідно наведеній вище формульовці другої властивости аритметичної середньої обчислити пересічний врожай так:

$$\frac{/2 \times 20/ + /3 \times I0/ + /5 \times 6/}{I0} = \frac{I00\,q}{I0} = I0\,q$$

Так обчислена аритметична середня й зветься зваженою аритметичною середньою - Media ponderata.

Таким чином, таке "важення" мусимо завжди робити тоді, коли окремі члени статистичного ряду квантитативно не мають однозначної "ваги" або значіння. Маємо, наприклад, встановити пересічну продажну ціну пшениці на ринку, на якому було продано:

300 Kg. по 70 коп. за кілограм
400 " по 60 " " "
I00 " по 50 " " "

Маємо тут три величини для ціни пшениці і кожна з них стосується до різної кількости проданої пшениці. Тоді наше обчислення пересічної ціни мусить бути таким:

70 x 300 = 2I.000
60 x 400 = 24.000
50 x I00 = 5.000
800 50.000 : 800 = 62,5

Одержане число 62,5 коп. і є зважена пересічна ціна пшениці.

На підставі сказаного, можемо зважену аритметичну середню формулювати так: коли означимо окремі дані, що стосуються до вартостей ознаки через X, а їх вагу літером f, зважена аритметична пересічна Mp /media ponderata/ буде рівною $\frac{\Sigma X \cdot f}{\Sigma f}$, в той час як звичайна аритметична середня є рівною $M = \frac{\Sigma X}{N}$. Розуміється, що обидві формули будуть тотожніми тоді, коли всі f є рівними між собою, себто f є константним.

Мусимо тут зазначити, що ми вживаємо назви "пересічна" і "середня", однаково для аритметичної середньої через те, що як ця остання, так і деякі інші види середніх, про які скажемо далі, вираховуються беручи на увагу всі одиниці певної сукупности, і через те, дійсно, є "пересічною величиною", в той час як інші види середніх, напр., модальна середня, не є пересічною величиною, а тому й назви "пересічна" для них вживати було б не оправданим. Для того, щоб пересічна з окремих даних була дійсно аритметичною пересічною з вартостей ознаки у всіх одиниць сукупности потрібно, як то зазначено вище, її обчислити яко зважену пересічну окремих даних. Одначе це не завжди буває можливим. Інколи маємо ряд даних про вартості ознаки, але не знаємо, до якої кількости одиниць певна дана стосується. Маємо, наприклад, лише дані про ціни, по яких продавали який-небудь крам у певну добу, але яка саме кількість продукту продавалась по тій або іншій ціні не знаємо. В таких випадках примушені обмежитись вирахованням звичайної пересічної аритметичної з окремих даних. Виникає питання, наскільки великі будуть в таких випадках відхилення звичайної пересічної від пересічної зваженої та за яких обставин допущені хиби будуть значними?

З прикладів, які ми подаємо вище, різниця між звичайною та зваженою пересічною випливає з того, що дані різного розміру себто вагости ознаки мали різну вагу. Коли в ряді даних значна частина їх мала по величині, а вага їх велика, ясно, що тоді зважену середню одержимо меншу ніж звичайну середню і навпаки, коли вищі дані будуть мати й вищу вагу, тоді зважена пересічна буде більшою від звичайної пересічної.

Коли візьмемо, одначе, велике число даних, в яких випадково комбінуються менші дані з більшою вагою і навпаки, то різниця між звичайною й зваженою пересічною може бути і невеликою, бо одні комбінації, так би мовити, урівноважуються другими /більші-меншими/.

Більша різниця між цими двома пересічними виникає тоді, коли менші чи нижчі дані більш-менш систематично мають певну вагу меншу або більшу себто, коли існує певна тенденція споєння вищих зглядно нижчих даних з більшою зглядно нижчою вагою, інакше, коли є певний зв'язок /кореляція/ між висотою даних /даною вартістю ознаки/ та їх вагою

Більша різниця між обома пересічними може статись і тоді, коли з цілої маси лише декілька даних мають виключно велику вагу, тоді зважена пересічна й буде під впливом цих виключно великих випадків. Але взагалі, чим більше число даних, тим більше можна чекати вирівняння випадковостей, а першорядне значіння буде мати основна причина різниці між звичайною й зваженою середньою це - наявність певної кореляції /зв'язку/ між вагою й висотою даних. Отже відхилення будуть тим більші, чим гостріше виявляється зв'язок між

висотою вартости ознаки та її вагою, а крім того тим більші, чим більшими є взагалі хитання даних та їх ваги в певній сукупності.

Важливим є також інше питання, а саме: чи завжди виведення зваженої середньої аритметичної є бажаним і чи не маємо таких випадків, коли обрахування зваженої недоцільне, а звичайна пересічна є вигіднішою. Наприклад, маємо певну кількість даних про врожай в хліборобських господарствах. Ці дані звичайно одержуємо від кореспондентів, які дуже часто добровільно надсилають відомості статистичним урядам. Ці дані стосуються не до всіх господарств, а тільки до певної їх частини, і хоч уряди стараються сітку кореспондентів організувати так, щоб вони рівномірно охоплювали певні господарські райони в краю, але й в такому випадку не маємо певности, що дані кореспондентів охоплюють більш-менш рівномірно господарства з різною висотою врожаю. Ще менше маємо певности в тому, що розподіл посівної площі господарств за величиною врожаю за даними кореспондентів, відповідає конкретному розподілу площі за розміром урожаю для цілої посівної площі даного краю. Ця непевність у дійсній репрезентативності даних кореспондентів позбавляє нас підстави вжити в таких випадках зваженої середньої арифметичної, бо той приблизний образ урожаю на даній території, який дасть нам звичайна пересічна з окремих даних, зважена пересічна уточнити чи удосконалити навряд чи зможе.

Звичайно, що не кожне репрезентативне спостереження тягне за собою негативне рішення про використання зваженої середньої, але вірне розв'язання цього питання обов'язково вимагає в кожному випадку серйозного продумання. Крім випадків репрезентативного спостереження можуть бути й інші випадки коли користання з середньої зваженої є недоцільним. Ці випадки характерні тим, що в них рівноцінними є для нас не окремі одиниці сукупности, а якраз окремі дані про них. Маємо, наприклад, дані про пересічні ціни пшениці за ряд років, при чому нас цікавлять зміни цін протягом часу. Щоб орієнтуватися в загальній тенденції цінові, яка виявляється до певної міри в нерівномірному руху цін од року до року, з'єднуємо окремі роки, наприклад, по пятирічних групах і для них обчислюємо аритметичні пересічні з пересічних цін за кожний окремий рік, що входить в пятиріччя. Чи треба в такому випадку брати на увагу значіння /вагу/ тих щорічних цін в залежності від того, чи був оборот продажу пшениці більший чи менший? Ні! Висота цін даного року є для нас рівнозначною одиницею, не дивлячись на те, чи більший, чи менший оборот пшеницею був у даному році на ринку. В такому досліді ми лише стежимо за зміною цін від одного року до другого, а в пересічній за кожне пятиріччя шукаємо якоїсь загальної течії в змінах цін, що витворилась з щорічних змін в даному пятиріччі; усунути границі між окремими роками та розглядати кожну одиницю проданої пшениці без огляду на рік, в якому вона була продана, не входить у наше завдання. Через те в таких випадках належить обчислювати пятирічну пересічну, яко звичайну, себто діленням пересічних щорічних цін на пять. Коли б нас цікавили ціни з іншого погляду, наприклад, у випадку бажання з'ясувати загальну висоту цін за даний період, тоді б з'єднання окремих років у пятиріччя мало б метою збільшення спостережень, на які спирається пересічна; на окремі одиниці проданої пшениці за пятиріччя дивилися б як на рівнозначні одиниці без огляду на те, в якому році вони були продані, а границі між окремими роками усунули б тим, що з одиниць пшениці проданих за пятиріччя склали б одну загальну сукупність. Отже мусили б обчислити пересічну з щорічних цін зважену, числячись з оборотом в кожному році.

Ті ж самі міркування треба прикласти й до тих випадків, коли студіюємо зміни явищ у просторі. Наприклад, коли нас цікавлять усі ринки певної области, як один загальний ринок цін якогось краму, тоді доцільно створити з цілої маси товару, проданого на всіх ринках, одну загальну сукупність, а на окремі одиниці товару дивитись як на рівнозначні, без огляду на те, де були продані. В такому випадку треба важити ціни окремих ринків розміром

обороту товару на окремих ринках. Словом, питання тут розв'язується в залежності від того, що мусимо вважати за рівнозначні індивідуальні одиниці, чи одинокі дані, чи окремі одиниці маси, до частин якої ці дані стосуються.

Зазначимо дані, що вагою не мусять завжди бути абсолютні числа одиниць відповідних частей сукупностей, а можуть бути й релятивні частоти. Це випливає з самої формули зваженої середньої аритметичної $\sum Xf / \sum f$. Дійсно:

$$\frac{\sum Xf}{\sum f} = \sum \left[X \cdot \frac{f}{\sum f} \right] / \sum \frac{f}{\sum f} ,$$

бо сума релятивних ваг $\sum \frac{f}{\sum f}$ буде рівною одиниці. З цього випливає також що можемо важити пересічну взагалі рядом чисел співмірних тим вагам, якими хочемо користатись. Наприклад, треба зважити пересічну цін різних міст, беручи на увагу оборот товарів відповідних міст, але обороту товару не знаємо, тоді можемо взяти яко співмірну вагу число людности у відповідних містах.

Коли змінна є тяглою, себто таблиця розподілу частоти має ряд з класовими інтервалами, в яких заключені величини змінної $x_1, x_2, x_3 - x_n$, але точного значіння їх в межах інтервалу не маємо тоді, щоб мати елементи для виведення середньої, звичайно допускається, що всі випадки в кожному інтервалу розміщуються рівномірно. Себто допускаємо, що дійсне значіння величини x лежить по середині інтервалу, а тоді вона є середньою аритметичною між границями інтервалу. Це дає можливість змінну тяглу перетворити в змінну розривну, відповідно до чого й робити обчислення.

Наприклад, маємо таку таблицю:

Таблиця 18.

Розміри заробітньої платні в крб.	Число робітників, що цю платню дістають	Перетворена змінна тягла в змінну розривну
1	2	3
1,70 - 1,71	6 /f_1/	1,705
$v_0 - v_1$		
1,71 - 1,72	12 /f_2/	1,715
$v_1 - v_2$		
1,72 - 1,73	13 /f_3/	1,725
$v_2 - v_3$		
1,73 - 1,74	10 /f_4/	1,735
$v_3 - v_4$		
1,74 - 1,75	5 /f_5/	1,745
$v_4 - v_5$		
Mp	=	1,7241

Таким чином, коли зазначимо границі клясових інтервалів через v_0, v_1, v_2 ..... v_n, то зважена середня аритметична може бути формульована так:

$$Mp = \frac{\frac{v_0 + v_1}{2} \cdot f_1 + \frac{v_1 + v_2}{2} \cdot f_2 + \ldots \frac{v_{n-1} + v_n}{2} \cdot f_n}{\sum /f/} = \frac{1}{\sum /f/} \cdot \sum \left[f \cdot \frac{v_{n-1} + v_n}{2} \right]$$

29. Зважена середня аритметична від довільно взятої основи.

Зважену середню аритметичну так само, як і звичайну середню аритметичну, можемо обчислити від довільно взятої основи.

Зазначимо довільно взяту основу /константу/ літерою "K", різниці між нею й дійсними величинами ряду через "x'", а частоту /фреквенцію/ через "f". Тоді маємо такий ряд:

$$K + r_1 \ldots\ldots f_1$$
$$K + r_2 \ldots\ldots f_2$$
$$K + r_3 \ldots\ldots f_3$$
$$\ldots\ldots\ldots\ldots$$
$$K + rn \ldots\ldots f_n$$

а середня пондерата

$$Mp = K + \frac{I}{\sum/f/} \cdot \sum/fr/$$

себто пересічна аритметична зважена є рівною довільно взятій основі плюс сума добутків різниць від неї на відповідні частоти поділена на суму всіх частот.

Візьмемо для прикладу попередню таблицю /I8/.

K	r	f	fr
I,7	+0,005	6	+0,030
I,7	+0,0I5	I2	+0,I80
I,7	+0,025	I3	+0,325
I,7	+0,035	I0	+0,350
I,7	+0,045	5	+0,225
		46	+I,II

$$Mp = I,7 + \frac{I,II}{46} = I,7 + 0,024I = I,724I.$$

Можна обчислення перевести й в інший спосіб. Беремо за довільну основу I,725, тоді

r	f	fr +	fr −
- 0,020	6		0,I2
- 0,0I0	I2		0,I2
0	I3	-	-
+ 0,0I0	I0	0,I	
+ 0,020	5	0,I	
	46	0,2	0,24

$$\sum/fr/ = /+0,2/ + /-0,24/ = -0,04$$

$$Mp = I,725 - \frac{0,04}{46} = I,725 - 0,00087 = I,724I$$

$\frac{\sum/fr/}{\sum/f/}$ Цей спосіб має назву с п о с о б у м о м е н т і в . Величина зветься першим моментом - v_1 , а K нулевим моментом - v_0

$Mp = K + v_1 \cdot i$, себто

середня аритметична зважена є рівною довільно взятій основі плюс добуток першого момента на величину інтервала / i /.

30. Типові й атипові пересічні

Величина аритметичної пересічної дає нам досить зручну характеристику сукупности в зв'язку з висотою змінної ознаки, одначе вона не дає нам уяви про форму розподілу частоти цієї ознаки в сукупності. Можемо мати цілком різні форми розподілу частоти декількох змінних, а аритметичні пересічні кожної з них будуть однаковими. Маємо, наприклад, наступні розподіли частоти в сукупностях:

Таблиця I9.

Вартість ознаки	частота		
	I сукупність	II сукупн.	III сукупн.
0	2	4	I4
I	8	8	I8
2	I6	20	20
3	28	40	24
4	56	32	I0
5	28	24	22
6	I6	I8	30
7	8	I2	I0
8	2	6	I4

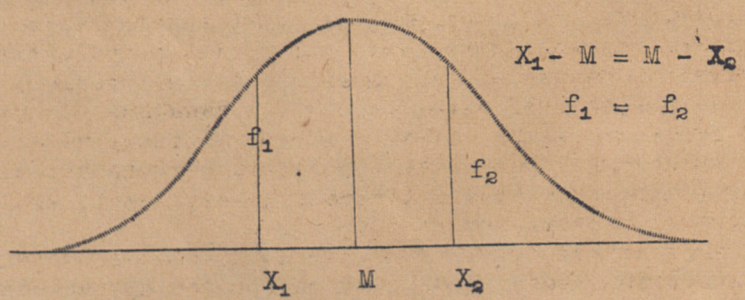
Перша сукупність має розподіл частоти стисло симетричної форми, друга малосиметричної, а третя має форму двовершинної кривої. Одначе аритметична пересічна для всіх цих сукупностей буде рівною 4-себто однаковою. З наведеної таблиці розподілу частоти бачимо, що лише в першій сукупності, яка є строго симетричної, співмірної форми, аритметична середня репрезентує вартість ознаки, до якої стосується найбільше число спостережень, а в асиметричній формі найчастіша вартість ознаки міститься по стороні аритметичної пересічної.

Таким чином, лише в співмірних, симетричних формах розподілу частоти сукупностей, аритметична пересічна найліпше характеризує цю сукупність і завжди відповідає тій вартості ознак, яка дає найбільше число спостережень, - є найчастішою величиною ознаки в сукупності, Це підтверджується також і тою властивістю аритметичної пересічної, що сума відхилень від неї завжди є рівною нулю. Дійсно, коли сума відхилень вартостей ознаки в сукупності від якоїсь вартости "К" є рівною нулю, тоді ця вартість і є аритметичною пересічною. Бо коли

$$\sum / X - K / \quad 0 \text{ тоді } \sum X - K \quad 0; \quad K = \frac{\sum X}{N} = M$$

Отже при симетричній формі розподілу частоти аритметична середня є найчастішою величиною вартости ознаки, вона поділяє площу, окреслену симетричною кривою й віссю абсциси, завжди на дві рівні половини, а відхилення вартостей ознаки розподіляються біля середньої рівномірно як по одній, так і по другій стороні кривої, як то можемо бачити на рисунку I3-му.

Рисунок I3.



Таким чином, коли помножимо кожне відхилення від найчастішої вартости /М/ на відповідні частоти й обрахуємо альгебричну суму одержаних добутків, то ця сума буде рівнятися нулю, бо суми відхилень, що мають знак плюс

і тих, що мають знак минус, будуть рівними. Отже при співмірному /симетричному/ розподілу частоти якоїсь змінної, найчастіша вартість ознаки є разом з тим і аритметичною пересічною. У випадку асиметричного розподілу частоти, чи буде він мати лівосторонню асиметрію /рис.I4/, чи правосторонню /рис.I5/ найчастіша вартість не буде відповідати аритметичній пересічній "М", а буде лежати або на ліво від неї, або направо, як то ясно з рисунків I4 і I5.

Рис. I5.

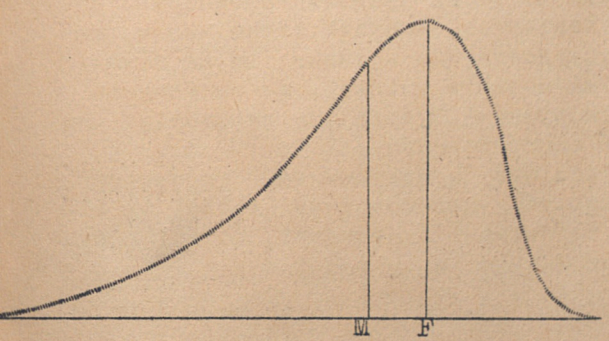
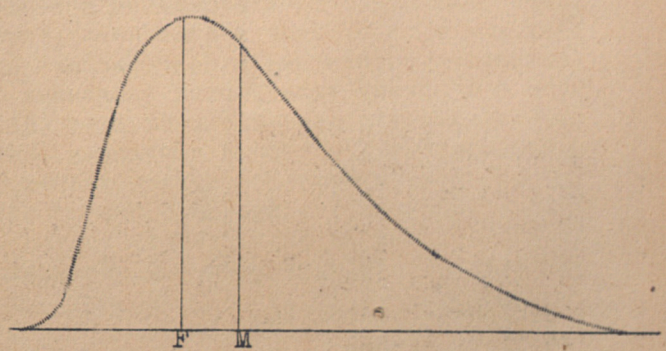


Рис. I4.



В цих випадках альґебрична сума відхилень вартостей ознаки від найчастішої їх величини не буде рівнятися нулю, бо відхилення з плюсом і мінусом будуть нерівними - їх суми не будуть компенсуватися.

Таким чином, при асиметричному одновершинному розподілу частоти аритметична пересічна не збігається з найчастішою вартістю ознаки й буде відсунута від неї на ту сторону, на яку більше простягається крива частоти і, розуміється, тим більше відсунута, чим більша асиметрія є в розподілу частоти. При дуже асиметричному розподілу частоти, коли максімум знаходиться на одному кінці розхилу кривої, розділ між аритметичною пересічною й найчастішою вартістю є дуже великим, себто аритметична пересічна не може бути величиною, яка добре репрезентує цілий розподіл частоти змінної.

Отже значіння аритметичної середньої яко характеристики статистичної сукупности є в залежності від форми розподілу частоти змінної; на цій підставі існує й поділ аритметичних середніх на типові й атипові /Кетле/. Аритметична середня є типовою тоді, коли зображує в явищі /репрезентує/ якийсь тип, до якого окремі випадки цього явища прямують яко до своєї найчастішої вартости, а відхилення від цього типу є лише випадковими та зустрічаються тим рідше, чим є більшою їх величина. Звичайно, що найбільш типовою буде аритметична середня у тих випадках, коли вона одночасно є й "об'єктивною" середньою,/див.лекція I2,стор.96/, але такі випадки маємо в статистичній практиці рідко. Навпаки, дуже часто маємо діло з суб'єктивними середніми, а тоді скійкість середньої, її типовість або нетиповість у зазначеному розумінні набуває великої ваги, особливо у випадках різних форм асиметрії, що потребують відповідної оцінки для себе з огляду на більшу або меншу типовість аритметичної пересічної. Коли аритметична середня не є типовою, в таких випадках треба шукати іншого виду середньої, яка б ліпше репрезентувала розподіл частоти певної змінної.

Деякі теоретики вважають аритметичну пересічну типовою пересічною у випадку лише ідеально співмірного /симетричного/ розподілу частоти, себто у випадку так званої нормальної кривої розподілу частоти, або кривої Гавса, з якою зустрінемося далі, але для того не має вистарчаючих підстав. Основним критерієм типовости аритметичної пересічної буде той факт, чи досить наближена вона до найчастішої вартости в одновершинному розподілу частоти якоїсь змінної і також у тих випадках, коли цей розподіл не є

"нормальним".

З наведеного вище випливає, що слабість аритметичної пересічної яко репрезентанта /типа/ сукупности виявляється тоді, коли спостерігаємо в окремих випадках певної сукупности хоч і не багато, але декілько екземплярів з дуже великов, або дуже малов вартістю ознаки. Наприклад, хочемо характеризувати в даному районі розмір селянських хліборобських господарств пересічнов їх величиновю в гектарах, або врожай пшениці цих господарств пересічнов величиновю на I гектар. Коли серед цих господарств зустрінемо декілька дуже великих господарств, дуже багатих людей, або ряд господарств з надзвичайним урожаєм через видатну й господарську якість, то вони можуть дуже вплинути на порушення типового розміру господарств, або їх урожайности цього району. Себто аритметична пересічна не дасть правдивого типового образу.явища та не буде добрим його репрезентантом. Виникає,таким чином, необхідність доповнення характеристики сукупности чимсь іншим, що не підлягало б великому впливу отих виключних екземплярів змінної, що вискакують дуже далеко від звичайного їх ряду. Ці порушення будуть тим впливовіші,чим менше одиниць сукупности входить у наше спостереження.

Отже мусимо спробувати знайти, подібно до аритметичної середньої, і інші величини, що давали б нам правдиву уяву про положення кривої на вісі абсцис, про загальну висоту вартостей ознаки в сукупності, на які б були також середніми величинами. Важливішими з них є м е д і я н а й м о д а.

30. Медіяна - центральна вартість ознаки. Коли серед цілого ряду вартостей змінної маємо, або знайдемо таку вартість цієї змінної, від якої сума частот вартостей ознаки вгору, себто менших від неї, і сума частот вартостей ознаки вниз, себто вартостей більших від неї, будуть рівними, то таку середню вартість змінної називаємо вартістю центральнов або медіянов. У відношенні до фреквенційної кривої /кривої розподілу частоти/, можна медіяну визначити як таку вартість змінної, через яку опущена на вісь абсцис прямовісна поділить поверхню окреслену кривов на дві рівні частини.

Маємо,наприклад, такий розподіл частоти вікових груп 27,7 мілійонів населення /змінна тягла/

Вікові групи населення	Число осіб кожної групи в мільйонах	
до I0 років	6,9	I2,5 міл.
I0 - 20 "	5,6	
20 - 30 "	4,8	
30 - 40 "	3,7	I0,4 міл.
40 - 50 "	2,8	
50 - 60 "	2,0	
60 - 70 "	I,3	
70 - 80 "	0,5	
80 - 90	0,I	
вище 90 "	0,0	
	27,7 мільйонів.	

Сума частот цієї сукупности складає 27,7 міл. осіб, а половина

$$/\frac{\sum f}{2}/ = \frac{27,7}{2} = I3,85 \text{ міл.}$$

Для характеристики пересічного чи середнього віку даної сукупности візьмемо медіяну. Вона мусить, як вартість центральна, знаходитись безперечно в клясі від 20-30 років, бо вгору від цієї кляси й вниз од неї сума частот не перевищує кожна половини частот цілої сукупности. Щоб верхня від медіяни сума частот дорівнювалась сумі частот нижче медіяни, себто, щоб вона була дійсно центральною вартістю ознаки, треба частоту, що відповідає ксясі, в якій знаходиться медіяна /4,8 міл./, пропорційно розподілити між частотами вартостей, що знаходяться вище й нижче цієї кляси. Беручи на увагу, що інтервал клясовий у даній таблиці є рівний для кожної кляси I0 рокам, будову пропорції маємо таку:

I3,85 - I2,5 = I,35 X_1 : I0 = I,35 : 4,8
I3,85 - I0,4 = 3,45 X_2 : I0 = 3,45 : 4,8

$$X_1 = \frac{I0 . I,35}{4,8} = 2,8I2 \qquad X_2 = \frac{I0 . 3,45}{4,8} = 7,I88$$

Додавши 2,8I2 / X_1 / до нижчої границі інтервалу кляси, в якій мусить знаходиться медіяна, себто до 20 - маємо 22,8I2 років і віднімавши від верхньої границі інтервалу цієї кляси - 7,I88 / X_2 / /30 - 7,I88/, одержимо також 22,8I2 років, себто медіяна є 22 роки іI0 місяців / 22 роки 9 місяців, 26 днів, 9 годин і 7 хвилин/.

Формулювання це в загальний спосіб найліпше подав Ф е х н е р, а саме: коли зазначимо через M_e - медіяну, літером g_1,- нижчу границю інтервалу клясового, в якому мусить знаходитись медіана, через V - суму частот вартостей до g_1 , через f_0 - частоту тієї кляси, в якій межить медіяна, а через i - величину інтервалу, то медіяна буде рівною:

$$Me = g + \frac{\left(\frac{\sum f}{2} - V\right) i}{f_0}$$

Користуючись конкретними даними поданої вище таблиці будемо мати:

$$M_e = 20 + I0 \frac{\left(\frac{27}{2} - I25\right)}{48} = 20 + \frac{I0 . I,35}{4,8} = 22,8I2$$

У випадках, коли маємо довгий ряд серій змінної тяглої, можемо користуватись припущенням, що розподіл вартостей ознаки в кожній клясі в межах інтервалів є рівномірним і обчислення медіяни з задовольняючим наближенням не зустрічає перешкод.

Наприклад, візьмемо довгий ряд серій змінної тяглої, що подана в таблиці 7-й розподілу частоти /стор.88 /. Приймемо на увагу також, що подані в ній класи змінної в формі "57 -", "58 -" і т.д. мусять вважатись за 57 I5/I6 дюйма - 58 I5/I6 дюйма, 58 I5/I6д- 59 I5/I6д. і т.д., бо вимір зросту переводився конкретно за таков саме точністю.

Тоді, користуючись конкретними даними цієї таблиці, можемо обчислення медіяни зробити в наступний спосіб:

Половина суми всіх частот сукупности $\frac{\sum f}{2}$ буде рівною 4292,5 $\frac{8585}{2} = 4292,5$/. Сума частот вартостей, що є нижчі 66 I5/I6 дюйма є рівною 3589. Різниця між півсумою всіх частот й сумою частот нижчих вартостей дає 703,5. Частота вартостей, що знаходиться в клясі 67 -, в якій мусить бути медіяна, є рівною I329.

Медіяна буде рівною: $66\frac{I5}{I6} + \frac{7035}{I329} = 66,47$ дюйма. Різниця між середнею аритметичною й медіяною для цієї сукупности складає всього 0,0I дюйма. Це через те, що конкретний розподіл частоти дорослих чоловіків за ознакою висоти тіла дає зовсім симетричну форму кривої, себто майже нормальну /Гав-

сову/ криву, а в таких випадках, розуміється, аритметична середня цілком збігається з медіяною.

В теоретичних студіях користуватись медіяною вигідно й легко, але не треба думати, що ця зручна форма середньої є корисною в усіх випадках і може цілком заступити аритметичну пересічну. Медіяна дає добрий вихід у тих випадках, коли остання кляса змінної точно не окреслена інтервалом, себто не має його вищої границі /вище 90 .... і тільки/ і через те аритметичною середньою користуватись не можна. Медіяну можна давати перевагу перед аритметичною середньою тоді, коли в ряді вартостей сукупности маємо значну кількість виключно малих або виключно великих вартостей. Великан по зросту при обчисленні медіяни не впливає на розмір середньої зросту певної кількости осіб більш, ніж яка будь особа в пересічному зрості. Коли середній річний прибуток певної кількости осіб виносить, наприклад, 500 карбованців, то - зміна, що наступить у медіяні при доданню до цієї групи осіб ще однієї особи, не буде залежати від того, що ця особа має прибуток 50.000 карбованців річно чи 5.000 карбованців або менше, в той час як на зміну аритметичної середньої особи з виключним прибутком можуть мати великий вплив. Отже, коли серед спостережень можуть траплятись випадково надмірні вартості - однаково чи наслідком помилок чи річево - медіяна дає середню точнішу, більш типову.

Разом з вигодами має одначе медіяна й певні дефекти, які торкаються передовсім того, що в деяких випадках медіяну визначити незручно, а інколи й зовсім не можна.

Дійсно, коли під медіяною розуміємо вартість, від якої вартості ознаки вищі й нижчі розміщені в сукупності в рівній кількості, виникає питання, як бути з тими сукупностями, які мають париste число одиниць. Коли маємо число одиниць сукупности $2n + I$, тоді тією одиницею, від якої вище й нижче найдемо рівне число одиниць, буде, розуміється, одиниця, що займає в зростаючому ряді $n + I$-не місце. Одначе при паристому числі членів, себто $2n$, не існує члена, від якого в зростаючому ряді вгору й вниз нарахуємо однакове число членів. Цю перешкоду можемо усувати тим, що в таких випадках будемо вважати за медіяну аритметичну пересічну з вартостей ознаки в n-му і $n + I$-му члені зростаючого ряду.

Друге утруднення випливає з того, що дуже часто, особливо в випадках розривної /переривної/ змінної одна й та сама вартість ознаки буде повторюватись в декількох одиницях сукупности. Наприклад, в ряді: 2,3,5,6,6,6, 7,9,I0,I2, маємо аж три шістки. Знайти величину від якої вгору й вниз було рівне число вартостей ознаки не можемо, бо коли візьмемо 6, то з нижчими вартостями будемо мати 3 випадки, а з вищими 4, коли візьмемо 7 то одержимо з вищими 3, а з нижчими 6. В цих випадках вартости, яка б відповідала дефініції в сукупності не знайдемо.

Не має медіяна й тієї важливої альґебричної властивости, яку має аритметична пересічна, а саме, що альґебрично не можемо обчислити медіяну цілої сукупности на підставі медіян поодиноких частей цієї сукупности.

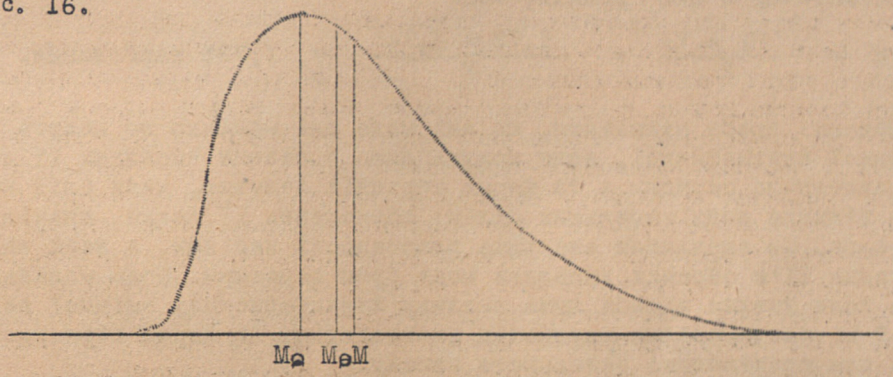
Л е к ц і я ч о т и р н а д ц я т а.

3I. Модальна середня - Moda.

Важливою такою середньою, що вживається поруч з аритметичною пересічною є **м о д а**, або найчастіша вартість ознаки в сукупности. Саме визначення цієї величини свідчить про те, що вона доповнює корисно аритметичну середню в тих випадках, коли розподіл частоти має асиметричний характер настільки, що аритметична середня не може служити типом явища. Цей найчастіший тип явища й дає нам мода.

У відношенню до фреквенційної кривої мода визначається, яко така точка на поземій вісі, якій відповідає максимальна прямовісна. Мода "Мо" є найбільша середня величина після аритметичної й медіяни, як то видко з рис. I6.

Рис. I6.



В випадку симетричного розподілу частоти мода буде центром співмірности й збігається з пересічною аритметичною.

Не має сумніву, що мода є важливим видом середньої в статистичних рядах асиметричної форми, але визначення її має й значні труднощі. Справа в тому, що при клясифікованому ряді /змінна тягла/ з інтервалами, найчастіша вартість змінної не є точно визначеною, а знаходження її в тій або іншій клясі буде залежати від того, які інтервали клясові встановимо для даної змінної. При великих інтервалах мода, так би мовити, заховується в значній кількості частот певного інтервалу, але й встановлення дрібних інтервалів клясових не досягає мети. через те, що матеріял роздрібнюється, кількість вартостей ознаки, що стосуються до кожного інтервалу, значно зменшується, а ціле угрупування сукупности втрачає ясність - правдивість.

Завдання є в тому, щоб відшукати середню вартість інтервалу, в якому частоти дадуть максимум - себто модальну вартість, такою будовою інтервалів клясових, яка б забезпечила й правильне розміщення частот розподілу середніх і вистачаюче число спостережень. Понеже ми обмежені практично в досягненні цієї мети в клясифікуванні змінних тяглих, мусимо часто задовольнятись приблизним усталенням вартости модальної.

Правда, є ще можливість знайти вихід з цього утруднення в той спосіб, що збудувавши ідеальну криву певного рівняння, яка б відповідала даному конкретному полігону частоти, визначити моду, яко вартість ознаки, що відповідала б максимальній прямовісній для збудованої кривої, але такий спосіб вимагає більшого знання математичної аналізи[x] й ми на ньому зупинятися не будемо. Практично найчастіше доводиться вживати наступного способу обчислення моди: нехай ряд, наприклад, заробітньої платні буде $v_0 - v_1$; $v_1 - v_2$ ..... $v_{n-1} - v_n$, а число робітників кожної кляси f_1; f_2; f_3 ..... f_n; тоді найбільша величина f і дасть моду:

Платня в карбованцях	число робітників.
I,7 - I,7I /$v_0 - v_1$/	I20 /f_1/
I,7I - I,72 /$v_1 - v_2$/	I89 /f_2/
I,72 - I,73 /$v_2 - v_3$/	730 /f_3/
I,73 - I,74 /$v_3 - v_4$/	987I /f_4/
I,74 - I,75 /$v_4 - v_5$/	I090 /f_5/
	I2000

---

x/ Дивись Czuber: Die statistischen Forschungsmethoden стор.7I.

Тут більшість випадків відповідає клясі $v_3 - v_4$ і через те модальна вартість лежить в межах цього інтервалу, себто вона $>$ I,73 і $<$ I,74. Припускаючи рівномірне розміщення випадків у цьому інтервалі,x/ беремо середню аритметичну з цих границь, тоді

$$M = \frac{v_3 + v_4}{2} = \frac{I,73 + I,74}{2} = I,735 \text{ карбованця.}$$

Нарешті треба відмітити, що ані мода ані медіяна не можуть заступити середньої аритметичної, вони можуть лише в певних випадках її доповнювати. Аритметична середня є та форма середніх величин, якою слід користуватися для всякого роду загальних цілей; обчислення її легке, значіння її завжди певне. Ми зазначили вже вище властивості медіяни, з яких видно, що використання її в окремих випадках може бути корисним. Щодо модальної середньої, то вона уявляє з себе дуже важливу характеристику змінної величини, але в той же час, завдяки труднощам знаходження, ця форма середньої менш придатна для практичного загального вжитку.

Між наведеними вище трьома видами середніх - аритметичною, медіяною й модою існує певний зв'язок. Аритметична, мода й медіана будуть д у ж е б л и з ь к и м и в е л и ч и н а м и, коли розподіл змінної є с и м е т р и ч н о ї ф о р м и; взаємна близькість цих середніх в таких випадках завжди характеризує типовість, стійкість колективного атипового явища. Середня аритметична буде вище медіяни, коли навіть невелика частина змінної розподілу частоти цілої змінної/ буде мати дуже великі покажчики частоти, вона буде нижчою за медіяни, коли високі покажчики частоти мало зустрічаються в таблиці розподілу, а, навпаки, частоти нижче середньої складають в ньому значну групу. Мода також буде нижче аритметичної середньої, коли мається хоч невелика група з високими покажчиками в цілому розподілі частоти змінної. Моду дуже тяжко визначати в тих випадках, коли ціла сукупність не є однородною щодо свого складу.

Таким чином, у тих випадках, коли розподіл частоти змінної набуває форми мало-асиметричного типу, або є близьким до співмірної форми розподілу частоти /Див. рис. 4 і 5 на стор. 88-89 /, можемо сміло користуватись наступною загальною формулою, що визначає зв'язок між наведеними трьома середніми:

$$Mo = M - 3\,/M - Me/$$

Середня модальна /мода/ рівна середній аритметичній без потроєної різниці між середньою аритметичною й медіяною статистичного ряду. Ця формула свідчить також про те, що в графічно зображеному розподілі частоти /полігона/ мода буде знаходитися на тій же стороні від аритметичної середньої, що й медіяна, але в віддалі в три рази більшій ніж медіяна. З цієї формули випливає, що:

$$I/\ M = \frac{3Me - Mo}{2} \qquad 2/\ Me = \frac{Mo + 2M}{3}$$

Геометрична середня. Геометричною середньою з n вартостей ознаки змінної є корінь n-ого ступеня /$\sqrt[n]{\ }$/ здобутку з тих n вартостей. Хай ознака X набуває вартостей $X_1, X_2, X_3 \ldots\ldots X_n$ /з яких деякі вартості можуть бути й однаковими/, тоді геометрична середня, яку звичайно визначають літерою G є рівною:

$$G = \sqrt[n]{X_1 . X_2 . X_3 \ldots . X_n} = /X_1 . X_2 . X_3 \ldots\ldots X_n/^{1/n}$$

---

x/ Звичайно таке припущення не завжди можливе, як то ми зазначали в викладі про середні.

Формула ця вказує між іншим на те, що при значній кількості одиниць сукупности, при обчисленні геометричної середньої треба користуватись логатмами, а саме:

$$\lg G = \frac{I}{n} / \lg X_1 + \lg X_2 \ldots + \lg X_n / = \frac{\sum \lg X}{n}$$

себто логаритм геометричної середньої є аритметичною середньою з логаритмів окремих вартостей ознаки. Через те, найшовши в таблицях /логаритмічних/ логаритми окремих вартостей ознаки можемо з ними проробити ті операції, які провадимо при обчисленні середньої аритметичної, себто в випадку великої сукупности можемо скласти таблицю розподілу частоти й вжити наведених нами вище скорочених і приблизних способів обчислення. Одержавши таким чином логаритм геометричної середньої відшукаємо відповідне для нього число в логаритмічних таблицях.

Важливо усвідомити собі, що геометрична середня, на розділ од трьох зазначених вище середніх, не може обчислюватись у тих випадках, коли навіть одна одиниця сукупности є рівною нулю або має від'ємний знак. В першому випадку геометрична середня перетворилася б у нуль, в другому - може набути вартости зовсім невідповідної. Через те в практиці, а також через тяжке обчислення, середньої геометричної вживається рідко.

Одначе геометрична середня має декілька цінних альгебричних властивостей, а саме: ця середня для цілої сукупности може легко обчислюватись через обчислення геометричних середніх окремих частей цієї сукупности. Себто, коли логаритми геометричних середніх для окремих частей сукупности будуть:

$$\lg G_1 = \frac{I}{n_1} \sum / \lg X_1 /, \quad \text{тоді} \quad \sum / \lg X_1 / = n_1 \lg G_1$$

$$\lg G_2 = \frac{I}{n_2} \sum / \lg X_2 / \quad " \quad \sum / \lg X_2 / = n \lg G_2$$

$$\lg G_3 = \frac{I}{n_3} \sum / \lg X_3 / \quad " \quad \sum / \lg X_3 / = n_3 \lg G_3$$

Через те логаритм геометричної середньої для цілої сукупности, що рівняється сумі логаритмів усіх окремих вартостей, поділені на кількість цих вартостей можемо формулювати так:

$$\frac{I}{n} [\sum / \lg X_1 / + \sum / \lg X_2 / + \sum / \lg X_3 / + \ldots], \text{ а також}$$

$$\lg G = \frac{I}{n} / n_1 . \lg G_1 + n_2 \lg G_2 + n_3 \lg G_3 \ldots . /$$

Себто логаритм геометричної середньої цілої сукупности є рівним сумі добутків з кількости окремих частей сукупности на логаритми геометричних середніх цих частей, поділений на цілу частоту сукупности.

Другою важливою властивістю геометричної середньої є те, що коли маємо для кожної одиниці сукупности не одну, а дві вартості ознаки, або дві квантитативні ознаки й цікавимось геометричною середньою з відношення /поділу/ вартостей цих ознак, - наприклад, з відношення довжини руки до висоти тіла в сукупності людей - тоді геометричну середню можна зручно вивести за допомогою геометричних середніх з вартостей обох ознак. Коли, наприклад, візьмемо ознаки X і Y та їх відношення X/Y , тоді геометрична середня цього відношення з n одиниць сукупности буде рівною відношенню геометричних середніх величини X і величини Y , або G_x/G_y , бо $\lg \frac{X}{Y} = \lg X - \lg Y$ і коли підсумуємо всі n таких рівнянь та поділимо на число n , то одержимо:

$$\frac{\lg \frac{X}{Y}}{n} = \frac{\lg X}{n} - \frac{\lg Y}{n} \quad \text{себто} \quad \lg G_{x/y} = \lg X - \lg Y$$

відціль $G_{x/y} = \frac{G_x}{G_y}$

себто геометрична середня відношення /поділу вартости ознаки/ є рівною відношенню геометричних середніх цих вартостей. Аналогічно може бути доказанним, що геометрична середня добутків є рівною добутку геометричних середніх.

Ці властивості геометричної середньої мають значіння в деяких спеціяльних завданнях, а головним чином для розв'язання проблем індексних чисел, про які буде мова далі.

Геометрична середня для даного ряду неоднакових чисел є завжди меншою ніж аритметична середня з того ж ряду чисел.

Через те, що геометрична середня вживається найчастіше в часових статистичних рядах, себто таких, в яких студіюється зміни явища в часі, їх динаміка, цю середню звуть також д и н а м і ч н о ю с е р е д н ь о ю.

Гармонійна середня. Вважаємо потрібним зазначити ще один вид середньої - гармонійну середню. Вона буде для якогось ряду чисел рівною зворотній вартості аритметичної середньої з зворотніх вартостей цих чисел. Коли поодинокі числові вартості ознаки зазначимо як X_1, X_2, X_3.....X_n, то гармонійна середня з них буде

$$H = \frac{I}{\frac{I}{n}\sum\left(\frac{I}{X}\right)}$$

Зворотню вартість гармонійної середньої $\frac{I}{H}$ можемо вважати як аритметичну середню з зворотніх вартостей I/X себто:

$$\frac{I}{H} = \frac{I}{n}\sum\left(\frac{I}{X}\right)$$

Як приклад користання цією середньою можемо взяти випадок, коли, наприклад, ціни товарів наведені, як то інколи буває, у формі кількости штук /або вагових одиниць/ за одну грошову одиницю. Обчислення аритметичної середньої з ряду таких даних означає по суті обчислення гармонійної середньої /точнішої зворотньої величини $\frac{I}{H}$ / з цін, поданих у звичайній формі - себто скільки грошових одиниць відповідає одиниці /штука/ товару.

Взагалі одначе в статистичній практиці цей вид середньої вживається дуже рідко.

Нарешті мусимо зазначити, що не тільки аритметична середня, але й всі інші середні можуть бути обчислені я к о з в а ж е н і с е р е д н і. Для геометричної середньої досить помножити зворотні вартості /величини/ ознаки на якусь вагу. У зваженій геометричній середній на вагу множиться логаритм вартостей ознаки - з них аритметична середня є, як знаємо, логаритмом геометричної середньої. Отже геометрична пондерата.

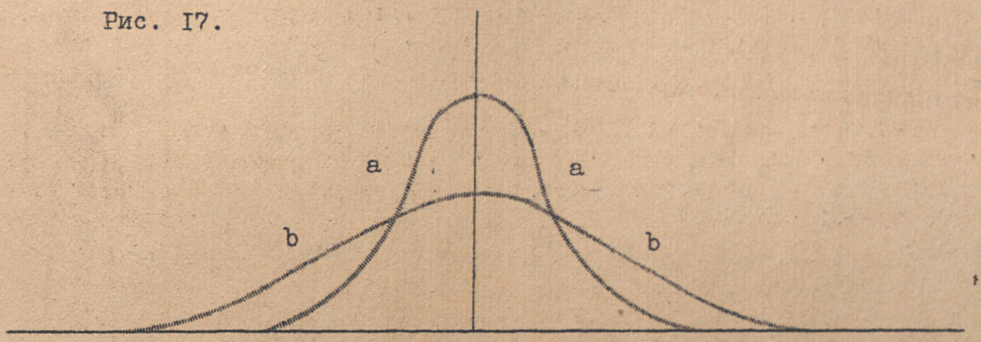
$$\text{Lg } G_p = \frac{\sum f \, \text{lg} X}{\sum X} \quad \text{або} \quad G_p = \sqrt[\sum f]{X_1^{f_1} \cdot X_2^{f_2} \cdots X_n^{f_n}} \quad \text{де "f" є вагою.}$$

34. Міра дисперсії /розхил/ змінної ознаки. а/ Загальні завваження. Всі середні величини є влучними характеристиками загальної висоти вартостей ознаки в сукупності - положення фреквенційної кривої на вісі абсцис. Але крім цього положення є ще інші властивості розподілу частоти якоїсь змінної, що вимагають також доброї числової характеристики. Будуючи таблицю розподілу частоти для якої будь змінної ознаки, ми спостерігаємо, що для кожної одиниці сукупности вартості ознаки міняються від одної одиниці до другої. Це коливання вартостей може набувати різних форм в залежності від того як близько купчаться окремі вартості ознаки біля аритметичної середньої величини даного ряду. Маємо

наприклад, дві таблиці розподілу частоти, які обидві дають форму симетричної кривої. В обох цих розподілах частоти аритметична середня, медіяна й мода в кожному будуть однакові ,але розподіли ці з погляду варіяційности ознаки /хитання її частоти/ можуть бути вельми різними.

В наступному рисунку - подаємо два розподіли частоти, в яких варіяція або розхил вартостей ознаки будуть зовсім різними:

Рис. I7.



В першому розподілі "а" як бачимо, вартості ознаки держаться дуже близько біля аритметичної середньої, в той час як в другому "в" хоч і купчаться теж коло аритметичної середньої, яко найчастішої величини ряду, одначе в цілому розкидані на більшому розхилі й більші відхилення від аритметичної середньої тут значно частіше зустрічаємо, ніж у першому розподілі частоти. Мусимо знайти певний числовий покажчик, числову характеристику цієї дисперсії ознаки в сукупності.

б/ Квадратичне /штандартне/ відхилення . З рисунку ми бачили, що більша варіяційність статистичного ряду /у формі кривої/ виявляється поруч з більшою частотою великих відхилень від середньої аритметичної, а через те пересічна величина відхилень вартостей ознаки від середньої величини й є покажчиком, мірою коливання ознаки або дисперсії статистичного ряду. Чим меншою є пересічна з відхилень усіх вартостей ознаки від аритметичної середньої, тим меншою є - дисперсія - розхил вартостей ознаки - і навпаки.

Нам одначе вже відомо, що альгебрична сума відхилень від середньої завжди є рівною нулю, тому обчислення пересічного відхилення від середньої не дасть нам можливости встановити розмір дисперсії. Через те вимір цього розміру необхідно переводити, виходячи або з абсолютних величин відхилень, себто брати їх без огляду на їх знак /додатній чи від'ємний/, або ще ліпше з квадрату відхилень, бо він /квадрат/ дає й для додатніх і для від'ємних відхилень завжди величину додатню,,а через те й сума квадратів ухилень од середньої аритметичної буде величиною, що різниться від нуля та яка набуває різних вартостей в залежності від більшої або меншої степені дисперсії ряду.

Візьмемо для прикладу сукупність з десяти одиниць наступних вартостей ознаки: /Дивись таблицю I8 на слід.сторінці/.

$\frac{\sum x^2}{n}$ Аритметична пересічна квадратів відхилень визначається відношенням але для кращого порівняння з іншими мірами /покажчиками/, в яких відхилення не є в квадраті, а також з самою аритметичною пересічною з вищенаведеного відношення витягнемо $\sqrt{\ }$ /корінь квадратовий/, тоді наша міра дисперсії, що звичайно означається грецькою літерою "б" /sigma/, буде формулуватися так

$$б = \sqrt{\frac{\sum x^2}{n}}$$

Таблиця I9.

Вартості ознаки	Відхилення від серед. аритметичної	Квадрат відхилень
X	Z - M = $\dot{x}$	$\dot{x}^2$
7	+ 3	9
2	- 2	4
5	+ I	I
4	0	0
3	- I	I
I	- 3	9
8	+ 4	I6
0	- 4	I6
6	+ 2	4
4	0	0
N = I0	M = $\frac{40}{I0}$ = 4	Сума $\dot{x}^2$ = 60

Отже в наведеному прикладі $б = \sqrt{\frac{60}{I0}} = \sqrt{6} = 2{,}45$.

Ця величина й є покажчиком дисперсії та зветься квадратовим або штандартним відхиленням[x]/

с/ Середнє квадратове відхилення. Операція обчислення квадратового відхилення ускладняється одначе тоді, коли сукупності й частоти вартостей ознаки великі. В таких випадках можемо вжити приблизного способу обчислення, що базується на певній альгебричній властивості квадратичного відхилення.

Означимо окремі вартості змінної як раніш через X, а літерою A певну величину /довільну основу/, від якої будемо встановляти відхилення, себто X - A буде означати відхилення вартости ознаки від цієї довільної величини. Це відхилення означимо через ξ . Коли підсумуємо квадрати відхилень від величини A , а одержану суму поділимо на суму частот тавитягнемо корінь квадратовий - будемо мати:

$$\sqrt{\frac{\Sigma\xi}{N}}$$

Цей вираз зветься "середнім квадратовим відхиленням від величини A"; означимо його літерою s . Це середнє квадратове відхилення різниться від штандартного відхилення тим, що окремі відхилення обчислені не від аритметичної середньої, а від довільної основи - величини A . Штандартне відхилення можемо одначе вважати за окремий випадок середнього квадратового відхилення, а то тоді, коли величина A буде цілком відповідати аритметичній пересічній. Інакше кажучи, штандартне відхилення є середнім квадратовим відхиленням від аритметичної пересічної. Між штандартним відхиленням "$б$" і середнім квадратовим відхиленням "s" існує певна залежність:

$s^2 = б + d^2$ тут $d = M - A$ /ріжниця між аритметичною середньою й довільною основою A/.

Таким чином квадрат середнього квадратового відхилення від довільно взятої основи A є рівним квадрату штандартного відхилення плюс квадрат ріжниці між пересічною й довільною основою.

x/ В літературі часто зустрічаємо й інші назви: Standard deviation, Standard. Abweichung, "квадратическая ошибка" або просто "середня похибка", "meanerror", "mittlerer Fehler".

Існуючий зв'язок між штандартним відхиленням і середнім квадратовим відхиленням можна висловити ще й інакше, коли підставимо замість s^2 та σ^2 їх значіння /вартості/, що встановлені були нами вище, тоді:

$$\frac{\sum \xi^2}{N} = \frac{\sum x^2}{N} + d^2$$

Цей зв'язок дуже важливий; через те пояснимо його далі.

$M - A = d$, себто $A = M - d$

Через це відхилення кожної даної вартости ознаки від A, яке означаємо "ξ", може бути визначено так:

$$\xi = X - / M - d / = X - M + d$$

Але $X - M$ є відхилення від аритметичної середньої, себто $= x$, тоді:

$\xi = x + d$, себто відхилення кожної вартости ознаки від довільної основи A є рівним відхиленню від аритметичної пересічної плюс різниця між аритметичною пересічною й довільною основою. Піднесемо обидві частини останнього рівняння в квадрат і тоді одержимо:

$$\xi^2 = x^2 + 2xd + d^2$$

це рівняння будемо мати для кожної вартости ознаки. Коли напишемо ці рівняння для всіх вартостей ознаки в сукупності й підсумуємо ліву й праву сторони рівняння - маємо:

$$\sum \xi^2 = \sum x^2 + 2d\sum x + Nd^2$$

понеже сума відхилень від аритметичної середньої $\sum x$ є рівною нулю, одержимо:

$$\sum \xi^2 = \sum x^2 + Nd^2, \text{ відціль}$$

$$\frac{\sum \xi^2}{N} = \frac{\sum x^2}{n} + d^2 \quad \text{замінимо}$$

$$\sqrt{\frac{\sum \xi^2}{N}} = s; \qquad \sqrt{\frac{\sum x^2}{N}} = \sigma \quad \text{будемо мати} \quad s^2 = \sigma^2 + d^2 \ /I/$$

Звернемо увагу на важливий наслідок, що випливає з останнього рівняння, а саме, що штандартне відхилення є меншим ніж середнє квардатове відхилення від довільної основи, коли остання не є рівною середній аритметичній. Дійсно, коли

$s^2 = \sigma^2 + d^2$, тоді $\sigma^2 = s^2 - d^2$, але d^2, яко кожний квадрат є завжди величиною додатньою. Таким чином, квадрат штандартного відхилення завжди на певну додатню величину є меншим, ніж квадрат середнього квадратового відхилення від довільної основи в тих випадках, коли ця основа не є рівною аритметичній пересічній.

З рівняння $\sigma^2 = s^2 - d^2$ випливає ще одна важлива властивість штандартного відхилення. Понеже $d = M - A = \sum \xi / N$ /див. стор. I02/, можемо написати, що

$$\sigma^2 = \frac{\sum \xi^2}{N} - \left[\frac{\sum \xi}{N}\right]^2$$

Коли візьмемо A = o, будемо мати замість відхилень ξ просто вартості ознаки / X /, тоді

$$\sigma^2 = \frac{\sum X^2}{N} - \left[\frac{\sum X}{N}\right]^2$$

Перший член правої сторони рівняння є аритметична середня з квадратів вартостей ознаки, а другий член є квадрат аритметичної середньої з вартостей ознаки. З цього випливає, що квадрат штандартного відхилення є рівним пересічній з квадратів минус квадрат пересічної з вартостей ознаки.

§ 40. Обчислення штандартного відхилення. Зв'язок, що з'ясований нами вище між штандартним відхиленням і середнім квадратовим відхиленням / $s^2 = \sigma^2 + d^2$ / дає можливість нам облегчити й обчислення конкретної величини штандартного відхилення. Можемо замість безпосереднього обчислення σ обчислити s^2, себто квадрат пересічного квадратового відхилення від конкретної довільної основи A, а вже від нього обчислити d^2, себто квадрат різниці між аритметичною пересічною та A; відціль одержимо σ^2, а далі й σ.

Облегчення є в тому, що можемо взяти довільну основу A таку, від якої відхилення було б легко обчислювати. Це облегшення є аналогичним тому, яке наводилось при обчисленні аритметичної середньої.

Таким чином, в першу чергу, треба обчислити квадрат пересічного квадратового відхилення від довільної основи A. Крім того, треба обчислити d - - різницю між A та аритметичною середньою. До встановлення цієї різниці можемо дійти й без обчислення аритметичної середньої, бо як вже було з'ясовано вище

$$M - A = d = \frac{\sum \xi}{N}$$

Себто різниця між аритметичною середньою та довільною основою, яка середній не рівна, є рівною альгебричній сумі відхилень од величини A, поділеній на цілу частоту сукупности, а цю суму відхилень в таблиці розподілу частоти обчислити дуже легко.

Візьмемо для прикладу таку таблицю розподілу частоти / Дані Юла /:

Таблиця 20.

Процент осіб, що одержують допомогу в округах Англії й Валісу		Частота / число округ /	Відхилення від довільної основи A	Добуток відхилень на частоту	Добуток квадратів відхилень на частоту
у вигляді змінної тяглої	у вигляді змінної розривної	f	ξ	fξ	fξ^2
I		2	3	4	5
0,75-I,25	I,0	I8	- 5	- 90	450
I,25-I,75	I,5	48	- 4	- I92	768
I,75-2,25	2	72	- 3	- 2I6	648
2,25-2,75	2,5	89	- 2	- I78	356
2,75-3,25	3	I00	- I	- I00	I00
3,25-3,75	3,5 = A	90	0	- 776	-
3,75-4,25	4	75	+ I	+ 75	75
4,25-4,75	4,5	60	+ 2	+ I20	240
4,75-5,25	5	40	+ 3	+ I20	360
5,25-5,75	5,5	2I	+ 4	+ 84	336
5,75-6,25	6	II	+ 5	+ 55	275
6,25-6,75	6,5	5	+ 6	+ 30	I80
6,75-7,25	7	I	+ 7	+ 7	49
7,25-7,75	7,5	I	+ 8	+ 8	64
7,75-8,25	8	-	+ 9	-	-
8,25-8,50	8,5	I	+ I0	+ I0	I00
	Разом	682		+ 509	
				$\sum f\xi$ = - 267	$\sum f\xi^2$ = 400I

Отже за довільну основу A беремо 3,5. Відхилення від основи A подаємо в графі 3-й для всіх кляс змінної. Кожне відхилення помножимо на частоту, а добутки підсумуємо /альгебрично/. Ця сума відхилень вартостей ознаки у всіх одиниць сукупности від величини A, себто $\sum \xi$ є рівною - 267. Ділимо її на частоту й одержимо:

$$\frac{\sum \xi}{N} = \frac{-267}{632} = -0,4225$$

Це число, як знаємо буде d , себто ріжниця між аритметичною середньою M й величиною A . Піднесемо його в квадрат і одержимо:

$$d^2 = 0,I785$$

Обчислимо далі $s^2 = \sum \xi / N$

В 4-й графі таблиці маємо добутки відхилень на частоту. Помноживши ці добутки на відхилення, одержимо добутки квадратів відхилень на частоту /графа 5/, які будуть, розуміється, числами додатними. Підсумуємо ці добутки й одержимо цілу суму квадратів відхилень од A, що є рівною - 400I. Поділимо її на частоту сукупности й одержимо

$$s^2 = \frac{\sum \xi^2}{N} = \frac{400I}{632} = 6,3307$$

Відіймемо далі d^2 від s^2, одержимо

$$\sigma^2 = 6,3307 - 0,I785 = 6,I522;\ \text{відціль}\ \sigma = \sqrt{6,I522} = 2,48$$

Але цю величину мусимо ще помножити на розмір інтервалу клясового, себто на 0,5 процента. Тоді σ = 248 x 0,5 = I,24 процента.

Як би наша змінна в таблиці розподілу частоти мала вигляд змінної розривної з інтервалом в одиницю, себто I, 2, 3, 4, 5 і т.д., тоді, розуміється, не потребували б рахуватись з інтервалами клясовими й наше обчислення штандартного відхилення було б цілком точним, не потребуючи такого множення конечної величини σ на розмір інтервалу клясового, як то робимо в даному прикладі. Одначе, у випадках змінної тяглої ми мусимо припускати, що пересічна вартість ознаки в одиниць кожного інтервалу клясового буде рівною середній вартості відповідного інтервалу, себто, що вартості ознаки в одиниць кожної кляси рівні пересічній величині відповідного інтервалу / 0,75 - I,75 =I, I,25 - I,75 = I,5 і т.д./. Таке припущення, як то ми бачили й при обчисленні аритметичної середньої дає тим більш точний результат, чим вірніше було наше припущення, що пересічна вартостей ознаки в екземплярів /одиниць/ кожної кляси є рівною пересічній вартості відповідного інтервалу клясового.

Через те, що таке припущення в дійсності не є ідеально точним і обчислення наше штандартного відхилення завжди буде лише приблизним. Тут ситуація наша буде трохи гіршою, ніж при обчисленню аритметичної середньої, бо коли йде справа про визначення пересічної з відхилень, не може бути ніякої компенсації помилок в обох частинах таблиці розподілу частоти, як то ми бачили при обчисленню аритметичної середньої. Тут нашим припущенням ми як правило, переоцінюємо відхилення як у верхній, так і спідній частині таблиці, а через те обчислене. таким способом штандартне відхилення буде завжди декілька більшим, ніж дійсне, докіль, звичайно, є оправданим припущення, що в дійсності частоти внутрі інтервалів клясових спадають в тому ж напрямі, в якому цей спад іде від інтервала до інтервалу.

Англійський статистик W.F. Sheppard, на підставі математичних міркувань, про які тут не будемо говорити, рекомендує робити виправлення до обчисленого вищезазначеним /приблизним/ способом квадрату штандартного відхилення, зменшуючи його на I/I2 = 0,0833. В нашому прикладі тоді будемо мати σ^2 = 6,I522 - 0,0833 = 6,0689, себто $\sigma = \sqrt{6,0689} = 2,46$ інтерва-

лу клясоврго, або-І,23 процента. Бачимо, що виправлення це в даному випадку є дуже незначним. Цей спосіб виправлення дає тим ліпший наслідок, чим ближчим є розподіл частоти до симетричної його форми.

Наведений вище спосіб обчислення штандартного відхилення дає нам можливість також і спосіб зручної контролі переведеного обчислення, бо взявши іншу довільну основу, мусимо прийти до однакового висліду, себто одержати тотожню величину штандартного відхилення.

Подібно до того, як при обчисленню аритметичної середньої, можемо й тут вжити способу обчислення, основаному лише на простому сумуванню.

Візьмемо ту саму таблицю розподілу частоти, як на стор.I04, в якій обчислення перших 3-х граф є тотожні.

Таблиця 2I.

Ріст у см.	частоти	суми перші	суми другі
I	2	3	4
I00 - I04	2	2	2
I05 - I09	7	9	II
II0 - II4	30	39	50
II5 - II9	69	I08	I58
I20 - I24	I03	2II	369
I25 - I29	I49	360 S_1^-=729	
I30 - I34	I94	S_0^-= 554	S_2^-= 590
A = I37 I35 - I39	229		
I40 - I44	250	S_0^+= 7I2	
I45 - I49	2I5	462 S_1^+= 905	S_2^+= 750
I50 - I54	I30	247	443
I55 - I59	64	II7	I96
I60 - I64	32	53	79
I65 - I69	I4	20	26
I70 - I74	6	3	6

Подібно до того, як в 3-й графі, виведемо інші суми /графа 4/ так, що будемо підсумовувати за порядком дані третьої графи; зпочатку згори аж до інтервалу I20-I24, який є віддалений на два місця від інтервалу довільної основи / I35 - I39 /, а потім знизу аж до інтервалу I50 - I54, який також віддалений на два місця /кляси/ від інтервалу довільної основи. Одержані так числа 4-ої графи, підсумуємо зокрема про верхню й зокрема про нижню частини таблиці. Ці суми означимо S_2^- і S_2^+. В нашій таблиці S_2^- = 590 а S_2^+ = 750.

Є доказаним,[x] що сума квадратів відхилень од довільної основи A є рівною:

$$\sum\xi^2 = /S_0^+ + S_0^-/ + 3/S_1^+ + S_1^-/ + 2/S_2^+ + S_2^-/ \quad /2/$$

Через те, щоб обчислити $\sum\xi^2$ треба проробити наступні операції:

$S^+ + S^-$	$3/S^+ + S^-/$	$2/S^+ + S^-/$	Разом
+ 554	+ 729	+ 590	I.266
7I2	905	750	+ 4.902
I.266	I634x3 = 4902	I.340x2 = 2.680	2.680
			$\sum\xi^2$ = 8.848 інтервалів

Поділимо це на цілу частоту сукупности й одержимо:

$$\frac{\sum\xi^2}{N} = S^2 = \frac{8848}{I495} = 5,9I84 \text{ інтервалу}$$

x/ Доказ цей наведений у Czubera "Die Statistischen Forschungs methoden" I вид. стор. 92-93, ми не наводимо з метою скорочення викладу.

$\Sigma\xi^2/N$, або $d = -\frac{333}{1495} = 0{,}02334$ інтервалу, а $d^2 = 0{,}0499$.

Віднімаючи d^2 від s^2 одержимо σ^2, який буде рівнятись:

$5{,}9184 - 0{,}0499 = 5{,}8685$, а $\sigma = \sqrt{5{,}8685} = 2{,}42$ інтервалу.

Помножимо одержану величину на розмір інтервалу = 5 см., тоді

$\sigma = 2{,}42 \times 5$ см. $= 12{,}10$ см.

§ 41. Обчислення штандартного відхилення цілої сукупности на підставі штандартних відхилень її частей. В практиці досить часто зустрічаємо необхідність визначення штандартного відхилення цілої сукупности на підставі штандартних відхилень частей цієї сукупности. Зазначений нами раніше зв'язок, що існує між "σ" і "s" дає можливість такого визначення.

Припустимо, що частота цілої сукупности є N , а частота двох часткових сукупностей /частей цієї сукупности/ N_1 і N_2 , себто $N = N_1 + N_2$ Аритметична пересічна цілої сукупности є M , а часткових M_1 і M_2 . Відповідні штандартні відхилення будуть σ, σ_1 і σ_2 . Дані $M_1 - M = d_1$ а також $M_2 - M = d_2$

Нам відомі штандартні відхилення σ_1 і σ_2 - треба знайти σ . Як це зробити ?

Відомі нам σ_1 і σ_2 є штандартні відхилення від аритметичних середніх M_1 і M_2 . Для того, щоб одержати σ є перш за все необхідним переобчислити їх від іншого початку /основи/ відхилень, а саме від M. На підставі відомого нам рівняння $s^2 = \sigma^2 + d^2$, мусимо додати до σ_1^2 та σ_2^2 відповідні d_1^2 та d_2^2 . Обидві ці одержані величини $\sigma_1^2 + d_1^2$ та $\sigma_2^2 + d_2^2$ пересічними квадратовими відхиленнями вартостей ознаки у відповідних часткових сукупностях від звичайної пересічної M. Коли означимо відхилення окремих вартостей ознаки від M, як завжди, через "x", дані суму квадратів відхилень од M у першій сукупності через $\Sigma_1 x^2$, а таку ж суму про другу сукупність через $\Sigma_2 x^2$, тоді будемо мати:

$$\sigma_1^2 + d_1^2 = \Sigma_1 x^2/N_1$$
$$\sigma_2^2 + d_2^2 = \Sigma_2 x^2/N_2$$
$$\Sigma_1 x^2 = N_1/\ \sigma_1^2 + d_1^2\ /$$
$$\Sigma_2 x^2 = N_2/\ \sigma_2^2 + d_2^2\ /$$

Дані, помноживши одержані величини $\sigma_1^2 + d_1^2$ та $\sigma_2^2 + d_2^2$ на відповідні частоти N_1 і N_2 , будемо мати суми квадратів відхилень від M, себто $\Sigma_1 x^2$ та $\Sigma_2 x^2$. Коли складемо обидві суми, дістанемо суму квадратів відхилень від M у цілій сукупності, себто

$$N_1/\ \sigma_1^2 + d_1^2/ + N_2/\ \sigma_2^2 + d_2^2/ = \Sigma x^2$$

Щоб одержати $\sigma = \sqrt{\Sigma x^2/N}$ залишається наведену суму поділити на N та витягнути корінь, тоді:

$$\sigma = \sqrt{\frac{N_1/\ \sigma_1^2 + d_1^2/ + N_2/\ \sigma_2^2 + d_2^2/}{N}}$$

Розуміється, що тотожній зв'язок існує й для сукупности, що складається з більшої кількости окремих сукупностей, а саме:

$$\sigma = \sqrt{\frac{N_1/\sigma_1^2 + d_1^2/ + N_2/\sigma_2^2 + d_2^2/ + N_3/\sigma_3^2 + d_3^2/\cdots}{N}} \qquad /3/$$

Для з'ясування цього обчислення скористаємось наступним прикладом /дані С.Кона/. Маємо 243 сільських господарств, що складаються з чотирьох

часткових сукупностей А, В, С, Д, Ознаков, що студіюємо, є посівна площа жита. Частоти окремих сукупностей /N_1, N_2 і т.д. загально N_i / наведені в графі 2 нижчеподаної таблиці, арітметичні середні ознаки в окремих сукупностях / M_i / наведені в графі 3-й цієї ж таблиці, а штандартні відхилення ознаки в окремих сукупностях в графі 4-й. Цілий образ будемо мати:

Таблиця 22.

окремі сукупності	число господарств N_i	пересічна посівна площа жита M_i	штандартне відхилення σ_i	цілі посівна площа жита $N_i M_i$	різниця між пересічною окремих сукупностей звичайною пересічною M d_i	квадрат цієї різниці d_i^2	квадрат штандартного відхилення σ_i^2	квадрат пересічного відхилення від звичайної середньої $\sigma_i^2 + d_i^2$	сума квадратів відхилень від звичайної середньої $N_i(\sigma_i^2 + d_i^2)$
I	2	3	4	5	6	7	8	9	I0
А	98	0,782	0,273	76636	-0,029	0,0008	0,0745	0,0753	7,3794
В	II0	0,843	0,292	92730	0,032	0,00I0	0,0853	0,0863	9,4930
С	3I	0,769	0,322	23839	-0,042	0,00I8	0,I037	0,I055	3,2705
Д	4	0,970	0,259	3880	-0,I59	0,0253	0,0671	0,0924	0,3696
	243			I97085					20,5I25
			M =	I97085 / 243	= 0,8II		σ^2 =	20,5I25 / 243	=0,0844
								σ = 0,29I	

§ 42. Правило "шости sigma". Щодо значіння штандартного відхилення, то є дуже важливим мати на увазі загальне /приблизно вірне/ правило, яке підтверджується в більшості випадків не дуже асиметричного розподілу частоти змінних. Це правило каже, що в межах шостикратного штандартного відхилення знаходяться майже всі вартості ознаки, що маються в цілій сукупності.

Так у прикладі, який ми подали вище /табл.20 стор.I22 /, штандартне відхилення σ визначено нами в розмірі I,24 процента. Помноживши його на шість будемо мати 6σ = 7,44 процента,- себто число досить близьке до вищої межи розхилу змінної /Процент осіб, що одержують допомогу/. В дійсності,як бачимо з таблиці, найвищий процент осіб, що одержують допомогу, складає 8,5. НА чім основано це правило зараз з'ясовувати не будемо, але зупинимось на ньому у своїх дальших викладах.

§ 43.Варіяційний коефіцієнт / коефіцієнт дисперсії /. Величина штандартного відхилення, як бачимо є о с н о в н о ю мірою варіяції вартостей ознаки змінних, себто статистичних рядів. Вона є найважливішою оцінкою того, оскільки в двох рядах, що мають однакові арітметичні середні, кожна з цих середніх є більш типовою, є більш надійною загальною характеристикою цілої сукупности. Той ряд, який має менше штандартне відхилення є більш стійким,дає більш нормальне коливання вартостей ознаки біля свого типу - середньої арітметичної. Одначе, при оцінці величини штандартного відхилення,треба мати на увазі, що ознака,набуваючи в ряді великих вартостей,природно мають коливання /хитання/ абсолютно більші, сильніші,ніж ознаки,вартості яких є малими.Відхилення,наприк.,висоти тіла окремих осіб від пересічного росту цілої їх сукупности виявляється в сантіметрах,але відхилення довжини пальця від його середньої довжини в цілій сукупності людей виявиться лише в міліметрах. Через те поруч з абсолютною мірою варіяції /дисперсії/, або штандартного відхилення, мусимо відшукати міру релятивної варіяції . За таку міру чи покажчик релятивної варіяції й береться звичайно відношення /звичайно процентове/ штандартного відхи-

лення до аритметичної середньої для кожного ряду зокрема, це відношення й зветься варіяційним коефіцієнтом або коефіцієнтом дисперсії.

Цей коефіцієнт означається частіше літеров V та формулюється так:

$$V = \frac{100\,\sigma}{M} \qquad /4/$$

Середнє відхилення. Найбільш важливою мірою варіяції ряду, як уже зазначалось вище, є штандартне відхилення. Але поруч з ним вживають і інших менш важливих мірників, а з цих другорядних найчастішим є середнє відхилення. Як нам відомо, обчислювати пересічну відхилень від середньої аритметичної, виходячи з алґебричної суми цих відхилень, поділеної на частоту сукупности, ми не можемо бо алґебрична сума цих відхилень завжди є рівною нулю.

Можливо, одначе, виходити з суми відхилень, не зважаючи на їх від'ємний або додатній знак, просто з суми абсолютних величин відхилень. Тоді мірою варіяції буде сума відхилень од аритметичної середньої, поділена на частоту сукупности; зазначаючи її через m, будемо мати:

$$m = \frac{\sum |x|}{N} \qquad /5/$$

Ця міра, що зветься "середнім відхиленням", не має тих корисних алґебричних властивостей, як штандартне відхилення, і аритметична середня, які дають можливість з ними зручно оперувати. Середнє відхилення можна обчислити як від аритметичної середньої, так і від медіяни, через те, що легко доказати, що середнє відхилення від медіяни є величиною меншою ніж середнє відхилення від якої будь іншої серії статистичного ряду /іншої величини тієї самої змінної/.

§ 44. Квартильне відхилення /quartil deviation/. Нарешті, для визначення варіяції вживається інколи також так званих к в а р т и л і в. Подібно до того, як медіяна в розподілі частоти розділює статистичний ряд на дві половини, так само й квартиль відмежовує одну четверту частину членів ряду; при тому квартиль, що відділює чверть ряду знизу зветься "нижнім", а що відділює чверть ряду вгорі - "верхнім". Перший означається звичайно Q_1, а другий через Q_2. Різниця між медіяною до кожного з квартилів може до певної міри також служити покажчиком варіяції /розкиданість/ вартостей ознаки. Через те, що при досить асиметричному розподілу частоти обидві ці різниці можуть бути нерівними одна одній, беруть, звичайно, половину суми цих різниць, тоді

$$Q = \frac{|Me - Q_1| + |Q_3 - Me|}{2} = \frac{Q_3 - Q_1}{2}$$

Коли $Q = I$, маємо нормальну дисперсію, коли $Q > I$, зверхнормальну, $Q < I$ піднормальну.

Абсолютний покажчик розкиданости вартостей ознаки будемо мати, поділивши на дві рівні частини різниці між верхньою та нижньою квартилями, а відносний покажчик дисперсії визначиться відношенням Q/Me.

Англійський статистик Боулі для обчислення розкиданости ряду пропонує іншу формулу, а саме:

$$Q = \frac{Q_3 - Q_1}{Q_1 + Q_3} \qquad /6/$$

Цей спосіб має перевагу тим, що дає безпосередню релятивну величину, якою користуємось як покажчиком дисперсії окремих членів статистичного ряду.

§ 45. Міра асиметрії або скошеності кривої. У випадку симетричної кривої /симетричної форми розподілу частоти/ всі члени вартостей ознаки, як зазначалось вище, майже повністю вкладаються в межах шостикратного штандартного відхилення / 6б /. Через це, коли спостерігаємо збільшення відносного числа випадків, що виходять поза ці межі, то це є доказом того, що асиметрія в розподілі частоти вартостей ознаки збільшується. Але це збільшення ще надає нам визначення ступеня асиметрії.

Необхідно знайти таку зручну характеристику, яка б одним числом зображувала ступінь асиметрії розподілу частоти. Найчастіше користуються для того формулою /Пірсон / наступною:

$\frac{M - M_o}{б}$, де M є аритметична середня, M_o - мода, а б - штандартне відхилення. Це й є так звана "міра асиметрії", "міра скошеності" / skewness/ фреквенційної кривої. У випадку симетричної кривої, як вже знаємо, M і M збігаються між собою; коли ж дві ці середні величини розходяться, тоді існує асиметрія, що збільшується просто пропорційно до різниці між двома порівнюваними величинами. Коли асиметрія правостороння, то завжди $M_o > M$, тому й різниця буде від'ємною, коли лівостороння, то, навпаки, різниця буде додатньою. Отже, коли розкиданість вартостей ознаки змінної дає форму присадкуватої кривої, тоді розходження середньої аритметичної й моди дає менший ступінь скошености, аніж таке розходження, яке дає форму стислої кривої, з піднятим її вершком. Наведена формула й має в собі елементи для порівняння форм асиметрії й її величини.

Таблиця.x/

Дорослі мужчини	
зріст у цалях	число осіб
57-	I
58-	3
59-	I2
60-	39
6I-	70
62-	I28
63-	320
64-	524
65-	740
66-	88I
67-	9I8
68-	886
69-	753
70-	473
7I-	254
72-	II7
73-	48
74-	I6
75-	9
76-	I
77-	I
Разом	6.I94

ВПРАВИ до лекцій I2, I3, I4 і 14-а.

До поданої ліворуч таблиці розподілу частоти визначіть наступні характеристики цієї сукупности:

I. Середню аритметичну - M
2. Медіяну - M_e /використавши формулу Фехнера/.
3. Моду - M_o.
4. Середнє квадратове відхилення /s /
5. Штандартне відхилення / б /.
6. Коефіцієнт дисперсії /варіяції/ - V
7. Міру асиметрії /скошености/.

x/ Майте на увазі, що інтервали ряду: "57-", "58-" і т.д. в дійсности є $56\frac{I5}{I6} - 57\frac{I5}{I6}$; $57\frac{I5}{I6} - 58\frac{I5}{I6}$. I цаля рівна 2,54 см.

Лекція п'ятнадцята.

Все, що ми виклали досі про статистичні характеристики сукупностей, є тим необхідним, що мусить увійти до основ теоретичної статистики з погляду опису явищ статистичнов методов. Але в перших моїх лекціях було вже звернено увагу на те, що крім статистичної ідеографії, перед кожним науково працювчим статистиком стоять і інші завдання - з'ясувати причинові зв'язки, що властиві колективним атиповим явищам та керувть ними. Ці наукові завдання і складавть предмет студій статистичної номографії.

Наведені нами характеристики статистичних сукупностей вживаються як при ідіографічних завданнях статистичного опису конкретної дійсности, так і при номографічних, себто при студіювванні причинових зв'язків. Але при першого роду дослідженнях, наведені раніше характеристики сукупностей є самі по собі завданням і метов, для студій номографічних вони є лише необхідним первісним ступенем дослідження, з якого наука виходить з тієв метов, щоб дальше, через порівняння статистичних характеристик, відшукати причинові зв'язки між явищами в тій або іншій сфері конкретної дійсности.

В другій частині викладів і будемо цікавитись дальшим використанням статистичних характеристик сукупностей, з метов студіювання причинових зв'язків між явищами за поміччв статистичної методи. Раніш, одначе, ніж приступимо до виконання поставленого перед нами завдання, мусимо ще зупинитись на спеціяльній важливій проблемі, що є тісно зв'язана з теоретичними основами середніх величин, а саме на проблемі індексних чисел або індексів, яким за останні часи, особливо з розвитком студіювання господарської коньюнктури, статистична наука й практика присвячує багато уваги.

§ 46. Індексні числа.

Особливу категорію середніх величин найчастіше зважених середніх складавть так звані індексні числа або індекси. Загально кажучи, це такі числові покажчики, які по своєму внутрішньому змісту дають рівнодійчу для цілого ряду чисел, з яких кожне характеризує якусь властивість певної категорії явищ. Нас можуть цікавити, наприклад, зміни в часі, а також і в просторі, такого явища, яке складається з цілого ряду окремих спеціяльних явищ, або виявляється назовні як ряд спеціяльних явищ, при тому емпірично можемо досліджувати лише ці спеціяльні явища. Щоб перейти від змін цих спеціяльних явищ до змін основного явища, що нас цікавить як щось ціле, і користуються індексами. Такий загальний індекс завжди синтезує цілий ряд явищ, як щось ціле, спільне всім тим спеціяльним явищам. Наприклад, рух цін кожного товару визначається рядом щорічних або щомісячних цифр, хитання яких для різних товарів можуть і збігатись і розходитись, себто ці хитання, можуть бути й однаковими й неоднаковими. Індекс таких цін і дає можливість ясно визначити нам як хитаються ціни всіх цих ріжних товарів в з а г а л і. Те ж саме можна визначити й для цілого ряду товарів якоїсь окремої їх категорії. Наприклад, хитання врожаїв окремих хлібів характеризуються дуже розбіжними рядами цифр. Індекс дає можливість одним числом визначити хитання врожаю в с і х в з а г а л і х л і б і в - виказати який взагалі був врожай певного року. Крім цього завдання, загальні числа - покажчики /індекси/ можуть мати й іншу мету: вони можуть дати вимір змін в якій будь величині, якої ми не можемо спостерігати безпосередньо, але яка, як ми добре знаємо, має певний вплив на багато інших величин, що їх ми можемо спостерігати, при тому вплив цієї основної величини маскується різними причинами, що різноманітно впливають на окремі величини. З цього погляду індекс є середня узагальнювча сукупність отих симптоматичних величин, що дозволяє судити про зміни тих основних причин, які є дійсним предметом нашого дослідження. Наприклад, цінність грошей, сама по собі, безпосередньо не піддається виміру, але ми можемо

зміни її визначити через зміну товарових цін. Проте, зміни в цінах як окремих, так і всіх взагалі товарів обумовлюються не тільки хитанням цінности грошей, як купівельного засобу, але також різними іншими причинами,що криються в умовах виробництва та збуту товарів. Наприклад, зміни в рівні заробітньої платні, що можуть цікавити нас, як наочна ознака стану ринку праці, залежать одначе не тільки від змін в загальних умовах ринку праці, але й від спеціяльних обставин виробництва й збуту в кожній зокрема галузі промисловости. Індекс і дає можливість синтезувати в одному числі вищезгадану різноманітність. Візьмемо такий приклад для обчислення індекса.

Маємо середні ціни за одиницю різних товарів для ряду років в абсолютних величинах - в карбованцях, франках і т.п.

Роки	Товари		
	A	B	C
I926	4,0	30	I5
I927	4,3	32	I4
I928	4,8	28	I6
I929	4,2	26	I2
I930	4,9	33	I8

Щоб простежити рух цих цін, мусимо порівняти їх з таким попереднім роком або періодом часу, який по серьозним міркуванням може вважатись вихідним,[x] основною базою для цікавого нам руху цін. Припустимо такою базою буде попередній п'ятирічний період перед війною /I909-I3 р.р./, а середні ціни за цей період для товарів A, B, C будуть наступні:

A - 3,8 = I00

B - 27 = I00

C - I3 = I00

Беручи кожну з цих цін /3,8; 27; I3/ за I00, обчислимо для кожного сучасного року релятивне значіння ціни для кожного року й товару сучасного п'ятиріччя, тоді будемо мати /відкидаючи десятичні знаки/

Роки	A	B	C
I926	I05	I06	II5
I927	II4	II8	I08
I928	I26	I04	I22
I929	II0	9I	92
I930	I28	I22	I38

Одначе ці релятивні ціни товарів A, B і C не будуть цілком задовольняти порівнянню тоді, коли ступінь обороту їх на ринку є різним, коли наприклад A є мука, B - пшоно, а C - перець, то вага їх у ринковому обороті не однакова, а для того, щоб зробити їх більш-менш однородними в цьому відношенні, необхідно порахуватись з вагою обороту кожного товару. Цю статистичну вагу й треба відшукати. Допустимо, що ця вага для товару A є рівною I0, для B - 5, а для C - I, тоді зважена середня для основи /бази все рівно будуть I00/ Мр рівняється:

$$\frac{I00 \times I0 + I00 \times 5 + I00 \times I}{I0 + 5 + I} = I00$$, але для сучасного

x/ Часто, наприклад, економісти для порівння сучасних цін беруть за базу І9І3 рік, або пятирічний період перед самою війною.

п'ятиріччя одержимо середні зважені, які дадуть такі величини:

Index numberus

Для 1926 року / Мр = $\frac{105 \times 10 + 106 \times 5 + 115 \times 1}{16}$ / = 106

1927 " . . . . . = 102

1928 " . . . . . = 119

1929 " . . . . . = 103

1930 " . . . . . = 127

Останні величини й будуть індекси, що визначають загальний рух цін всіх розглянутих товарів.

Подібні індекси можна обчислювати й для різних інших явищ. Наприклад, для змін загального розміру безробіття на підставі розміру безробіття в окремих галузях виробництва; для оцінки змін урожаїв зернових культур на підставі чисел про розмір урожаїв окремих зернових рослин; для виміру змін загального рівня курсів цінних біржових паперів, на підставі змін курсів окремих цінних паперів і т.д.

З вищенаведеного прикладу випливає, що для обчислення індексу, що характеризує зміну цін товарів, є необхідним вважати на вагу в товаровому обороті окремих товарів та встановивши таку вагу, користуватись для обчислення індексів media ponderata. Але в залежності від об'єкту, з яким оперуємо, і мети індексу не завжди є необхідним вживати зваженої середньої.

Як зазначає проф, Р. О р ж е н ц ь к и й в деяких випадках індекс доцільно обчислювати, не зважаючи на вагу.

Дійсно, хочемо, наприклад, знайти загальний індекс товарових цін з метою встановити зміни цінности грошей. Ці зміни мусять обумовлюватись змінами цін усіх товарів без виключення, при тому зовсім однаково як у тих товарах, що продаються в невеликій кількості, так і в тих, що мають великий оборот на ринку. Тут вага не потрібна, важливо лише взяти якнайбільше число товарів, бо цім збільшуємо можливість виявлення дійсно загальної для всіх товарів причини зміни ціни товарів. Зважена середня в даному випадку була б не відповідною, навіть шкідливою, бо в дійсності зміни в цінах товарів, що залежать від грошей, для окремих товарів замасковується тими змінами, що викликаються умовами виробництва і збуту отих окремих товарів і через те вага тільки збільшить значіння цих зайвих для нашої студії впливів. Навпаки, вага для нас обов'язкова, коли хочемо, наприклад, знайти загальний рівень дорожнечі життя, бо зміна цін кожного окремого товару відбивається в бюджеті споживача тим сильніше, чим в більших кількостях цей товар споживається.

Щодо формального складу /структури/ індексів, то звичайно розрізняються три їх форми, а саме:

1/ пересічні з окремих індексів, себто цінових відношень для двох різних періодів часу[x]/

2. відношення пересічних цін для двох різних періодів часу.

3. відношення створених будь яким способом загальних сум видатків або сум оборотів товарів для двох різних періодів часу.

В першій формі маємо безпосередньо основний зміст загального індексу такий, як ми його характеризували вище, себто як числовий покажчик, синтезуючий ряд поодиноких цінових відношень для окремих товарів. В цьому випадку першою головною операцією буде встановлення цінового відношення для кожного поодинокого товару, а потім з цих відношень обчислюється пересічна /зважена або не зважена/.

---

x/ або двох різних місцевостей, районів, міст, при чому це відноситься до всіх трьох форм тут наведених.

В другій формі першою операцією обчислення буде встановлення пересічної ціни всіх товарів за перший і другий період часу, а потім уже обчислення відношення цих пересічних. У третьої форми першою операцією буде встановлення загальних сум видатків для кожного періоду, а потім, так само як і в другому випадку, обчислення відношення цих сум.

Між цими трьома формами не можна провести цілком точну границю, бо досить часто певними змінами ці форми можуть бути перебудовані одна в другу

Коли через p_1 означимо ціни товарів у спостереженому періоді /рік, місяць і т.д./, а через p_0 ціни їх за період базисний /основний/, та через q_0 відповідну кількість товарів /спожитих або проданих/ за базисний період, то можемо обчислити індекс у формі третій:

$$\frac{\sum p_1 q_0}{\sum p_0 q_0}$$

Себто індекс показує, яким є відношення загальної вартости всіх товарів, кількість яких спожита в базисному періоді, але по цінах спостереженого періода-до загальної вартости тієї самої кількости товарів по цінах базісного періоду. Одначе, поділенням чисельника й знаменника наведеної формули на $\sum q_0$ можемо індекс написати в формі відношень пересічних /зважених/ з цін за обидва періоди, себто в формі другій.

$$\frac{\sum p_1 q_0}{\sum q_0} \Big/ \frac{\sum p_0 q_0}{\sum q_0}$$

Нарешті, цей ідекс може бути написаний способом , відповідаючим першій формі.

$$\frac{\sum \frac{p_1}{p_0} \cdot p_0 q_0}{\sum p_0 q_0}$$

себто в формі зваженої пересічної з поодиноких індексів p_1/p_0, де значіння ваги мають величини $p_0 q_0$, що суть суми видатків /або обороту/ базисного періоду.

В цій переміні формулювання індексу величини, які ми одержали яко вагу, мають логічний сенс, але не завжди це буває так. Візьмемо, наприклад, індекс примітивної форми, що вважається сучасними теоретиками за несправну, а саме відношення суми цін окремих товарів в період спостережений до суми цін базисного періоду /індекс третьої форми/:

$\frac{\sum p_1}{\sum p_0}$. Цей індекс можна написати також $\frac{\sum \frac{p_1}{p_0} \cdot p_0}{\sum p_0}$ себто в формі зваженої пересічної з цінових відношень, але за значіння ваги будемо тут мати ціни окремих товарів. Одначе, ціна певного товару залежить від одиниці, якою її міряємо. Ціна за тонну буде в тисячу разів більша, ніж ціна за кілограм. Таким чином, вага залежить тут від одиниць, в яких товар зміряно, які є цілком свавільними, а крім того вага не має розумної основи, бо не має сенсу надавати більшого значіння тому товару, ціна якого за одиницю є вища. Вважаємо необхідним це завважити через те, що цієї форми індексу вживається й досі, хоч вона є цілком несправною. Між іншим відомий американський індекс Bradstreets обчислюється за такою методою. Правда, там стараються усунути свавілля в виборі одиниць товарів тим, що ціни всіх товарів визначаються за тією самою одиницею ваги. Але й цим, навряд чи можна виправити хиби такої методи, бо існує така система одиниць товарів, для якої існуючі ринкові ціни не можуть бути приведені, так би мовити, до "одного знаменника". Вугілля оцінюється, наприклад, у франках за метричний центнар або за тонну, цукор за кілограм, а діяманти за карат. Коли б перевели всі товари до однієї вагової одиниці, виникає небезпека, що для деяких з них одержали б надзвичайно високі, а для деяких надзвичайно низь-

кі вартості цін і в той же час дуже мінливі, розумно необґрунтовані, величини мусили б нам служити вагою для обчислення зваженої пересічної.

Взагалі слід відмітити, що наведена вище друга форма індексу по відношенню до першої й третьої має підрядне значіння. Щодо відносного значіння цих двох останніх форм / І-ої і 3-ої/, які досить часто можуть бути замінені одна одною, то всеж таки важливіше значіння треба признати першій формі; як вже було зазначено вище ця форма безпосередньо визначає основу загального індексу, яко числового покажчика, синтезуючого поодинокі індекси або цінові відношення. Лише в цій формі можемо провести добре розподіл індексів на зважені й незважені та ясно визначити, що саме ми беремо за вагу, яка обумовлює вплив цін окремих товарів на загальний індекс.

Нарешті, лише в цій формі можна досягнути ясного порівняння індексів різної структури як серед цінових індексів між собою, так і між індексами цін та іншими індексами - заробітньої платні, розміру виробництва і т.п., в яких не маємо діла з сумами обороту або споживи, але в яких легко розпізнаємо простоту індексів з певною вагою або без неї.

Конкретних форм індексів відповідно їх змісту чи меті досить багато. Наприклад, І р в і н г Ф і ш е р [x] наводить їх більше ста тридцяти. Ми зупинимося на характеристиці лише головних з них, перш за все на індексах, що є незваженими пересічними цінових відношень окремих товарів. Тут треба завжди пам'ятати, що для вирішення питання чи є індекс зваженим чи ні, головним для нас критерієм є, що маємо пересічну з окремих цінових відношень /окремих індексів/ окремих товарів, а не з чогось іншого. В таких випадках ці цінові відношення будуть для нас вартостями ознаки X, в обох формах індексів як простій, так і в зваженій.

$$\text{а/}\quad \frac{\sum X}{n} \qquad \text{та} \qquad \text{б/}\quad \frac{\sum XV}{\sum V}$$

при тому в тих випадках, коли йде про аритметичну пересічну величина V буде вагою лише тоді, коли вона є у взаємному зв'язку з X, як то подано у взорі б/, та що числа, які є X є в інших взаємозалежностях, ніж величина V в наведеній формі, не є вагою.

В залежності від того, який вид пересічної візьмемо для індексу, останній також буде мати видові відміни - аритметичний, геометричний, гармонійний, а також медіяна й мода. Відміна індексу, збудована на підставі аритметичної пересічної, буде виглядати:

$$I_{01} = \frac{I}{n}\sum\frac{p_1}{p_0} \quad / I /$$

Тут літ. I є індекс, а "OI" означає, що взято відношення цін базисного періоду / O / до цін періоду, який ми спостерігаємо / I /. Навпаки, I_{10} буде означати індекс, де взято відношення цін спостереженого періоду до базисного. Під p_1 і p_0 розуміємо, як і раніш, ціни товарів відповідних періодів, а n є число товарів.

Цей індекс має ту основну ваду, що незадовольняє к р и т е р і ю з а м і н и ч а с у.

Під останнім І.Ф і ш е р розуміє такий критерій придатности взору індексу, який задовольняє умові:

$$I_{10} = \frac{I}{I_{01}}, \text{ або } I_{01} \times I_{10} = I$$

Себто, коли замість індексу визначаючого відношення цін періоду спостереженого / I / до цін періоду базисного / O /, обчислимо індекс, визначаючий зворотнє відношення, то цей другий індекс має бути рівним зворотній вартості першого, - іншими словами, що добуток прямого й зворотнього індексу буде рівним одиниці. Якраз дуже часто вживається в практиці індекс обчислений як проста /незважена/ аритметична пересічна з цінових відношень.

---

x/ Дивись його праць: The Making of Index Numbers

$I_{01} = \frac{\sum \frac{p_1}{p_0}}{n}$ а між тим він і не відповідає зазначеному вище критерію.

Для ілюстрації цього критерія візьмемо такий приклад.

Маємо, припустимо, три товари, ціни яких в різних періодах були

	період 0	період I
товар A	30	60
" B	20	30
" C	20	IO

Цінові відношення /окремі індекси/ будуть

	$\frac{p_1}{p_0}$	$\frac{p_0}{p_1}$
товар A	$\frac{62}{30} = 2$	$\frac{30}{60} = 0,5$
товар B	$\frac{30}{20} = I,5$	$\frac{20}{30} = 0,7$
товар C	$\frac{IO}{20} = 0,5$	$\frac{20}{IO} = 2.$

$\sum \frac{p_1}{p_0} = 4,0$ $\qquad$ $\sum \frac{p_0}{p_1} = 3,2$

Відціль $I_{01} = \frac{\sum \frac{p_1}{p_0}}{n} = I,33$, а $I_{10} = \frac{\sum \frac{p_0}{p_1}}{n} = I,07$

Таким чином $\frac{I}{I_{01}}$ буде рівною 0,75, але не I,07, як би те вимагалось критерієм.

Можна спостерегти далі таку комбінацію цін товарів в різних періодах, коли, при обчисленні по вищенаведеній формулі індексу, одержали б зовсім абсурдний висновок, а саме, що ціни пересічно збільшились від базисного періоду до спостереженого й, навпаки, від спостереженого до базисного, себто, коли б замість періодів часу мали б два райони або два міста A й B, то ціни в місті A були б пересічно вищими, ніж в місті B, а одночасно ціни в місті B були б вищі ніж в місті A. При великому числі товарів, звичайно, зрідка зустрічаємо такий абсурдний висновок, але сама можливість таких випадків вказує на те, що ця форма індексу має велику ваду.

Вада зазначеної вище формули індексу підтверджується також і тим, що вона не відповідає й другому критерію справности взору індексу чи так званому к р и т е р і ю і н т е р к а л я ц і ї /Борткевич/. Істота цього критерія є такою: коли маємо не два, а три різні періоди часу /0, I, 2/, то, обчисливши індекс другого періоду по відношенню до першого, себто I_{01} та помноживши його на індекс третього періоду по відношенню до другого, себто I_{12}, одержимо ту ж саму величину, яку б здобули, встановивши безпосередньо індекс третього періоду по відношенню до першого - I_{02} себто $I_{02} = I_{01} \times I_{12}$.

Це правило можемо узагальнити в той спосіб, що воно є дійсним і для більшого числа періодів. Дійсно, коли справедливі рівняння $I_{02} = I_{01} \times I_{12}$ та рівняння $I_{03} = I_{02} \times I_{23}$, то буде вірним і рівняння $I_{03} = I_{01} \times I_{12} \times I_{23}$ при n періодах формула має вигляд

$$I_{on} = I_{01} \times I_{12} \times I_{23} \times \dots I_{n-1\ n};$$

можемо додати ще один член: індекс першого періоду, встановлений на основі останнього /n /, себто I_{no}, тоді маємо таку загальну формулу:

$$I_{00} = I_{01} \times I_{12} \times I_{23} \times \dots I_{n-1,n} \times I_{no}$$

Беручи на увагу, що I_{00} завжди є одиниця, маємо, що й

$$I_{01} \times I_{12} \times I_{23} \times \dots I_{n-1,n} \times I_{no} = I$$

Таким чином, індекс відповідає критерію інтеркаляції тоді, коли добуток ряду індексів, в якому базисним періодом кожного наступного індексу є спостережений період попереднього, а спостереженим періодом останнього індексу є базисний період першого індексу - буде рівним одиниці. Відділ цей взаємний зв'язок може бути дійсним лише тоді, коли індекс задовольняє критерію інтеркаляції. Коли існує така взаємозалежність, мусить бути вірним і рівняння:

$$I_{01} \times I_{12} \times \dots I_{ik} \times I_{ko} = I$$

$$I_{01} \times I_{12} \times \dots I_{ik} \times I_{kl} \times I_{lo} = I$$

Ці рівняння не були б одночасно правдивими, коли б не було вірним те, що $I_{ko} = I_{kl} \times I_{lo}$, а це й є вислів критерія інтеркаляції. Ця властивість індексів створювати замкнуте коло вищенаведеного взору є в своїй основі лише іншим виразом критерія інтеркаляції, з якимизустрічаємось в статистичній літературі під назвою "колового /ланцюгового/ критерія" /circulator test/.

Цей критерій дуже важливий, бо дає нам доказ також того, що індекс незадовольняючий "критерію інтеркаляції", не задовольняє й критерію заміни часу "Критерій інтеркаляції" означає, що $I_{kl} \times I_{li} = I_{ki}$ та що $I_{ik} \times I_{kl} \times I_{li} = I$. Але поділивши друге рівняння на I_{ik} , переконаємось, що одночасна правдивість обох рівнянь є можливою лише тоді, коли

$$I_{ki} = \frac{I}{I_{ik}} \text{ , або } I_{ki} \times I_{ik} = I \text{, себто}$$

коли індекс задовольняє "критерію заміни часу". Таким чином, формули критерія заміни часу є лише особливим випадком для формули критерія інтеркаляції або, інакше, критерія ланцюгового. З цього випливає, що коли індекс задовольняє критерію інтеркаляції, то він задовольняє й критерію заміни часу. Навпаки, коли індекс задовольняє критерій часу, то з цього ще не випливає, що він обов'язково відповідає також і критерію інтеркаляції.

Наведений вище критерій має велике практичне значіння тоді, коли хочемо перевести якийсь раніш обчислений індекс на нову основу, себто змінити базисний період, з якими порівнюємо спостережений період часу. Маємо наприклад, індекс цін 1933 року встановлений у відношенні до 1913 року, але хочемо перейти до виміру рівня цін року 1933 по відношенню до року 1920. Це означає, що маючи I_{02} хочемо знайти I_{12} , або маючи I_{il} , знайти I_{kl}. Згідно з критерієм інтеркаляції $I_{il} = I_{ik} \times I_{kl}$, тоді $I_{kl} = \frac{I_{il}}{I_{ik}}$ себто, знаючи I_{il} , можемо діленням його на I_{ik} одержати I_{kl}.

Таким чином, знаючи індекс цін по відношенню 1933 року до 1913 року, і бажаючи знайти індекс 1933 до 1920 року, мусимо перший індекс поділити на індекс встановлений у відношенню цін 1920 до 1913 року. Через те, що звичайно, опубліковуються цінові індекси до якоїсь однієї базисної основи, найчастіше, до 1914 або 1913 років, нам часто буває корисним мати можливість взаємно порівнювати довільно взяті два роки, себто одержати індекс про довільний рік 1 з базою в другому довільному році k на підставі ряду вже опублікованих індексів з загальною базою в році "i". Це є можливість лише в таких індексів, які задовольняють критерію інтеркаляції.

Цей критерій й переконує нас, що наведена вище формула індекса /1/, основана на аритметичній середній не є справною.

Цей самий критерій підтверджує й те, що для таких індексів найліпше користуватись геометричною середньою:

$$I_{01} = \sqrt[n]{\frac{p'_1}{p'_0} \cdot \frac{p''_1}{p''_0} \cdot \frac{p'''_1}{p'''_0} \dots} \qquad /2/$$

Індекс, побудований на цій середній задовольняє обом згаданим вище критеріям. Дійсно, на підставі відомих нам /стор. 117-18/ властивостей геометричної середньої, будемо мати:[x]

$$I_{01}^{(2)} = \sqrt[n]{\frac{p_1'}{p_0'} \cdot \frac{p_1''}{p_0''} \cdot \frac{p_1'''}{p_0'''} \cdots\cdots} = \frac{\sqrt[n]{p_1' \cdot p_1'' \cdot p_1''' \cdots\cdots}}{\sqrt[n]{p_0' \cdot p_0'' \cdot p_0'''}}$$

$$I_{12}^{(2)} = \sqrt[n]{\frac{p_2'}{p_1'} \cdot \frac{p_2''}{p_1''} \cdot \frac{p_2'''}{p_1'''} \cdots\cdots} = \frac{\sqrt[n]{p_2' \cdot p_2'' \cdot p_2''' \cdots\cdots}}{\sqrt[n]{p_1' \cdot p_1'' \cdot p_1''' \cdots\cdots}}$$

$$I_{02}^{(2)} = \sqrt[n]{\frac{p_2'}{p_0'} \cdot \frac{p_2''}{p_0''} \cdot \frac{p_2'''}{p_0'''} \cdots\cdots} = \frac{\sqrt[n]{p_2' \cdot p_2'' \cdot p_2''' \cdots\cdots}}{\sqrt[n]{p_0' \cdot p_0'' \cdot p_0''' \cdots\cdots}}$$

Коли помножимо I_{01} на I_{12}, а після скоротимо на $\sqrt[n]{p_1' \cdot p_1'' \cdot p_1'''}$, одержимо рівняння:

$$I_{01}^{(2)} \times I_{12}^{(2)} = I_{02}^{(2)}$$

яке задовольняє як критерію інтеркаляції, так і критерію заміни часу. В користь геометричної середньої для простих індексів цінових свідчить ще й те що, як показують дослідження, розподіл частоти окремих цінових індексів набирає в більшості випадків значно асиметричної форми, при чому асиметрія досить зменшується, коли від окремих звичайних індексів цін перейдемо до їх логаритмів, себто складемо таблицю розподілу частоти логаритмів окремих індексів біля їх аритметичної середньої. Співмірність розподілу логаритмів означає також і співмірність розподілу релятивних відхилень індексів од геометричної середньої, інакше, що відхилення n-разів донизу є однако часті, як і відхилення n -разів вгору. Це свідчить про те, що геометрична середня з окремих індексів цінових є більш типовою характеристикою, ніж аритметична пересічна. Деякою вадою геометричного індексу є те, що він більше складнішої конструкції та потрібує користання логаритмичними таблицями, але з погляду методологічного він є бездоганним, в тих випадках, коли не можемо або не потребуємо обчислювати зваженого індексу.

Треба також зазначити, що ані гармонійна середня для індексу $I_{01} = \frac{I}{\frac{I}{n}\sum\frac{p_0}{p_1}}$ ані медіяна та мода не мають ніякої переваги перед аритметичною середньою й через те практично не мають серйозного значіння при обчисленню такого роду індексів.

Перейдемо далі до розв'язання питання: чим треба керуватись при обчисленні зважених індексів?

Щоб важення було логічно правильним, з огляду на мету індексу, треба в формі ваги відбивати значіння певного товару в споживанні /коли будуємо індекс видатків на життя/, або його значіння в обороті на ринку даної країни./коли індекс має іншу мету/. Чим же слід міряти таке значіння? Не можемо його міряти безпосередньо кількістю товарів /спожитих, або існуючих в обороті/, бо маси різних товарів є величини взаємно непорівняльні. Крім того таким способом зважені індекси, оскільки б ми мусили мати діло з їх кількістю, залежали б на одиницях, в яких товари міряні, а про незручність різних систем міряння товарів /тонн, кілограм, карат/ ми вже говорили вище.

Отже, ліпшою вагою для цінових індексів можуть бути лише суми видатків на річні товари, або відповідні суми оборотів різних товарів, інакше кажучи, добутки кількости товарів на їх ціну.

Понеже в кожному індексі порівнюються завжди два періоди часу, виникає питання для якого періоду треба брати оті суми видатків або оборотів - до базисного чи спостереженого? Можна брати й для одного й для другого.

Індекс можемо конструювати за взором зваженої аритметичної середньої так:

---

x/ Число /(2)/ в дужках вгорі індексу вказує на те, що беремо другу по черзі формулу, а не першу.

$$I_{01}^{(3)} = \frac{\sum \frac{p_1}{p_0} \cdot p_0 q_0}{\sum p_0 q_0}$$ але можемо подати й в іншій формі: $$I_{01}^{(4)} = \frac{\sum \frac{p_1}{p_0} \cdot p_1 q_1}{\sum p_1 q_1}$$

У першій формулі множники ваги / p та q / взяті з базисного періоду, в другій - з періоду спостереженого. Але не буде зовсім шкідливим, коли б ми взяли ціни з основного періоду, а кількість товару з спостереженого, або навпаки:

$$I_{01}^{(5)} = \frac{\sum \frac{p_1}{p_0} \cdot p_0 q_1}{\sum p_0 q_1} \qquad I_{01}^{(6)} = \frac{\sum \frac{p_1}{p_0} \cdot p_1 q_0}{\sum p_1 q_0}$$

Замітимо, що формули $I^{(3)}$ і $I^{(5)}$ можуть бути легко написані й в третій з дефінованих вище форм індексів / дивись стор. 131 /, зокрема в формі відношень сум видатків.

$$I_{01}^{(3)} = \frac{\sum \frac{p_1}{p_0} \cdot p_0 q_0}{\sum p_0 q_0} = \frac{\sum p_1 q_0}{\sum p_0 q_0}, \text{ а}$$

$$I_{01}^{(5)} = \frac{\sum \frac{p_1}{p_0} \cdot p_0 q_1}{\sum p_0 q_1} = \frac{\sum p_1 q_1}{\sum p_0 q_1}$$

В першій формулі маємо відношення вартости маси товарів базисного періоду по цінах спостереженого періоду - до вартости маси товарів базисного періоду по цінах цього періоду. В другій маємо відношення вартости маси товарів спостереженого періоду по цінах цього періоду - до вартости маси товарів спостереженого періоду по цінах базисного періоду. Обидві формули цілком влучні для характеристики ступеня подорожчання товарів. Форми $I^{(4)}$ та $I^{(6)}$ з погляду "третьої форми" індексів не дуже ясні для викладання, а через те й практичне значіння їх менше.

Треба також мати на увазі, що в формулах $I^{(3)}$ та $I^{(5)}$ вага $p_0 q_0$ та $p_0 q_1$ знаходяться не в прямому зв'язку з ціновими відношеннями p_1/p_0, бо величина p_0, яка є чисельником останніх, є множником ваги. При непрямому зв'язку між вагою та вартістю ознаки зважена аритметична середня є нижчою, ніж проста аритметична середня /незважена/. Через те формула $I^{(3)}$ та $I^{(5)}$ дають взагалі менші величини індексів, ніж по формі незваженої аритметичної середньої $I^{(1)}$. У формулах $I^{(4)}$ та $I^{(6)}$, навпаки, залежність між ціновими відношеннями p_1/p_0 та вагою $p_1 q_1$ та $p_1 q_0$, що викликана величиною p_1, яка є в обох спільним множником,- є прямою і через це ці формули, як правило, дають більший результат як незважена пересічна.

З наведених чотирьох формул жадна не задовольняє критерію заміни часу, а через те й критерію інтеркаляції, і в цьому їх основна вада. Крім того, вони потрібують даних про масу товарів / у споживанні чи обороті/ за кожний період, себто конкретно, за кожний рік, місяць або й тиждень, бо публікуються й тижневі індекси. Одержання таких даних взагалі є тяжким, а інколи й зовсім неможливим. В зв'язку з цим треба мати на увазі дуже поширену форму зваженого індексу, що висунута проф. Б о р т к е в и ч е м. В цій формулі кількісті товарів - q є константні, себто незмінні від одного періоду до другого:

$$I_{01}^{(7)} = \frac{\sum \frac{p_1}{p_0} \cdot p_0 q}{\sum p_0 q} = \frac{\sum p_1 q}{\sum p_0 q}$$

Ця форма задовольняє обом критеріям, бо дійсно

$$\frac{\sum p_1 q}{\sum p_0 q} \times \frac{\sum p_2 q}{\sum p_1 q} = \frac{\sum p_2 q}{\sum p_0 q}$$

Маючи дані про кількість товарів у споживанні чи обороті за окремі періоди, можемо брати пересічні за цілий ряд періодів і ці пересічні вважати вагою q. На випадок, коли б не мали всіх необхідних чисел, можемо обмежитись і пересічною чисел з неповного числа періодів, або даними лише з певного поодинокого періоду, для якого маємо дані.

Цей спосіб важення індексів є особливо влучний для індексів, що показують рівень видатків на життя, при яких, звичайно, спеціяльним дослідженням споживчих бюджетів, за певний період часу, встановлюється певна сукупність товарів /продуктів/ в певних кількостях і вона для цілого ряду років береться, звичайно, яко основа для обчислення індексів.

Нарешті, яко константні, можемо брати не тільки множники q у вазі, але й саму вагу. Форма цього зваженого індексу мала б в такому випадку наступний вигляд:

$$I_{01}^{(8)} = \frac{\sum \frac{p_1}{p_0} . pq}{\sum pq}$$

або зазначаючи вагу /pq/ одною літерою V, будемо мати:

$$I_{01}^{(8)} = \frac{\sum \frac{p_1}{p_0} . V}{\sum pV}$$

Такого роду важення вживається найчастіше для індексів гуртових цін, про які не має часто точних даних відносно розміру обороту й вихід з такого положення знаходимо в встановленні константих ваг.[x] Одначе й цей спосіб має ту вади, що не задовольняє критерію заміни часу, а через те й критерію інтеркаляції, бо

$$\frac{\sum \frac{p_0}{p_0} . V}{\sum V} \times \frac{\sum \frac{p_1}{p_0} . V}{\sum V}$$ не є рівною одиниці.

Щодо відношення індексів $I^{(7)}$ та $I^{(8)}$ до незваженого аритметичного індексу, то через ті самі умови як і в індексах $I^{(3)}$ та $I^{(5)}$, вони мають тенденцію давати менші величини ніж $I^{(1)}$.

З **аритметичних** зважених індексів, індекс форми $I^{(7)}$ має методологічну перевагу над рештою. Має він перевагу також і над зваженими **неаритметичними** індексами - геометричним, гармонійним, а також індексами з медіяною та модальним. Зважений геометричний індекс задовольняє критерію інтеркаляції лише при умові константности ваги /pq/, одначе, є менш вигідним, ніж $I^{(7)}$ через складність своєї конструкції.

Не зважаючи на важливе практичне значіння теорії індексів, ми змушені тут лише коротко зупинитись на основах методології індексів. Сучасних визначних теоретичних праць по теорії індексів є досить багато, проте можемо **тут звернути увагу тих, хто хотів би доповнити викладені дуже коротко** теоретичні основи індексів. Ці праці наступні: 1/ Irving Fischer "The Making of index Numbers", 2/prof.Bortkewicz "Zweck und Structur einer Preisindexzahl / Nordisk Statistik Tidskrift 1923,1924/ 3/ prof.W.C.Mitchell "Index Numbers" в збірнику Business Statistics /ed by M.T.Copeland, Cambridge 1917/, та в Bulletin of the United Bureau of Labor Statistics

---

x/ Наприклад, Комісією Брітанської Асоціяції Наук було запропоновано взяти такі норми ваги: м'ясо 10, вугілля 10, пиво 9, молочні пролукти 7,5, пшениця 5, ячмінь 5, залізо 5, риба 2,5, нафта 1 і т.п.

Лекція шістнадцята.

ЧАСТИНА II.

ДОСЛІДЖЕННЯ ПРИЧИНОВИХ ЗВ'ЯЗКІВ.

39. Номографічні завдання статистичної методи. Першими кроками наукового статистичного дослідження, як ми вже знаємо, є відповідне підготовлення матеріялу спостережених масових явищ. Ми стараємося підмітити в спостережених масах п р а в и л ь н о с т и співіснування, послідовності розвитку явищ та, підслідивши, точно їх констатувати. Коли одначе встановлені правильності при дальшому дослідженні повторюються в просторі й часі, себто виявляють певну сталість,- то маємо право говорити вже не про правильності, а про з а к о н о м і р н о с т і, про встановлення п е в н и х е м п і р и ч н и х з а к о н і в.

Статистичні узагальнення, що мають характер емпіричних законів, можуть належати до одного з наступних видів:

І. З а к о н о м і р н о с т і с п і в і с н у в а н н я, які констатуються в масі, що знаходиться або мислиться в стані спокою. Такими будуть, наприклад, співвідношення між різними частинами тієї самої маси: розподіл населення по статі, віку, співвідношення народжень чоловічої й жіночої статі і т.п.

2. З а к о н о м і р н о с т і ч а с т о т и /інтенсивности/, що висловлюють залежність дослідженого явища від створившого це явище оточення.

3. З а к о н о м і р н о с т і р у х у, або розвитку, що висловлюють зміни масових явищ у часі; наприклад, зміни смертности людей на протязі століття, або зміни її в зв'язку з віком людей.

Установлівчи вказані закономірності, ми залишаємося в полі лише н а у к о в о г о о п и с у я в и щ, роблячи тим дуже корисну й необхідну наукову працю, що збільшує людське знання та підготовлює індуктивний матеріял для з'ясування явищ, що відбуваються в суспільстві, або важливі для нього.

Наукове завдання не обмежується, одначе, одним констатуванням існуючих правильностей. Людський розум намагається звести ці правильності й закономірності до зв'язків причин і наслідків, звести складні комплекси причин до простіших принципів, та встановити закони діяння цих причин, себто з'ясувати явище.

Індуктивні висновки, за помічю яких статистика констатує зв'язок одних явищ з другими, яко зв'язок причини й наслідку, і складає четверту важливу групу с т а т и с т и ч н и х з а к о н і в п р и ч и н о с т и. До них слід, наприклад, віднести висновки про зв'язок цін на предмети першої необхідности зо злочинність, зв'язок житлових умов з захоруванням і смертність, про вплив розміру врожаю на товаровий обмін і т.под.

Згадане вище завдання статистичної методи й зветься с т а т и с т и ч н о ю н о м о г р а ф і є ю, що є завершенням наукового оброблення статистичного матеріялу.

Головним засобом до віднайдення й встановлення причинових зв'язків двох або декількох явищ статистикою є зіставлення статистичних рядів, в які зводиться плянова кожна статистична сукупність, при тому у своїх міркуваннях статистик завжди користується методами індуктивної логіки. Вихідною точкою цих метод є властива людині, як результат вікового досвіду, певність у тому, що все на світі, а значить і суспільні та інші явища, відбувається в певному строго закономірному порядку і що рівні причини викликають й рівні наслідки, себто, що наслідки пропорційні причинам. Без цієї певности не може бути якогось наукового дослідження.

Дуже важливо також правильно розуміти терміни "причина" й "наслідок".

Простого співіснування або хронологічної послідовности двох чи декількох явищ ще не досить для того, щоб визнати існування причинової взаємозалежности між ними. Приклади переконують нас, що часто між двома або декількома спостереженими фактами стверджується причиновий зв'язок тоді, коли між цими фактами в дійсності існує тільки звичайна хронологічна послідовність. Небезпека ця дуже велика для статистика і через те в своїх висновках він мусить бути надзвичайно обережним.

Є й друге джерело помилок, небезпечне для статистика, яке гарно схарактеризовано Л а п л а с о м: "Наші страсті, забобони й пануюча опінія, перебільшуючи імовірност , що їм сприяють, та зменшуючи значіння протилежних імовірностей, є поважним джерелом небезпечних ілюзій".

Отже треба пам'ятати, що те, що ми звемо причиною, є не тільки попереднє в часі тому, що звемо наслідком; причина відноситься до наслідку так, що коли б її або рівнозначної їй події не відбулось, то не було б і самого наслідку. Та умова, при відсутности якої явище не відбулось би так, як воно відбулось в дійсності, і є причина"./Мінто/

З поняттям причини ми зв'язуємо одначе й дещо більше. Ми припускаємо, що дана причина, якщо сукупність інших умов залишиться без змін, і в будучому незмінно буде викликати той самий наслідок. Треба також мати на увазі, що в філософському розумінні кожна причина, чи діє вона на протязі величезного часу, чи тільки одну мить, о д н а к о в о н е о б х і д н а для того, щоб явище відбулось так, як ми його спостерігаємо в дійсности, а не якось інакше. Випадковости, філософськи кажучи,- не має. Випадок є фікція, міра нашого незнання. Ми звемо випадковим те, що не зможемо з'ясувати.

Далі, категорії постійних і перемінних причин мають лише умовне значення. Вони лише допомагають дослідникові в тяжкій праці ананлізи складного комплексу причин і наслідків та в оцінці релятивного значення тої або іншої причини в створенні результата.

В цьому умовному значенні й треба розуміти категорії постійних і перемінних причин, себто, що перші /постійні/ є ті, які діють постійно й завжди в однаковому напрямі та обумовлюють вияв статичних /від слова "статика" - Л.Ш./ серій, а другі є ті, що діють постійно, але змінюють свою інтенсивність і напрям.

Як приклад, подамо тут одну з влучних схем, що розрізняє три роди причин або факторів:

I. фактори фізичні: клімат, ґрунт, тепло, вогкість, географічна конфіґурація, орографія, атмосферні зміни і т.под.

2. фактори антропологічні: стать, вік, фізичні й психічні властивости, дідичність, раса.

3. фактори соціяльні: ґущавина, скупченість /міста, села/ і розподіл населення, розподіл багатства й ступінь нерівности станів, законодавство, політична конструкція, ступінь духовної культури, релігія, громадський стан, професія, індустріялізм і т.і.

Більшу частину факторів третьої категорії слід віднести до причин перемінного характеру.

Не будемо зупинятись на формулюванні метод індуктивної логіки /п'ять/, вони мусять бути читачеві відомі,[x] але важливо тут відмітити, що висновки зроблені прикладенням одного або декількох індуктивних логічних принципів мають силу повного логічного доказу тільки тоді, коли дослідник певен, що він взяв на увагу всі можливі обставини, які можуть мати вплив на створення дослідженого результату. "Можливість встановити закономірності явищ, чи то в вигляді емпіричних законів, чи то в вигляді законів, що розкривають кавзальний /причиновий/ зв'язок, і таким чином з'ясувати окремі явища, в їх конкретних умовах, обмежена тою областю, в якій існуючі перед нашими очима явища залежать од умов, що дозволяють точно висловити їх в загальних поняттях і виділити кожну окрему умову за поміччю досвіду або хочби шляхом

---

x/ Дуже систематично й гарно викладені методи індуктивної логіки в праці Джона Стюарта Мілля "Система логіки".

логічної аналізи; обмежена далі області, в якій можливе знання всіх без виключення фактів, впливаючих на склад і зміну явище, та в якій сукупність невідомих фактів залишається безумовно незмінно-постійною /Зігварт/.

В емпіричній дійсності ці необхідні умови майже ніколи нездійсними. Відсіль і випливає необхідність ставити явище, що досліджуємо, в точні умови досвіду, переводити за певним пляном штучну комбінацію між явищами й їх елементами, виходячи з більш-менш імовірних припущень про існування між ними причинової залежности - це й означає робити науковий досвід. При вдало поставленому досвіді дослідник може врахувати всі без виключення елементи такої комбінації й певний в тому, що він не проглядів жадної обставини даного причинового ланцюга, може спільно оперувати методами індуктивної логіки.

Але навіть і в царині природознавства /фізіолонічні, метеорологічні явища і т.п./ вищенаведеної вимоги логіки додержати не можна.

"Множність причин" і "переплутаність наслідків" часто не можна усунути,- залишається якась решта, що невкладається в додержані рямки індуктивних метод. Відсіль випливає необхідність масового спостереження метеорологічних явищ і широке прикладення статистичної методи взагалі до колективних, атипових явищ. У всіх випадках де, "множність причин" і "переплутаність наслідків" не може бути усунута, прикладається система правил і засобів, що об'єднується поняттям статистичної методи.

Через комбінацію й зіставлення рядів ми маємо можливість елімінувати /усунути/ цілі групи причин, виділити окремі причини й дослідити їх вплив, себто до певної міри заміняти таким чином досвід. Зустрічаючись з певною сукупністю невідомих фактів, ми інколи користуючись студіюванням стійкости статистичних рядів, одержуємо можність твердити, що сукупність цих невідомих нам фактів не змінилася.

Дуже важливо є знати, що статистичні закони причинности ніколи не досягають ступеня повної певности. Вони завжди залишаються більш-менш приблизними імовірностями.

Звичайно, з цього не випливає якоїсь безнадійности, бо й факт, що сонце завтра зійде над обрієм не є певний, а лише в високому ступені імовірний.[x]/

Таким чином, встановити закон якого-будь явища щастить лише в рідких, виключних випадках. Знайти закон явища - це значить знайти такий зв'язок між цим явищем і деякими умовами, передуючими цьому явищу або йдучими в парі з ними, по якому можна, виходячи з цих умов, встановити кожний вид і кожний ступінь цього явища. В своїй закінченій формі такі закони висловлюються в вигляді точних математичних /функціональних/ зв'язків між умовами й явищем, або між причиною й наслідком.

В більшості випадків причини явищ настільки складні, що не має можливости не тільки встановити певні зв'язки між ними й явищами, але й просто виділити їх. У всіх подібних випадках треба або відмовитись од дослідження явища, або поставити завдання в іншій формі, а саме: дослідити або виявити закони наставання явища в великому числі випадків, у кожному з яких можливо, але не обов'язково, таке наставання. Розуміється, що в такій постановці питання відносно дослідження даного явища ставиться цілком на статистичний ґрунт і завдання наукового дослідження зводиться до студіювання відповідної статистичної сукупности.

Другим науковим завданням номографії є передбачити ненаступивше ще явище або відновити бувше по знайдених законах. Передбачити явище або відновити його на підставі певних умов даного моменту, часто буває неможливим

---

x/ Коли згідно з християнським літочисленням уважати I933 рік 744I-им од сотворення світу, то факт сходу сонця I січня I933 року мав би імовірність: $\frac{7440 \times 365 + I}{7440 \times 365 + \frac{I}{2}} = 0{,}9999996$, а імовірність, що воно не зійде рівнятиметься 0,0000004.

і тоді навіть, коли відомі закони, по яких явище відбувається. Це буває в усіх тих випадках, коли наслідки відомих нам умов дуже складні для того, щоб ми могли на підставі їх конструювати очікуване явище.

Ми можемо, наприклад, знати по якому шляху буде рухатись крапля дощу, що падає при певних умовах, але з'ясувати на підставі цього знання, як буде вкрита цими краплями певна площа на протязі певного часу, завдання над силу для нас через свою складність. Ми можемо, наприклад, знати закони руху молекул газу, але на підставі лише них і стежучи за рухом кожної молекули, ми не можемо знайти тиснення газу на стінки посудини, в якій цей газ знаходиться.

В подібних випадках не має іншого виходу ніж користуватись методом статистичного дослідження.

Таким чином, при розв'язанні основних завдань науки сама природа явищ в більшості випадків приводить нас до статистичної методи дослідження.

40. Статистична метода й теорія імовірностей. Головною рисою всіх статистичних сукупностей є завжди те число, що показує скільки разів спостережена певна ознака, певне явище або предмет, себто частота цього явища, предмету, властивости, ознаки. Наукове й практичне значення статистичної методи базується на тому факті, властивому майже всім статистичним сукупностям, що висновки, які будуються на частоті ознак однієї сукупности мають більшу чи меншу силу й для інших сукупностей, однородних з першою /дослідженою/.

Цей факт і дає нам можливість знаходити закономірності та дозволяє нам статистично передбачати явища.

При загальному теоретичному студіюванні сукупностей дуже важливо мати сукупності, в яких зв'язки між частотами їх ознак обумовлювались би простими й ясними для простеження причинами, а не складними та заплутаними комплексами причин недоступними для аналізи, як це бачимо в дійсно існуючих конкретних сукупностях. Тільки в таких математично складених сукупностях ми можемо простежити вплив причин на частоту ознак і здобути надійні загальні результати. Далі, студіюючи такі різні по обсягу, але однородні щодо складу сукупності, ми можемо також виявити через що результати, одержані для однієї сукупности, маємо право переносити на другі однородні з першою.

Теорія імовірностей і дає нам можливість побудувати такі сукупності, ясні по своїй структурі. Вона студіює частоти певних властивостей на основі простих припущень, при чому ці частоти тут мають назву імовірностей. З теоретичних основ рахування імовірностей статистична метода й користується для знаходження в конкретних сукупностях тих взаємозалежностей і кореляційних /статистичних, подібних до причинових/ зв'язків, які є головним завданням статистики,- статистичної номографії.

41. Основні правила рахування імовірностей. У величезному числі випадків ми не можемо певно твердити про появлення даної події або явища А, а можемо лише сказати, що А з'явиться, відбудеться або не з'явиться не відбудеться.

Виконання ряду умов, після яких питання про появу чи не появлення А має вже певне рішення, звичайно називаємо спробою /іспитом/. Припустимо, що в якійсь спробі мислимо здійснення однієї з можливостей або випадків. Тоді кажемо, що вони є д и н о м о ж л и в і тоді, коли один із випадків обов'язково мусить здійснитися. Будемо називати також н е с у м і с н и м и такі випадки, коли здійснення любого з них виключає можливість появлення кожного іншого одночасно з ним у спробі. Нарешті, назвемо їх р і в н о м о ж л и в и м и, коли появлення одного з випадків є настільки ж законним, як і появлення кожного з решти. Припустимо, далі, що з n єдино можливих, несумісних і рівноможливих випадків m буде таких, що появлення кожного з

них дасть наслідком не появлення А. Тоді імовірність явища чи події А називається дріб m/n. Зазначаючи імовірність А літерою P/A/, тоді:

$$P\ /A/\ = \frac{m}{n}$$

Тут m буде число випадків, що сприяють появленню А, а n число всіх можливих випадків, задовольняючих згаданим вище трьом умовам.

На підставі цього поняття імовірности можемо дати таке визначення: Імовірність події або випадку є відношенням числа однаково можливих взаємно виключаючих випадків, що сприяють даному наслідку, до цілого числа випадків, однаково можливих і взаємно себе виключаючих.

В кошику 6 білих і 8 чорних куль. Кулі перемішані й виймається навмання одна куля. Імовірність витягти білу кулю буде рівною 6/I4, бо всіх можливих випадків, які єдино можливі несумісні й які можемо вважати рівноможливими є I4, а сприяючих появленню білої кулі 6. Імовірність витягнення чорної кулі буде 8/I4.

Кидаємо монету підряд 3 рази. Як велика імовірність, що при цьому, все рівно в якому порядку, випаде 2 рази Орел, а раз решка?

Всіх можливих випадків 8:

ООО /три орли/, ООР /два орли, решка/, ОРО /орел, решка, орел/, ОРР, РОР, РРО, РРР, РОО.

Сприяють випадку З: ООР, ОРО, РОО.

Коли монети кидаються навмання, наші 8 випадків треба вважати рівноможливими. Через це імовірність появлення в 3 киданнях двох орлів і одної решки буде 3/8.

Коли всі з можливих випадків сприяють А й вони є єдино можливі, то А певне, бо обов'язково появиться. Тут імовірність є $n/n = I$.

Коли всі з можливих випадків не сприяють А й вони є єдино можливі, то А не появиться, себто воно не можливе і $n/n = 0$.

Коли ми маємо число N всіх можливих шансів, а n_1, n_2, n_3, ... n_n - числа шансів, сприяючих появленню випадків, то

$$P_1 = \frac{n_1}{N};\quad P_2 = \frac{n_2}{N};\quad P_3 = \frac{n_3}{N};\quad \ldots\ldots\quad P_n = \frac{n_n}{N};$$

$$P_1 + P_2 + P_3 + \ldots\ldots P_n = \frac{n_1 + n_2 + n_3 + \ldots\ldots n_n}{N} = I$$

Сума всіх імовірностей є рівною одиниці. Таким чином кожний вираз імовірности лежить між границями 0 та I, себто > 0 й $< I$ або правильна дріб.

Чим ближче імовірність до нижчої границі / 0 /, тим менше подія певна, чим ближче до верхньої границі / I /, тим більше вона наближається до певности, до здійснення.

4I. Теорема складання імовірностей. Імовірність появлення одного з несумісних випадків /подій/, без вказування якого саме, є рівною сумі імовірностей цих випадків.

Для прикладу візьмемо три несумісних події А, В, С. Хай зо всіх n єдино можливих, несумісних і рівноможливих випадків, випадки

a_1, a_2, ... a_k сприяють А

b_1, b_2, ... b_l сприяють В

c_1, c_2, ... c_m сприяють С

Себто: $P/A/ = \frac{k}{n}$; $P/B/ = \frac{l}{n}$; $P/C/ = \frac{m}{n}$

З'єднуючи випадки, що сприяють А,з випадками, що сприяють В,і випадками, що сприяють С, одержимо випадки, що сприяють або А, або В,або С, так що імовірність появлення однієї з цих подій, без указування, якої саме, є рівною $\frac{k + l + m}{n}$, але $\frac{k + l + m}{n} = \frac{k}{n} + \frac{l}{n} + \frac{m}{n} = P/A/ + P/B/ + P/C/$

себто теорема є вірна.

в кишику

Маємо 4 білих, 3 чорних і 9 синіх куль. Виймаємо навмання кулю. Знайти імовірність, що вона буде біла або черна. Імовірність білої кулі є рівна 4/I6, чорної 3/I6 таким чином по теоремі складання ймовірностей рішення завдання дасть 7/I6.

Н а с л і д о к. Назвемо події А і В протилежними, коли подія В складається через непоявлення А й навпаки. Хай імовірності цих подій рівні відповідно p й q. Тоді на підставі складання ймовірностей імовірність появлення однієї з них є рівною $p + q$. Але, понеже одна з них обов'язково мусить відбутись, то та ж сама імовірність є рівною одиниці. Таким чином $p + q = I$ себто сума імовірностей протилежних подій є рівною одиниці.

42. Теорема множення імовірностей. З'ясуємо спочатку поняття залежних і незалежних подій, важливі для дальшого. Події А, В,....М називаємо незалежними, коли імовірність кожної з них не залежить од появлення чи непоявлення решти подій. Події А, В .....М будуть залежними одна від одної, коли імовірність кожної з них змінюється в залежності від появлення чи не появлення решти з них.

Візьмемо кошик з 4 білими й 5 чорними кулями й будемо розглядати результати виймання куль по одній, при чому подією А назвемо вихід білої кулі при першому вийманні й подією В - вихід білої кулі при другому вийманні. Розглянемо два способи виймання куль.

Перший спосіб. Вийнята куля повертається знову в кошик. Отже при другому вийманні всі умови появлення білої кулі залишаються такими ж, як при першому вийманні, через те $P/B/ = 4/9$, незалежно від того чи витягнуто в перший раз білу чи чорну кулю. В цьому випадку подія В й є незалежною.

Другий спосіб. Витягнута куля знову до кошика не повертається. В цьому випадку після першого виймання число куль в кошику змінюється, бо коли при першому вийманні витягнуто білу кулю, то в кошику залишається 3 білих і 5 чорних куль, себто $P/B/ = 3/8$; коли при першому вийманні витягнуто чорну кулю, себто подія А не відбулась, то в кошику залишиться 4 білих і 4 чорних кулі, і $P/B/ = 4/4$; в цьому випадку ми бачимо, що при другому виймання куль імовірність В має різні значення в залежності від появлення чи непоявлення А, себто В залежить від А.

Пояснивши ці поняття, перейдемо до теореми множення імовірностей, яка звучить:

Імовірність відбутись двом подіям разом є рівною добутку імовірности одної з них на імовірність другої, обчисленому з припущенням, що перша подія відбулась.

Розглянемо події А й В й зазначимо протилежні їм події через "а" й "в" так, що "а" означає непоявлення А і в - непоявлення В. Припустимо, що в спробах, що мають наслідком появлення чи непоявлення А й В, можливі чотирі комбінації: АВ, Ав, аВ й ав, при тому з N єдиноможливих несумісних і рівноможливих випадків цим комбінаціям сприяють відповідно К, l, m і n випадків. Тоді події А, незалежно від того, відбудеться вона разом з В чи безнеї, сприяє $k + l$ випадків, а події "а" при тій же умо-

ві $m + n$ випадків. Так само подіям B й "в" незалежно від появлення чи непоявлення A сприяють відповідно $k + m$ і $l + n$ випадків. !!

Вказаний зараз розподіл всіх N випадків можливо вкласти в таку таблицю:

	A	a	
B	k	m	k + m
в	l	n	l + n
	k + l	m + n	N

Як бачимо, одночасному появленню A й B сприяє k з N випадків, а імовірність появлення A й B, яку ми означимо через $P/AB/$, є рівною k/N ; зазначаючи далі, через $P/A/$, імовірність A, обчислену незалежно від того чи появилось чи не появилось B, - будемо мати:

$$P/A/ = \frac{k + l}{N}$$

Зазначаючи далі через $P_{A/B/}$ імовірність B, обчислені з припущенням, що A відбулося, будемо мати

$$P_{A/B/} = \frac{k}{k + l} \qquad \text{понеже}$$

при вказуванню на появлення A з N випадків можливими залишаються тільки $k + l$ випадків, сприяючих A і з них k випадків сприяють B.

Але $P/A/ \cdot P_{A/B/} = \frac{k + l}{N} \cdot \frac{k}{k + l} = \frac{k}{N}$ і таким чином $P/AB/ = P/A/ \cdot P_{A/B/}$ /I/

Це рівняння й формулює теорему множення імовірностей. Аналогічним способом знайдемо, що також $P/AB/ = P/B/ \cdot P_{B/A/}$ /2/

Візьмемо такі приклади: I/ З кошика з 6 білими та 8 чорними кулями виймається два рази куля, при тому витягнута куля назад до кошика не повертається. Знайти імовірність виймання двох білих куль. Імовірність витягнення першої білої кулі є рівною 6/I4, другої-5/I3, а по теоремі множення імовірність, що обидва рази витягнемо білу кулю буде

$$\frac{6}{14} \cdot \frac{5}{13} = \frac{3}{7} \times \frac{5}{13} = \frac{15}{91}$$

2/ Маємо дві категорії кошиків: перша має 3 кошика, в кожному з яких міститься 5 білих і 7 чорних куль, друга - 5 кошиків, кожний з яких містить тільки червоні кулі. Кошики обох категорії переплутані й навмання з кожного взятого кошика виймається куля. Яка імовірність що вона буде біла?

Поява білої кулі, за умовами завдання, уявляє з себе сумісну появу двох подій: A - кошика першої категорії й B вихід з нього білої кулі, себто імовірність виходу білої кулі = $P/A/ \cdot P_{A/B/}$. Але $P/A/ = 3/8$, бо кошиків 8, а з них 3 першої категорії; далі, - $P_{A/B/} = 5/12$ і таким чином знайдена імовірність $3/8 \times 5/12 = 15/96$.

Теорема тягне за собою певні наслідки.

Наслідок I-й: коли події A і B незалежні, то імовірність появлення їх разом є рівною добутку їх імовірностей. Дійсно, в цьому випадку чи появилась подія A чи не появилась, імовірність B залишається незмінною, і через те: зазначаючи її через $P/B/$, будемо мати $P_{A/B/} = P/B/$, а

відсіль $P/AB/ = P/A/ \cdot P/B/$ / 3 /. З кошика з 5 білими й 7 чорними кулями витягається два рази куля, з поверненням її назад до кошика. Яка імовірність появлення білої кулі в обох вийманнях ?

Зазначимо через A вихід білої кулі в першому вийманні, через B- в другому. Тоді $P/A/ = 5/12$, а за рівнянням 3-м і $P/B/ = 5/12$, а $P/AB/ = 5/12 \times 5/12 = 25/144$.

Наслідок 2-й. Коли маємо $P/AB/ = P/A/ \cdot P/B/$ себто, коли імовірність сумісної появи A й B рівна добутку їх імовірностей, то події A й B незалежні.

Дійсно, коли б події A й B були залежні, то ми мали б:

$$P/AB/ = P/A/ \cdot P_{A/B/}$$

Порівнявши це рівняння з попереднім, знайдемо, що $P_{A/B/} = P/B/$ / α/ або, як указано вище / див.таблицю комбінацій на стор. 145 і далі/ $\frac{k}{k+1} = \frac{k+m}{N}$. Будуючи похідну пропорцію одержимо:

$$\frac{/k+m/ - k}{N - /k+1/} = \frac{k+m}{N}, \text{ або } \frac{m}{m+n} = \frac{k+m}{N}$$

бо $N = k + 1 + m + n$. Але $\frac{m}{m+n} = P_{a/B/}$

Таким чином з рівняння / / випливає рівняння $P_{a/B/} = P/B/$ /β/.

З рівнянь /α/ і /β/ бачимо, що в тому випадку, коли $P/AB/ = P/A/ \cdot P/B/$ імовірність B при вказанню на появлення або непоявлення A залишається однаковою, що виявляє незалежність B від A. Так само можна з'ясувати, що в такому випадку й A не залежить відд B.

Наведені два наслідки нас повчають: для того, щоб події A й B були незалежні, необхідно й досить, щоб справедливим було рівняння:

$$P/AB/ - P/A/ \cdot P/B/ = 0$$

Можемо тепер узагальнити теорему множення імовірностей. Хай дано декілька подій A, B, C, ... L, M. Розмістимо їх в якомубудь порядку на вибір, наприклад, в тому, як вони написані. Тоді

$$P/ABC.... LM/ = P/A/ \cdot P_{A/B/} \cdot P_{AB/C/} \cdot .... P_{ABD...L/M/} \quad /5/$$

Інакше кажучи, імовірність появи розглянутих подій разом є рівна добутку їх імовірностей, обчислених для кожної з них із припущенням, що передуші йому в вибраному нами порядку події дійсно відбулись.

Рівняння / 5 / можна встановити, розглядаючи подію ABC... M, як сумісну появу подій A й BC... M; далі, подію BCD... M, як сумісну появу подій B й C,d... M і т.д.

Коли події A, B, C, ... M незалежні, тоді рівняння / 5 / спрощується й приймає вигляд:

$$P/ABC... M/ = P/A/ \cdot P/B/ \cdot P/C/ \cdot .... P/M/ \quad /6/$$

Себто у випадку незалежности ряду подій імовірність їх сумісної появи рівна добутку їх імовірностей.

ВПРАВИ /До лекції 15-ої/:

1. Чи завжди треба вживати зваженої середньої при обчисленню індексу?
2. Для яких індексів і через що найліпше користуватись геометричною середньою?
3. Який спосіб важення дуже корисний для індексів, що показують рівень видатків на життя?

Лекція сімнадцята.

43. Імовірності у випадку численности причин.

Численність наслідків маємо тоді, коли загальну, менш конкретну причину поставимо в зв'язок з вужчим., більш конкретним наслідком, себто, коли не всі можливі максимально конкретні діяння причини ведуть до кожного цікавого для нас наслідку. Навпаки, численність причин будемо мати тоді, коли загальний наслідок, себто обсягом своїм ширший, буде поставлений в зв'язок з вужчою, конкретнішою причиною, інакше кажучи, коли зо всіх можливих конкретних випадків наслідку не всі випливають з даної причини. Наприклад, хтось помер; причини смерти цієї особи ми не знаємо; цих причин може бути банато, і кожна з них /хороба, нещасливий випадок/ не вичерпує цілий обсяг наслідку, кожна причина сягає лише на частину його можливих конкретних перебігів. Коли б ми детально простудіювали ознаки, що сприяють даній смерті, то знайшли б у них властивості того, що є наслідком дійсної певної причини. Із усіх можливих конкретних перебігів наслідку ми бачили б та загально характеризували ті, які випливають з певної даної, а не якоїсь іншої причини. Можемо, звичайно, переконатись на підставі певних характерних ознак, що смерть була заподіяна, наприклад,- тифом ... Одначе, доки цього довести не можемо, залишається наслідок, щодо обсягу ширшим, ніж окремі причини, які з ним ставимо в зв'язок, а з цього випливає, що є можливі різні припущення причин даного явища.

Як можемо характеризувати ступінь зв'язаности наслідку з кожною можливою причиною в випадках численности причин ?

Ступінь зв'язаности будемо міряти поділом /відношенням/, який мають у числі всіх можливих конкретних перебігів наслідку ті перебіги, що випливають з даної причини. Цей поділ буде імовірність того, що наслідок, який нас цікавить, відбувся через таку саме, а не іншу причину. Отже, імовірність того, що спостережена смерть відбулась через хоробу тифу, буде той поділ із всіх можливих конкретних перебігів смерти, за данних обставин /доки їх звичайно, спостережемо/, який вчинює ті перебіги з них, що випливають з тифу, яко з своєї причини. Чим більший поділ можливих перебігів наслідку випливає з даної причини, тим імовірнішим є, що цей наслідок відбувається під безпосереднім впливом даної причини.

Загальне правило для відшукання такого роду імовірностей може бути формульовано так:

Маємо n однаково можливих, взаємовиключаючих випадків досвіду. З них m_1 ,сприяє події /явищу/ E_1 , m_2 - події E_2 та m_3 події A. Подія. E_1 , як і подія E_2 обов'язково ведуть до події F , що нас цікавить, яко результат досвіду. Досвід переведений і ми спостерігаємо, що подія F настала. Вона могла одначе відбутись як у наслідок наставання події E_1 , так і в наслідок наставання події E_2. Виникає питання, як великою є імовірність, що цей результат /подія F / став наслідком через відбуття події E_1?

Ця імовірність буде рівнятись відношенню тих однаково можливих випадків події F , які випливають з події E_1 /є одночасно випадки події E_1/ до цілого числа однаковоможливих, взаємовиключаючих випадків, через які подія F може відбутись. Подія F настане через здійснення m_1 випадків, сприяючих події E_1 або m_2 випадків, сприяючих E_2 , тобто через $m_1 + m_2$ однаковоможливих випадків. До причини для нас цікавої /події E_1 / належить з тих $m_1 + m_2$ випадків m_1 . Таким чином імовірність того, що подія F настала в наслідок відбуття E_1, буде рівною

$$\frac{m_1}{m_1 + m_2}$$

Поділивши чисельник і знаменник цього виразу на загальне число однаковомож-ливих випадків досвіду n, себто з приєднанням також і m_3 випадків, які сприяють події A, що нас не цікавить - одержимо:

$$\frac{\frac{m_1}{n}}{\frac{m_1}{n} + \frac{m_2}{n}}$$

Таким чином, у чисельнику маємо імовірність події E_1 /цікава нам причина наслідку F / до переведення досвіду, себто, виходячи від всіх рів-номожливих випадків досвіду, а в знаменнику суму відповідних імовірностей для подій E_1 та E_2, себто суму відповідних імовірностей всіх можливих причин цікавого нам наслідку F. Подібно до цього імовірність причини E_2 формулюється так:

$$\frac{\frac{m_2}{n}}{\frac{m_1}{n} + \frac{m_2}{n}}$$

Наші міркування досі стосувались до подій, які могли відбуватись лише за двома причинами. Розуміється, що ці міркування є справедливими й для довільного числа можливих причин події або явища.

Коли імовірності різних можливих причин $E_1, E_2 \ldots E_s$ даного яви-ща /перед його наставанням/ означимо через $p_1, p_2 \ldots p_s$, тоді імовірність того, що явище F настало через дану причину E_1 /цю імовірність зазначимо через P_1 /, буде рівною:

$$P_1 = \frac{p_1}{p_1 + p_2 + \ldots + p_s} \qquad /I/$$

Також для причини E_2 одержимо:

$$P_2 = \frac{p_2}{p_1 + p_2 + \ldots + p_s} \qquad /I/$$

а для причини E_s:

$$P_s = \frac{p_s}{p_1 + p_2 + \ldots + p_s} \qquad /I/$$

Розуміється, що P буде мати найбільшу вартість для тієї причини, для якої "p" має найбільшу вартість, інакше кажучи, з різних можливих при-пущень про причину наставшого наслідку є найімовірнішим те, що стосується до причини, яка має сама по собі найбільшу імовірність.

44. Теорема Бейєса

В наведених вище міркуваннях ми виходили з того припущення, що даний наслідок міг настати через діяння різних причин, але кожна з цих при-чин викликає обов'язково даний наслідок. З числа однаковоможливих наслідків F, деякі випадки випливають з причини E_1 а деякі з причини E_2, але випадки наслідку F, випливаючи з даної причини, вичерпують усі однаково-можливі випадки цієї причини. Інакше кажучи, не може бути таких випадків

цієї причини, які б не вели до наслідку F. Через те й ціла кількість рівноможливих випадків наслідку F рівнялась сумі всіх однаковоможливих випадків причини E_1 /m_1/, далі всіх однаковоможливих випадків причини E_2 /m_2/ себто $m_1 + m_2$. Це й є дійсно чистий випадок численности причин.

В дійсності, одначе, дуже часто зустрічаємося з випадками, коли численність причин зв'язується з численністю наслідків, себто, коли даний наслідок може мати різні причини та коли одночасно кожна з цих причин не веде обов'язково до цього наслідку, але може привести також і до інших наслідків.

Наведений вище приклад про смерть якоїсь особи, що могла статися через занедужання на різні хвороби, стосується якраз до такої комбінації численности причин з численністю наслідків, але не до чистого випадку численности причин. А то тому, що кожна з хвороб, через яку міг статися цікавий для нас випадок смерти, не веде обов'язково в кожному випадку до смерти; хвороба може скінчитись також одужанням хорого. Закінчення хвороби смертю є лише дещо більше-менше імовірне й в цьому випадку не має тої точної певности, як, наприклад, з кількаразовим витягненням з різних кошиків білих і чорних куль.

Колиб ми хотіли з реального життя навести приклад випадку чистої численности причин, як то можна зробити з досвідом витягнення куль або карт з цілої колоди, мусіли б скористуватись також, скажемо, випадком смерти, про яку знаємо, що була заподіяна одною з отрут, яка напевно викликає смерть /напр., від ціянових сполук/ і є лише сумнів у тому, від якої саме з цих сполук настала смерть /напр., від ціян калі, або ціян-водневої кислоти/.

Отже, випадок численности причин і наслідків має своє джерело в тому, що з одного боку, цікавий нам наслідок є визначений так широко, що його обсяг охоплює конкретні перебіги різних причин, а, з другого боку, кожна з цих причин є своїм обсягом ширша, ніж конкретні перебіги, що ведуть до даного наслідку, бо охоплює також і ті перебіги, які до даного наслідку не ведуть.

Поняття імовірности й в випадках численности причин і наслідків є необхідним знаряддям для встановлення безпосередности зв'язку наслідка з кожною з можливих причин. Цей зв'язок міряється тут імовірністю того, що даний наслідок настав діянням своєї причини, себто міряється за допомогою умовної імовірности даної причини в припущенні, що наслідок стався.

Ця умовна імовірність і тут є рівною відношенню числа можливих максімально конкретних перебігів наслідку, випливаючих з даної причини, до цілої кількости можливих максімально, конкретних перебігів наслідку.

У випадку чистої численности причин імовірність того, що наслідок настав од даної причини, є рівною імовірності даної причини, поділеній на суму імовірностей всіх можливих причин. Тут же у випадку численности причин і наслідків у встановленню імовірности того, що даний наслідок був заподіяний даною причиною, не є важливим л и ш е т е, наскільки наставання даної причини було само по собі імовірніше або меш імовірне, ніж наставання інших причин, а також і те, наскільки великою була умовна імовірність даного наслідку в припущенні наставання кожної з причин, бо при наставанню кожної причини, даний наслідок може настати й не настати.

Як каже Зігварт, внутрішній логічний зміст теореми Б а й є с а є в тому, що "найбільш імовірним вважається тут завжди те припущення, при якому результат, що фактично відбувся, настає з найбільшою імовірністю", себто, що кожна з причин, якій можемо приписати спостережену подію, тим більш імовірна, чим більш імовірним є здійснення події при наявності цієї причини.

Таким чином, коли імовірності наставання різних причин / E_1, E_2 .... E_s / означимо літерою p_1 p_2 .... p_s, а умовні імовірності наставання наслідку в припущенні, що кожна дана причина діяла, означимо через p'_1 p'_2 .... p'_s /так що p'_2 означає, напр., умовну імовірність наслідку F в припущенню, що причина E_2 діяла/, тоді цікава нам імовірність P_1, що наслідок F настав через діяння причини E_1 можна висловити так:

$$P_1 = \frac{p_1 \cdot p_1'}{p_1 \cdot p_1' + p_2 \cdot p_2' + \ldots p_s \cdot p_s'} \qquad / 2 /$$

Також імовірність про причину E_2 буде:

$$P_2 = \frac{p_2 \cdot p_2'}{p_1 \cdot p_1' + p_2 \cdot p_2' + \ldots p_s \cdot p_s'} \qquad / 2 /$$

Про причину E_s

$$P_s = \frac{p_s \cdot p_s'}{p_1 \cdot p_1' + p_2 \cdot p_2' + \ldots p_s \cdot p_s'} \qquad / 2 /$$

Інакше кажучи: імовірність того, що даний наслідок настав через діяння певної з можливих своїх причин, є рівною дробу, що в чисельнику має добуток імовірностей цієї причини на умовну імовірність наставання наслідку, в припущенні, що ця причина діяла, а в знаменнику суму відповідних добутків про всі можливі причини наслідку.

Це й є теорема Б а й є с а /англ.математик Вауєв /.

Щож каже нам теорема Байеса відносно прикладу з конкретного життя себто відносно того, яка імовірність, що, наприклад, смерть даної особи настала від тифу ? Ця імовірність по теоремі буде рівною добутку імовірности занедужання тифом на імовірність того, що хто занедужав тифом, помре - поділеному на суму подібних добутків про всі хвороби, в яких, відповідно з обставинами даного випадку смерти, можемо вбачати причину смерти. Щоб розв'язати питання про те, яке припущення тут є найбільш імовірним, треба зважити також і те, яка імовірність захорування різними хворобами та небезпека кожної з них для життя /імовірність смерти після захорування цими хворобами/. Розуміється, для цього мусимо мати дані: імовірність захорування тифом обчислена по відношенню до загальної кількости населення даної місцевости; імовірність для захорівшого тифом померти - себто скількість випадків смерти буває серед всіх, що захворіли тифом; подібні дані треба мати й для інших хвороб.

З теореми Байеса випливають наступні наслідки:

П о - п е р ш е: Коли в рівнянні / 2 / першому / P_1 = / всі p' /себто $p_1' \; p_2' \ldots p_s'$ / зведемо до одного тоді взір буде мати вигляд:

$$P_1 = \frac{p_1}{p_1 + p_2 + \ldots p_s}$$

Але $p_1' \; p_2' \ldots p_s'$ є умовні імовірності наставання наслідку нам цікавого в припущенні, що кожна причина діяла. Зробимо ж цю імовірність рівною одиниці, себто дійсній певності, прийдемо до випадку, коли кожна з причин обов'язково веде до цікавого нам наслідку. Це й буде випадок чистої численности причин.

П о - д р у г е: Коли в рівнянні теореми Байеса зробимо всі p /себто $p_1 \; p_2 \ldots p_s$ / взаємно між собою рівними, то рівняння буде:

$$P_1 = \frac{p_1'}{p_1' + p_2' + \ldots p_s'} \qquad / 3 /$$

Але $p_1 \; p_2 \ldots p_s$ є імовірності наставання всіх можливих причин даного випадку, як таких /без огляду на їх наслідки/. Отже, коли припустимо, що всі можливі причини цікавого для нас наслідку перед наставання цього

явища є однаково імовірні ., тоді взір, що показує зв'язок наслідку з кожною з його причин, дуже упрощується. Залишається в ньому лише умовні імовірності наслідку, в припущенні діяння кожної з причин. Тут є дуже важливим, щоб припущення це було зовсім обґрунтованим.

Для статистики теорема Байєса має велике значення, бо дозволяє ясно схоплювати конструкцію імовірностей, що лежать в основі різного роду відносних частот, які студіюємо статистичною методою, - імовірностей, які міряють тісноту вільного зв'язку в випадках численности причин і наслідків.

ЗАКОН ВЕЛИКИХ ЧИСЕЛ.

З терміном "закон великих чисел" зв'язано дуже багато різноманітних правил і формул, значення головніших з них буде висвітлено в короткому систематичному викладі поняття цього закона, що є науковою основою розуміння статистичної методи в дослідженні масових /колективних/ атипових явищ.

Загальне поняття закона великих чисел можемо з'ясувати собі з наступного прикладу:

Будемо витягати значну кількість разів кулі з кошика, в якому знаходяться 3 білих і 7 чорних куль, при чому умови досвіду дозволяють нам вважати витягнення поодиноких куль за рівноможливі випадки. Зробимо, наприклад, 1000 тягів /повертаючи кожний раз після тягу знов кулю до кошика/ та обчислимо для цих 1000 тягів відносну частоту появлення білої кулі, себто відношення витягнених білих куль до цілої кількости витягнених куль,- тоді така емпірична відносна частота буде наближуватись до вартости імовірности виходу білої кулі при кожному тягу, яка за встановленими вище умовами досвіду буде рівною, як нам то вже відомо, 3/10. Отже, з 1000 витягнених куль буде число білих куль не дуже далеким від 300. Звичайно, може їх бути й 280 або 320 але мусіли б свій досвід повторити дуже багато разів, щоб хочби в одному з цих досвідів число витягнених білих куль дорівнювалось, скажемо, 210, себто таке далеке число від 300 було б виключенням.

Загальним формулюванням наведеного вище факту конкретної дійсности й є, так званий, закон великих чисел. Він свідчить, що коли ті або інші явища /події/ мають певні імовірності, або теж саме, коли вони знаходяться при певному комплексі умов у вільному причиновому зв'язку з певним ступенем тісноти, то при досить великій кількості спостережень або досвідів, емпірично спостережені відносні частоти явищ /подій/ є близькими до своїх імовірностей і то тим ближчими, чим більша кількість спостережень зроблена.

Важливість цього твердження набуває для статистики тим більшого значення, що лише для невеликої частини явищ конкретної дійсности та для штучно створених умов при лотореях та досвідах основаних на випадку, ми можемо a priori /наперед/ встановити імовірності явищ /подій/. Це через те, що лише тоді можемо спиратися твердо на припущення "однаковоможливих /рівноможливих/ випадків". Але при студіюванню реальної природної дійсности найчастіше цих "рівноможливих" випадків не можемо виділити або обчислити й про імовірності явищ можемо судити тут лише a posteriori /після досвіду/ на підставі емпіричної спостереженої частоти цих явищ або подій.

Закон великих чисел й дає можливість своєї аплікації на явища реальної дійсности та оправдує таке поступування через те, що емпіричні відносні частоти явищ при задовольняючім великому числі спостережень дійсно є близькими до імовірностей явищ. Це й дає нам право прикладення правил рахування імовірностей до конкретної дійсности.

Дякуючи відомому математикові Я к о в і Б е р н у л л і, закон великих чисел одержав і своє математичне формулювання в формі теореми числення імовірностей.

В цій теоремі, загально кажучи, йде річ про явище або подію - скажемо A - яка може настати при кожному спостереженню або досвіду, при тому існує певна імовірність / p /, що подія настане. Подією A може, наприклад, бути витягнення білої кулі з кошика, появлення якої можна очікува-

ти.з імовірністю 3/10 при кожному тягу. Виконаємо n тягів кулі з кошика /цілу серію їх/; відбудеться в такому досвіді подія A припустимо m разів, себто число повторення події A в серії тягів буде m, а відносна частота її буде m/n. Повстає питання: як велика імовірність, що відносна частота m/n набуде певної вартости в серії n тягів, себто, що будемо мати в ній певну кількість повторення події A? Відповідаючи на це математична формула закона великих чисел встановлює; що найімовірнішою вартістю цієї відносної частоти m/n є вартість, яка рівняється імовірності події A /себто p/, і коли збільшуємо досить число спостережень, то навіть зовсім маленьке відхилення /похибка/ відносної частоти події від її імовірности стає надзвичайно неймовірним.

У математичному формулюванню закона великих чисел зустрічаємось з подвійного роду імовірностями, які мусімо строго розрізняти, а саме:

а/ з імовірностями яких-будь простих /однозначних/ подій /подія A/, конкретні відносні частоти які нас цікавлять /напр., імовірність витягнення білої кулі з кошика при кожному тягу/.

б/ з імовірностями того, що відносна частота події в масі спостережень набуде певної вартости - властиво, що не відхилиться від імовірности події більш ніж на певну величину.

Найімовірніша вартість відносної частоти й математичний доказ закона великих чисел. З наведеного випливає, що в доказі закону великих чисел ставимо собі завдання з'ясувати чи встановити імовірність частоти повторення /відбуття/ спостереженої події в серіях /сукупностях/ певної суми досвідів /іспитів/.

Той факт, що подія A в даній серії іспитів повториться певне число разів, є сам по собі с к л а д н о ю п о д і є ю, імовірність якої може бути встановлена тим, що вийдемо від імовірности простіших /нескладних/ подій, що творять складну подію. Це ми робимо за допомогою відомих нам правил /теорем/ додавання /складання/ й множення імовірностей.

Візьмемо приклад: в кошику є 4 білі й 6 чорних куль. Вважаючи витягнення поодиноких куль однаковоможливим випадком, твердимо, що імовірність витягнення білої кулі при кожному тягу є рівною 4/10. Понеже витягнення білої або чорної кулі є фактом единоможливих вислідів іспиту /тягу/, сума імовірностей їх є рівною одиниці, а імовірність витягнення чорної кулі є рівною 1 - 4/10 = 6/10. По кожному тягу витягнена куля знову кладеться до кошика, так що іспити є н е з а л е ж н і /вислід одного іспиту не змінює імовірности висліду другого іспиту/ а імовірність витягнення білої кулі, як також і чорної залишається сталою бігом цілого досвіду.

Питаємо: а/ які будуть імовірності поодиноких можливих частот появлення білої кулі при 2 тягах /серія 2 іспитів/?

Ці частоти можуть бути 0, 1 та 2 /разів/, відповідні відносні частоти білої кулі будуть: 0, 1/2, або 1 /2/2/. Розглянемо кожну з цих відносних частот спостережної події, користуючись правилом додавання й множення імовірностей.

Імовірність того, що біла куля при 2 тягах не появиться ані разу, є імовірністю того, що в обох тягах буде витягнена чорна куля, себто імовірність одноманітного наставання 2 незалежних подій, з яких імовірність кожної зокрема є рівною 6/10. Така імовірність за теоремою множення імовірностей визначиться як: 6/10 x 6/10 = 0,36. Імовірність того, що біла куля появиться при 2 тягах один раз є імовірність наставання одної з двох подій: появлення білої кулі лише при першому тягу та появлення білої кулі лише при другому тягу. Така імовірність визначається, за теоремою додавання /складання/ імовірностей, яко сума імовірности витягнення білої кулі лише при першому тягу та імовірности витягнення її лише при другому тягу. Але кожна з цих імовірностей є сама по собі імовірністю події складної, бо імовірність витягнення білої кулі лише при першому тягу є імовірністю одночасного нас-

тавання двох незалежних подій, себто витягнення білої кулі при першому тягу й невитягнення її при другому. Така імовірність за теоремою множення імовірностей, буде рівнятися добутку імовірности витягнення білої кулі при першому тягу /що рівна 4/10/ на імовірність витягнення чорної кулі при другому тягу /що рівна 6/10/, себто 4/10 х 6/10. Однаково імовірність витягнення білої кулі лише при другому тягу визначається яко добуток двох імовірностей але в іншому порядку, себто 6/10 х 4/10. Отже, щоб одержати імовірність того, що біла куля буде витягнена один раз при 2 тягах, все одно в якому з них, треба обидві взаємно собі рівні імовірності скласти, себто взяти два рази добуток 4/10 х 6/10. Цікава нам імовірність буде тоді рівною 0,48.

Нарешті, імовірність того, що біла куля появиться два рази при двох тягах, визначається за теоремою множення імовірностей, яко добуток імовірности витягнення білої кулі при першому тягу та імовірности витягнення її при другому тягу, себто 4/10 х 4/10 = 0,16.

Питаємо: б/ яка імовірність всіх тих частот появлення білої кулі за тих же умов при трьох тягах? Може бути, що біла кудя при всіх трьох тягах не вийде ані разу, або вийде 1, 2, 3 рази.

Імовірність того, що біла куля не появиться ані разу, себто, що всі три рази вийде чорна куля, за теоремою множення імовірностей дасть 6/10 х 6/10 х 6/10 = 0,216. Витягнення білої кулі один раз бігом 3 тягів може виявитися в 3-х взаємновиключаючих видах, себто біла куля може вийти при першому, або другому, або третьому тягу. Імовірність наставання любої з цих, взятих окремо подій буде рівною сумі імовірностей цих подій. Але імовірність витягнення білої кулі при першому тягу й її невитягнення при другому й третьому тягу є рівною за теоремою множення імовірностей добутку імовірности витягнення білої кулі на двічі взяту імовірність витягнення чорної кулі, себто 4/10 х 6/10 х 6/10. При умові сталої імовірности буде однаковій величині, як вище, рівнятись і витягнення білої кулі при другому тягу, а також і при третьому тягу. У всіх трьох випадках треба взяти яко множник один раз імовірність витягнення білої кулі - 4/10, а два рази яко множник імовірність чорної кулі. Таким чином в результаті одержимо: 3 х 4/10 х 6/10 х 6/10 = 0,432

Імовірність витягнення білої кулі двічі в трьох тягах є імовірність того, що одержиться якась з таких комбінацій: появлення білої кулі при першому й при другому тягу, або при першому й третьому, або при другому й третьому. Імовірність кожної з цих комбінацій встановиться, за теоремою множення імовірностей, яко добуток 4/10 х 4/10 х 6/10 /два рази біла й один раз чорна/. Відціль імовірність подвійного витягнення білої кулі в любій з цих трьох комбінацій буде рівною: 3 х 4/10 х 4/10 х 6/10 = 0,288.

Нарешті, імовірність витягнення тричі білої кулі в трьох тягах, за теоремою множення імовірностей, дасть: 4/10 х 4/10 х 4/10 = 0,064.

в/ Перейдемо до імовірностей різних частот появлення білої кулі в 4 тягах. Тут вона може появитись: 0, 1, 2, 3, 4 рази.

За теоремою множення імовірностей, імовірність непоявлення зовсім білої кулі буде рівною: 6/10 х 6/10 х 6/10 х 6/10 = 0,1296.

Імовірність одноразового витягнення білої кулі в чотирьох тягах, визначається яко сума 4 імовірностей, з яких кожна є добуток імовірности білої кулі взятої множником один раз на імовірність чорної кулі взятої множником 3 рази, себто буде рівною: 4 х 4/10 х 6/10 х 6/10 х 6/10 = 0,3456.

Витягнення білої кулі два рази в 4 тягах може відбутись в шости взаємно виключаючих видах, які схематично можемо написати так /означаємо витягнення білої кулі через "б", а чорної - через "ч"/.

б	б	ч	ч
б	ч	б	ч
б	ч	ч	б
ч	ч	б	б
ч	б	ч	б
ч	б	б	ч

До визначення числа цих взаємновиключаючих форм цікавої для нас події можемо дійти простішимиспособом за допомогою числення комбінацій. Йде про те, кількома різними способами можуть бути обсаджені білою кулею два місця з 4-х, що дані нам до диспозиції, себто до обчислення числа комбінацій з 4 елементів по 2. На підставі формули комбінацій з n елементів по

$$m = /\,{}^{n}_{m}\,/ = \frac{n/n-I/\,/n-2/\ldots n-m+I}{I.2.3.\ldots\ldots m}$$

одержимо:

$$/\,{}^{4}_{2}\,/ = \frac{4.3}{I.2} = 6.$$

Імовірність кожної з цих комбінацій визначається, за теоремою множення імовірностей, яко імовірність витягнення білої кулі, взятої два рази, за множник на імовірність витягнення чорної кулі, взятої також два рази за множник, себто: 6 x 4/I0 x 4/I0 x 6/I0 x 6/I0 = 0,3456.

Появлення білої кулі три рази в 4 тягах може здійснитись у 4 різних формах, себто:

$$/\,{}^{4}_{3}\,/ = \frac{4.3.2}{I.2.3} = 4.$$

Імовірність кожної з них буде рівною: 4/I0 x 4/I0 x 4/I0 x 6/I0, а імовірність, що яка-будь з них здійсниться, буде рівною 4 x 4/I0 x 4/I0 x 4/I0 x 6/I0 = 0,I536.

Наресті, імовірність, що біла куля появиться 4 рази визначається добутком: 4/I0 x 4/I0 x 4/I0 x 4/I0 = 0,0256.

г/ Перейдемо тепер до I0 тягів /іспитів/ та аналогічно до викладеного вище одержимо вартості імовірностей можливих частот появлення білої кулі, а разом з тим напишемо в виді таблички знайдені величини при 4 тягах, тоді будемо мати такі вартості:

При 4 тягах		При I0 тягах	
Частота появлення білої кулі.	Імовірність	Частота появлення білої кулі.	Імовірність
0	0,I296	0	0,0060
I	0,3456	I	0,0403
2	0,3456	2	0,I209
3	0,I536	3	0,2I50
4	0,0256	4	0,2508
	I.0000	5	0,2007
		6	0,III5
		7	0,0425
		8	0,0I06
		9	0,00I6
		I0	0,000I
			I.0000

Розуміється, що тим же способом могли б одержати імовірності різних частот появлення білої кулі для довільного числа тягів - I00, I000, I0000 і т.д.

Спосіб обчислення може бути визначений для цього в загальній формі.

Коли означимо літерою p імовірність витягнення білої кулі при кожному тягу / 4/I0 в числовому прикладі/, при чому імовірність протилежної події /появлення чорної кулі/ мусить рівнятись $I - p$ / 6/I0 в числовому прикладі/, коли далі означимо цілу кількість іспитів /тягів/ літерою n, а цікаве для нас число появлення білої кулі через m, тоді імовірність кожного даного порядку, в якому біла куля може вийти m-разів у n тягах, буде рівною p, взятому m-разів, яко множник /себто $p.m$/, а $I - p$, взятому /$n - m$/ разів, яко множник, себто

$$p^m / I - p /^{n-m}$$

Цей вираз треба взяти яко складник стільки разів, в кількох різних порядках може біла куля появитись m разів при n тягах. Число цих порядків вираховуємо за взором комбінацій $/ \frac{n}{m} /$. Тоді імовірність витягнення білої кулі m-разів у n тягах /без розділу, в якому порядку/ визначиться формулою:

$$/ \frac{n}{m} / p^m / I - p /^{n-m} \qquad / I /$$

Цей взір має загальне значення й за його поміччю можемо визначити імовірність, що подія відбудеться m разів при n незалежних іспитів для довільних вартостей m і n.

Наведений нами вище приклад обчислення імовірностей навіть декількох частот появлення білої кулі в десяти іспитах, вже дає можливість спостерігати різні правильності, важливі для уяви доказу закона великих чисел Через те обмежимось цим прикладом з невеликим числом частот та імовірностей.

Звернім увагу на імовірності/ подані в таблиці/ різних частот появлення білої кулі. Бачимо, що не всі числа появлення білої кулі мають однакову імовірність. Найбільша імовірність 0,2508 має появлення білої кулі чотири рази, при чому на обидві сторони від цієї імовірности частоти появлення кулі, убувають ці імовірності вгору й вдолину правильно, закономірно/.

Появлення білої кулі чотири рази в I0 іспитах означає її відносну частоту, що є рівною 4/I0. Але знаємо, що в нашому прикладі чотирьом десятим рівняється також і імовірність вийти білій кулі при кожному тязі - p. Такими чином, найбільшу імовірність тут має те число появлення білої кулі, при якому її відносна частота рівняється імовірності її виходу в кожному тязі.

Крім того, правильне убування від цієї відносної частоти на обидві сторони величин імовірности показує, що чим більше відхиляється відносна частота білої кулі від найімовірнішої своєї вартости, що дорівнюється імовірності свого появлення при кожному тязі - тим імовірність тієї частоти є меншою.

Це твердження може бути доведено в загальній формі. Коли означимо число іспитів /тягів/ рівним n, число появлення цікавої нам події рівним m, а імовірність наставання події при кожному іспиті рівним p, то, спираючись на загальну формулу / I / для обчислення імовірности кожного даного числа появлення події, можемо за допомогою досить довгих, але порівнявчи простих альгебричних висновків, довести,x/ що найбільша імовірність буде мати те число появлення події /така вартість m/, при якому відносна частота події m/n буде рівною імовірності її наставання при кожному іспиті p. Інакше, коли означимо найімовірнішу відносну частоту події m_0/n, тоді:

$$\frac{m_0}{n} = p \qquad /2 /$$

---

x/ Дивись "Елементарний курс математичної статистики" проф. В.Романовського, а також у конспекті "Математичної статистики" Др.С.Романовського.

А найімовірнішим абсолютним числом появлення /найімовірніша вартість m / буде:

$$m_0 = np \qquad /3/$$

себто найімовірніша вартість m є рівною цілій кількості іспитів, помноженій на імовірність події при кожному іспиті.

Може статись, що np не буде цілим числом, а між тим число появлень події природно не може бути дробом. В таких випадках, як свідчать математичні міркування, найімовірніша частота появлення події не буде рівнятись np , але буде до нього досить близьком.

Ті ж самі математичні міркування приводять також до того, що найімовірніше число появлення події / m_0 / знаходиться в межах

$$np - /1 - p/ < m_0 < np + p \qquad /4/ \quad +)$$

Коли np + p є дробом, тоді np - /I - p/ , як рівно ж і np + p -I буде також дробом, бо обидва ці числа різняться між собою лише на I.

Л е к ц і я в і с і м н а д ц я т а.

В наведеному вище прикладі 4-х тягів кулі найбільшу імовірність - 0,346 - мали дві частоти появлення білої кулі, а саме частоти I й 2 . Це випливає з того, що в цьому прикладі np / 4 x 4/I0 = I,6 / було дріб, а найімовірнішими були дві вартості m :

$$np + p = I,6 \quad 0,4 = 2$$
$$np - /I - p/ = I,6 - 0,6 = I$$

Таким чином" найімовірнішою буде та абсолютна частота появлення або відбуття події, при якій її відносна частота є рівною /або зовсім близьком/ імовірності події.

54. Біноміяльний розподіл частоти.

В попередніх прикладах появлення білої кулі при 4-х і I0-ти тягах, ми в результаті одержали таблиці, в яких в першому рядку величини частоти появлення події поставлені поруч з імовірностями цієї події. Можемо також уявити собі й таку таблицю, в якій наведені імовірності вказувть, що в серії іспитів спостережена подія - скажемо B , появиться або відбудеться - I, 2, 3 ... n разів. Така таблиця, звичайно, нагадує нам звичайну таблицю розподілу частоти з тим лише розділом, що в другому рядку, замість частот. , наведені імовірності різних вартостей змінної /якоїсь ознаки/, при чому змінною тут буде число появлення /відбуття/ події в серії іспитів.

Складемо такого роду таблицю в альгебричній формі, виходячи постійно зі взору / I / наведеного нами вище на сторінці I55-й, а саме:

$$/ \binom{n}{m} / p^m /I - p/^{n-m}$$

В ньому для упрощення I - p замінимо літером q . Себто, коли p означає імовірність відбуття явища B при окремих /поодиноких/ іспитах, тоді q буде означати імовірність зворотнього явища, себто того, що явище B не відбудеться, інакше кажучи, q є імовірність, доповнююча p до I і таким чином p + q== I. Після цієї зміни взір / I / буде мати вигляд:

---

x/ Дивись А.А.Марков: Исчисление вероятностей. 4 видання стор.37 і далі

$$/ \binom{n}{m} / p^m q^{n-m} \qquad / 5 /$$

Для окремих вартостей m, себто числа разів відбуття явища B в серії n іспитів, одержимо наступні величини:

Число разів відбуття події /m/.	Імовірність відповідного числа	
0	$/ \binom{n}{0} / p^0 q^n$	
I	$/ \binom{n}{1} / p^1 q^{n-1}$	
2	$/ \binom{n}{2} / p^2 q^{n-2}$	/ 6 /
.....	....	
n	$/ \binom{n}{n} / p^n q^0$	

Коли підставимо замість $/ \binom{n}{m} /$ його поодинокі вартості, на підставі відомого взору

$$\frac{n\,/n-1/\,/n-2/\ldots\,/n-m+1/}{1.2.3.\ldots\ldots m}$$

одержимо:

Число разів відбуття події /m/	Імовірність відповідного числа.	
0	q^n	
I	npq^{n-1}	
2	$\frac{n/n-1/}{\sqrt{1+2}} p^2 q^{n-2}$	/ 7 /
.....	.....	
n	p^n	

Припустимо далі, що переведемо s серій іспитів по n іспитів у кожній серії та для кожної серії зазначимо скільки разів у ній відбулась подія B, себто вартість величини m. Далі, клясіфікуємо /розкладаємо/ s серій за числом відбуття в кожній з них події B /за величиною m/ і складемо з них таблицю розподілу. В таблиці поруч з кожною вартістю m напишемо число серій, в яких відбулась ця вартість m. Коли поділимо кожне таке число серій на їх цілу кількість s, одержимо таблицю розподілу, подану в відносних числах, або в відносних частотах. Коли далі припустимо, що відносна частота серій з кожною даною вартістю m /з кожним даним числом відбуття події B/ буде рівною точно імовірності цієї вартости m, тоді така таблиця розподілу буде тотожньою в усіх своїх числах з вищенаведеною таблицею імовірностей / 7 /. Інакше кажучи, за цим припущенням, досить було б члени другого рядку таблиці / 7 / помножити на s /на число серій/, щоб одержати звичайну таблицю розподілу s серій іспитів за вартістю m, визначеною в абсолютних частотах серій.

З цього випливає, що таблицю імовірностей / 7 / можемо вважати за і д е а л ь н у таблицю розподілу ряду серій іспитів за числом відбуття або появлення даної події. Ідеальною вона буде в тому розумінні, що відповідає такому ідеальному випадку, коли відносні частоти серій кожним даним числом відбуття події точно б дорівнювалися апріорним імовірностям кожного даного числа появлення події.

Можемо цю таблицю назвати т е о р е т и ч н о ю, або а п р і о р н о ю таблицею розподілу на-розділ від таблиць розподілу з емпіричними

величинами, які характеризують д і й с н і с т ь. В таблиці / 7 / легко спостерігаємо, що, йдучі одна за одною вартості імовірностей, уявляють із себе йдучі члени розкладення біному $/p + q/^n$, себто:

$$/p + q/^n = q^n + npq^{n-1} + \frac{n/n - I/}{1,2} p^2 q^{n-2} + \ldots + p^n$$

Понеже $p + q = I$, розуміється, і сума всіх виразів на правій стороні рівняння, як також і в рядку 2 таблиці / 7-ої / є рівною I.

Такий ідеальний розподіл серій іспитів за числом появлення в них певної події й при умові незалежности іспитів, зветься через це б і н о м і я л ь н и м р о з п о д і л о м ч а с т о т и.

Цей розподіл має свої властивості. З одною з них ми вже ознайомлені, а саме, що найбільшу імовірність має та кількість появлення події, при якій відносна частота події рівняється /або дуже близька/ її імовірності. Імовірності по обох сторонах від цієї кількости спадають. Як, одначе, спадають - с и м е т р и ч н о чи а с и м е т р и ч н о? Себто, - має біноміяльний розподіл характер співмірний чи неспівмірний ? Це залежить від імовірности події p і від числа іспитів n в серії. Коли $p = q = I/2$, тоді розподіл є співмірний і незалежний на числі іспитів в серії. Наприклад, шукаємо імовірності появлення білої кулі, як то ми робили вище, в 3 або 4 тягах з кошику, що має однакове число білих і чорних куль, тоді одержимо співмірний розподіл:

При 3 тягах:		При 4 тягах:	
Число появлення білої кулі	Імовірність	Число появлення білої кулі	Імовірність
0	0,I25	0	0,0625
I	0,375	I	0,250
2	0,375	2	0,375
3	0,I25	3	0,250
		4	0,0625

Коли p не є рівним I/2, а число іспитів n є невелике, тоді розподіл буде неспівмірним, як видно з прикладів, коли імовірність появлення білої кулі рівнялась 4/I0 . Але при збільшенні числа тягів /наприклад з I0 тягами/, себто, по мірі того, як число n росте, ця неспівмірність, асиметрія зменшується.

Як для кожного розподілу частоти, так і для біноміяльного, можемо знайти аритметичну середню й штандартне відхилення.

Обидві ці величини для наведеної вище таблиці біноміяльного розподілу легко знайти відомим вже нам способом /див.лекції I2 й I4/. Альгебрична форма цієї таблиці ні в чим засадничо тому не перешкоджає. Нічого не заважає також і те, що маємо тут справу з відносними частотами /апріорними/, бо для переходу до абсолютних чисел треба лише всі відносні частоти помножити на одну й ту саму величину s /число серій/, а потім для одержання останнього висліду мусіли б суму відхилень та суму їх квадратів поділити на тех саме число. Через те не потрібно робити ані першого, ані другого, бо вислід од того не зміниться.

Початком відхилень /який, звичайно, ми раніше брали від довільної основи A / в цій таблиці візьмемо в пункті O , себто тут відхилення від величини A будуть рівнятись самим вартостям ознаки /змінної/, а їх середня дасть середню з вартостей ознаки. Добутки, що повстали через множення відхилень та їх квадратів на відповідні відносні частоти будуть виглядати так:

m	f/s	$\xi f/s$	$\xi^2 f/s$
0	q^n	0	0
I	$n\,p\,q^{n-1}$	$n\,p\,q^{n-1}$	$n\,p\,q^{n-1}$
2	$\frac{n\,/n-I/}{I.2}p^2\,q^{n-2}$	$-n\,/n-I/\,p^2q^{n-2}$	$2\,n/n-I/\,p^2q^{n-2}$
3	$\frac{n\,/n-I/\,/n-2/}{I.2.3.}p^3q^{n-3}$	$\frac{n\,/n-I/\,/n-2/}{I.2.}p^3q^{n-3}$	$\frac{3n\,/n-I//n-2/}{I.2.}p^3q^{n-3}$
...	.....	.....	.....

Підсумуємо рядок $\xi f/s$, витягнувши np за дужки, одержимо:

$$\frac{\sum\xi f}{s} = np\left\{q^{n-1}+/n-I/\,pq^{n-2}+\frac{/n-I//n-2/}{I.2.}p^2q^{n-3}+....\right\}$$

Много член в дужках є біном $/p+q/^{n-1}$, розвинутий в ряд після потенції, себто:

$$\frac{\sum\xi f}{s} = np\,/p+q/^{n-1}$$

Але $p+q=I$, також і $/p+q/^{n-1}=I$. Таким чином

$$\frac{\sum\xi f}{s} = np$$

Себто шукана нами пересічна для такої таблиці розподілу буде рівною np. Раніш ми бачили, що найімовірніша вартість m , себто modus поглянутої нами апріорної таблиці розподілу, також рівнявся np . Зараз бачимо, що тій самій величині рівняється й аритметична середня такої таблиці розподілу, себто в біноміяльному розподілі аритметична середня і мода однакові.

Підсумуємо далі рядок $\xi^2 f/s$, але перед тим поділимо кожний член цього рядку на 2 складник , а саме:

$$0$$

$$npq^{n-1} \quad + \quad 0$$

$$n\,/n-I/\,p^2q^{n-2} \quad + \quad n\,/n-I/\,p^2q^{n-2}$$

$$\frac{n\,/n-I//n-2/}{I.2.}p^3q^{n-3} \quad + \quad \frac{2n\,/n-I//n-2/}{I.2.}p^3q^{n-3}$$

$$.....$$

$$.....$$

Перший рядок так зміненого ряду є цілком тотожній з рядом $\xi f/s$, сума якого, як вище ми переконалися, є рівною np . В рядку інших складників можемо витягнути за дужки $n\,/n-I/\,p^2$, тоді одержимо:

$$n\,/n-I/\,p^2\left\{q^{n-2}+/n-2/\,p\,q^{n-3}+.....\right\}$$

Бачимо, що сума в дужках є в ряд розвинутий біном $/p+q/^{n-2}$ який рівняється I /бо $p+q=I$ /. Через це, ця частина суми рядка $\xi^2 f/s$ буде рівнятись $n\,/n-I/\,p^2$, а разом з першим складником $/np/$ дасть:

$$\frac{\sum\xi^2 f}{s} = np+n\,/n-I/\,p^2 = np\left\{I+/n-I/\,p\right\}$$

Згадаємо, що $\sigma^2 = \frac{\sum \xi^2 f}{s} - \left\{\frac{\sum \xi f}{s}\right\}^2$, таким чином наше апріорне /теоритичне/ σ^2 рівняється в даному випадку: $\sigma^2 = np\{1 + /n - 1/ p\} - n^2p^2 =$

$= np\{1 + np - p - np\} = np/1 - p/ = npq$, себто $\sigma = \sqrt{npq}$

Таким чином, аритметична середня з частот появлення події в усіх серіях нашої апріорної таблиці розподілу рівняється np, а штандартне відхилення є рівним $\sqrt{npq}$, де n означає число іспитів в серії.

Коли за ознаку при клясіфікації /ряду/ візьмемо не абсолютні частоти вияву події в серіях / 0,1, 2 ... /, але відносні частоти події / 1/n, 2/n /, тоді аритметична середня відносних частот події буде рівнятись:

$$\frac{np}{n} = p$$

себто, буде рівною імовірності події, а штандартне відхилення відносних частот події буде рівним:

$$\frac{\sqrt{npq}}{n} = \sqrt{\frac{pq}{n}}$$

Наведені вище зв'язки є цілком правильними, розуміється, лише для апріорного розподілу імовірностей, хоч вони ні в якому разі не є тотожніми для розподілу емпіричних серій іспитів, одначе, зв'язки ці, як побачимо далі, мають дуже важливе значення для статистики.

66. Рівняння Лапласа й теорема Бернуллі. У вищенаведених міркуваннях ми з'ясовували як відшукується імовірність кожної даної частоти події /абсолютної чи відносної/ та до яких вислідів веде визначення імовірности поодиноких частот у загальній формі. Одначе, необхідно є розв'язати питання й про те, якою є імовірність того, що відносна частота події не відхилиться від імовірности цієї події більше, ніж на певну величину ?

Лише обчисленням імовірности кожної поодинокої відносної частоти події цікава для нас проблема не вичерпується, через те, що при великому числі іспитів імовірність кожної даної відносної частоти події буде релятивно дуже невеликою, а при безконечному числі іспитів, /limes / буде безконечно малою. Це випливає з того, що сума імовірностей всіх можливих частот появлення події при даній кількості іспитів мусить бути рівною одиниці, понеже ті частоти цілковито вичерпують явище /наш масовий іспит/ та є єдиноможливими його вислідами. Але для n іспитів можливих частот цікавої для нас події /абсолютних або відносних/ буде $n + 1$, при чому, крім чисел од 1 до n треба взяти на увагу й 0. Через те чим більшим є n /число іспитів/ тим буде меншою частка одиниці, що визначає імовірність кожної даної відносної частоти. З прикладів ми бачили, що імовірність найімовірнішої відносної частоти білої кулі при 3 тягах була рівною 0,432, при 4 тягах 0,346, а при 10 тягах 0,253, себто абсолютно зменшується. При багатьох тисячах тягів ця імовірність також буде найбільшою, але своєю абсолютною величиною, розуміється, буде зовсім маленькою. Через те є необхідним відповісти також не тільки на питання, як великою є імовірність того, що відносна частота події набуде певної вартости, але й на питання, яка є імовірність того, що відносна частота події не відхилиться від імовірности події більше, ніж на певну величину, себто, що відхилення її від імовірности буде триматися в певних межах.

Отже, беручи приклад з появленням білої кулі в 10 тягах, мусимо відповісти не тільки на те, яка є імовірність кожної з можливих частот появлен-

ня білої кулі, але й про те, як великою є імовірність того, що частота появлення кулі не відхилиться від своєї найімовірнішої величини, себто 4, більше, ніж, скажемо, на I , себто що відносна частота білої кулі не відхилиться від імовірности 4/10 більш, ніж на I/10, як у більшу, так і в меншу сторону. Відповідь на це будемо шукати в наступних міркуваннях:

Імовірність того, що число появлення білої кулі не відхилиться від числа 4 більше, ніж на I, є імовірністю того, що число появлення білої кулі буде або 3, або 4, або 5. Така імовірність за теоремою складання імовірностей визначиться яко сума імовірности триразового, чотириразового та п'ятиразового появлення білої кулі, себто:

$$0{,}2150 + 0{,}2508 + 0{,}2007 = 0{,}6665$$

Але дуже велику практичну вагу має завдання з а г а л ь н о г о обчислення /формульовання/ того, що відносна частота події не відхилиться від імовірности події на певну величину. Отже, знайти загальну формулу, що дозволяє про кожну імовірність події p і про кожне число іспитів n винайти імовірність того, що відносна частота події m/n не відхилиться від її імовірности p, більше, ніж на певну величину Δ, себто, що буде вірним твердження нерівности, що

$$p + \Delta > m/n > p - \Delta$$

Логічно це завдання не є складне, воно розв'язується підсумуванням імовірностей всіх абсолютних /або відносних/ частот події, які припадають на границі $p + \Delta$ і $p - \Delta$, як то ми зробили в числовому прикладі вище. Але рішення цього завдання в такий спосіб, щоб знайти аналітичний вираз залежности між відхиленням Δ та імовірністю, що відносна частота події вийде до розмежжя $p \pm \Delta$ є математично досить складний і лише не так давно французькому математикові Лапласові пощастило знайти загальну формулу для розв'язання цього завдання[x]/.

Ми наведемо тут лише конечний вислід, до якого прийшов Лаплас. Він довів, що при серії n іспитів імовірність того, що відносна частота події m/n не відхилиться від її імовірности p більш ніж на величину Δ - буде функцією виразу

$$u = \frac{\Delta}{\sqrt{\frac{2pq}{n}}} \qquad /8/$$

Тут літери Δ, p і n мають означення, яке ми вже подали вище, а q, так, як і раніш, є рівним $I - p$. Приблизним аналітичним виразом цієї функції є певний інтеграл, який означимо через $\Phi/u/$. Для статистика-практика, який не знає вищої математики, можливо користуватись цим інтегралом через те, що існують таблиці, в яких для кожної вартости величини u наведена відповідна вартість інтегралу $\Phi/u/$, що приблизно визначає цікаву для нас імовірність. З таких табличок, знавши u, можемо без якихбудь утруднень відшукати $\Phi/u/$. Наводимо тут в цілях практичних таку таблицю /дивись на слідуючій сторінці/.

Наведена вище формула /8/ дає можливість для кожної даної кількости іспитів визначити імовірність того, що відносна частота певної події не відхилиться від імовірности цієї події більше, ніж до певної межи, яку можемо встановити. Наприклад, зробимо n тягів кулі з кошика й ставимо собі питання, яка є імовірність того, що відносна частота білої кулі в n тягах,- імовірність витягнення якої при кожному тягу є рівною p - не відхи-

---

x/Не маємо можливости наводити тут цілий математичний аналіз Лапласа, а для тих, хто цікавиться б цим аналізом, можу рекомендувати виклад його, наприклад, в конспекті "Математичної статистики" стор.7,8 і далі Др.С.М. Р о м а н о в с ь к о г о. Подєбради. 1928. Це рівняння часто називають рівнянням Моавра-Лапласа, бо ще в 1733 році пощастило вченому De Moivre розв'язати завдання для менш загальної умови.

Таблиця вартостей $\Phi / u / = \frac{2}{\sqrt{\pi}} \int_0^u e^{-z^2} dz$

u	Ф / u /	u	Ф / u /	u	Ф / u /	u	Ф / u /
0,0	0,00000	I,0	0,84270	2,0	0,99532	3,0	0,99998
0,I	0,I245	I,I	0,88020	2,I	0,99702	3,I	0,99999
0,2	0,22270	I,2	0,9I03I	2,2	0,998I4	3,2	0,99999
0,3	0,32863	I,3	0,9340I	2,3	0,99886	3,3	0,99999
0,4	0,42839	I,4	0,95229	2,4	0,9993I	3,4	0,99999
0,5	0,52050	I,5	0,966II	2,5	0,99959	3,5	0,999999
0,6	0,60386	I,6	0,97635	2,6	0,99976	3,6	0,999999
0,7	0,67780	I,7	0,98379	2,7	0,99987	3,7	0,9999999
0,8	0,742I0	I,8	0,98909	2,8	0,99992	3,8	0,9999999
0,9	0,7969I	I,9	0,99279	2,9	0,99996	3,9	0,9999999

$\Phi / \frac{3}{\sqrt{2}} / = 0,99728$ Log e = 0,4342945 $\text{Log}\sqrt{\pi} = 0,497I499$ $\text{Log}\frac{I}{\sqrt{\pi}} = I,75I425I$

литься від імовірности p більше, ніж на певну величину Δ ?

Встановимо в формулі / 8 / відомі нам вартості n, p, q та Δ. знайдемо u , а потім в наведеній вище таблиці вартостей функції Ф / u / знайдемо відповідну u імовірність. Зробимо, наприклад, I00 тягів з кошика, в якому знаходиться 4 білих і 6 чорних куль. Імовірність витягнення білої кулі / p / = 4/I0, імовірність витягнення чорної кулі / q / = 6/I0. Найімовірнішим є, що в I00 тягах біла куля появиться 40 разів. Але нас цікавить не ця імовірність появлення білої кулі, але імовірність того, що білих куль появиться, скажемо, не більш 45 й неменше 35, себто, що відносна частота появлення білої кулі буде не більшою, ніж 45/I00 й не меншою, ніж 35/I00, інакше кажучи, що відносна частота відхилиться від імовірности появлення білої кулі не більш ніж на 5/I00 / I/20 /.

Цікава нам імовірність, яка рівняється сумі імовірностей всіх частот появлення білої кулі, починаючи 35 і кінчаючи 45, відшукується приблизно за формулою Лапласа, а саме: замість n поставимо в формулу I00 ; замість p - 4/I0, замість q - 6/I0, замість Δ - I/20, тоді

$$u = \frac{\frac{I}{20}}{\sqrt{\frac{2 \cdot 4/I0 \cdot 6/I0}{I00}}} = 0,72$$

По таблиці Ф/ u / знайдемо, що величині u розміром 0,72 відповідає вартість Ф/u/ 0,678. Таким чином, імовірність того, що число білих куль в I00 тягах не буде більшим, ніж 45 і меншим, ніж 35, рівняється приблизно 0,678 .

Розглядаючи уважно обсяг формули / 8 /, можемо констатувати також наступне: з інтеграла Ф/u/ випливає, що зі збільшенням u росте й імовірність Ф/u / . Але величина Δ , що означає межі відхилення відносної частоти m/n від імовірности p , знаходиться в чисельнику величини u , через те, чим більшим є границі відхилень, які собі поставимо, тим більшим

буде u , а тим самим і $\Phi(u)$. Це цілком зрозуміла річ, бо ця імовірність $\Phi(u)$ означає суму імовірностей ряду частот, спадаючих до перемежку /границь/, яке ми намітили. Чим ширшим є перемежок, тим більше частот припадає до нього й тим більшою буде сума їх імовірностей.

Величина n /число іспитів/ входить знаменником до дробу під коренем, а цей дріб є знаменником рівняння / 8 /; через те, чим більше є n, тим меншим є дріб під коренем і тим більшим є чисельник формули / 8 /, а через те й величина u. Отже, при даній границі відхилень, чим є більшим число іспитів, тим є більшою імовірність, що відносна частота m/n не відхилиться від імовірности p більше, ніж на дану величину Δ.

Збільшуючи число іспитів, можемо збільшувати й імовірність того, що відхилення відносної частоти від імовірности не переступить певних меж. Коли досягнувши цим шляхом якоїсь цілком задовольняючої нас імовірности, будемо далі збільшувати число іспитів, можемо, розуміється, зменшувати межі відхилень, не зменшуючи тим імовірности того, що відносна частота не переступить цих меж. Таким чином, зі збільшенням числа іспитів, звужуються межі відхилень, на які можемо розраховувати з певною імовірністю.

Це є математична основа твердження, що частоти тим ближче репродукують імовірності, виразом яких вони є, чим більшим є число спостережень, на які частоти оперті. З таблиці $\Phi(u)$ бачимо, що зі зростом u, $\Phi(u)$ значно наближується до одиниці, так що при $u = 4$, $\Phi(u)$ різниться від одиниці всього на 2 стомільйонових.

Таким чином, збільшуючи n при даному Δ можемо одержати для u поступово вартости I, 2, 3, 4 і т.д., знаючи, що вже при $u = 4$ імовірність $\Phi(u)$ є зовсім близьком до певности /правди/, бо математично вона міряється одиницею. Зробивши $\Phi(u)$ досить близьком одиниці, можемо дальшим збільшенням іспитів /спостережень/ і залишаючись при тому $\Phi(u)$, зменшувати Δ /межі відхилень/ до довільно малих границь.

Відціль випливає, що, коли маємо можливість безконечно збільшувати число іспитів чи спостережень, то при досить великому числі їх настане імовірність того, що відносна частота події буде довільно мало різнитись від імовірности цієї події, себто буде довільно близьком до повної певности.

Це твердження й є відомою теоремою Яківа Бернуллі, який його математично довів в кінці XVII століття та про що було опубліковано лише в 1713 році в посмертному виданню його твору " Ars conjectandi".

56. Модуль. Вищенаведена теорема Бернуллі стосується тих випадків, коли маємо можливість збільшувати число спостережень до безмежности. Формула Ляпласа практично має значно більше значення, бо дозволяє для кожного даного конечного числа спостережень визначити імовірність, з якою можемо очікувати, що відносна частота даної події не відхилиться від імовірности цієї події більше ніж на певну величину, або навпаки, визначити ту межу відхилення відносно частоти від імовірности, на яку можемо покладатись з певною імовірністю. В попередньому прикладі ми знайшли імовірність того, що відносна частота витягнення білої кулі при 100 тягах з кошика, з числом 4 білих і 6 чорних куль, не відхилиться від імовірности витягнення білої кулі більше, ніж на 1/20.

Розв'яжемо зараз інше завдання, а саме: знайти ту межу відхилення відносної частоти від імовірности події, на яку можемо покладатись з певною, наперед встановленою імовірністю. Для того ж прикладу нехай нас буде задовольняти та імовірність розміром 0,9, себто $\Phi(u)$ буде 0,9. Таким чином, імовірність появлення білої кулі є 4/10, $\Phi(u) = 0{,}9$, $n = 100$, $p = 4/10$, $q = 6/10$, - обчислимо Δ.

З таблиці "інтеграла імовірности" знайдемо, що u приблизно є рівним 1,20. З рівняння / 8 / випливає, що

$$\Delta = u\sqrt{\frac{2pq}{n}} \qquad / 9 /$$

Підставимо наші величини й одержимо, що

$$\Delta = 1{,}20 \cdot \sqrt{\frac{2 \cdot 4/10 \cdot 6/10}{100}} = 1{,}20 \times 0{,}069 = 0{,}08$$

Себто з імовірністю 0,9 можемо очікувати, що відносна частота білої кулі не відхилиться від імовірности появлення білої кулі / 4/10 / більше, ніж на 0,08 як у бік збільшення, так і зменшення. Чи то, що ця відносна частота не буде більшою, ніж 0,4 + 0,08 = 0,48 і меншою, ніж 0,4 - 0,08 = 0,32. Отже біла куля не появиться при 100 тягах більше, ніж 48 разів і менше, ніж 32 рази.

Одначе, практично найчастіше при обчисленню меж відхилення не виходимо від довільної імовірности, але від деяких певних вартостей імовірности $\Phi/u/$, відповідаючих певним вартостям u і то цілим вартостям - I, 2, 3, і т.д.

В цих випадках, як випливає з рівняння / 9 / межі відхилення Δ є рівними $\sqrt{2pq/n}$, взятому I, 2, 3 і т.д. разів, згідно з тим, яке є u. Себто обчислення меж відхилення, на які можемо покладатись з імовірністю $\Phi/u/$, відповідаючою одній з вартостей u = I, 2, 3 разів і т.д., є дуже простим, бо в кореню $\sqrt{2pq/n}$ знаходимо немов готову міру /одиницю/ для відхилення, яку лише мусимо взяти I, 2, 3, зрідка більше, разів.

$$\Delta = u\sqrt{\frac{2pq}{n}}; \quad \text{коли} \quad u = 1, \text{ тоді} \quad \Delta = \sqrt{\frac{2pq}{n}}$$

Коли u = I, тоді по таблиці вартість $\Phi/u/ = 0{,}843$. З цією приблизно імовірністю можемо чекати, що відхилення відносної частоти не переступить межи один раз взятого виразу $\sqrt{2pq/n}$. Коли u = 2, тоді $\Phi/u/ = 0{,}995$, тоді $\Delta = 2\sqrt{2pq/n}$; коли u = 3, тоді $\Phi/u/ = 0{,}99998$. З цією імовірністю /великою/ можемо твердити, що відхилення відносно частоти від імовірности події не переступить межи три рази взятого $\sqrt{2pq/n}$.

Таким чином вираз $\sqrt{2pq/n}$ є, начебто одиницею, мірилом можливого відхилення. Вираз цей і зветься **м о д у л е м**, та означається літерою C.

Модуль є мірою межних відхилень відносних частот події від їх імовірностей.

ВПРАВИ до лекції I6-ої:

I. Що значить знайти закон явища ?

2. Поясніть - зовсім коротко - наступні категорії **м о ж л и в и х в и п а д к і в**: I/ єдиноможливі; 2/ взаємновиключальні; 3/ рівноможливі /однаковоможливі/ ?

3. В кошику знаходиться a білих, b чорних, c червоних і d синіх куль. Виймаємо наосліп одну кулю. Яка імовірність, що витягнена куля буде або білою, або чорною, або синьою ?

4. В кошику маємо a - білих куль з № I, b - теж білих куль з № 2, c - червоних куль з № I і d - теж червоних куль з № 2. Тягнемо наосліп одну кулю. Яка імовірність, що витягнемо кулю червону з № 2 ?

5. Кидаємо n однакових кісток до гри, боки яких позначені числами I, 2, 3, 4, 5. 6. Яка імовірність, що при такому киданню появиться наперед визначене число ?

Л е к ц і я д е в'я т н а д ц я т а.

Імовірність кожної події, в тому числі і такої, що уявляє з себе відхилення відносної частоти події від її імовірности не більше, ніж на певну величину, є репродукована приблизно у відносній частоті цих відхилень. Через те, коли припускаємо з імовірністю 0,843, що відносна частота явища нам цікавого не відхилиться від імовірности явища більш, ніж на модуль, мусимо знати, що в випадку переведення тисячі спостережень /іспитів, спроб/, або при створенні тисячі відносних частот явища, приблизно в 843 випадках з цієї тисячі спіткаємо відносну частоту явища, що лежить у межах модуля і лише в І57 приблизно випадках зустрінемо відносну частоту, що переступить межу модуля.

При тисячі серій по І00 тягів з кошика з 4 білими і 6 червоними кулями, відносна частота появлення білої кулі приблизно в 843 серіях буде знаходитись в межах модуля, який дорівнюється в цьому прикладі

$$\sqrt{\frac{2 \times 0{,}4 \times 0{,}6}{100}} = 0{,}069,$$ себто в 843 випадках /серіях/ з тисячі не буде більшою, ніж 0,40 + 0,069 = 0,469 і меншою, ніж 0,4 - 0,069 = 0,33І.

Як бачимо, таке мірило, як модуль є запорукою певного більшого чи меншого ступеня рідкости відхилень, що можуть переступити певну межу і, розуміється, не дає нам запоруки в тому, що взагалі не спіткаємо тих чи інших відхилень відносної частоти події від її імовірности. Само собою, певний ступінь рідкости явищ вже дає нам можливість практично з нею не рахуватись, і в цьому саме й полягає велика вага модуля яко критерія. В багатьох випадках і одноразовий модуль дає вже надійне мірило граничних відхилень відносної частоти події від їх імовірностей, себто один раз узятий модуль є межою відхилень, на яку можемо покладатись з правдоподібністю 0,843. Це цілком зрозуміло, бо потрійний модуль дає нам межу, з переступленням якої можемо спіткатись лише в двох випадках зі ста тисяч випадків. В практичному житті часто не рахуємось з ухилами відносно значно частішими. Врати одноразовий чи подвійний модуль завжди залежить уже від самого дослідника в залежності від суті речі /явища/, з якою він має до діла.

У формулі Л а п л а с а знаходимо засіб, як обчислити такі межі відхилень відносних частот від імовірности, з переступленням яких не маємо потреби практично рахуватись. Одначе може статись, що межі відхилень обчислені для невеликого числа спостережень або спроб, уявляючи з себе для нас досить надійну імовірність, фактично будуть остільки широкими, що цей розхил їх позбавить їх практичного значення, Припустимо, що задовольняємось потрійним модулем і хочемо знайти межі відхилень відносної частоти появлення білої кулі від її імовірности 0,4 при І00 тягах. Одноразовий модуль, як ми бачили вище, є рівним 0,069, а потрійний 0,069 x 3 = 0,207. Це дасть межі для відносної частоти білої кулі 0,607 / 0,4 + 0,207/ і 0,І93 / 0,4 - 0,207/, себто з імовірністю 0,99998 можемо твердити, що біла куля при І00 тягах появиться не більше, ніж 60 разів і не менше, ніж 20. Таке передбачення можливих вартостей цікавої нам частоти дає занадто широкі межі і втрачає таким чином практичне значення. В таких випадках можемо досягнути вужчих меж відхилень, збільшивши число спостережень або спроб. Отже, збільшивши число спроб для вищенаведеного прикладу до тисячі, будемо мати модуль

$$c = \sqrt{\frac{2 \times 0{,}4 \times 0{,}6}{1000}} = 0{,}022,$$ а потрійний 0,066.

Таким чином, для відносної частоти білої кулі одержимо вужчі межі 0,466 та 0,334, себто при тисячі спроб можемо передбачити з імовірністю 0,99998 появлення найменше 334 і найбільше 466 білих куль.

87. Міра точности відносної частоти. Зупинимось також на величині зворотній модуля. С відноси-

ної частоти якоїсь події A. Вона важливою є для нас через те, що служить мірою точности відносної частоти. Зазначивши її через h, будемо мати

$h = \frac{1}{c}$ /модуль/ / II /, себто, коли підставимо значення модуля $\sqrt{\frac{2pq}{n}}$

одержимо $h = \sqrt{\frac{n}{2pq}}$ / I2 /.

Далі математично можна довести, що імовірність того, що відносна частота події A в n іспитах буде відхилюватись від імовірности події A в той або другий бік менше, ніж на u-разовий модуль, є рівною Φ / u /.

Це твердження формулюється так:

$$P / -uc < \frac{m}{n} - p < cu / = \Phi / u / \qquad / I3 /$$

Спираючись на взори /II/ та /I3/, можемо вияснити й значення h, а саме:

Коли n буде рости, c буде убувати - це виявляє рівняння /I3/. Будемо далі при збільшенню n зберігати однакову вартість u, тоді рівняння /II/ покаже, що рівно-імовірні відхилення відносної частоти $\frac{m}{n}$ від імовірности p події A будуть тим менші, чим більшим буде n, понеже c зі збільшенням n убуває. Але, коли c убуває, тоді за рівнянням/II/ h росте, через те більшим вартостям h відповідають менші рівно-імовірні відхилення m/n від p. Коли дані вартости m/n, що дають нам всілякі ряди спроб,/всіляке число іспитів/, будемо вважати за наближене значення /вартости/ числа p - імовірности події A,- тоді вищенаведене твердження можемо висловити ще й так: ростучим вартостям величини h відповідають рівно-імовірні наближення відносної частоти m/n до імовірности p. Через те величину h і звуть мірилом точности відносної частоти m/n, себто міряють нею точність, з якою m/n наближується до p.

Відмітимо далі властивості міри точности / h /. Рівняння /I2/ виявляє зразу, що міра точности відносної частоти події A при постійній імовірності A /при постійній вартості p / пропорційна квадратовому кореню з числа іспитів. Візьмемо два ряди спроб, в одному n спроб, а в другому n_1, хай імовірність p події A для обох рядів є та сама, а h і h_1 - міри точности відносної частоти A для першого й другого ряду.

Тоді

$$h = \sqrt{\frac{n}{2pq}} \qquad h_1 = \sqrt{\frac{n^1}{2pq}}, \quad /I4/, \quad \text{а} \quad \frac{h}{h^1} = \frac{\sqrt{n}}{\sqrt{n^1}} \quad /I5/.$$

Отже міри точности відносної частоти A для двох рядів спроб, відносяться як квадратові корені з числа спроб рядів. Через те, коли один ряд іспитів /спроб/ буде складатись з числа спроб більшого в 4,9,25 ... разів, ніж другий, тоді міра точности відносної частоти A для першого ряду буде в 2,3 і 5... разів більшою,ніж для другого

58. Обернення теореми Бернулі й взору Лапласа.

Роглянута нами вище теорема Бернулі дозволяла по даній імовірності судити про частоти, себто, про дійсні результати іспитів /спроб/, або спостережень, при чому вихідною імовірністю могла бути як апріорна, так і емпірична імовірність. В останньому випадку лише потрібно замісць p і q підставити в формулах одержані при спостереженню або іспиті частоти m/s і n/s. Виходячи з теореми Байєса, що дає загальний принцип для обчислення імовірности причин при наявности певних, з експеримента або практики відомих, наслідків, є можливим розв'язати і зворотнє питання, а саме: по даних наслідках, себто по даних результатах спостережень /іспитів, спроб/ обчислити невідому нам наперед апріорну імовірність, а також і ступінь наближення цієї обчисленої імовірности до одержаної в спробі або іспиті частоти.

Із звичайної теореми Бернулі виходить, що чим більше є кількість виконаних досвідів n, тим меншою є імовірність зробити значну помилку, беручи замісць

чи замість с відносну частоту m/n , а замість $\sqrt{\frac{2pq}{n}}$ корінь $\sqrt{\frac{m/n.\left[1-\frac{m}{n}\right]}{n}} = \sqrt{\frac{2m[n-m]}{n^3}}$

Через те, коли імовірність p є невідома, то при великих n ми можемо бути певними, що ця невідома імовірність p не буде різнитись од відносної частоти m/n більше, ніж на

$$3\sqrt{\frac{2m[n-m]}{n^3}}$$

Для зрозуміння доцільности оберненої теореми Бернулі треба пам'ятати, що імовірність того, що даний наслідок наступить через діяння даної причини, при умові багатьох причин і наслідків, встановлюється за теоремою Байєса, а також після цієї ж формули можемо обчислити імовірність того, що спостережена нами відносна частота події виявилась у зв'язку з даною імовірністю події, себто, що невідома нам імовірність події має певну вартість.

Від цієї, з'ясованої за формулою Байєса імовірности того, що невідома нам імовірність події має певну вартість, мусимо перейти до імовірности того, що вона не різниться від спостереженої відносної частоти більше, ніж на певну величину. Для цього треба перевести додавання імовірностей поодиноких вартостей невідомої нам імовірности, що знаходяться в певних межах. Альгебричний вираз такого додавання, як показує математичний розгляд, може бути приблизно заступлений відомим нам уже інтегралом /імовірности/ $\Phi/u/$, що є функцією величини u , але u в цьому випадку буде рівним

$$u = \frac{\Delta}{\sqrt{2\frac{\frac{m}{n}\left[1-\frac{m}{n}\right]}{n}}} = \frac{\Delta}{\sqrt{\frac{2m[n-m]}{n^3}}} \qquad /16/$$

В цій формулі Δ є певна границя розділу між спостереженою частотою події та її імовірністю нам невідомою, n ціла кількість досвідів /іспитів/ m спостережена кількість повторення події, а також m/n спостережена нами відносна частота цієї події. Щоб обчислити імовірність того, що невідома нам імовірність події не різниться від її відносної частоти m/n , спостереженої підчас іспиту, більше ніж на Δ, треба підставити до взору /16/ відповідні вартості m, n та, Δ, обчислити u , а також з відомої нам таблиці функції $\Phi/u/$ для даного u знайти $\Phi/u/$, себто, проблизну вартість пошукуваної імовірности.

Навпаки, знайшовши певну задовольняючу імовірність, з якої хочемо збудувати наше припущення, можемо винайти імовірність границі розділу між спостереженою частотою події та її імовірністю, виходячи з певного $\Phi/u/$, знайдемо u , а змінивши взір /16/ на

$$\Delta = u\sqrt{2\frac{\frac{m}{n}\left[1-\frac{m}{n}\right]}{n}} \qquad /17/$$

та підставивши вартості величин u, m та n - знайдемо пошукувану границю цю Δ Звернемо увагу, що вираз /16/ різниться від формули Лапласа /формули 8/ лише тим, що замість імовірности події / p /, яка за умовами зараз поставленого завдання є нам невідома, тут нас цікавить емпірично спостережна її відносна частота / m/n /. Отже практично таке "обернення" взору Лапласа для випадку, коли за емпірично спостережених відносних частот події є потрібним зробити висновок про їх імовірності буде дуже простим: в основі вживаємо того ж взору, лише невідому нам імовірність події p заміняємо спостереженою її відносною частотою m/n , а далі поводимось як у випадку простого завдання /звичайної, а не "оберненої" формули/.

Треба мати на увазі, що до формул 16 - 17 приходимо на підставі того припущення, що імовірності p_1 , p_2, p_3 ....у формулі Байєса є однаковими.

В даному випадку такими імовірностями є імовірності того, що пошукувана імовірність нашої події набуде певної вартости, але так, як ці імовірності уявляються нам a priori перед переведенням спроби.

Таке припущення майже неможливо обґрунтувати, особливо тоді, коли маємо до діла з природними або соціяльними явищами, що студіюємо статистичним методом. Це є найтяжча проблема в розв'язанні питання про те, оскільки теорія імовірностей та її формульовки можуть прикладатись до колективних, атипових явищ реальної дійсности. Умови рівноможливости і незалежности події в теорії, або в спробах з витягненням куль чи в грі є щось інше, ніж в складних явищах реального життя.

Одначе поки йде справа про ті випадки, коли взір /I6/ з досить задовольняючою імовірністю дає досить вузькі межі для відхилення Δ, що обумовлено великими числом спостережень, склад апріорних імовірностей p_1, p_2 p_3 ..... у взорі Байєса для результату не грає ролі і при таких умовах можемо вважати "обернений" взір Лапласів, як приблизний і математично доцільний. Для використання наведеного вище взору можемо навести такий конкретний приклад. При тисячі тягів кулі з кошика вийшла біла куля 600 разів. Як великою є імовірність того, що невідома нам імовірність витягнення білої кулі не різниться від спостереженої відносної частоти 0,6 більше, ніж на 0,05? До взору I6-го підставимо n= I000, m =600, Δ 0,05, тоді одержимо:

$$u = \frac{0{,}05}{\sqrt{\frac{2 \times 0{,}6 \times 0{,}4}{1000}}} = 2{,}28$$

В таблиці вартостей Φ/u / для u =2,28 знайдемо, що Φ/u / = 0,999, себто з такою імовірністю можемо припускати, що імовірність витягнення білої кулі не різниться від одержаної відносної частоти 0,6 більше, ніж на 0,05, інакше, що ця імовірність не більша, ніж 0,65 і не менша, ніж 0,55. Аналогічно до цього можемо знайти і межі розділу між спостереженою відносною частотою та її імовірністю, виходячи з певної імовірности Φ/u / /звичайно виходять з імовірности 0,843, або 0,995, або 0,99998/, при цьому взір /I6/ служить тут модулем.

59. Нормальний розподіл частоти або крива Гавса. Розглядаючи таблицю інтеграла імовірностей, наведену на стор. I62, можемо спостерегти, що вона є теоретичним розподілом частоти, яка в колонці вартостей ознаки має одначе : не частоти події /про що ми говорили при біноміяльному розподілу частоти стор.I56/, а відхилення відносних частот події від її імовірности. Ці відхилення обчислені яко частини модуля / c / так, що величина "u" в першій колонці таблиці є рівною

$\frac{\Delta}{\sqrt{\frac{2pq}{n}}}$, себто вона є відхиленням відносних частот від імовірности,

поділена модулем. Нпр., число 0,I0 в цій колонці означає відхилення рівне одній десятій модуля. Крім того ця таблиця є кумульована тим, що про кожну вартість величини u наведено схему імовірностей, що припадають всім вартостям цієї величини від 0 аж до даної вартости. Напр., число 0,I245 проти u = =0,I0 є сума імовірностей всіх можливих відхилень від нулі до 0,I0 модуля і ця сума подає власне складну імовірність того, що відносна частота події не відхилиться від імовірности події більше, ніж на 0,I модуля. Треба також мати на увазі, що числа тої самої колонки, означають відхилення відносної частоти і в одну й в другу сторону від імовірности, себто уявляють з себе $\pm \Delta$ висловлену в модулю. Таке скомпонування таблиці є можливим через те, що розподіл інтегралу Φ/u / імовірностей є точно симетричний.

Біноміяльний розподіл, про який ми подавали вище /стор.I56-I59/, що є основою інтеррала імовірностей, у випадку коли p є нерівним I/2, буде більш-менш симетричним лише при тій умові, що p не буде дуже близьким до

0 або I і що число спроб /спостережень/ є досить великим. Відціль випливає, що взіром Лапласа не слід користатись для рідких явищ, себто таких, що мають дуже малі імовірності. В таких випадках користуються іншою формулою, про яку не вважаємо доцільним тут говорити.

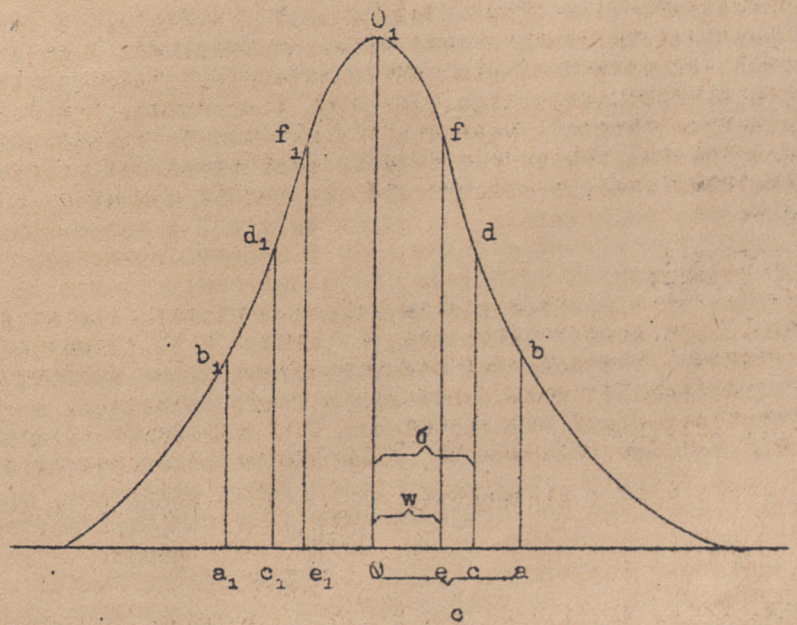
Нормальний розподіл або Г а в с о в а [x] крива, що має такий саме точно-симетричний розподіл як і величини інтегралу імовірностей, є таким чином особливою формою симетричного розподілу взагалі. Рівняння нормальної кривої звичайно визначається так:

$$y_x = y_0 \, e^{\frac{-x^2}{2\sigma^2}} \qquad /\ 18/$$

В ньому літерою X означено відхилення вартости ознаки від середньої аритметичної, яка одночасно є й модою й медіяною; y_x означає ординату кривої для кожного даного x, себто частоту відповідної вартости x, σ означає штандартне відхилення; e -основу натуральних логаритмів /2,71828 /; y_0 або ордината, відповідаюча нулевій вартості x означає максімальну ординату, бо при $x = 0$ буде y_x завжди найбільшим. Це випливає з того, що при $x = 0$

$$e^{-\frac{x^2}{2\sigma^2}} = \frac{1}{e^{\frac{x^2}{2\sigma^2}}} = 1,$$ а при який будьінший вартості x буде $\frac{1}{e^{\frac{x^2}{2\sigma^2}}} < 1$

Нормальний розподіл геометрично виглядає так:



В цій кривій, величина x, при аналогії з біноміяльним розподілом відносних частот події в серіях спроб, коли цей розподіл буде нормальним,- буде означати те, що раніше ми означали Δ, себто відхилення відносної частоти події від імовірности цієї події, y_x - імовірності відповідних вартостей x, найбільша ордината y_0 відповідає найімовірнішій вартості x, яков буде 0, бо найімовірнішою є відносна частота події, що відповідає імовірності події.

Не тяжко пересвідчитись далі, що σ в рівнянні /I8/ кривої означає штандартне відхилення відносних частот; в біноміяльному розподілі воно є рівним $\sqrt{\frac{pq}{n}}$ Коли помножимо штандартне відхилення на $\sqrt{2}$, одержимо $\sqrt{\frac{2pq}{n}}$.

---

x/ Німецький математик, що збудував криву розподілу похибок у масовому спостереженні.

що як бачимо є формула модуля, себто межа відхилення, на яку можемо розраховувати з імовірністю 0,843. Таким чином в біноміяльному нормальному розподілі штандартне відхилення помножене на $\sqrt{2}$ є рівним модулю. На підставі цього можемо твердити, що показчик ступеня для e у рівнянні кривої нормальної

$$\frac{x^2}{2\sigma^2} = \left(\frac{x}{\sigma\sqrt{2}}\right)^2$$

є не щось інше, як квадрат відхилення відносної частоти від імовірности / Δ або x /, обчисленого в модулях, яко одиницях, інакше квадрат відомої нам величини u лише зі знаком мінус. Отже і в рівнянні нормальної кривої відповідні імовірності y_x визначаються яко функції величини u. Розділ є лише в тому, що в нормальній кривій маємо до діла з імовірностями відповідними окремим відхиленням, в той час як в таблиці інтеграла Φ/u / з сумами однакових імовірностей для певних меж відхилень.

В рівнянні нормальної кривої окремі вартості y_x означають поодинокі її ординати, між тим вартості інтеграла Φ/u / означають площі оточені відповідними ординатами кривої. Дійсно, на рисунку 18 нормальної кривої відрізок на вісі абсцис Oa означає модуль c, імовірність відхилення рівного модулю уявляє з себе ордината ab та визначається відповідною вартістю y_x в рівнянні /18/. Імовірність, що плюс-мінус відхилення не переступить межі модуля, що визначена відповідною вартістю інтеграла Φ/u /, зображена частиною площі кривої, оточеною в пунктах $abO_1b_1a_1$. Понеже ця частина складає 0,843 цілої площі кривої, через це модуль служить за границю відхилення, на яку можемо покладатись з імовірністю 0,843. В границях штандартного відхилення $\pm\sigma = \frac{c}{\sqrt{2}}$ лежить біля 0,683 площі кривої / cdO_1c_1 / а через те σ може служити за межу відхилення, на яке можемо покладатись лише з імовірністю 0,683. Нарешті, в межах величини, яка зветься імовірною похибкою або помилкою та творить 0,47694 модуля або 0,67449σ і означається літерою W, знаходиться половина площі кривої / $efO_1f_1e_1$ / а через те з імовірністю 0,5 можемо вважати, що відхилення не переступить межі імовірної помилки W

Зазначимо далі коротко властивості нормальної кривої:

1. Є це крива унімодальна / з одним верхом / і точно симетрична /співмірна/, в якій аритметична середня, мода й медіяна однакові,

2. 0,683 всіх одиниць сукупности, що характеризується цією кривою, мають вартості ознаки, що різняться від середньої небільше, ніж на $\pm\sigma$

3. 0,843 площі кривої знаходиться в межах $\sigma\sqrt{2}$. В межах імовірної помилки $W = 0{,}67449\sigma$ / знаходиться половина цілої площі кривої; це означає що 0,843 числа одиниць цілої сукупности мають вартости ознаки, що не різняться від аритметичної середньої більше, ніж на $\pm c = \pm\sigma\sqrt{2}$, а половина одиниць сукупности має вартості ознаки, що не відхиляються від середньої більше, ніж на $\pm W = 0{,}67449\,\sigma$. При цьому треба добре пам'ятати, що наведене значення і модуль і імовірна помилка заховують лише для нормального розподілу, або близького до нормального.

4. Площа оточена нормальною кривою і віссю абсцис /x-ів/ є рівною об'єму цілої сукупности N. Це твердження може бути доведено лише на підставі вищої математичної аналізи, через те приймемо його без доказу.

5. Подібно тому, як квадрат штандартного відхилення $\sigma^2 = \sum x^2/N$ що є сумою квадратів відхилень від середньої, поділеною на цілу частоту сукупности, - можуть бути обчислені величини $\sum x^3/N$ /сума кубів відхилень ділена на частоту/, $\sum x^4/N$, $\sum x^5/N$ і т.д. Ці величини звуться м о м е н т а м и, себто $\sum x/N$ - перший момент, $\sum x^2/N$ - другий момент і т.д.; вони відповідно ступені x найчастіше означаються: μ_1 μ_2 μ_3 .....

Моменти такої форми, де початком, від якого відхилення обчислюються є середня аритметична звуться "моментами біля пересічної або центральними". Розуміється, що моменти можуть бути створені і біля довільного початку. Нормальний розподіл має ту важливу властивість, що при ньому: а/ всі непаристі центральні моменти / μ_1 μ_3 μ_5 і т.д./ є рівними 0, що випливає з факту співмірности нормального розподілу й б/ всі паристі моменти можна обчислити на підставі μ_2. Досить тут навести лише зв'язок між другим і четвертим

моментом, $\mu_4 = 3\mu^2$ а через те, що другий момент є квадратом штандартного відхилення, то $\mu_4 = {}^2 3\sigma^4$, себто відношення $\mu_4 / \sigma^4 = 3$. Поскільки перша властивість пояснюється співмірністю /симетрією/ нормального розподілу, постільки друга є в зв'язку з певною взаємною залежністю між середніми і скрайніми частинами кривої. Коли ця певна взаємозалежність буде порушена на користь скрайніх частин площі кривої, тоді четвертий момент у порівнянні з другим релятивно зросте і тоді μ_4/σ^4 буде більшим 3, коли навпаки, то меншим 3. Англійський біометрик P e a r s o n /Пірсон/, розглядаючи типи кривих визначив, яко критерій подібности до нормальної вираз $\mu_4/\sigma^4 = 3$ і означив його літерою β_2. Для співмірности встановив яко критерій вираз $\beta_1 = \mu_3^2/\mu_2^3$, що в нормальному розподілі рівняється нуль.

Нормальні розподіли зустрічаються головним чином при досліджуваннях антропометричних та біометричних /зріст, довжина деяких частей організмів і т.д./, а також при обчисленнях фізичних і астрономічних величин /розподіл випадкових помилок/, але істота їх дуже важлива як зразок розподілу, з яким можна порівнювати інші, більш-менш симетричні розподіли вартостей ознак.

Вимір відхилень у статистичних рядах має велике значення. Воно дає дослідувачеві певний об'єктивний маштаб за поміччю якого розв'язується два важливих питання: I/ про ступінь об'єктивної точности даної середньої, себто наскільки спостережені відхилення від середньої є наслідком впливу випадкових причин та 2/ про ступінь релятивної точности даних двох або декількох середніх при порівнянні статистичних рядів.

За поміччю кратних від імовірної помилки можна легко обчислити детальний розподіл членів серії ряду. При цьому, коли відхилення відповідають закону випадкових помилок /по нормальній кривій/, то розподіл членів серії буде однаковий при яких завгодно конкретних вартостях. Через те, щоби не робити обчислень теоретичної /імовірної помилки/ та її кратних для кожного окремого випадку можна покористуватись заранє обчисленою помічною таблицею, яка в скороченому вигляді виглядає так:

Величина відхилень від середньої по відношенні до W.	Відносна частота відхилень при I000 спостережень		РАЗОМ
	+	-	
Від 0 - 0,5	I32	I32	264,0
0,5 - I,0	II8	II8	235,9
I,0 - I,5	94,2	94,2	I88,3
I,5 - 2,0	67,2	67,2	I34,4
2,0 - 2,5	42,8	42,8	85,6
2,5 - 3,0	24,4	24,4	48,8
3,0 - 3,5	I2,4	I2,4	24,8
3,5 - 4,0	5,6	5,6	II,2
> 4,0	3,5	3,5	7,0
	500	500	I000

Наведемо далі таблицю конкретних вимірів зросту італійських рекрутів і частоту розподілу: /див. на слід. сторінці/.

Обчислення теоретичної частоти робилось так:

Емпірична імовірна помилка обрахована наступно: середній зріст з I000 вимірів одержався рівним I метру 63I міліметрам / M = I,63I м./.

Одна четвертина рекрутів /250/ мали зріст між I,68I і I,674 м. - відхилення = 0,043 м.

Одна четверта частина /250/ мала зріст між I,68I - I,588 м. - відхилення = 0,043. Такими чином половина випадків розподілилась в межах M + +0,043 і M - 0,043. Решта випадків поза цими межами. Через те за емпіричну величину імовірного відхилення /W/ було взято 43 міліметра /0,043 м./. Дальше зіставлення з теоретично обчисленими величинами не уявляє вже жадних труднощів. Коли таблиця частоти дається з вагов, то розуміється, обчислення робиться по правилах зваженої середньої, і в їх обрахунку виходить зручніше з довільно взятої основи.

Групи /кляси/	Зріст рекрутів у міліметрах. I	Відносна частота кожної групи/дійсних випадків на I000 /. 2		Частота обчислена теоретично. / а пріорі / 3		Різниця між 2 і 3 4
I	I803	6,9		3,5		+ 3,4
2	I782,5-I803	7,4		5,6		+ I,8
3	I760 -I782,5	I3,8		I2,4		+ I,4
4	I738,5-I760	25,2		24,3		+ 0,9
5	I7I7 -I738,5	4I,I		42,8		- I,7
6	I695,5-I7I7	65,4		67,2		- I,8
7	I674 -I695,5	90,6		94,2		- 3,6
8	I652,5-I674	II9,7		II8		+ I,7
9	I63I -I652,5	I30,8		I32		- I,2
I0	I63I -середня					
II	I63I -I609,5	I32,3		I32		+0,3
I2	I609,5-I588	I20,9		II8		+2,9
I3	I588 -I566,5	98,8		94,2		+4,6
I4	I566,5-I545	76,3		67,2		
I5	I545 -I523,5	33,8	I25,4	42,8	I34,3	-8,9
I6	I523,5-I502	I5,3		24,3		
I7	I502 -I480,5	8,4		I2,9		
I8	I480,5-I459	5,I	22,6	5,I	2I,5	-I,I
I9	< I459	9,I		3,5		
		I000		I000		

Отже, антрометричні виміри показують, що розділи частоти в цій області знання часто-густо симетричні й близькі до нормального розподілу й користування теоретичним критерієм цього розділу для конкретних вимірів антропометрії є дуже важливим і корисним.

ВПРАВИ до лекції I7-ої.

I. У випадках численности причин як можемо характеризувати зв'язок наслідку з кожною можливою причиною і якою загальною формулою цей зв'язок характеризується ?

2. Якою загальною формулою користуємося для характеризування зв'язку численности причин з численністю наслідків ?

ВПРАВИ до лекції 18-ої:

I. Від чого залежить співмірність /симетричність/ або асиметричність біноміяльного розподілу ?

2. Доведіть чи є однаковими чи ні в біноміяльному апріорному розподілі величини аритметичної середньої й моди ?

3. Через що для статистика є важливим значення інтеграду Лапласа?

ВПРАВИ до лекції 19-ої:

I. Зроблено 10.000 тягів куль з кошика і витягнено 3.000 білих куль. Як великою є імовірність того, що невідома нам імовірність витягнення білої кулі не відрізниться від спостереженої відносної частоти цієї кулі більше, ніж на 0,1 ?

2. В чому суттєва різниця між біноміяльним, симетричним розподілом частоти взагалі і нормальним розподілом або нормальною кривою ?

3. Що означає імовірна похибка /помилка/ в нормальному розподілі ?

4. Яка різниця величин модуля й імовірної похибки ?

Лекція двадцята.

52. Математичне очікування /надія/. Інтерпретуючи закон великих чисел, ми досі виходили з частот явищ та зв'язку цих частот з імовірністю явищ.

Релятивна частота, як знаємо, характеризує сукупність предметів, явищ, лише з погляду наявности або відсутности тої чи іншої ознаки, але не з погляду квантитативних /кількісних/ змін самої ознаки. Одначе статистикові дуже часто доводиться мати до діла зо студіями зв'язків, що виникають або існують у квантитативних ознак явищ.

Певний причиновий зв'язок у випадках квантитативно-мінливих явищ визначається тим, що за наявности даної причини є можливим лише єдиний квантитативний вираз, єдина числова вартість явища - наслідку x, себто, всі конкретні випадки явища - причини A ведуть до одного кількісного виразу явища - наслідку x.

Отже, коли певний зв'язок означає, що за наявности причини можемо з повною певністю очікувати лише єдину кількісну вартість ознаки прийнятої за наслідок, тоді логічно відціль випливає, що при так званих "вільних причинових зв'язках" кожна дана вартість явища або ознаки, прийнятої за наслідок має лише певну імовірність, мірилом якої є кількість конкретних перебігів причини, що до цієї вартости приводять.

Припустимо, що якась змінна кількісна ознака /явище/, яку беремо за наслідок якогось комплекса /сполуки/ умов може набути n вартостей - x_1, x_2 x_3 ,...x_n, при чому імовірності цих вартостей, відповідно до кожної з цих вартостей, будуть рівнятися P_1, P_2, P_3 ....P_n ; а сума цих імовірностей буде рівною одиниці. Тоді будемо мати певну систему всіх кількісних вартостей та імовірностей, що їм відповідають. Така система розподілу має назву закона розподілу квантитативної ознаки або явища. Доки можемо вважати або приймати конкретні частоти такої ознаки за приблизні вирази імовірностей, доти, і в цьому саме розумінні, можемо й закон розподілу в цілому вважати за теоретичну характеристику емпіричної таблиці розподілу якоїсь квантитативної ознаки.

Подібно до того, як для емпіричних розподілів частоти ми відшукува-

ли раніш /див. пересічні величини/ таку характеристику, яка б єдиною величиною, єдиним числом визначала б нам характер розподілу, при чому важливішою з таких величин вважали аритметичну середню - так і для закона розподілу мусимо визначити таку величину, винайти відповідну апріорну /теоретичну/ характеристику.

Нам вже відомо, що аритметична середня з вартостей кількісної ознаки /якоїсь змінної/, що складається з n одиниць, коли кожна з цих вартостей / x_1, x_2, x_3.... x_n / може певне число разів m повторюватись / m_1, m_2, m_3 .... /, може бути визначеною так:

$$x_1\frac{m_1}{n} + x_2\frac{m_2}{n} + x_3\frac{m_3}{n} + \dots x_s\frac{m_s}{n_s}$$

Тут множники при іксах $\frac{m_1}{n}$, $\frac{m_2}{n}$, $\frac{m_3}{n}$ і т.д. є ні що інше як релятивні частоти відповідних вартостей змінної x. Для переходу від емпіричного розподілу до теоретичного закону розподілу, можемо, замість релятивних частот m/n вартостей змінної, взяти імовірності їх p_1, p_2, p_3 ....p_n і т.д. Тоді, взявши за множник ці імовірності, одержимо $x_1p_1+x_2p_2+ x_np_n$ себто $\Sigma\, x_ip_i$

Таким чином, ця теоретична характеристика буде складатись з суми добутків можливих числових вартостей ознаки на імовірності цим вартостям відповідаючі. Ця теоретична характеристика вільного причинового зв'язку, що має свій емпіричний вираз в аритметичній середній, називається математичним очікуванням квантитативної ознаки , та найчастіше означається через E /від слів: esperance, expectation, Erwartung /.

Таким чином, математичне очікування $E/x/ = \sum x_ip_i$ [x/]. Воно визначає нам очікувану типову величину квантитативної ознаки, що вигливає з імовірностей всіх можливих її вартостей. Вона є середня з можливих вартостей ознаки зважена через імовірності цих вартостей, бо при обчисленню її надаємо кожній вартості тим більшу вагу, чим більшою є її імовірність.

Конкретно уявити математичні очікування досить легко, бо в житті ми часто з ними зустрічаємось. Припустимо, що якась особа в певній грі може виграти 200 карбованців і програти 100 карб., імовірність виграти є рівною 0,6, а імовірність програти 0,4. Тоді математичне очікування суми, яку ця особа може виграти, буде рівним: 0,6 x 200 + 0,4 /-100/ = 80 карб.

Припустимо далі, що ця особа повторює гру багато разів N і виграє її відбулася m разів, а програ n разів, тоді середня сума, яку вона одержить буде рівною:

$$\frac{200\,m - 100n}{N} = \frac{m}{N} \times 200 - \frac{n}{N} \times 100$$

Згідно з теоремою Якова Бернуллі ми можемо очікувати з імовірністю довільно близькою до певности, що m/N і n/N будуть різнитись довільно мало від імовірностей 0,6 і 0,4, аби тільки N було досить великим. Отже, з цією імовірністю можемо очікувати, що середня сума одержана особою після буде довільно мало різнитись од суми 0,6 x 200 - 0,4 x 100 , себто від математичного очікування вигри особи. Виходить, що наперед з імовірністю довільно близькою до певности можемо передбачати, якою буде середня вигра особи, коли число перебігів гри буде досить великим, складаючи математичне очікування вигри; з такою ж імовірністю можна передбачити й суму, яку особа одержить після певного повторення гри і для цього треба математичне очікування вигри помножити на число повторень гри.

Для математичного очікування маємо також о с н о в н і т е о р е м и:

I. Математичне очікування суми величин є рівним сумі математичних очікувань цих величин.

Візьмемо дві величини x і y , що приймають, наприклад, по три вартості кожна:

---

x/ $\Sigma\, x_ip_i$ означає суму $x_1p_1+ x_2p_2+ \dots x_np_n$

x_1 і y_1 з імовірністю p_1
x_2 і y_2 " p_2
x_3 і y_3 " p_3

Тоді $E/x/ = x_1p_1 + x_2p_2 + x_3p_3$ та $E/y/ = y_1p_1 + y_2p_2 + y_3p_3$. Сума цих величин приймає вартості $x_1 + y_1$, $x_2 + y_2$, $x_3 + y_3$, а імовірності їх будуть p_1, p_2, p_3, себто $E/x + y/ = p_1/x_1 + y_1/ + p_2/x_2 + y_2/ + p_3/x_3 + y_3/$. отже, $E/x + y/ = /p_1x_1 + p_2x_2 + p_3x_3/ + /p_1y_1 + p_2y_2 + p_3y_3/ = E/x/ + E/y/$, що й доводить теорему. Аналогічно можемо довести, що й $E/x + y + \ldots t/ = E/x/ + E/y/ + \ldots E/t/$. / I /.

II. Математичне очікування добутку незалежних величин є рівним добутку математичних очікувань цих величин.

Незалежними величинами будемо вважати такі, імовірність кожної з яких набирати любу з своїх можливих вартостей не залежить від того, які вартості набирають решта величин.

Припустимо, що дві величини x і y набирають, наприклад, по три вартости: x_1, x_2, x_3 та y_1, y_2, y_3 з імовірністю p_1 p_2 p_3 і q_1 q_2 q_3

При умові незалежности величин, добуток величини x і y може набрати вартостей:

x_1y_1 x_1y_2 x_1y_3
x_2y_1 x_2y_2 x_2y_3
x_3y_1 x_3y_2 x_3y_3

Їх імовірності за теоремою множення імовірностей відповідно будуть рівними:

p_1q_1 p_1q_2 p_1q_3
p_2q_1 p_2q_2 p_2q_3
p_3q_1 p_3q_2 p_3q_3

Отже математичне очікування xy буде рівним:

$$E/xy/ = p_1q_1x_1y_1 + p_1q_2x_1y_2 + p_1q_2x_1y_3 + p_2q_1x_2y_1 + p_2q_2x_2y_2 + p_2q_3x_2y_3 + p_3q_1x_3y_1 + p_2q_2x_3y_2 + p_3q_3x_3y_3$$

Витягнувши за дужки p_1x_1 в першому рядку, p_2x_2 - в другому і p_3x_3 - в третьому, залишимо в дужках у всіх трьох рядках ту саму величину $q_1y_1 + q_2y_2 + q_3y_3$. Таким чином $E/xy/ = /p_1x_1 + p_2x_2 + p_3x_3/./q_1y_1 + q_2y_2 + q_3y_3/$, або

$$E/xy/ = E/x/ . E/y/ \qquad / 2 /$$

Коли маємо три незалежних величини x, y, z, то за рівнянням / 2 / та беручи yz як одну величину можемо написати:

$E/xyz/ = E/x/.E/yz/$, але $E/yz/ = E/y/.E/z/$

отже, $E/xyz/ = E/x/.E/y/.E/z/$.

Так само можна довести справедливість теореми і для довільного конечного числа множників.

З вищенаведених теорем випливають такі наслідки:

I. Математичне очікування суми змінної і сталої величини / c / є рівним математичному очікуванню змінної величини плюс стала величина.

$$E/x + c/ = E/x/ + c$$

Цей наслідок випливає з першої теореми, бо математичне очікування константи є рівним самій константі, а це випливає з того, що константа може бути прийнятою за змінну величину, яка набирає лише однієї вартости з імовірністю одиниця.

2. Математичне очікування добутку змінної величини й константи є рівним математичному очікування змінної величини, помноженому на константу.

$E/x.c/ = c.E/x/$, що є наслідком другої теореми.

III. Коли величина x може набувати лише додатніх вартостей і A буде сталим додатнім числом, то імовірність того, що x набуде вартости, яка перевищить A, є більшою, ніж $\frac{E/x/}{A}$, де E/x/ буде математичним очікуванням величини x.

Хай будуть x_1, x_2, ... x_i, x_{i+1} .... x_n всі єдиноможливі вартості x та p_1, p_2 .... p_i, p_{i+1} .... p_n відповідні імовірності цих вартостей.

Розглянемо три можливі тут випадки:

а/ Жадна з вартостей x_1, x_2 .... x_n не перевищує сталої величини A, тоді імовірність, що одержана вартість x не перевищить A, є рівною одиниці: $P/x \leq A/ = I$. Математичне очікування $E/x/ = p_1x_1 + p_2x_2 + \ldots + p_nx_n$ є певне додатнє число /всі вартості x по умові додатні/. Через те, яке б не було A. - $P/x \leq A/ = I > I - E/x/ : A$

б/ Кожна з вартостей змінної x перевищує A, себто $x_1 > A$, $x_2 > A$ і т.д. Тоді $E/x/ = p_1x_1 + p_2x_2 \ldots + p_nx_n > p_1A + p_2A \ldots + p_nA$; $E/x/ > A/p_1 + p_2 \ldots + p_n/$. Якщо ряд вартостей x вичерпує собою всі можливі вартості x, то $p_1 + p_2 + \ldots + p_n = I$ і $E/x/ > A$; через те

$$\frac{E/x/}{A} > I, \text{ а } I - \frac{E/x/}{A} \text{ буде від'ємною величиною.}$$

Крім того імовірність $P/x \leq A/$ в цьому окремому випадку є рівною нулю.

в/ Стала величина A знаходиться поміж двома вартостями ряду чисел x_1, x_2, ... x_k, x_{k+1} ..... x_n. Отже нехай $x_1 \leq A$, $x_2 \leq A$, ... $x_k \leq A$, але $x_{k+1} > A$, $x_{k+2} > A$ ..... $x_n > A$. Математичне очікування можемо написати: $E/x/ = /p_1x_1 \; p_2x_2 \ldots \; p_nx_n/ + /p_{k+1}x_{k+1} \ldots p_nx_n/$. Можемо припустити, що приймаємі одно з чисел x_1, x_2 .... x_n є більшим, ніж A, бо в протилежному разі мали б окремий випадок перший / а / і що приймаємі одно з цих чисел не перевищує A, бо в противному разі мали б випадок другий / б/ Відціль випливає, що ми можемо припустити, що дві окремі суми, на які ми розподілили E/x/ дійсно існують. Кожна з цих окремих сум є якесь додатнє число, а E/x/, що є рівним сумі двох додатніх чисел, буде більшим кожного з них. Зокрема будемо мати що:

$E/x/ > p_{k+1}x_{k+1} + p_{k+2}x_{k+2} + \ldots + p_nx_n$, але по умові: $x_{k+1} > A$, $x_{k+2} > A$ .... $x_n > A$, тому $E/x/ > p_{k+1}A + p_{k+2}A + \ldots + p_nA$, інакше: $E/x/ > /p_{k+1} + p_{k+2} + \ldots + p_n/.A$. Далі, сума $p_1 + p_2 + \ldots + p_k + p_{k+1} + p_n = I$, через те $p_{k+1} + p_{k+2} + \ldots + p_n = I - /p_1 + p_2 + \ldots + p_k/$ і $E/x/ > A/I - /p_1 + p_2 + \ldots + p_n/ /$. Далі

$$\frac{E/x/}{A} > I - /p_1 + p_2 + \ldots + p_k/, \text{ або } p_1 + p_2 + \ldots + p_k > I - \frac{E/x/}{A}$$

Сума $p_1 + p_2 + \ldots + p_k$ є ні що інше як імовірність того, що x прийме якусь одну з вартостей x_1, x_2 .... x_n, які не перевищують A,

$$p_1 + p_2 + \ldots p_k = p/x \leq A/.$$

Таким чином і в третьому окремому випадку наведена вище теорема справедлива.

53. Математичне очікування і теорема Чебішова.

З вищенаведеної теореми / III / математичного очікування випливає дуже важлива теорема російського математика Ч е б і ш о в а, що дозволяє твердження теорії імовірностей зв'язати з середньою величиною статистичного ряду, яку можна очікувати в певних наперед поставлених межах. Істоту теореми Чебішова коротко можемо з'ясувати так. Хай кількість n незалежних величин x_1, x_2, x_3 .... x_n може збільшуватись до безмежности. Далі уявім собі, що кожна з цих величин набирає в наслідок певного спостереження або досвіду Д якусь одну з своїх єдиноможливих вартостей. Припустимо, що ці вартості будуть:

$$\begin{array}{lll} \text{для} & x_1 \ldots\ x_1', & x_1'', \quad x_1''' \ldots\ . \\ " & x_2 \ldots\ x_2', & x_2'', \quad x_2''' \ldots\ . \\ " & x_i \ldots\ x_i', & x_i'', \quad x_i''' \ldots\ . \\ " & x_n \ldots\ x_n, & x_n, \quad x_n \ldots\ . \end{array} \qquad / \text{I} /$$

Означимо відповідні імовірності цих вартостей для

$$\begin{array}{lll} \text{для} & x_1 & \text{через} \quad p_1', \quad p_1'', \quad p_1''' \\ " & x_2 & " \quad p_2', \quad p_2'', \quad p_2''' \\ " & x_i & " \quad p_i', \quad p_i'', \quad p_i''' \\ " & x_n & " \quad p_n', \quad p_n'', \quad p_n''' \end{array}$$

Себто перед виконанням досвіду I є імовірність p_1', що x_1' набере вартости x_1', імовірність p_1'', що x_1 набере вартости x_1'', імовірність p_2', що x_2 набере вартости x_2' і т.д.

Чебішов і поставив далі питання про те, яка ж буде по виконанню досвіду / I / середня аритметична всіх вартостей, що одержать n величин ?

Середня аритметична по виконанню досвіду може бути одержаною з відношення

$$\frac{x_1 + x_2 + x_3 + \ldots\ . + x_n}{n}$$

коли замість x_1, x_2, ... x_n підставимо вартості, які ці величини набрали в нашому припущенні / I /. Перед виконанням досвіду ця середня є невідомою і ми можемо поставити питання відносно імовірности, що ця середня визначиться в певних наперед поставлених межах.

Означимо через e_1, e_2, e_3, ... e_n математичні очікування величин x_1, x_2, x_3, ... x_n і т.д., тоді:

$$\begin{aligned} e_1 &= E/x_1/ = p_1'x_1' + p_1''x_1'' \ \ldots\ . \\ e_2 &= E/x_2/ = p_2'x_2' + p_2''x_2'' \ \ldots\ . \\ e_n &= E/x_n/ = p_n'x_n' + p_n''x_n'' \ \ldots\ . \end{aligned}$$

Поруч з величинами x_1 x_2 x_n розглянемо також їх квадрати, себто x_1^2 x_2^2 .... x_n^2. Отже досвід I тягне за собою визначення вартостей і величин x_1 x_2 .... x_n і квадратів цих вартостей. Відносно можливих вартостей квадратів величин ми можемо також означити їх математичне очікування, себто:

$$\begin{aligned} E_1 &= E/x_1^2/ = p_1'/x_1'/^2 + p_1''/x_1''/^2 + \ldots\ . \\ E_2 &= E/x_2^2/ = p_2'/x_2'/^2 + p_2''/x_2''/^2 + \ldots\ . \\ E_n &= E/x_n^2/ = p_n'/x_n'/^2 + p_n''/x_n''/^2 + \ldots\ . \end{aligned}$$

Тоді Чебішов довів, що коли кожне з очікувань E_1, E_2 .... E_n, який би не був індекс n, не перевищує якої сталої межі M, тоді, чим більше буде n, тим з більшою імовірністю можемо чекати, що середня аритметична отриманих вартостей змінної x_1, x_2 .... x_n буде мало різнитись від середньої аритметичної математичних очікувань $e_1 + e_2 + \ldots\ . + e_n$. Таким чином, значення виразу $\Sigma x_i/n$, що одержимо після досвіду I, буде мало різнитись від середньої $\Sigma e_i/n$.

В більш точному формулюванні теорема Чебішова висловлюється так: Коли математичне очікування квадратів величин якоїсь змінної залишаються обмеженими в міру того, як кількість їх росте до безмежности, тоді як би малі не були додатні числа ε і η, імовірність того, що дріб $\Sigma x_i/n$ не вийде поза перемежок

$$\frac{\Sigma e_i}{n} - \varepsilon \ , \quad \frac{\Sigma e_i}{n} + \varepsilon$$

перевищує $1 - \eta$, як що n є більшим певного вистачавче великого індексу N.

Отже треба довести, що коли $n > N$ то

$$P\left/ -\varepsilon < \frac{\sum x_i}{n} - \frac{\sum e_i}{n} < +\varepsilon \right/ > 1 - \eta \qquad /2/$$

Нерівності: $-\varepsilon < \frac{\sum x_i}{n} - \frac{\sum e_i}{n} < \varepsilon$ та $\left/ \frac{\sum x_i}{n} - \frac{\sum e_i}{n} \right/^2 < e^2$ однаково мають тут місце, одночасно й імовірності їх є однакові. Через те досить лише довести, що при вистачаючe великому n імовірність:

$$P\left[\left/ \frac{\sum x_i}{n} - \frac{\sum e_i}{n} \right/^2 < \varepsilon^2\right] > 1 - \eta \qquad /3/$$

Вираз $\left/ \frac{\sum x_i}{n} - \frac{\sum e_i}{n} \right/^2$ може набирати лише додатні вартості, через те, прикладаючи попередню теорему / III /, де / $A = \varepsilon^2$ /, напишемо:

$$P\left[\left/ \frac{\sum x_i}{n} - \frac{\sum e_i}{n} \right/^2 < \varepsilon^2\right] > 1 - \frac{E\left[\left/ \frac{\sum x_i}{n} - \frac{\sum e_i}{n} \right/^2\right]}{\varepsilon^2} \qquad /4/$$

де $E\left[\left/ \frac{\sum x_i}{n} - \frac{\sum e_i}{n} \right/^2\right]$ /5/ буде математичне очікування величини $\left/ \frac{\sum x_i}{n} - \frac{\sum e_i}{n} \right/^2$ перед виконанням досвіду Д та коли невідомо ще, які вартості наберуть величини $x_1 \quad x_2 \ldots x_n$ /змінної x /. Якщо нам можливо довести, що, починаючи з певного $n > N$, дріб $E\left/ \frac{\sum x_i}{n} - \frac{\sum e_i}{n} \right/^2$ зробиться меншим, ніж η, тоді з нерівности / 4 / випливає нерівність / 3 / і теорема буде доведена.

Розглянемо математичне очікування / 5 /, яке напишемо так:

$$E\left[\left/ \frac{\sum x_i}{n} - \frac{\sum e_i}{n} \right/^2\right] = E\left[\frac{I}{n^2} \left/ \sum x_i - \sum e_i \right/^2\right] \qquad /6/$$

З суті визначення математичного очікування взагалі випливає, що в разі якась величина x може набирати якісь вартості з певними імовірностями і C буде якесь постійне число, тоді $E\,/Cx/ = CE\,/x/$.

При кожному n, при якому має виконуватись досвід Д, сочинник I/n^2 вважається постійним, через це рівняння / 6 / можемо послідовно переписати:

$$= \frac{I}{n^2} E\left[\left/ \sum x_i - \sum e_i \right/^2\right] = \frac{I}{n^2} E\left[\left/ x_1 + x_2 + \ldots + x_n \right/ - \left/ e_1 + e_2 + e_n \right/\right]^2 = \frac{I}{n^2} E\left[\left/ x_1 - e_1 \right/ + \left/ x_2 - e_2 \right/ + \ldots + \left/ x_i - e_i \right/ + \ldots + \left/ x_n - e_n \right/\right]^2$$

$$= \frac{I}{n^2} E\left[\left/ x_1 - e_1 \right/^2 + \left/ x_2 - e_2 \right/^2 + \ldots + \left/ x_n - e_n \right/^2 + 2\left/ x_1 - e_1 \right/ . \left/ x_2 - e_2 \right/ + 2\left/ x_1 - e_1 \right/ . \left/ x_3 - e_3 \right/ + \ldots + 2\left/ x_{n-1} - e_{n-1} \right/ . \left/ x_n - e_n \right/\right]$$

За теоремою, що математичне очікування суми величин є рівним сумі їх математичних очікувань напишемо:

$$E\left[\left/ \frac{\sum x_i}{n} - \frac{\sum e_i}{n} \right/^2\right] = \frac{I}{n^2}\left[E\left/ x_1 - e_1 \right/^2 + E\left/ x_2 - e_2 \right/^2 \ldots + 2E\left/ x_1 - e_1 \right/ \left/ x_2 - e_2 \right/ + \ldots + 2E\left/ x_{n-1} - e_{n-1} \right/ . \left/ x_n - e_n \right/\right]$$

Математичне очікування всіх добутків таких як $/x_1 - e_1/$, $/x_2 - e_2/$ є нулі, бо величини $x_1, x_2, \ldots x_n$, а з ними і різниці $x_1 - e_1$, $x_2 - e_2$ ..... $x_n - e_n$ є незалежні і прикладаючи теорему, що математичне очікування добутку незалежних величин є рівним добуткові їх математичних очікувань, напишемо:

$$E/x_1 - e_1/ \cdot /x_2 - e_2/ = E/x_1 - e_1/ \cdot E/x_2 - e_2/ = E\left[x_1 + /-e_1/\right] \cdot E\left[/x_2 + /-e_2/\right] = \left[E/x_1/ + E/-e_1/\right] \cdot \left[E/x_2/ + E/-e_2/\right]$$

За насділком 2 з теореми II математичне очікування константи є рівним цій константі /сталій /, через те останній добуток буде:

$$/e_1 - e_1/ \cdot /e_2 - e_2/ = \text{o.o.} = 0, \text{ а тому}$$

$$E\left[\left/\frac{\sum x_i}{n} - \frac{\sum e_i}{n}\right/^2\right] = \frac{I}{n^2} \sum_{c=1}^{n} E/x_i - e_i/^2 \text{ , але}$$

$$/x_i - e_i/^2 = x_i^2 - 2e_i x_i + e_i^2 = x_i^2 + /-2e_i/x_i + e_i^2 \text{, а через те:}$$

$$E/x_i - e_i/^2 = E/x_i^2/ - 2e_i E/x_i/ + E/e_i^2/ - 2e_i + e_i + e_i^2 = E_i - e_i^2 \text{ ,}$$

бо математичне очікування x_i^2 ми умовилися означати через E_i.

Остаточно: $$E\left[\left/\frac{\sum x_i}{n} - \frac{\sum e_i}{n}\right/^2\right] = \frac{I}{n^2} \sum_{i=1}^{n} /E_i - e_i^2/$$

Повертаючись до нерівности / 4 / бачимо, що імовірність

$$P\left[\left/\frac{\sum x_i}{n} - \frac{\sum e_i}{n}\right/^2 < \varepsilon^2\right] > I - \frac{\sum_{i=1}^{n} /E_i - e_i^2/}{n^2 \varepsilon^2} \qquad /7/$$

Кожна з ріжниць $E_i - e_i^2 < E_i$, а тому

$$\sum / E_i - e_i^2 / < E_1 + E_2 + \ldots + E_n$$

По умові теореми кожне математичне очікування $E_i < M$, тому

$$\sum / E_i - e_i^2 / < nM$$

З цього випливає, що

$$P\left[\left/\frac{\sum x_i}{n} - \frac{\sum e_i}{n}\right/^2 < \varepsilon^2\right] > I - \frac{M}{n\varepsilon^2}$$

При вибраному ε вистачить взяти N настільки великим, щоб $M/N\varepsilon^2 < \eta$, тоді при кожному $n > N$ буде $M/N\varepsilon^2 < \eta$ і напевно

$$P\left[\left/\frac{\sum x_i}{n} - \frac{\sum e_i}{n}\right/^2 < \varepsilon^2\right] > I - \eta$$

Таким чином, теорема Чебішова цілком правдива. По кожних ε і η, які б які б малі вони не були, ми можемо знайти індекс N з нерівности

$$N > \frac{M}{\varepsilon^2 \eta}$$

Велике значення нерівностей Чебішова і його теореми є, таким чином, в тому, що маємо можливість більшого узагальнення закона великих чисел, себто під його формулювання підвести також і середні величини, виходячи з поняття математичного очікування. Кожна емпірична величина - яко зріст або вік людини, посівна площа селянського господарства - двора, як окремого підприємства, число робітників у промисловому підприємстві і т.п. можемо розглядати як ви-

раз якогось математичного очікування, в рівній також мірі й середній аритметичній, з яких хочемо емпіричних величин, відповідає своє математичне очікування, яке уявляє з себе аритметичну також середню з математичних очікувань, що складають середню окремих величин /змінної/.

Ця середня з математичних очікувань в загальному формулюванні закона великих чисел і грає ту саму ролк, яка належиться імовірності p в теоремі Бернуллі.

Спираючись на математичну істоту закона великих чисел, ми можемо собі добре з'ясувати ряд важливих питань, що тісно зв'язані з істотою масового статистичного спостереження, для якого закон великих чисел є, таким чином, принциповою основою.

При доказі теореми Чебішова було знайдено, що $E/x_i - e_i/^2 = E_i - e_i^2$ Пригадаємо, що $e = E/x_i/$, тоді можемо наведене вище рівняння написати так:

$$E\left[x_i - E/x_i/\right]^2 = E/x_i^2/ - \left[E/x_i/\right]^2$$

Взагалі, коли x є величина, що може набирати якісь вартості з певними імовірностями, тоді

$$E\left[x - E/x/\right]^2 = E/x^2/ - \left[E/x/\right]^2 \qquad /8/$$

Витягнемо корінь квадратовий з лівої частини цього рівняння одержимо:

$\sqrt{E\left[x - E/x/\right]^2}$ цей вираз є квадратичне ухилення величини x і часто означається через $\Delta/x/$, отже $\Delta/x/ = \sqrt{E/x^2/ - \left[E/x/\right]^2}$

54. Теорема Пуассона й Бернуллі. Будемо виконувати n досвідів відносно певної події A, імовірність відбуття якої є для різних досвідів різною та не залежить від результатів попередніх досвідів. Досвіди означимо $Д_1$, $Д_2$, ... $Д_n$. Імовірність відбуття події A через p_1, p_2 .... p_n. Далі визначимо величин x_1, x_2, x_3 .... x_n. При тій умові, що, коли в i - ім досвіді подія A відбудеться, то величина x набере вартости I, а якщо в цьому досвіді подія A не відбудеться, то x набере вартости 0. Відносно величин змінної x_1, x_2, ... x_n можемо ствердити:

I/ вони незалежні, бо по умові на відбуття чи невідбуття події /A/ не впливають результати попередніх досвідів.

$$2/\ E/x_i/ = p_i I + /I - p_i/\ 0 = p_i$$

$$3/\ E/x_i^2/ = p_i I^2 + /I - p_i/\ 0 = p_i \leq I$$

Математичні очікування квадратів величин x є обмежені /$\leq I$/. Величини змінної x задовольняють всім умовам теореми Чебішова, а тому, які б малі не були ε і η, знайдеться таке N що, коли $n > N$, то

$$P\left[-\varepsilon < \frac{\sum x_i}{n} - \frac{\sum e_i}{n} < \varepsilon\right] > I - \eta \qquad /9/$$

де e_i буде математичним очікуванням величини x_i; рівним, як вказано вище p_i. Після переведення n досвідів деякі величини змінної x наберуть вартости одиниці, а інші вартості нулі. Через це після закінчення досвідів сума $x_1 + x_2 + \ldots + x_n$ буде рівною кількості разів відбуття події A протягом цих досвідів. З цього випливає, що перед виконанням досвідів різні можливі вартості суми $x_1 + x_2 + \ldots + x_n$ є ні що інше як різні можливі абсолютні частоти m відбуття події A протягом n досвідів. Через це рівняння /9/ можна написати:

$$P\left[-\varepsilon < \frac{m}{n} - \frac{\sum p_i}{n} < \varepsilon\right] > I - \eta \qquad /I0/$$

Отже коли кількість досвідів росте, то все більше росте імовірність, що відносна частота відбуття події A буде все менше відрізнятися від середньо аритметичної імовірностей p_i. Це твердження й зветься теоремою П у а с с о н а.

Якщо імовірності p_1, p_2 .... p_n є однакові, себто $p_1 = p_2 = ...$ ... p_n, тоді й $\Sigma p_i / n = p$ і ми матимемо теорему Бернуллі

$$P\{-\varepsilon < \frac{m}{n} - p < \varepsilon\} > 1 - \eta$$

Вистачить для цього, щоби було $n > N$, де $N > \frac{1}{\varepsilon^2 \eta}$

Таким чином, теорема Пуассона твердить: Коли для ряду незалежних відносно події A іспитів або спроб, імовірність появлення її буде рівною в порядку спроб p_1, p_2, p_3 ...., то при досить великому числі іспитів з імовірністю довільно близькою до певности можна чекати, що відносна частота A в переведених іспитах буде довільно мало різнитись од середньої імовірности цієї події A.

55. Закон великих чисел і поняття "маси". Даючи дефініцію статистичної методи на самому початку свого курсу, ми однією з характерних ознак її визнали "масове спостереження". Що ж таке буде маса /масове явище/? Поверхове розуміння закона великих чисел може привести до висновку, що "маса" буде тим більшою "масою", чим більшим є число спостережень. Такий помилковий погляд приводив не раз дослідників до Massensuht, як кажуть німці, до настирливого прямування накопичити як мога більше число спостережень, щоб зробити висновки з можливо більшої маси цифр.

Одначе робити загальні висновки, визначати пересічні величини маємо сенс лише для таких сукупностей, де маємо підставу знайти певні типовості. Коли в погоні за великим числом будемо брати лише величезні маси, не ураховуючи впливів зовнішнього природного й соціяльного оточення, ми будемо мати лише аритметичні висновки, позбавлені внутрішнього змісту. При таких умовах великі числа не виявляють, а стирають соціяльний тип. Такого роду пересічні типові, так би мовити, притягають життя за волосся. Під один числовий вираз підведемо речі суттєво різні й в результаті будемо мати не з'ясування, а затемнення життєвих реальних відношень і закономірностей.

Виходячи з математичного трактування закона великих чисел, легко переконатись, що збільшення за всяку ціну маси спостережень зовсім не випливає з єства закона великих чисел: Збільшення числа спостережень необхідно постільки, поскільки воно здібно дати важливе звищення точности і певности результата. В чому тут ця важливість-можемо побачити з такого прикладу: протилежні події народження хлопчика й дівчинки мають однакову імовірність 0,5. Візьмемо послідовно число спостережень 20.000, 80.000, 320.000, 2.000.000. Тоді теоретична межа відхилення, на підставі формули модуля $/ \sqrt{2pq} / n /$ буде змінюватись так: при 20.000 спостережень модуль буде $= \sqrt{\frac{2 \times 0{,}5 \times 0{,}5}{20000}} = \sqrt{\frac{1}{40000}} = \frac{1}{200}$, при 80.000 = I/400, при 320.000 = I/800, при 2.000.000 = I/2000. Відповідно цьому теоретичні межі відхилення частот від наперед даної імовірности будуть такими: при 20.000 спостережень - від 0,495 до 0,505, при 80.000 - від 0,4975 0,5025, при 320.000 - від 0,49875 до 0,50I25, при 2.000.000 - від 0,4995 до 0,5005.

Таким чином, перехід од такого невеликого порівнюючи числа спостережень як 20 тисяч до більшого 80 тисяч, дає дуже невелике збільшення точности результата, не кажучи вже про 2 мільйона спостережень. При дослідженнях соціяльно-економічних явищ мусимо визнавати доцільним значно менші точності. Не в накопиченні великих мас фактів, а в широкій їх диференціяції полягає завдання статистики, а закон великих чисел лише вимагає, щоб окремі групи маси, які розглядаємо, в достатній мірі зберегали характер маси, щоб відшукати в них ступінь вияву постійно діючих причин. На цій основі мусять будуватись і комбінаційні таблиці статистичного матеріялу.

Лекція двадцять перша.

56. Апріорна середня похибка. Ми вже знаємо з попередніх лекцій, що емпіричний розподіл вимагає, крім пересічної /середньої/ величини ознаки, ще характеристики, що з'ясовує варіяцію /коливання/, яку ми називали штандартним відхиленням /$\bar{\sigma}$ /. Подібну характеристику можемо мати і для теоретичного /апріорного/ "закона розподілу". Штандартне відхилення /похибка теж/ визначається як корінь квадратов з середньої квадратів відхилень якоїсь змінної величини від її середньої величини. Пристосовуючи це й до закону розподілу, візьмемо відповідно квадрати відхилень од математичного очікування та витягнемо корінь квадратовий з них. Припустимо, наприклад, що величина x набуває вартостей x_1, x_2 x_3 .... x_s з імовірностями p_1, p_2, p_3 ... p_s і зазначимо, що математичне очікування $E/x/ = a_1$. Тоді $/x_1 - a_1/$, $/x_2 - a_1/$ .... $/x_s - a_1/$ будуть відхиленнями величини x від її математичного очікування, а тоді $/x_1 - a_1/^2$, $/x_2 - a_1/^2$.... $/x_s - a_1/^2$ будуть квадрати цих відхилень.

Математичне очікування такого квадрату буде:

$$E\left[/x - a_1/^2\right]$$ та дасть нам міру коливання /варіяції/; але щоб визначити апріорну величину цілком аналогічну штандартному відхиленню, витягнемо корінь квадратовий і тоді одержимо

$$\bar{\sigma} = \sqrt{E\left[/x - a_1/^2\right]}$$, себто теоретичну середню похибку /відхилення/, означивши її літером $\bar{\sigma}$, при чому знак над сігмою даємо для відрізнення теоретичної величини від відповідної емпіричної. Цю величину інколи означають і через $\Delta/x/$, про що ми згадували на стор. I80.

Далі, на підставі дефініції математичного очікування /див.стор. I74/, можемо подати, що

$$\sqrt{E\left[/x - a_1/^2\right]} = \sqrt{\sum_1^s /x_i - a_1/^2 \cdot p_i}$$

З цього виразу бачимо, що замість емпіричних відносних частот окремих відхилень тут взято їх імовірності p_i , а відхилення обчислені не від емпіричної середньої, а від математичного очікування a_1 .

Розглядаючи біноміяльний розподіл, як ідеальну форму розподілу, себто як "закон розподілу" серії іспитів відносної частоти /див.стор.I60 /, ми вже і там зустрілись з штандартним відхиленням та визначили його як $\sqrt{pq/n}$, яке є також апріорним відхиленням , та яке ми зараз називаємо середньою похибкою /помилкою/.

Відносно середньої похибки можемо твердити,

I/ Що квадрат середньої похибки змінної величини є рівним математичному очікуванню квадрату величини мінус квадрат її математичного очікування.

Дійсно, $E\left[/x - a_1/^2\right] = E/x^2 - 2xa_1 + a_1^2/$, або згідно з теоремами про математичне очікування та що a_1 є константним:

$$E\left[/x - a_1/^2\right] = E/x^2/ - 2a_1E/x/ + a_1^2$$

Підставимо замість $E/x^2/ = a_2$, тоді можемо написати:

$$E\left[/x - a_1/^2\right] = a_2 - 2a_1^2 + a_1^2, \text{ себто } \bar{\sigma}_x^2 = E\left[/x - a_1/^2\right] = a_2 - a_1^2 \quad /11$$

а середня похибка буде рівною:

$$\bar{\sigma}_x = \sqrt{a_2 - a_1^2} \quad /12/$$

2/ Можемо зробити відносно середньої похибки інше, неменш важливе твердження, що квадрат середньої похибки суми двох або більше незалежних величин є рівним сумі квадратів середніх похибок цих величин.

Дійсно, візьмемо крім x другу величину y та припустимо, що $E/y/=b_1$ $E/y^2/ = b^2$, тоді одержимо:

$$\bar{\sigma}^2_{x+y} = E\left[/x + y/^2\right] - E\left[/x + y/^2\right] = E/x^2 + 2xy + y^2/ -$$
$$-\left[E/x/ + E/y/\right]^2 = a_2 + 2a_1b_1 + b_2 - a_1^2 - 2a_1b_1 - b_1^2 = /a_2 - a_1^2/ + /b_2 - b_1^2/$$

а згідно з рівнянням / II / $\bar{б}^2_{x+y} = \bar{б}^2_x + \bar{б}^2_y$ та $\bar{б}_{x+y} = \bar{б}^2_x + \bar{б}^2_y$
Цілком аналогічно цьому можемо доказати, що $\bar{б}_{x-y}$ /середня похибка різниці між двома не залежними величинами/ буде рівною

$$\bar{б}_{x-y} = \sqrt{\bar{б}^2_x + \bar{б}^2_y}$$

Це твердження можемо узагальнити для довільного числа незалежних величин, себто

$$\bar{б}_{x+y+z...} = \sqrt{\bar{б}_x + \bar{б}_y + \bar{б}_z} \qquad / 13 /$$

З цієї властивости середньої похибки випливає той важливий висновок, що від середньої похибки незалежних величин, можемо перейти до середньої похибки середньої аритметичної з ряду незалежних величин.

Отже середня похибка середньої аритметичної з ряду незалежних величин є рівною кореню квадратовому з суми квадратів середніх хиб /відхилень/ величин, поділеній на число цих величин.

Приймемо під увагу, що середня похибка добутку змінної величини на константу є рівною константі, помноженій на середню похибку змінної величини, бо $\bar{б}^2_{cx} = E\left[/cx - ca/^2\right] = E\left[c^2/x - a_1/^2\right] = c^2 . E\left[/x - a_1^2/\right] =$
$= = c^2\bar{б}^2_x$, відціль $\bar{б}_{cx} = c.\bar{б}_x$

Отже коли в рівнянні / 13 / помножимо $x + y + z...$ на константу $1/n$, то одержимо:

$$\bar{б}_{1/n\ /x+y+z.../} = \bar{б}_м = \frac{\sqrt{\bar{б}^2_x \quad \bar{б}^2_y \quad \bar{б}^2_z \ .....}}{n} \qquad / 14 /$$

Коли б всі величини мали одну ту ж саму середню похибку:

$$\bar{б}_x = \bar{б}_y = \bar{б}_z = ... \bar{б}_o \quad \text{тоді б мались} \quad \bar{б}_M = \frac{\sqrt{n\bar{б}^2_o}}{n} = \frac{\bar{б}_o}{\sqrt{n}} \qquad / 15 /$$

себто, що середня похибка аритметичної пересічної буде рівнятись загальній середній похибці величин, поділеній на корінь квадратовий з суми цих величин.

На підставі вищенаведеного, можемо зробити нарешті ще один дуже важливий для теорії статистики висновок, а саме:

Коли припустимо, що з якоюсь величиною x , яка може набувати ряд вартостей з певними імовірностями, переведемо цілий ряд серій іспитів /спроб/ і всі ці величини числом n є незалежні, тоді математичне очікування аритметичної середньої M з вартостей величини x в серії іспитів буде рівним математичному очікуванню самої величини. Середня похибка аритметичної середньої /в припущенні, що стала середня похибка величини про всі окремі іспити є рівною $\bar{б}$ / буде рівною

$$\bar{б}_M = \frac{\bar{б}_x}{\sqrt{n}} \qquad / 16 /$$

Таким чином, середня похибка аритметичної середньої з n незалежних величин є рівною середній похибці самої величини, поділеній на $\sqrt{n}$.

Відціль випливає важливий для статистичної методи висновок, що очікуване коливання /хитання/ вартостей аритметичної середньої з ряду спостережень зменшується по відношенню до хитання змінної величини в окремих спостереженнях /іспитах/ пропорційно кореню квадратовому з числа спостережень.

В зазначеному вище відношенні зменшуються можливі відхилення аритметичної середньої з ряду спостережень од математичного очікування величини, в порівнянні з відхиленнями вартостей величини в окремих спостереженнях. Отже, коли з певної цілої маси /сукупности/ робиться вибір "наосліп" /механичний/ одиниць цієї маси та визначається для них середня вартість якоїсь ознаки, то очікувані відхилення середніх вартостей ознаки в ряді можливих таких вибірань наосліп од правдивої /дійсної/ середньої величини для цілої маси /сукупности/ будуть у середньому по відношенню до $\sqrt{n}$ меншим, ніж хитання ознаки в цілій масі.

Ми бачили раніше /стор. 160/, що в ідеальній формі біноміяльного розподілу ряду серій незалежних іспитів відповідно відносній частоті події шан-

дартне відхилення рівнялося $\sqrt{pq/n}$., себто відповідає середній похибці /$\bar{б}$/, але різниця в тому, що раніше ми говорили про середню похибку відносної частоти події, яка може відбутись або не відбутись при кожному даному іспиті, в той час як зараз говоримо про середню похибку середньої аритметичної ознаки, що може при окремих іспитах набувати різних вартостей. Але якісну ознаку /яка може бути лише наявною або відсутньою/ можна вважати за спеціяльний випадок кількосної ознаки, при тому такої, яка може набувати лише вартості I /випадок наявности/ або 0 /випадок відсутности/. Відносна частота події з цього погляду є аритметичною середньою з вартостей ознаки /нулів і одиниць/, яких вона набуває в ряді іспитів. Чому ж є рівною середня похибка такої спеціяльної середньої аритметичної ?

Математичне очікування вартости ознаки в поодиноких іспитах, коли ознака набуває вартости I з імовірністю p , а вартості 0 з імовірністю q = I - p, - буде рівним I x p + 0 x q = p. Відхилення від цієї вартости ознаки будуть I - p та 0 - p = - p. Середня похибка вартости буде рівною

$$\sqrt{/I - p/^2 . p + /-p/^2 . q} = \sqrt{/I - 2p + p^2/ p + p^2/I - p/} = \sqrt{p/I-p/} = \sqrt{p}$$

Щоб перейти до середньої похибки аритметичної середньої для n іспитів / в даному випадку до відносної частоти/ є потрібним /див.рівняння I6/ величину $\sqrt{pq}$ поділити на величину $\sqrt{n}$, тоді середня похибка відносної частоти події в серії n незалежних іспитів /спостережень/ буде рівною $\sqrt{pq/n}$

57. Інтеграл Лапласа й імовірність відхилень при означеному числі спостережень.

З вищенаведеного можна бачити, що для інтерпретації закона великих чисел, в тій загальній формі, в якій він формульований теоремою Бернуллі та в якій припускається необмежене зростання числа спостережень, висновки Чебишова цілком задовольняють. Одначе при розв'язанню питання, якою є імовірність тих чи інших відхилень /похибок/ при даному означеному числі спостережень ці висновки не можуть заступити інтеграл Лапласа, бо формулування Чебишова залишають широкі границі щодо очікуваних величин. Завдяки працям акад. А. Маркова можемо одначе твердити, що інтеграл імовірностей /Лапласа/ може вживатись і бути корисним не тільки в тих випадках, де маємо діло з відносними частотами, але й при середніх з кількосних ознак. Тут так само інтеграл є лише приблизним, але досить точним виразом імовірности того, що емпірична середня не відхилиться від математичного очікування більше, ніж на певну величину.

Отже з імовірністю рівною приблизно інтегралу Ф/u/, що обчислюється по відомій нам таблиці інтегралу, можемо очікувати, що відхилення середньої аритметичної з вартостей змінної величини при n незалежних іспитах /спробах або спостереженнях/ від математичного очікування змінної величини не переступить межі

$$\pm \Delta = u . \frac{\bar{б}\sqrt{2}}{\sqrt{n}}$$

Таким чином, встановивши собі певну граничну похибку Δ , мусимо знайти, чому рівняється u , себто кільки разів перевищує відхилення величину $\bar{б}\sqrt{2}/\sqrt{n}$, що є в даному випадку модулем. З таблиці інтегралу звичайним способом знайдемо імовірність Ф /u/. Величина $\bar{б}\sqrt{2}/\sqrt{n}$, що є модулем, є в той же час середньою похибкою /відхиленням/ аритметичної середньої /$\bar{б}/\sqrt{n}$ /, помноженою на $\sqrt{2}$.

Отже не тільки для випадку відносних частот, але й для більш загального випадку середніх величин, середня похибка зберегає значення так само "константи", в нормальному розподілі, себто межу похибки, на яку можемо покладатись з імовірністю 0,683 та яка при помноженні на $\sqrt{2}$ дає модуль або межу відхилення, на яку можемо покладатись з імовірністю 0,843. Однаково заховується при цьому також значення імовірности, яка належиться решті мір відхилень в нормальному розподілі зокрема для "імовірної похибки", що є рівною 0,67449 середньої похибки. Середня похибка аритметичної пересічної $\bar{б}/\sqrt{n}$ є рівною середній похибці самої змінної величини /в індивідуальних іспитах/ - $\bar{б}$ -, поді-

лений на корінь квадратовий з числа спостережень. Отже, коли знаємо середню похибку змінної, можемо діленням на $\sqrt{n}$ не тільки визначити середній очікуваний розхил /варіяцію/ середніх, що має свій вираз і середній похибці середньої, але й знати, що на межу відхилення рівну цій середній похибці $\bar{б}\sqrt{n}$ можемо розраховувати з імовірністю 0,683. Помноживши цю межу відхилення на числа: $\sqrt{2}$, $2\sqrt{2}$, $3\sqrt{2}$ і т.д. можемо подібно таблиці "інтеграла імовірностей" дати своєму висновку бажаний ступінь точности.

Припустимо, що в кошику знаходиться певне число білетів, на яких є числа x_1 , x_2 .... x_s , при тому число x_1 повторюється в тому числі білетів, яке творить p_1 в цілому числі білетів, так що імовірність витягнення білета з числом x_1 є рівною p_1 , відповідна імовірність для числа x_2 є рівною p_2 і т.д. Математичне очікування числа x на білеті, витягнутому з кошика, буде

$$a_1 = x_1p_1 + x_2p_2 + \ldots + x_sp_s$$

Робиться n тягів, при тому кожний витягнений білет кожний раз повертається до кошика. Ставимо питання. як далеко може відхилитись аритметична середня з чисел на n витягнених білетах од зазначеного вище математичного очікування числа на білеті. Щоб відповісти на це, користуючись інтегралом Лапласа, мусимо обчислити середню похибку числа на білеті при поодиноких тягах за взором:

$$\bar{б}_x^2 = E\left[/x - a_1/^2\right] = /x_1 - a_1/^2 \cdot p_1 + /x_2 - a_1/^2 \cdot p_2 + \ldots /x_s - a_1/^2 p_s$$

Поділивши обчислене $\bar{б}_x$ числом $\sqrt{n}$, одержимо середню похибку аритметичної середньої $\bar{б}_x/\sqrt{n}$. З імовірністю 0,683 можемо очікувати, що аритметична середня з числа на n витягнених білетах не відхилиться від математичного очікування більше, ніж на цю величину. Як перейти від такого відхилення до такої межі відхилення, що мала б більшу імовірність, ми вже знаємо з викладеного трохи вище.

В статистичній практиці в більшості випадків зустрічаємось одначе зо зворотнім завданням, а саме, знаючи емпіричну середню, мусимо знайти імовірну величину математичного очікування, що є основою пересічної, або знайти приблизні межі, в яких може ця пересічна знаходитись. Подібно до того, як це було з'ясовано в випадку відносної частоти, теорія каже, що й при поставленому вище зворотньому завданні, коли число спостережень /іспитів/ досить велике, можемо без значної помилки користатись і тут з тих самих формул, лише з тою ріжницею що в них замінимо апріорні величини емпіричними. Отже, виходячи від емпіричної середньої замість од математичного очікування, мусимо замість середньої похибки змінної величини поставити емпіричне штандартне відхилення для того числа спостережень, яке маємо до розпорядимости. Таким чином середня похибка буде мати взір $\bar{б}/\sqrt{n}$, а модуль $\bar{б}\sqrt{2}/n$.

Наприклад, ми виміряли діяметр 625 горошин певного сорту гороху і встановили, що цей діяметр є рівним в середньому 8,2 мм., при чому $\bar{б}$ було рівним 0,36 мм. Як далеко від цієї пересічної може відхилитись невідоме нам математичне очікування діяметра горошин для цього сорту гороху? Підставимо до взору середньої похибки пересічної емпіричне штандартне відхилення змінної $\bar{б}$ = 0,36, тоді середня похибка буде:

$$\frac{0,36}{\sqrt{625}} = 0,0144$$

Відціль модуль буде рівним 0,02, а подвійний модуль 0,04, Себто з досить великою імовірністю можемо сподіватись, що невідоме нам математичне очікування діяметра цього сорту гороху відхилиться від емпіричної середньої 8,2 мм. в ту або другу сторону не більше, ніж на 0,04 мм.

Треба пам'ятати, що інтеграл імовірностей так само як при користанню ним у випадках відносної частоти, є лише приблизним виразом потрібних імовірностей, але визначає їх тим певніше, чим є більшим число спостережень. Подібно до того, як біноміяльний розподіл відносних частот стає близьким до "нормального" лише при певному числі спостережень або іспитів в серії їх, так і

розподіл пересічних стає близьким до нормального розподілу лише при певному числі спостережень, з яких пересічна обчислена, та стає до нього тим ближчим, чим більшим є число спостережень.

Цей методологічний висновок дуже важливий для статистики, бо в великому числі наукових статистичних досліджень дає їм тверду базу, та можливість при вистачаючому числі спостережень користуватись точною математичною методою для обґрунтування своїх висновків. В цьому й полягає велике наукове значення "закона великих чисел".

58. Емпірична перевірка закона великих чисел. Викладене вище обґрунтування закона великих чисел в теорії імовірностей підкреслює ту основну думку, що формулювання цієї теорії можуть прикладатись до емпіричного матеріялу, зібраного значною кількістю спостережень; теорія каже, що в випадку переведення багатьох серій іспитів /спостережень/, можемо в кожній тисячі таких серій лише декілька разів спіткатися з відхиленням відносної частоти, що перевищує межу подвійного модуля, а, наприклад, в I00 тисячах серій спіткаємось лише 2-3 рази з відхиленням, ідучим понад межу потрійного модуля.

Отже теорія не дає запоруки в тому, що для кожного поодинокого випадку відносна частота, не відхилиться від імовірности за межі потрійного модуля, Це викликає часто в тих, що починають студіювати теорію імовірностей якийсь скептицизм - мовляв, теорія одно, а в дійсності може статися дещо зовсім інше. При цьому звичайно, забувається про складну природу колективних атипових явищ, закономірності яких випливають назовні лише через масове спостереження.

Одначе і строго теоретично доведена можливість прикладання математичних способів аналізи матеріялу статистичного масового спостереження вимагала деякої іспитової контролі теоретичних положень і наукова думка давно вже старалась знайти таку емпіричну контролю закона великих чисел. Цілком природно, що спроби такі найлегше могли переводитись в іспитах з кулями та грах, основаних на випадку та в яких імовірності можемо визначати апріорно і порівнювати з ними емпірічні відносні частоти.

Таких іспитів було переведено багато, але для ознайомлення з ними досить навести декотрі з них. Данський статистик H. Westergaard зробив по сто тягів куль в ста серіях з кошика з білими і червоними кулями, при чому витягнуту кулю повертав знову до кошика. Цих сто серій розподілилося фактично відповідно появленню білої кулі так:

Число появлення білої кулі	Число серій	Число появлення білої кулі	Число серій
34	I	50	9
39	I	5I	5
40	2	52	I0
4I	2	53	4
42	2	54	3
43	3	55	3
44	3	56	5
45	4	57	4
46	5	58	4
47	6	6I	I
48	5	62	I
49	II	63	I

Як часто зустрічались у ряді серій відхилення відносної частоти білої кулі від її імовірности, не перевищуючи певної межі? Модуль для цього прикладу буде такий:

$$\sqrt{\frac{2.\ 0{,}5 \times 0{,}5}{100}} = 0{,}071$$

Себто з імовірністю 0,843 можна було сподіватись, що відносна частота білої кулі в кожній серії не відхилиться від її імовірности / 0,5 / більше, ніж на 0,071, інакше, буде дорівнюватись не більше, ніж 0,571 і не менше, ніж 0,429; це теж саме, що білих куль у серії не буде більше, ніж 57 і менше, ніж 43. Поглянемо на фактичні дані і побачимо, що число появлення білої кулі держалося в вищезазначених межах у 85 серіях із ста. Отже, емпірична спроба досить точно гармонує з теорією.

Однаково можемо перевірити як частими є серії, в яких відносна частота білої кулі держалася в межах апріорного штандартного відхилення або середньої похибки, яка в цьому прикладі буде рівною $\sqrt{\frac{0{,}5 \times 0{,}5}{100}} = 0{,}05$, себто де відносна частота не була більшою, ніж 55 і меншою, ніж 45. Таких серій фактично є 70 із ста. Імовірність не перевищити відхилення за межі середньої похибки, як знаємо, буде 0,683. І тут стверджується теоретичне передбачення.

Можемо розв'язати завдання контролю ще інакше, а саме: обчисливши для нашої таблиці розподілу 100 серій тягів, за одержаною відносною частотою білої кулі - штандартне відхилення /рахуючи відхилення поодиноких відносних частот од імовірности $p = 1/2$/ та звірити, як близьким є одержане відхилення до апріорного, або до теоретичної середньої похибки:

$$\sqrt{\frac{0{,}5 \times 0{,}5}{100}} = 0{,}05$$

Фактичне штандартне відхилення бу де рівним 0,0523, що є дуже близьким до теоретичної величини.

Відомий статистик K. Pearson використав записі, ведені про наслідки кожної гри на рулетці в Монте-Карло. Мав до розпорядимости записи про 16.141 гру, з зазначенням на якому полі рулетки куля застановилася. Відносна частота застоновлення кулі на червоному полі була рівною 0,5015 /при імовірності 0,5/. Відхилення знаходилося в межах подвійного модуля.

В іспитах переведених Бюффоном /підкидання монети/ також у Адольфа Кетле / тягі білих і чорних куль/, у Джевонса - що кидав десять монетами нараз дві тисячі сорок вісім разів, також було стверджено сподівання теорії.

Таким чином, Лапласова теорема дає нам можливість обчислити як межі відхилень відносних частот події від її імовірности, так і межі відхилень пересічних від математичного очікування, при тому такі, перевищення яких є малоімовірним.

Спираючись далі на тезу, що довільно мало імовірні події є дуже рідкими, можемо сподіватись, що в дійсності мало коли спіткаємося з відхиленнями емпіричних величин, що переступлять ці межі.

На сторінці 172. цього курсу для іншої мети ми подали також приклад Веніні відносно розподілу зросту італійських рекрутів, в якому апріорна частота ознаки мало різниться від фактичної; цей приклад до певної міри може служити і для емпіричної контролі теоретичних величин, що вживаються для інтерпретації закона великих чисел, хоч відноситься не до штучних спроб, а до явищ самого життя.

69. Коефіцієнт дивергенції. Для виявлення ступеня наближення між даними емпіричних іспитів або спостережень і теорією вживають інколи відношення штандартного відхилення до середньої похибки. Це відношення називають коефіцієнтом дивергенції /розходження / та означають літерою Q.x/ Коли через x означимо відхилення поодиноких

x/ Інколи коефіцієнт дивергенції називають коефіцієнтом дисперсії, що є неправильним, бо плутає його з поняттям варіяції. / S.Kohn."Zaklady teorie itd 230/.

ноких відносних частот од імовірности / p /, через n число іспитів в серії, а літерою s число серій, тоді коефіцієнт дивергенції формулюється так:

$$Q = \frac{\sqrt{\frac{\sum x^2}{s}}}{\sqrt{\frac{pq}{n}}}$$

Ступінь наближення міряється близкістю Q до одиниці.

Коли візьмемо дані наведеного вище іспиту Вестергарда то

$$Q = \frac{0,0523}{0,05} = 1,046$$

Як довів проф.Борткевич, ліпше з цією метою користуватись величиною Q^2, себто відношенням квадрата емпіричного штандартного відхилення до квадрату середньої похибки, бо лише для величини Q^2, а не для Q може бути доведеним, що її математичне очікування є рівним одиниці / Ко стор.230/. В даному прикладі

$$Q^2 = \frac{\sum x^2}{s} : \frac{pq}{n} = \frac{0,2737}{0,25} = 1,095$$

Для критерія того, що Q^2 відхилюється від одиниці настільки, що не може вже служити міром згідности теорії з дійсністю встановлено для Q^2 його власну середню похибку, яка є рівною приблизно $\sqrt{2/s}$, де s є число іспитів а середня похибка $Q = \sqrt{1/2s}$.

В іспиті Вестергарда s = 100, тоді середня похибка Q^2 буде рівною $\sqrt{2/100} = 0,141$. Модуль $Q^2 = 2/\sqrt{100} = 0,2$. Фактичне відхилення Q^2 від 1, що дорівнюється 0,094, знаходиться в межах середньої похибки та далеко не досягає модуля. Перевірка в багатьох випадках цього коефіцієнта, яко покажчика згідности дійсности з теорією, показала, що в великому числі випадків, в умовах вільних причинових зв'язків, цей коефіцієнт служить добрим мірником близькости теорії з практикою і що теоретичні формулювання зовсім не є зайвою вигадкою, але відповідають правдивим процесам емпіричного життя. Це через те, що існують певні царини дійсного життя, в яких можемо апріорі творити обґрунтувані гипотези для вартостей імовірности явищ. До таких явищ можемо, наприклад, віднести явища схрещування рас тварин і рослин. що студіюється на основі вчення Менделя про розщеплення нащадків гібридів. Тут, на підставі теоретичних міркувань, вдається нам визначити імовірності появлення нащадків з тою чи іншою ознакою та порівняти з цими імовірностями також емпіричні відносні частоти. В цих випадках закон великих чисел цілком успішно витримує іспит. Факт цей сам по собі свідчить про те, що поняття числово визначеної імовірности не обмежується лише колом штучних іспитів або спроб, але може бути аплікованим на певні явища дійсного життя.

Відносно ряду таких інших явищ, явищ соціяльного характеру, можемо навести емпіричне свідоцтво того, що закон великих чисел також є дійсним і для такого ряда явищ.

ВПРАВИ до 20-ої лекції:

1. Що таке є математичне очікування і яка характеристика емпіричної статистичної сукупности йому відповідає ?

Лекція двадцять друга.

60. Стабільність статистичних чисел і міра дисперсії В.Лексіса.

Статистичні відносні числа, як нам уже відомо, можуть, хоч і незавжди, інтерпретуватись, як імовірності виявлення чи появи якоїсь "одиниці" певного характеру в сукупностях, що їх студіюємо. Через те саме статистичне число набуває ясности й певности, а для статистичної аналізи набуваємо можливість користуватись готовим апаратом рахування імовірностей. В цьому спочиває головне значення рахування імовірностей для статистики. Зокрема за поміччю теореми Бернуллі-Пуассона можемо встановляти міру "припущення", або інакше "релятивне довір'я" статистичних вартостей та подати хоч приблизно число одиниць, потрібних, щоб статистичне відносне число досягло певної стабільности./стійкости/. Одначе поступування наше було б неправильним, коли б старались завжди, в кожному випадку, втиснути статистику до "виру" теорії математичної імовірности. Навпаки, мусимо шукати критерія, за яким можемо судити, чи припустимо прикладення конструкцій теорії імовірностей до конкретного випадку, чи ніколи.

Беручи на увагу вплив різних порушень умови незалежности спроб або подій на розділи між відносними частотами та на величину середньої похибки, є дуже важливим зорієнтуватися в тому, чи є порушеною ця умова в певному колі реальної дійсности, що студіюється статистичною методою, і в якій формі порушена. В противному разі, коли оцінюємо можливі похибки /відхилення/ відносних частот від імовірностей /або середніх від математичного очікування/, можемо вжити несправного формулювання середньої похибки, а через те хибно оцінити й можливу похибку. Це має значення як для оцінки якоїсь однієї відносної частоти, так і до порівняння двох або декількох відносних частот, а це останнє завдання якраз досить часто мусить статистиком бути розв'язане. Коли саме ходить про певне розходження між двома відносними частотами, найчастіше питаємо: до якої міри можемо вважати це розходження за "невипадкове". Себто, чи це розходження було викликано розходженням в імовірностях, а не лише випадковими ухилками обох відносних частот від тієї ж самої імовірности, яка є їх основою. Певну відповідь на це питання, в зв'язку з даним числом спостережень, дають відомі нам формули межних випадкових похибок від імовірности. Коли розходження між відносними частотами перевищує ці межі, можемо з більшою або меншою імовірністю вважати його важливим. Формули ці виводились одначе при умові незалежности іспитів /спроб/. Коли характер явища є такий, що ця умова є порушеною, можемо помилитися в оцінці невипадковости розходження у відносних частотах. Наявність розходження, що здається нам неможливою за припущенням, що основою обох відносних частот є та ж сама імовірність, може якраз випливати з позитивного зв'язку між окремими спробами, який збільшує силу хитання відносних частот.

Таким чином, раніше, ніж користатись формулами, треба дуже уважно розглянути умови, за яких дана подія або явище відбувається щодо можливости порушення умови незалежности іспитів /спроб/.

В деяких випадках це питання є неспірним. Можемо; наприклад, вважати що з погляду імовірности народження дитини тієї або іншої статі однієї пари батьків можемо приймати за "незалежні спроби" і, навпаки, з погляду імовірности заразливих хвороб ясно, що окремі індивіди не можуть вважатись "незалежними спробами", але до якої міри залежність спроб, в даному випадку має вплив на середню похибку - сказати тяжко. Бувають і такі випадки, коли не можна взагалі твердо вирішити чи є спроби залежні чи незалежні.

Статистичні відносні величини не можна, як зазначено вище, просто ідентифікувати з математичними імовірностями. Досвід також учить, що як правило вони строго не задовольняють принципам незалежности, рівноможливости і стабільности, наявність яких необхідна для аплікації математичної імовірности. Один дослідник, наприклад, виясновчи вплив добробуту та віку на смертність населен-

ня, поділив його в певному великому місті на три групи: I/ немаєтні, 2/ з вистарчаючими прибутками і 3/ з добрими прибутками. З цих груп померло в певному віці /на I0.000 осіб/ наступне число осіб:

В і к	I	II	III
25 - 35	96	73	58
35 - 45	I9I	I02	92
45 - 55	356	I73	I59

З одержаними тут відношеннями, одначе, не можемо оперувати математичними імовірностями через те, що безперечно числа смертности не задовольняють принципу рівноможливости, бо смертність є очевидно не просто пропорційна добробуту. Коли взагалі кажучи принцип незалежности й рівноможливости та стабільности в статистиці майже не існують, виникає питання, чи не можуть вони прикладатись в окремих випадках з таким наближенням, щоб статистичні дані могли бути міряні математичними імовірностями? Подруге, коли і як далеко статистичні дані можуть мірятись цими імовірностями?

Давала б відповідь нам на це найліпше крива сукупности, побудована на такому великому числі спостережень, щоб була б стаціонарною і при цьому Гавсовою. В такому випадку ми, хочби формально, мали підставу вважати дані вимірними математичним імовірностям. На жаль, у великій більшості випадків не маємо такої кількости спостережень, а з другого боку, робити це з великою стратою часу й праці з тією метою, щоб пізніше переконатись, що крива не є нормальною, було б мало доцільним.

Мусімо шукати іншого способу, іншого практичного і ясного критерія, що давав би можливість наперед судити чи подані статистичні числа міряються математичними імовірностями чи ні.

В сумнівних випадках можуть нам бути корисними критерії "нормальної", "піднормальної" і "наднормальної" дисперсії,[x] встановлені визначним німецьким теоретиком статистики В. Лексісом / W.Lexis /, що допомагають емпіричному розпізнаванню випадків незалежности спроб /спостережень/. Ця метода основана на вимірі дисперсії /коливання/ відносних частот певних подій, що студіюються статистиком в рядах серій /років, місяців і т.д./ та в з'ясуванні того чи відповідає це коливання такому, яке б мало настати за умовою незалежности спроб /в широкому розумінні/ та сталости імовірности від серії до серії. З цією метою порівнюється емпіричне штандартне відхилення ряду відносних частот з апріорною середньою похибкою /при умові незалежности іспитів/. Відношення цих величин - коефіцієнт дивергенції Q, з яким ми вже ознайомились вище, служить нам критерієм оцінювання умов, за яких відбувається подія. Коли Q є близьким до одиниці / в межах своєї власної середньої похибки/, маємо підставу припустити незалежність іспитів в даній категорії явищ /подій/ та таку дисперсію вважаємо нормальною. Коли Q значно менше одиниці маємо дисперсію с у б н о р м а л ь н у /піднормальну/, коли більше одиниці н а д н о р м а л ь н у. Треба мати на увазі, що величина Q дає нам необхідний, але не цілком певний критерій аплікації рахування імовірностей до емпіричного матеріялу. Головна ціна цієї величини полягає в тому, що коли вона є значно меншою одиниці, себто коли маємо субнормальну дисперсію, статистичні дані з великою імовірністю не можуть підлягати закону математичної імовірности, ясніше кажучи, законам випадковости подій.

При обчисленню штандартного відхилення відносних частот, що знаходиться в чисельнику Q, в тому випадку, коли апріорна імовірність p нам відома, обраховуємо відхилення від цієї імовірности. Коли p нам невідома /найчастіше в соціяльних явищах/, обраховуємо відхилення від середньої відносної частоти для всіх серій загалом, при чому рекомендується ділити $\sum x^2$ не s

---

x/ Під дисперсією /розхилом/ розуміється, як уже було наведено на стор.II8, характер /спосіб/ розміщення статистичних вартостей біля середньої своєї вартости.

/числом серій/, а $s - 1$. Середню відносну частоту можемо також обчислити, яко середню зважену, частотами окремих серій. Так само в знаменнику Q /себто $\sqrt{pq/n}$ у випадку невідомої p беремо замість неї середню відносну частоту для всіх серій загалом. Щодо числа n /число іспитів у серії/, то, коли воно міняється від серії до серії /що часто буває в явищах реального життя/ можемо замість n дати середнє число іспитів із всіх серій, - бо, коли коливання n не досить велике, така заміна не викликає значної неточности.

Як уже раніш було зазначено, ліпше обчислювати Q^2, ніж Q, бо якраз математичне очікування Q^2 за умовою незалежности іспитів є рівним точно одиниці. Коли при студії коливань певного кола явищ за ряд років, одержимо в різних часових періодах і в різних країнах вартості Q^2 близькі до 1, то можемо припускати, що дане явище відбувається за умовою незалежности іспитів. При порівнянні двох певних відносних частот з того ж кола явищ, можемо вжити звичайної формули середньої похибки, або модуля для вирішення питання чи є розходження у цих відносних частотах невипадковим.

В дійсності дуже часто трапляється, що вартість Q^2, буде перевищувати одиницю більше, ніж на середню похибку самої величини Q^2. Така вартість цієї величини може означати, що між іспитами існує зв'язок збільшуючий апріорну середню похибку відносної частоти, а також і те, що основна імовірність події від серії до серії /від року до року, або від місяця до місяця/ не залишається однаковою. Одначе, для розв'язання питання маємо полегчення тоді, коли вартість Q^2 буде в даному колі явищ більш-менш сталою, себто для нього характерною, бо тоді можна припускати, що це зв'язано з якоюсь специфічною формою залежности між іспитами /спостереженнями/, а не з часовими змінами імовірности явища від серії до серії. Тоді вартість чисельника Q можемо з певною підставою вважати за міру випадкових похибок у даному колі явищ.

Треба зупинитися також над поняттям стабільности, стійкости статистичних рядів, як його розуміє Лексіс в своїму дослідженні дисперсії.

Візьмемо такі два ряди цифер: перший: 30 - 32 - 31 - 34 - 32 - 32 - 30 - 34 іт.д. і другий: 28 - 30 - 36 - 29 - 32 - 34 - 28 - 33 - итд. На перший погляд немає сумніву, що перший ряд має менші коливання чисел і через те мусить бути означений як більше стабільний, більш стійкий. З погляду Лексісового розуміння стабільности перший ряд може трактуватись як більш і в тій же мірі й менше стабільним. Ступінь стабільности в цьому специфічному розумінні визначається зовсім не абсолютною великістю або невеликістю коливань, а ступенем їх відповідности тим, що обчислені апріорно з додержанням закона нормальних похибок. Коливання першого ряду не виходять з досить вузьких меж 30 - 34, межі коливань другого ряду значно ширші - 28 - 36, але коли нормальні межі для першого ряду будуть припустім, 31,5 й 33,5, для другого - 27 й 38, то в специфічному розумінні Лексіса другий ряд буде стабільнішим од першого. Вимір стабільности робиться зіставленням штандартного відхилення, одержаного в статистичного ряду, з модулем або іншою, математично формульованою на підставі закона великих чисел, міром теоретичної дисперсії. Формулювати це можемо так: у випадках, коли апріорна імовірність p нам відома

$$Q = \frac{\sqrt{\frac{\sum x^2}{s}}}{\sqrt{\frac{pq}{n}}}$$

В явищах, де апріорна імовірність p нам невідома

$$Q = \frac{\sqrt{\frac{\sum x'^2}{s-1}}}{\sqrt{\frac{p'q'}{n}}}$$

де x' означає відхилення, обчислені від пересічної з відносних частот /зваженої або простої/, p' означає цю пересічну з відносних частот, а q' як звичайно, означає її доповнення до I.

Щоб з'ясувати поступ користування міроюдисперсії, візьмемо приклад відповідних конструкцій статистичних рядів.

Припустимо, що явище A має ряд z серій з обсягом s статистичних даних. Нехай в окремих серіях цього ряду явище A відбудеться m_1, m_2... m_z - разів; тоді дроби m_1/s, m_2/s ....m_z/s створять ряд статистичних відносних частот p_1, p_2 .... p_z, які можемо вважати за емпірично встановлені вартості невідомої імовірности p появлення факта A, найімовірніша вартість якого, за умовою, що величини p_k задовольняють принципам стабільности, рівноможливости й взаємної незалежности, - буде рівною артметичній середній

$$p_0 = \frac{[p_z]}{z} = \frac{[m_z]}{zs} \ldots\ldots$$

Середня теоретична похибка вищенаведених членів ряду за формулою Бернуллі буде

$$\mu_1 = \sqrt{\frac{pq}{s}} = \sqrt{\frac{p[1-p]}{s}}$$

Практично ту саму вартість можемо обчислити з формули x/

$$\mu_2 = \sqrt{\frac{[\lambda^2]}{z-1}}, \text{ де } \lambda_k = p_0 - p_k;\ k = 1,2\ldots z$$

Відціль Q буде рівним: $Q = \frac{\text{дійсність}}{\text{теорія}} = \frac{\mu_2}{\mu_1}$; коли $Q = 1$, тоді $\mu_2 = \mu_1$ дисперсія є нормальною.

П р и к л а д I. Для періода I89I-I900 р.р. подає саксонська статистика наступні дані числа народжених по роках /серіях/; z = I0.

р і к	чоловічої статі	взагалі	p_k	$p_0 - p_k = \lambda$	$\lambda\lambda$
I89I	75530	I47.480	0,5I2I4	- 49	240I
92	7325I	I42.527	0,5I349	+ I3I	I7I6I
93	74853	I46.I58	0,5I2I8	- 50	2500
94	74339	I45.66I	0,5I036	-227	5I529
95	74854	I46.I60	0,5I2I4	- 49	240I
96	78090	I52.2I7	0,5I30I	+ 38	I444
97	7880I	I53.663	0,5I283	+ 20	400
98	80340	I56.962	0,5II85	- 78	6084
99	8I332	I58.579	0,5I290	+ 27	729
I900	8I638	I58.563	0,5I487	+224	50I76
	773628	I.507.970	5.I2627		I34285
		[I0s]	[p]		$[\lambda^2]$ x I0000

x/ у випадку, коли $p_k = p$, себто, коли імовірність є величиною сталою $\mu_2 = \sqrt{\frac{pq}{s}}$

З таблиці маємо:

$$p_0 = \frac{[p]}{z} = \frac{5.I2627}{I0} = 0,5I263$$

$$I - p_0 = I - 0,5I263 = 0,48737; \quad s = I50797$$

Отже

$$\mu_1 = \sqrt{\frac{0,5I263 \times 0,48737}{I50797}} = \frac{I}{777}$$

$$\lambda^2 = 0,0000I34285, \quad z = I0$$

тоді $$\mu_2 = \sqrt{\frac{0,0000I34285}{9}} = \frac{I}{877}$$

Відділь $$Q = \frac{\mu_2}{\mu_1} = \frac{777}{877} = 0,95$$

Тут є дисперсія майже нормальна

П р и к л а д 2. З якого видно, як міняється вартість Q . В таблиці a означає число померлих чоловіків на I000 померлих жінок.

вік /роки/	a	Q	вік / роки /	a	Q
I - 2	I028	I,5	45 - 50	II43	3,4
2 - 3	990	I,I	50 - 55	II24	4,3
3 - 5	947	I,2	55 - 60	I055	4,3
5 - I0	878	I,6	60 - 65	962	3,5
I0 - I5	7I3	2,5	65-- 70	9I3	4,3
I5 - 20	770	2,I	70 - 75	906	4,I
20 - 25	I095	I,7	75 - 80	903	2,I
25 - 30	965	I,5	80 - 85	866	I,3
30 - 40	826	2,I	85 - 90	800	I,3
40 - 45	943	2,3			

В перших 5 роках, як і в останніх I0 роках / 80-90 / маємо невелику наднормальну дисперсію, що найчастіше зустрічаємо в статистичній практиці. В ці періоди помирають обидві статі рівномірно. Від 80-90 років чоловіки помирають інакше, ніж жінки. Ще значно більшою буде вартість Q , коли спостерігаємо відношення самогубств у чоловіків і в жінок, тут Q досягає 5.

П р и к л а д 3. В Голяндії за даними I880-I889 років дані смертности по вікових групах були наступні /літера V означає вік/:

V	Q	V	Q	V	Q	V	Q
0 /до року/	6,5	5	2,5	20	I,2	60	I,I
I	4,6	7	I,8	30	I,3	70	I,2
2	3,7	9	I,2	40	I,4	80	I,0
3	3,8	I0	0,8	50	I,6	90	0,8

З цього випливає, що при смертності в дитячому віці та аж до 20 років. не можна припускати, що смертність вібдувається в умовах випадковости, більшменш стабільність явища наступає пізніше в старших роках. Через те, наприклад, не можна статистичні числа віку до 20 років опрацювати на підставі математичної імовірности та на них базувати прогнозу необхідну для асекурації.

Таким чином, нормальну дисперсію можемо очікувати всюди, де імовірність p бігом часу не змінюється. При зміні p перестає бути дисперсія нормальною. Лексіс також показав, що й в випадку змінности цієї величини p відповідає принципам математичної імовірности, дисперсія стає наднормальною, одначе протилежного твердження зробити не можна, бо коли $Q = I$, то ця умова є лише релятивно необхідною, але не є цілком вистарчаючою умовою для існування законів випадку. Щодо коефіцієнта дисперсії піднормальної /субнормальної/, товін на практиці зустрічається у виключних випадках. Отже переважаючою формою дисперсії для статистичних рядів треба визнати **наднормальну дисперсію** і в той же час **піднормальну стабільність**.

Важливим є також орієнтуватись в тому, з якою саме мірою близькости до одиниці / $Q =$ менше одиниці, $Q =$ більше одиниці/ мусимо підходити до оцінки коефіцієнта Q. Коли б було точно встановлено, що при нормальній дисперсії штандартне відхилення мусить точно рівнятись модулю, тоді б саме найменше ухилення Q від одиниці в ту або іншу сторону давало вже б наднормальну або субнормальну дисперсію. На практиці, де не маємо безмежного числа спостережень, однаковість модуля і штандартної похибки буде лише приблизною. Як же оцінити цю приблизну рівність цих величин? Проф. **Борткевич** встановив таке точне мірило припустимого ухилення величини Q від одиниці, визначивши його наступним рівнянням:

$$\sqrt{\frac{\sum / I - Q_i /^2}{n}} = \frac{I}{\sqrt{2z}}$$

де Q_i означає різні, одержані з даного матеріяла значення Q_i, а z число серій, по яких підраховувались спостереження.

Значення коефіцієнта дисперсії не мусимо занадто переоцінювати. Він є лише характерною величиною для з'ясування можливости прикладання рахування імовірностей до емпіричного матеріялу. Апріорі навіть є допустимим, що Q, хоч і значно різниться від одиниці, одначе для емпіричного матеріялу може бути прикладено рахування імовірностей. Стає це тоді можливим, коли число спостережених вартостей є малим і коли серед цих вартостей зустрічаються окремі вартості, дуже великі або дуже малі, та виключенням яких із статистичного ряду коефіцієнт дисперсії негайно знижується. Треба мати також на увазі, що теорія завжди виходить з дуже великого числа спостережень /спроб/, одначе цей ідеал в дійсності майже недосяжний. Розміри цього курсу не дозволяють зупинитися на цілому ряді важливих критичних завважень, які б були зроблені теоретиками для оцінки значення коефіцієнта дисперсії в аналізі статистичних рядів. Для тих, хто бажав би більше поцікавитись оцінкою коефіцієнта дисперсії вважав би дуже корисним ознайомлення з третім розділом праці А.А.Кауфмана: "Теорія і методы статистики" / 4 видання, 1922 р./.

§I. Репрезентативне статистичне спостереження /Вибіркова метода спостереження/. Одним з важливих правил, на яких базується правильне переведення статистичного спостереження є, як нам відомо /стор.8/, додержання принципу вичерпливого числення одиниць певного об'єкту.

Одначе, відомо, що досягти ідеалу вичерпливого спостереження незавди буває можливим. Відсутність фінансових засобів, необхідність переведення перепису та його розроблення на протязі якнайскорщого часу, часто стоять тому на перешкоді. Через те виникла потреба розв'язати теоретичне питання про те, чи не є можливим при переведенню масового статистичного спостереження певного об'єкту, перевести його не в розмірі цілої його вичерпливої маси одиниць на

певному просторі, а обмежитись у певних границях лише частиною цієї маси, цього об'єкту, зберігаючи при тому вимогу діяння закона великих чисел /принципу масовости/, що є основою статистичної методи дослідження колективних атипових явищ. Першим, хто простудіював цю проблему і вніс на обговорення Міжнародного Статистичного Інституту в 1895 році, був норвезький статистик А. К і є р /A.N.Kiaer /. В своїому дальшому науковому розвитку проблема ця студіювалася цілим рядом теоретиків і в останній час найліпше обґрунтована проф. А. Боулі та проф. А.А. Чупровим, в результаті праць яких, XVI сесією Міжнародного Статистичного Інституту в 1925 р. в Римі, була прийнята важлива резолюція, про яку скажемо пізніше.

Першим питанням, яке повстало для теоретика при розв'язанні проблеми можливости використання репрезентативної методи в статистиці, було питання про те, як за даними лише вибірки /частини маси/ з якоїбудь певної сукупности можна зробити висновки про властивості цілої сукупности, а також - чи можна науково зробити порівняння двох сукупностей на підставі лише вибіркових даних, одержаних із цих сукупностей. Теоретичне розв'язання цієї проблеми з практичним освітленням результатів досвіду подано проф. Боулі в його дуже цінному підручнику " Elements of statistics "[x]/ і базується на конструкціях рахування імовірностей.

Отже, маємо якусь цілком певну статистичну сукупність, що складається з досить великого числа N / приблизно декілька десятків тисяч одиниць об'єкта/, з яких pN має певну окрему властивість, тоді число p можемо знайти з задовольняючою точністю за допомогою репрезентативної методи.

Як розподілені в просторі наші об'єкти - рівномірно чи гніздами, це не грає ролі, але ми можемо число p визначити як математичну імовірність, коли понумеруємо всі N об'єктів, а потім, змішавши нумера, відповідні даній сукупності, станемо витягати нумера навмання. Тоді p буде математичною імовірністю витягти нумер, відповідаючий об'єкту з тою властивістю, яка нас цікавить. Цікаво, що при цьому не є істотним, щоб n об'єктів, що студіюється вибірковою методою, складало визначну частину цілої сукупности, бо точність результата залежить тільки від самого n /вона зростає пропорційно $\sqrt{n}$ /, через що, наприклад, при n = 1000, ми знайдемо з однаковою точністю p основної сукупности, що має десять тисяч елементів, як і тієї, що має сто тисяч елементів, аби тільки pn не було дуже малим та дорівнювалось хочби декільком десяткам.

З цього випливає, що практичному дослідникові необхідно наперед точно обмежити основну сукупність, що він студіює, і не розповсюджувати своїх висновків на елементи, які до неї не ввійшли. Себто, на основі дослідження, наприклад, однієї округи галицького Поділля робити висновки для цілого Поділля. Після того, як всі елементи головної сукупности зафіксовані, увага мусить бути звернена на те, щоб кожний елемент мав рівні вигляди на дослідження / Не можна допускати суб'єктивного вибору/. Ця вимога, одначе, не буде порушеною, коли ціла сукупність розкладається за певними ознаками на декілька частин, при умові, що пропорція механічно вибраних об'єктів для всіх частин та ж сама /наприклад, кожне десяте, або п'яте господарство вибирається для всіх частин господарств/; це дозволяє, одночасно з завданням загального дослідження, перевести порівняння й частин між собою. Ні в якому разі не можна по міркуванням практичности та спрощення техніки спостереження направляти вибір до найбільш легких чи сприятливих дільниць чи територіяльних відтинків. Розуміється, що матеріял одержаний від добровільних анкет не може розглядатись як правильно виконана вибіркова метода.

Необхідно забезпечити при виборі також те, щоб по можливості жаден з елементів /одиниць/, на який упаде жеребок /напр., кожне п'яте, кожне десяте господарство/, не минув перепису. Коли ж цього не буде через щось досягнуто із n об'єктів а відпаде з дослідження, а ознаком, що студіюємо, характеризується P' /n' - a / елементів, тоді, крім теоретичної помилки / $б=\sqrt{\frac{P'Q'}{n-a}}$ /

---

x/ Виклад теоретичних основ репрезентативної методи проф.А.Боулі дуже добре зроблений і в російській мові під ред. та поясненнями проф.С.Н.Бернштейна. Вид. Центр.Статист.Управл.УССР,1924 р.

/ при обчисленні p для цілої сукупности треба брати на увагу й те, що недосліджена /пропущена/ частина могла мати також цікаву для нас ознаку і через те в дійсності треба позитивне ухилення /похибку/ збільшити на ap'/n а негативне на aq'/n ; це виправлення незначне, порівнюючи з теоретичною похибкою поки a/n мале. Напр.,при $n = 1000$, пропущення $a = 5$ намічених переписом елементів мало б деяке значення тільки при p' або q' дуже невеликих, коли $б \leq 0{,}01$ /див. С.Бернштейн/.

При порівнянню середнього значення якоїбудь ознаки в двох різних сукупностях на підставі відповідно вибраних з них частин важливо мати на увазі те, що величина певної ознаки в даній сукупності /напр.,вік чоловіків, що женяться, кількість орної землі в окремому господарстві і т.п./ не підлягає закону нормального розподілу частоти, через те визначення середньої величини ознаки і штандартної похибки ще замало без доповнюючої аналізи, щоб охарактеризувати сукупність, а через те й користуватись просто таблицею функцій випадкових похибок $F/z/$ для обчислення який відсоток всіх одиниць сукупности має розміри ознаки, що знаходяться в даних границях; наприклад, цілком можливо, що в одній окрузі середній земельний наділ менше, ніж у другій, але в цій останній більшість населення гірше забезпечена землею, ніж у першій. Питання про те, в якій мірі середня величина ознаки характерна для сукупности є окреме і могло б бути трактовано, коли б дослідник поставив собі метою відтворити всю криву розподілу ознаки за репрезентативним дослідженням. Але це безпосереднього значення немає для обоснування способів приблизного визначення середньої через вибір частини одиниць сукупности та порівняння цих величин в різних групах. Це цікаве теоретичне завваження проф.С.Бернштейна вважаємо потрібним тут підкреслити.

Для того, щоб з'ясувати важливість використання репрезентативної методи для практичної мети взагалі та щоб засвідчити, зокрема,яке не тільки теоретичне, а й суто практичне значення має рахування імовірностей для дослідження навіть складних соціяльно-економічних явищ статистичною методою, наведемо тут успішне оброблення матеріялів сільсько-господарського перепису 1926 року в Болгарії репрезентативною методою. Мотивом використання цієї методи було те, що за відсутністю фінансових засобів перевести оброблення матеріялу електричними машинами Статистичний уряд не міг, а необхідні величезні обрахунки різних статистичних таблиць ручною роботою забрали б дуже багато часу,між тим Міністерство Хліборобства надало якмога скорше одержати результати перепису для необхідної орієнтації щодо стану й ситуації сільського господарства в державі. Для задоволення цього О.Андерсоном[x], професором статистики на високій торговельній школі у Варні був вироблений плян опрацьовання матеріялу репрезентативною методою, що був ухвалений статистичною радою в 1928 році.

В головних рисах плян праці переводився так: перш за все було перевірено всі формуляри перепису[xx]/щодо повноти й якости поданих даних,а потім клясифіковані на групи за розміром обробленої землі. В кожній такій групі і по кожній окрузі в країні було встановлено: число господарств, ціла площа обробленої землі й число осіб тривало працюючих з поділом на статі.

Далі було вирішено питання, який спосіб вибору господарств визнати більш доцільним - "обдуманий вибір"/ choix judicieux/ чи "механічний". Цілком раціонально плян зупинився на "механічному способі", бо лише при такому способі теоретично допустимо встановити межі можливих помилок.

В зв'язку з приняттям механічного способу необхідно було відповісти на такі питання:

I. Яку саме частину цілої сукупности /господарств/ треба взяти до вибору ?

---

x/ Один із визначних статистиків школи проф. А.А. Чупрова.

xx/ Формуляр мав у собі дані, що характеризували кожне окреме господарство в найширшому розумінні, себто ціла площа, власна й арендована, розподіл по вгіддям, площа посівна різних культур, робітництво, інвентар живий і мертвий і т.д., словом, повна програма основного перепису.

2. Яким способом перевести вибір окремих формулярів перепису, щоб уникнути важливих хиб ?
3. Як організувати опрацьовання вибраних формулярів ?
4. Яку дефінітивну форму мусять мати таблиці?
5. Як встановити межі можливої похибки для кожного числа таблиці?

Теоретичні уваги до цього пляну були наступні:

Коли вибір одиниць переводиться механічно, мусять відхилення відносних частот і аритметичних середніх, обчислених для даних вибору від подібних характеристик цілої сукупности, відповідати "закону великих чисел".

Закон розподілу цих похибок /відхилень/ визначається інтенгралом Лапласа /інтегралом імовірности/. Яко єдиний параметр, який дозволяє знайти в таблицях імовірність похибки, що не переступить даної межі, буде т.зв. модуль. Коли вибір формулярів виконується за схемою, при якій витягнені кулі / в даному втпадку формуляри/ повертаємо знов до кошика, буде модуль визначатись формулою: $\sqrt{\frac{2p\,/I - p/}{n}}$ / I /, де p є імовірність, до якої має наближуватись знайдена відносна частота, а n є число вибраних одиниць.

Коли вибір робиться за схемою, при якій кулі, власне формуляри, не повертаються до кошика, модуль є рівним:

$$\sqrt{\frac{2p\,/I - p/}{n\,\frac{N-I}{N-n}}} \qquad / 2 /$$ формула А.А. Чупрова

Понеже імовірність p в даному випадку, як правило, буде невідома - вона мусить бути заступлена емпіричною відносною частотою m/n, де m є число одиниць, що мають спостережену ознаку, а n число всіх одиниць вибору. Тоді формула модуля змінюється в формулу:

$$\sqrt{\frac{2\,\frac{m}{n}\,/\,I - \frac{m}{n}\,/}{n\,\frac{N-I}{N-n}}} \qquad / 3 /$$

Коли далі припустимо, що для границь можливих хиб ми вирішили взяти $\pm$ k - кратний модуль, тоді для меж похибок відносної частоти m/n дістанемо вираз:

$$\frac{m}{n} \pm k \sqrt{\frac{2\,\frac{m}{n}\,/I - \frac{m}{n}\,/}{n\,\frac{N-I}{N-n}}} \qquad / 4 /$$

або $$\frac{m}{n}\left/ I \pm k\,\frac{n}{m} \sqrt{\frac{2\,\frac{m}{n}\,/I - \frac{m}{n}\,/}{n\,\frac{N-I}{N-n}}}\right. \qquad / 5 /$$

Другий член у дужках є так звана релятивна помилка δ,

$$\delta = \pm k\,\frac{n}{m} \sqrt{\frac{2\,\frac{m}{n}\,/I - \frac{m}{n}\,/}{n\,\frac{N-I}{N-n}}} \qquad / 6 /$$

Відціль, після математичних змін, одержимо рівняння:

$$\frac{n}{N} = \frac{I}{I + \frac{/ N - I / \frac{m}{n} \delta^2}{2k^2 / I - \frac{m}{n} /}} \qquad / 7 /$$

Коли вартість k, δ та m/n відомі, можна обчислити n/N, себто яка частина з цілого N мусить бути призначеною для вибору, або в якому відношенню мусять бути між собою n та N. Залежить лише на тому, які вартості треба взяти за k, δ та m/n.

Було рішено, що вибір буде переведений для кожної округи зокрема, і по серйозній уваді було рішено взяти такі вартості: $k = 3/2$, $\delta = I/2$, а $m/n = I/50$. Значення цих вартостей треба добре знати. Те, що за коефіцієнт модулю було взято $k = 3/2 = I,5$, означало, що репрезентативне оброблення даних задовольняється імовірністю можливих похибок визначеною вартістю 0,966 / Замітимо, що в моїому викладі відповідно літері k коефіцієнт модуля означено літерою u, а при $u = I,5$ імовірність визначається вартістю Φ/u /, рівною теж 0,966 /. Це значить, що можемо очікувати, що межі $\pm$ I,5 модуля будуть переступлені лише в 34 випадках з тисячі або в 3,4 випадках із ста. Визначили $\delta = I/2 = 50\%$, себто, що можливі помилки, релятивно обчислені, можуть давати $\pm$ 50%. Вираз $m/n = I/50 = 2\%$ означає найменшу відносну частоту спостереженої ознаки, себто, що будуть взяті для опрацьовування такі ознаки, яких відносна частота в виборі буде дорівнюватись хоч би 2%. Підставимо значення цих величин до вищенаведеної формули та для зручности обчислимо замість N/n /замість n/N /, тоді одержимо

$$\frac{N}{n} = I + \frac{N - I}{882} \qquad / 8 /$$

Коли підставимо за N завжди ціле число господарств в окрузі, дістанемо n рівне числу господарств, які треба взяти для вибору. Розуміється, що зо зростанням N падає релятивна величина вибору, себто відношення n/N. Наприклад, за формулою 8-ю треба б було вибрати I/2 господарств округи, яка б мала всього 883 господарства, I/3 господарств треба б було вибрати для округи з I765 господарствами, I/4 з округи, що має 2647 господарств. Або, навпаки, при N рівному 2000 буде n/N рівне I/3,26, себто, в цьому випадку мусили б вибрати кожний третій формуляр, при $N = I6000$ буде $n/N = I/I9,I3$, себто вибирався б кожний I9-й формуляр. Було вирішено скористатись репрезентативною методою в округах, в яких вибір міг би складати I/3 цілого числа, себто в таких, де було переписано більше, ніж I765 господарств. Для округ з меншим числом господарств вже опрацьовувались всі формуляри, себто вичерпливо. Репрезентативною методою було таким чином опрацьовано 84 округи з 9I, решта вичерпливо. Розмір виб ру хитався від I/3 до I/24, а саме:

I/3 в I округі	I/8 в 9 округах	I/I3 в 7 округах	I/I8 в I окрузі
I/4 в 4 "	I/9 в 7 "	I/I4 в 4 "	I/I9 в I "
I/5 в 4 "	I/I0 в 5 "	I/I5 в 5 "	I/20 в 2 "
I/6 в 3 "	I/II в 8 "	I/I6 в I "	I/24 в I "
I/7 в I4 "	I/I2 в 4 "	I/I7 в 3 "	

Далі, коли був встановлений розмір вибору, було приступлено до вибірання формулярів. Вони перекидались один за другим, при тому витягались кожний 3-й, 4-й, 5-й і т.д. відповідно до того, як великим мусив бути за формулою / 8 / вибір для тої або іншої округи. Наприклад, в окрузі Aitos було 5937 господарств, за формулою мусило вибіратись I/7,7. Для більшої певности десяті долі були відкинуті і таким чином в цій окрузі витягався кожний 7-й

формуляр. В окрузі Aitos з ЇІ.0І3 господарств за формулою мусило братись I/I3,5 - а брався кожний I3-й формуляр. Строго додержувалося, щоби вибір робився цілком механічно; не смів вибиратися формуляр лише через те, що його легше було читати, або через те, що він був ліпше заповнений - це привелоб до "невипадкової"похибки".

Після переведення вибору всі вибрані формуляри опрацьовані за пляном, а результати занесені до таблиць. Розуміється, що одержаних таким чином абсолютних цифер безпосередньо використати не можна, бо вони характеривали лише самий вибір /вибіркову масу/. Через те треба було обчислити числа відносні, але в тих випадках, коли в певному полі виявлялось мале абсолютне число /наприклад, менше ніж 20/, або виявлялись якібудь ознаки з меншою відносною частотою, ніж 2%, мусили клясифікування виправити так, щоб абсолютна частота до кожного поля таблиці записувалась вище , ніж 20, а релятивна частота вище ніж 2%. Ці скорочені таблиці послужили основою для обчислення релятивних чисел /відсотків/, що визначали розподіл господарств за цілою площею, за площею обробленою, за числом осіб тривало зайнятих, за кількістю робочих тварин і т.п. Нарешті, треба було визначити межі похибки кожного релятивного числа та кожної пересічної. Отже, для розхилу меж похибок був взятий ± К-кратний, себто I I/2 кратний модуль. Такими чином кожне релятивне число m_i/n_i занесене до таблиці мусило бути становлено з границями своєї абсолютної похибки, а саме:

$$\frac{m_i}{n_i} \pm \frac{3}{2}\sqrt{\frac{2\frac{m_i}{n_i}/I - \frac{m_i}{n_i}/}{n_i\frac{N_i - I}{N_i - n}}} \qquad / 9 /$$

Для того, щоб таблиці були ясними, було вирішено скласти спеціяльні інтерполяційні таблиці для меж абсолютних похибок частоти типу m/n в окремих виборах, та опублікувати їх при методологічних поясненнях для таблиць. Наводимо тут одну з помічних таких таблиць для вибору в відношенню I:7 /дивись додаток ч.3 /

Лекція двадцять третя.

62. Репрезентативна метода /продовження/. В наведеній в окремому додатку 3-му до цієї лекції, інтерполяційній / про основи інтерполяції скажемо далі/ таблиці всі числа повстали зі взору для абсолютної похибки в формулі / 9 /, але в цій формулі подані границі похибки для релятивного числа m_i/n_i , а в таблиці йде про релятивні числа форми $I00\, m_i/n_i$ себто про відсотки. Щоб границі похибки були в таких самих розмірах себто, щоб вони були в відсотках, слід результативні числа для границь похибки помножити на сто.

Наприклад, перше число в таблиці, себто границя ± 2,75% для 2% при n = I00, є взятий I00 разів результат формули:

$$\pm \frac{3}{2}\sqrt{\frac{2\frac{2}{I00}/ I - \frac{2}{I00}/}{I00\frac{700 - I}{700 - I00}}}$$

Цей результат при малому вибору n = I00 був би одначе нісенітницею, бо нижня границя похибки йшла б під нуль: 2 - 2,75 = -0,75. Таким чином, вибірка в відношенні I:7 буде законною відповідно формулі / 9 / лише для округів маючих 5293 аж 6I74 господарств, себто, з яких вибірка буде хитатись між n= 75I до 882 господарств. В такому випадку треба в таблиці брати на увагу лише ос-

танні рядки передостанньої графи /стовбця/, в якому границя ± 3,45% для 44% при n = 800 є взятий сто разів результат виразу

$$\pm \frac{3}{2} \sqrt{\frac{2 \cdot \frac{44}{100} / 1 - \frac{44}{100} /}{800 - \frac{5600 - 1}{5600 - 800}}}$$

Такого роду інтерполяційні таблиці були складені для всіх вибірок по округах, себто, як подано раніше, для вибірок від I/3 аж до I/24. Користання з них можна пояснити так: наприклад, для округи Aitos переведена вибірка складала I/7 цілого числа господарств в цій окрузі, при чому число вибраних формулярів /господарств/ було рівним n = 844. З цих 844 вибраних господарств було знайдено 9,8% / 100 m/n / господарств розміром до I-го гектара. Яка є помилка цього відсотка, себто в яких межах може хитатись % таких господарств /до I гект./ серед цілого числа 5937 /=N / господарств в окрузі. Для нашого прикладу можемо користатись наведеною інтерполяційною таблицею, бо вона складена якраз для вибірки I:7. Отже в наведеній окрузі Aitos заокруглюємо вибірку I/7 до 800 господарств /замість 844/. В таблиці знайдемо в графі передостанній і в рядку, що найближче відповідає 9,8% себто в рядку для 10%, число, що показує границі похибки; це число буде рівним 2,08 /приблизно - 2,1/ і воно означає, що дійсне число господарств до I-го гектара в цілій окрузі Aitos знаходиться десь у границях: 9,8% ± 2,1%, себто між 7,7% і 11,9%.

Крім релятивних чисел /себто відсотків окремих груп господарств по їх розмірах серед цілого числа господарств, або відсотків площі окремих груп до цілої площі всіх господарств, а також число осіб затруднених в господарстві і т.д./, занесених до остаточних таблиць, складених для результатів репрезентативного оброблення перепису, було обчислено також аритметичні середні. Наприклад, було знайдено середній розмір господарств в окремих групах господарств та середній розмір господарства в цілій окрузі і т.п. Через те треба було для аритметичних середніх, одержаних на підставі даних виборки, також знайти границі можливих похибок. Понеже при переведенню вибору формулярів було додержано умов, які нам відомі з схеми витягнення куль з кошика, себто куля витягається механічно і не повертається назад до кошика, було використано формулу модуля /за А.Чупровим/ для визначення можливої похибки середньої, а саме:

$$\sqrt{\frac{2\mu^2}{n \frac{N - 1}{N - n}}} \qquad / 10 /$$

а сама середня з визначенням границь можливих помилок була формульована так:

$$X/n/ \pm \frac{3}{2} \sqrt{\frac{2\mu^2}{n \frac{N - 1}{N - n}}} \qquad / 11 /$$

В цій формулі $X/n/$ означає аритметичну середню якоїсь квантитативної ознаки X /напр.,розмір сільського господарства в гект./, із всіх одиниць /випадків/ у виборці / X_1, X_2, X_3,.... X_n /, а μ /що часто означають і літерою б / є штандартна /середня квадратова/ похибка від середньої X/n / у взорі

$$\sqrt{\frac{\sum /X_i - X/n/ /^2}{n}}$$

а інколи в такому взорі зо знаменником, зменшеним на одиницю, себто

$$\sqrt{\frac{\sum /X_i - X/n/^2}{n - 1}}$$

Витягнувши $X/n/$ за дужки у формулі / 11 / у вигляді $X/n//1 \pm \delta/$ буде δ релятивною похибкою в формулі:

$$\delta = \pm 3/2 \; \frac{I}{\bar{X}/n} \sqrt{n \frac{2\mu^2}{N-I} \cdot \frac{N-I}{N-n}} \qquad / I2 /$$

Для практичного обчислення всі ці формули можна зробити простішими. Особливо обчислення штандартного відхилення /похибки/ μ^2 можна визначити взором:

$$\mu^2 = \frac{\sum / X_i^2 /}{n} - \bar{X}^2/n/ = \frac{\sum /X_i^2 - n\bar{X}^2/n/}{n} / \quad / I3 /$$

або, зменшивши знаменник на одиницю $= \frac{\sum /X^2/ - n\bar{X}^2/n/}{n-I}$ / I4 /

Маємо на підставі емпіричного матеріялу таку таблицю:

розміри господарства	число господарств n_i	площа цих господарств p_i	середня площа одного господарства p_i/n_i
до I гектара	n_1	p_1	p_1/n_1
I - 2 "	n_2	p_2	p_2/n_2
2 - 3 "	n_3	p_3	p_3/n_3
.....	.....	.....	.....
.....	.....	.....	.....
Р а з о м	$\sum n_i = n$	$\sum p_i$	$\frac{\sum p_i}{\sum n_i} = \frac{\sum p_i}{\sum n} = \bar{X}/n/$

В такій таблиці опрацьовані результати репрезентативного вибору для поодиноких округ. Важливо тут мати на увазі, що тут ми не знаємо площі окремих господарств /напр., в першій групі X_1', X_1'', X_1''' .... X_1^n /, але суму їх площ для цілої групи p_1.

Подану вище формулу / I4 / для μ^2 можемо подати на підставі даних наведеної таблиці. Одначе для виразу $\sum /X_i^2/$ не маємо, як зазначили вище, поодиноких X, але лише суми p_1, p_2 і т.д. для окремих груп господарств. Замінимо тоді поодинокі величини X для кожної окремої групи господарств середньою величиною p_i/n_i так, що квадрат їх буде $/p_i/n_i/^2$. Ці квадрати треба взяти n_i разів, себто для кожної групи господарств треба взяти величину

$$/\frac{p_i}{n_i}/^2 . n_i = \frac{p_i^2}{n_i}$$

Підсумуємо ці величини для всіх груп, тоді одержимо $\sum /p_i^2:n/$, що й є приблизно рівним $\sum /X_i^2/$. Вираз $\bar{X}^2/n/$, себто квадрат середньої величини господарства з цілого вибору знайдемо також, бо з таблиці видко, що він є $/\sum p_i:n/^2$, так що

$$n.\bar{X}^2/n/ \quad \varepsilon \; \frac{/\sum p_i/^2}{n} n \qquad \text{Таким чином:} \quad \mu = \sqrt{\frac{\sum /\frac{p_i^2}{n_i}/ - \frac{/\sum p_i/^2}{n}}{n-I}} \qquad / I5 /$$

Щоб облегчити обчислення, беруться вартості p_i лише з першими трьома знаками, а решта замінюється нулями:, наприклад, ряд I3,I372, I489, 27869 перетворюється в ряд I3, I370, I490, 27900; така манипуляція результат змінює дуже мало.

Прикладення теоретичної схеми до конкретного матеріялу можемо бачити з таких числових даних по окрузі Відін:

Групи господарств по розміру площі	Число господарств n_i	Площі цих господарств p_i	p_i^2	$\frac{p_i^2}{n_i}$	$\frac{p_i}{n_i}$
		ha	ha	ha	ha
0 - I	82	46,9	2.I2 ,6	26,82	0,57
I - 2 "	II2	I67,0	27,889,0	249,0I	I,49
2 - 3	II0	266,0	70.756,0	643,24	2,42
3 - 4	90	3I3,0	97.969,0	I.088,54	3,48
4 - 5	85 x/	277,0x/	76.729,0	902,69	3,26x/
5 - 6	65	355,0	I26.025,0	Ix938,85	5,46
6 - 8	II3	780,0	608.400,0	5.384,07	6,90
8 - I0	70	626,0	29I.886,0	5,598,23	8,94
I0 - I5	87	I.040,0	I.08I.600,0	I2.432,I8	II,95
I5 - 30	4I	802,0	643.204,0	I5.687,90	I9,56
30 і більше	3	I06,0	II.236,0	3.745,33	35,33
Разом	858	4.778,9	3.037.823,6	47.696,86	5,57
	$\sum n_i = n$	$\sum p_i$	$\sum p_i^2$	$\sum p_i^2/n_i/$	$X/n/=\sum p_i/n$

Для цієї Відінської округи:

N = II066 число всіх господарств в окрузі.

n = 858 число вибраних господарств, себто до виборки взято кожний I3-й формуляр-господарство / N/n = I2,9 /.

$X/n/$ = 5,57 ha середній розмір господарства на підставі виборки.

μ = 44,95 ha штандартна похибка середньої X /n / за формулою /I5/,

себто $\mu = \sqrt{\frac{47.696,86 - 4778,9^2 \times 858}{857}} = \sqrt{24,57} = 4,95$ ha

δ = ± 0,06I57 = релятивна можлива похибка середньої X /n / за формулою /I2/, себто:

$$\delta = \pm \frac{3}{2} \cdot \frac{I}{5,57} \sqrt{\frac{2 \times 24,57}{858 - \frac{II065}{I0208}}}$$

δ .X/n / = ± 0,34 ha = абсолютна можлива похибка середньої X /n /.

Таким чином, середній розмір господарства в окрузі Відінській на підставі виборки буде 5,57 ha, але коли б опрацьовано було ввесь матеріял, всі II066 господарств в окрузі, то середній розмір господарства лежав би десь у межах 5,23 ha і 5,9I ha. Таким способом було опрацьовано матеріял в кожній окрузі.

Обраховані числа абсолютні, числа релятивні, а також середні для кожної виборки були зведені в таблиці /Ці таблиці були порівняні з таблицями, складеними для даних цілої округи/.

На підставі даних цих таблиць /для вибіркового матеріялу/ було складено нові таблиці з даними для цілих округ, а саме так, що абсолютні числа, одержані за виборком було помножено на вартість N/n , себто, коли вибір було

---

x/ Тут в оригіналі допущена якась хиба: або число господарств цієї групи почислено неправильно, або сума їх площі, бо середня 3,26 є меншою нижчої границі інтервалу групи від 4-5 гект.,чого бути не може; вона мусить бути більша.

зроблено в відношенню I:7, то кожне абсолютне число з виборки було помножено на сім. Таким способом на підставі репрезнетативних даних було вираховано величини, що приблизно мусили відповідати дійсному масовому /вичерпливому/ матеріялу перепису. Дані одержані репрезентативною методою при обробленню болгарського сільсько-господарського перепису були для 5 округ перевірені в той спосіб, що по цим округам було переведено також оброблення повного матеріялу перепису. Результати перевірки свідчать про те, що репрезентативна метода може цілком себе виправдати і в таких випадках, де маємо до діла з складними соціяльно-економічними явищами, аби використання методи робилося з додержанням теоретичних правил цієї методи та прикладалося до статистичного матеріялу задовольняючої якости. Про це може свідчити нижченаведена контрольна таблиця, складена для 5 округ Болгарії для перевірки можливости використання вибіркової методи для оброблення перепису x/

Групи господарств по розміру в гектарах	Число господарств, одержане:				Їх площа в гектарах			
	в абсолютних		в % %		в абсолютних		в % %	
	вичерпливо	репрезент. методою	вичерпливо	репрезант. метод.	вичерпл.	репрезантованою методою	вичерп.	репрез. метод.
0 - I	24.295	24.354	I0,9	I0,8	I0.999	II.003	0,7	0,6
I - 2	I9.994	I9.806	9,0	8,8	29.073	28.729	I,8	I,7
2 - 3	2I.645	2I.595	9,7	9,6	53.I34	53.036	3,3	3,0
3 - 4	2I.69I	2I.788	9,7	9,7	74,832	75.275	4,7	4,3
4 - 5	2I.I28	2I.350	9,5	9,5	93.894	94.756	5,9	5,4
5 - 6	I8.509	I8.780	8,3	8,4	I00.7I5	I02.I8I	6,3	5,9
6 - 7	I5.778	I5.8I3	7,I	7,0	I0I.725	I02.050	6,3	5,9
7 - 8	I3.254	I3.II3	5,9	5,8	98.760	97.760	6,2	5,6
8 - 9	II.528	II.797	5,2	5,3	97.424	99.887	6,I	5,8
9 - I0	9.2I2	9.383	4,I	4,2	87.I52	88.649	5,4	5,I
I0 - I5	27.475	27.778	I2,3	I2,4	330.964	332.409	20,6	I9,I
I5 - 20	I0.292	I0.280	4,6	4,6	I75.479	I74.926	I0,9	I0,I
20 - 30	6.005	6.686	2,7	3,0	I4I.474	I57.566	8,8	9,I
30 - 40	I.434	I.378	0,7	0,6	48.285	45.708	3,0	2,6
40 - 50	383	386	0,2	0,2	I6.767	I6.895	I,I	I,0
50 - I00	307	280	0,I	0,I	I9.57I	I7.I69	I,2	I,0
I00 - 200	63	82	0,0	0,0	8.407	II.883	0,5	0,7
200 - 300	22	33	0,0	0,0	5.II0	8.2I2	0,3	0,5
300 - 400	II	I3	0,0	0,0	3.734	4.2I2	0,2	0,2
400 - 500	8	9	0,0	0,0	3.604	4.343	0,2	0,2
500 і більше	43	36	0,0	0,0	I04.657	2I2.829	6,5	I2,2
Р а з о м	223.II7	224.470	I00	I00	I.605.760	I.739.478	I00	I00

x/ Користуємось тут таблицею проф. Др. Мраза, що, переробивши подані в оригіналі дані, показані в декарах / I0 декарів I гект./, подає в своїому огляді цифри відповідно в гектарах.

Як видно, дані, одержані через використання репрезентативної методи, досить близькі до тих, що вираховані з цілого емпіричного матеріялу перепису. Вибіркова метода щодо числа господарств дала розходження з дійсністю всього на 6%. В багатьох абсолютних і релятивних цифрах маємо вражаючу подібність. Більші розділи спостерігаються в групах більших господарств, особливо в останній групі, більше 500 гект., щодо % площі, який припадає на цю групу, але ці ріжниці дотикаються дуже невеликого числа господарств. Ми довше зупинилися на цій методі через велике її практичне значення, а також для освітлення й того значення, яке взагалі мають математично строго побудовані конструкції для використання їх у статистичній методі.

Поскільки питання про репрезентативну методу в 1895 році поперше вніс норвезький статистик A.N.Kiaer на обговорення сесії Міжнароднього Статистиного Інститута і з того часу використання її в практиці йде дуже поволі, вважаємо потрібним навести тут резолюцію XVI сесії цього Інститута /1925 р./ в Римі, приняту, після всебічно теоретичного освітлення цієї методи проф. А. Бовлі[x]та покійним проф.А.А.Чупровим. Ця резолюція каже:

"Беручи на увагу неможливість в багатьох випадках зібрати повні статистичні дані відносно об'єкта, що студіюємо, та необхідність через це робити загальні висновки на підставі часткового матеріялу; беручи далі на увагу, що навіть в тих випадках, коли мають повні дані, одержані вичерпливим спостереженням, можемо одержати додаткові відомості для окремих студій над ними з додержанням строгих правил спеціяльної методи і що економія праці, часу й грошей. яку дає вибіркова метода, дозволяє часто використати зібрані відомості більш повно, ніж це дозволяє праця над вичерпливим матеріялом, а тим самим більш досконалогпростудіювати питання, зв'язані з дослідженням,- Інститут ухвалив:

I. Нагадати резолюцію, ухвалену на сесії 1903 року в Берліні і звернути увагу на значні користі, які можемо мати через використання репрезентативної методи з додержанням наступних умов:

А. Механічний вибір. Береться певне число одиниць так, щоб одиниці, що складають ціле, мали б однакові шанси /підставу/ попасти до вибору. Ступінь точности при цьому залежить в значній мірі від числа одиниць, що ввійшли до вибору і яке мусить бути досить великим для того, щоб відхилення через діяння випадкових причин були незначними.

Б. Свідомий вибір. /choix judicieux/. Маса, що підлягає дослідженню, розподіляється на певні групи, з яких вибираються такі, що взагалі характеризуються тими ж ознаками, як і ціла маса. Щоб одержати уяву про точність оцінок, які випливають з такого вибору, число груп мусить бути задовольняючим для виміру коливань характерних елементів від однієї групи до другої. Понеже точність оцінок дуже залежить від обережности при переведенню вибору, рекомендуються такі правила остороги:

1/ Вибір за тим же самим принципом слід зробити два рази, або й більше; після порівняння результатів вибрані випадки з'єднати /цей спосіб рекомендується також і при механічному виборі/.

2/ Через повторне спостереження час од часу досконало простудіювати відношення між вибірком й цілою масою.

II. Інститут радить ставити досліди так, щоб по можливості дати спромогу обчислення ступеню точности результатів. Ці результати мусять вказувати на границі похибки, яка в них може мати місце.

III. Інститут нагадує й наголошує на важливості побажання /резолюція 1903 року/ про те, що до публікації результатів вибіркового дослідження мусить у всіх випадках обов'язково додаватися детальний опис способів вибору, прикладених у дослідженні".

---

x/ Дуже корисним вважав би в додаток до мого викладу ознайомлення з гарним перекладом відповідних глав підручника проф.А.Бовлі, зроблених під редакцією і вступним словом проф. С.Н. Б е р н ш т е й н а під назвою "Теоретические основания выборочного метода" Изд. Цент.Статист.Управл. УССР 1924 р.

63. Інтерполяція. Під інтерполяцією взагалі слід розуміти сукупність способів, за допомогою яких можна знайти деякі величини, що не одержуються безпосередньо з постереження. Математичні операції, що переводяться для такої мети, мусять давати в результаті обчислень найбільш імовірні величини для даного випадку. Завдання це не легке, бо майже всі способи інтерполірування базуються на математичній аналізі, знання якої ми не можемо припускати у слухачів УТГІ. З практичних, одначе, міркувань мусимо подати найпростіші випадки. Отже потрібуємо інтерполяції в таких випадках:

А. Маємо дані для якихбудь явищ, що характеризуються певною постійністю, одначе ці дані одержуються лише через періодичні спостереження, а не постійні в часі /кожнорічні, кожномісячні ітд./, тоді виникає потреба знайти ті дані, які відсутні в певні періоди часу за браком спостереження.

Наприклад, коли маємо дані переписів населення, що відбуваються лише через п'ять років, і хочемо встановити число населення в моменти між цими переписами.

Б. Хочемо групувати елементи статистичної серії в інші групи, ніж ті, що вже подані нам в закінченому зведенню матеріяла. У даних перепису маємо, наприклад, групування господарств по розміру з інтервалами в два або три гектари, таблиці населення по 5-тирічних групах, а потребуємо знати як розподіляються господарства в країні по групах з різницею в один гектар, а розподіл населення по однорічних вікових групах і т.п.

В. Нарешті, коли величини, одержані спостереженням ми замінюємо приблизними величинами з метою з'ясувати або визначити в більш простій, легкій для аналізи формі,- закономірність існуючу в основі явища, що студіюємо. Наприклад, складну ломану лінію руху смертности, в якійбудь країні за певний період часу, замінюємо математичною кривою, що підлягає в своїй будові математичному закону. Такого приблизно характеру є й ті величини інтерполяційної таблиці, що були визначені для опрацьовування даних болгарського перепису репрезентативною методою, та про які вже говорили ми вище.

З даних переписів ми знаємо, наприклад, що да 5-річчя з I900 до I905 року число населення збільшилось у пропорції I00:II0, яка була кількість населення в I90I, I902, I903 і I904 роках. Для рішення завдання скористуємося рівнянням простої лінії: $y = ax + b$. / I /.

Беручи I900 рік за вихідний, надамо незалежній змінній / x / послідовні значення від 0 до 5, тоді одержимо ряд вартостей залежної змінної / y /:

При вартості $x = 0$ $y_0 = 0a + b = b = 100$ / I900 /
$x = 1$ $y_1 = a + b =$ / I90I /
$x = 2$ $y_2 = 2a + b$ / I902 /
$x = 3$ $y_3 = 3a + b$ / I903 / / 2 /
$x = 4$ $y_4 = 4a + b$ / I904 /
$x = 5$ $y_5 = 5a + b$ II0 / I905 /

Для величин y маємо два певних рівняння $b = 100$; $5a + b = 110$. Відціль легко одержати вартості постійних коефіцієнтів a й b, себто:

$$5a = 10; \quad a = 2; \quad b = 100.$$

Таким чином, перше рівняння / $y = ax + b$ / буде мати вигляд: $y = 3x + 100$. Отже замінивши у формулах / 2 / вирази коефіцієнтів a й b, знайденими конкретними вартостями, одержимо:

$y_0 = 100$ /I900/
$y_1 = 2 + 100 = 102$ /I90I/
$y_2 = 4 + 100 = 104$ /I902/
$y_3 = 6 + 100 = 106$ /I903/
$y_4 = 8 + 100 = 108$ /I904/
$y_5 = 110$ /I905/.

Ми знайшли величини р е л я т и в н о г о щ о р і ч н о г о з р о с т у населення за пятирічний період. Знавчи, що по перепису 1900 року населення дорівнювалося, наприклад, 30 мільйонам, дуже легко знайдемо абсолютні цифри населення для решти років / x:30 = 102:100 = 30,6 мільйонам і т д./.

Наведений приклад характеризував інтерполірування між двома точками.

Припустимо далі, що нам треба зробити інтерполяцію через три точки й розв'язати таке завдання: по даних статистики за 1927 рік померло чоловіків на Україні /УССР/ у віці: від 21-23 років уключно 5.446 осіб, від 24-26 років - 4.942 особи і від 27-29 років 4.020 осіб. Треба встановити скільки осіб /чоловіків/ померло в кожному році за цей період?

Для розв'язання цього завдання беремо рівняння параболи

$$y = ax^2 + bx + c \qquad / 3 /,$$

означаючи через x - вік, а через y кількість осіб даного віку.

Нам треба далі встановити кількість померлих наступних дев'яти груп померлих, а саме: 21 року, 22 р.,23 р.,24 р.,25 р.,26 р.,27 р.,28 р.,29 р.

За вихідну точку беремо середню з дев'яти груп, себто 25 років, відповідному x надаємо вартости 0, а далі, нижчим від середньої групам послідовно вартості від -1 до -4, а вищим групам від +1 до +4. Тоді одержимо такий ряд вартостей для y :

Вікова група	X		
21	-4	$y_0 = 16a - 4b + c$	5.446
22	-3	$y_1 = 9a - 3b + c$	
23	-2	$y_2 = 4a - 2b + c$	
24	-1	$y_3 = a - b + c$	4.942
25	0	$y_4 = c$	
26	+1	$y_5 = a + b + c$	
27	+2	$y_6 = 4a + 2b + c$	4.020
28	+3	$y_7 = 9a + 3b + c$	
29	+4	$y_8 = 16a + 4b + c$	

Додавчи три верхніх, три середніх і три нижніх рівняння, знайдемо що:

$$29a - 9b + 3c = 5446$$
$$2A + 3c = 4942$$
$$29a + 9b + 3c = 4020$$

Маємо три рівняння з трьома невідомими звідки можемо знайти вартості коефіцієнтів a , b та c , а саме: a = -7; b = -79,2; c = 1645,3 і наше рівняння визначиться : $y = -7x^2 - 79,2x + 1645,3$

Підставлявчи знайдені вартості коефіцієнтів до рівнянь y , одержимо:

$16a - 4b + c$	$= -112 + 316,8 + 1645,3$	$= 1850,1$	5.445,7
$9a - 3b + c$	$= - 63 + 237,6 + 1645,3$	$= 1819,9$	
$4a - 2b + c$	$= - 28 + 158,4 + 1645,3$	$= 1775,7$	
$a - b + c$	$= - 7 + 79,2 + 1645,3$	$= 1717,5$	4.919,2
c	$=$	$= 1645,3$	
$a + b + c$	$= - 7 - 79,2 + 1645,3$	$= 1556,4$	

$$\left.\begin{array}{lll} 4a + 2b + c = \ldots\ldots\ldots\ldots\ldots & I458,9 & \\ 9a + 4b + c = \ldots\ldots\ldots\ldots\ldots & I344,7 & 4.020,I \\ I6a + 4b + c = \ldots\ldots\ldots\ldots\ldots & I2I6,5 & \end{array}\right\}$$

Щоб порівняти інтерполяційний ряд з дійсними цифрами померлих за відповідні роки, наведемо тут дійсні цифри померлих відповідного віку:

Вікова група	Дійсно померло	Інтерполяцією обчислено
2I рік	I863	I850,I
22 "	I889	I8I9,9
23 "	I694	I775,7
24 "	I669	I7I7,5
25 "	I6I9	I645,3
26 "	I654	I556,4
27 "	I592	I458,9
28 "	I373	I344,7
29 "	I055	I2I6,5

З наведеного прикладу ми можемо бачити, що розходження даних емпіричних і одержаних інтерполяцією в деяких групах віку є дуже великим. Прикладення більш досконалих математичних інтерполяційних формул, про які не можемо тут ми говорити, бо всі вони основані на знанні вищої математичної аналізи не в усіх випадках може нас задовольнити також. Все це свідчить про те, що інтерполяцією треба користуватися з надзвичайною обережністю не тільки через складність її методологічну, але й через властивість об'єкту, з яким, звичайно, оперуємо.

Хочемо, наприклад, знайти через інтерполяцію дані урожаю зернових хлібів на Україні за ті роки, за які не маємо емпіричних даних. Наприклад, маємо врожай в пудах з десятини за низку років так, що за десять років з I9II-20 р.р. ці цифри конкретно дані. За I92I р. і 1922 р. даних не маємо. За I923 - I928 р. знову дані маємо. Необхідно визначити еволюцію врожаю за всі ці роки. Виходячи з даних конкретних, через інтерполяцію можемо одержати величини розміру врожаю в р.р.I92I-і I922, але без сумніву, обчислені таким способом цифри для цих років будуть зовсім не відповідати дійсності і це через те, що I92I р. був роком виключним, роком страшного голоду й роком врожайче низького врожаю, в той час як величина його, обчислена інтерполяційно, базувалася б на даних урожаю та їх еволюції за роки нормальні. Отже дослідник, що не зверне уваги на виключність умов урожаю I92I р. зовсім не правильно висвітлив би еволюцію врожаю за I9II-I928 роки.

Дослідник при використанню способу інтерполяції завжди мусить мати на увазі з самого початку праці, чи не мається в частині того періоду даного об'єкта, яку хочемо поповнити через інтерполяцію, таких років, взагалі моментів, які мусять бути трактовані, як виключне явище в цілому динамічному процесі об'єкта за період дослідження.

Не завжди це буває можливим для дослідника, через те й користування методою інтерполяції вимагає попереднього досконалого знання характеру, умов і обставин, в яких відбувається явище, що досліджуємо.

ВПРАВИ /до 2I-ої лекції/:

I. Що таке "маса" з оглядом на статистичну методу ?

Лекція двадцять четверта.

64. Статистика й математика. В попередніх лекціях було висвітлено значення математично-символічних конструкцій і теорема конструкцій рахування імовірностей для наукової аналізи того об'єкта, тих статистичних сукупностей, що ними доводиться оперувати дослідникові статистичнов методов. Вказано було також і на те, з додержанням яких умов мусять прикладатися ці конструкції, щоб зберегти своє дійсно наукове значення.

Історія розвитку науки про статистичну методу яскраво вказує на те, що математичні формулювання і зокрема теорія функцій набирають все більшої ваги для суто-наукових статистичних досліджень значного кола явищ органічної природи, не виключаючи також і явищ соціяльно-економічних. Одначе сучасні теоретики статистичної науки далеко не з однаковим оптимізмом дивляться на можливість широкого прикладення математичних метод так званої математичної статистики до розв'язання проблем, що виникають в процесі статистичного дослідження явищ соціяльно-економічної природи, коли це дослідження ставить собі завдання не тільки ідіографічні, а й номографічні.

Дійсно, коли закономірності й причинові взаємозалежності в явищах, що пошукуються статистичнов методов, не можуть бути науково з'ясованими і визначеними за допомогов лише принципів індуктивної логіки, але потребують обґрунтування в принципах рахування імовірностей, то в випадку признання цих останніх практично недоцільними а може й принципово зайвими; висновки статистичної номографії залишаються "вісячими в повітрі".

Як бути, наприклад, з тим, що видатний теоретик і практик статистики А.Кауфман вважає, що "питання нашого права оперувати способами числення імовірностей для безконечно широкої сфери "загальної статистики" залишається в кожномуразі для нього відкритим". Академик С.Бернштейн недавно[x]/, з'ясовуючи поняття статистичної кореляції, каже: "Умови, за яких зберігаються принципи теорії імовірностей, так само трудно реалізувати, як лабораторні досвіди, де в чистому вигляді виявляються закони механіки та фізики. В конкретній дійсності суспільного та народньо-господарського життя ми ще більш далекі від математичних імовірностей та кореляцій, аніж в техніці від твердих тіл та ідеальних рідин теоретичної механіки. Якщо, підбираючи у відповідний спосіб індивідів певної риси в біології, можна утворити умови близькі до математичних схем теорії імовірностей і з певними підставами говорити про кореляцію між тими чи іншими ознаками, то в царині економіки статистичне дослідження має далеко грубіший емпіричний характер і ми апріорі не знаємо, наприклад, в чім повинна полягати схожість двох економічних станів для того, щоб приписати одну й ту саму величину математичній надії /або імовірному значенню/ ціни вугілля чи якого іншого певного товару. Так само не має математичного сенсу і коефіцієнт кореляції / зв'язку - Л.Ш./ між ціною вугілля та ціною чавуна".

З другого боку, Bowley твердить, що "передпосилкою прикладання формул рахування імовірностей є теоретичний розподіл похибки, що лежить в основі всіх явищ", а Forcher , що "в дійсності все життя і зокрема життя соціяльне розгортається в вигляді безперервних функцій і таким чином може досліджуватись за допомогою аналізи безмежно малих" і далі нам тільки здається, що життя і зокрема життя соціяльне розгортається не в вигляді безперервних функцій, бо ми тільки ще не бачимо промежних ланків у ланцюзі явищ, ми ще не досить нагострили для цього наш духовний зір:

Проф. хемії А.Щукарів ще в 1913 році в своїому докладі в підсекції статистики на 13. З'їзді російських природників і лікарів на тему: "Проблема

---

x/ Вістник Статистики України. 1928 . І. "Поняття кореляції між статистичними величинами.

застосування математичної аналізи до явищ соціяльного життя" не тільки дозволяє повну можливість, а й визнає необхідним користуватись математичними конструкціями т.зв. хемічної механіки та статистичної фізики, що вільно допускають перенесення цих конструкцій до ізопотенціяльних подій /інакше рівноможливих/ взагалі. Наводячи формули й приклади з хемічних реакцій /процес розчинення твердого тіла в певному об'ємі рідини/ та закона хемічної рівноваги, Щукарів далі каже: Таким чином існує царина ізопотенціяльних подій, в яких свобода не є протилежна закономірності. Як бачимо, всі наведені конструкції є дуже загальні: коли в них говоримо про частки, іхні взаємні заміни, з'єднання, переходи з одного середовища фази в друге, то сама природа цих часток цілком не входить у склад наведених рівнянь. На місце часток можемо підставити абищо: кулі, що свобідно котяться, врешті людей з хіними свобідними вчинками. Міркування та рівняння залишаться старі й ми одержимо: з а к о н и с у с п і л ь н и х п р о ц е с і в". Свої аналогії Щукарів далі ілюструє прикладами: наприклад, визначає процес скорости одруження, аналогічно до скорости хемічної реакції. "Детальніше опрацьовання цих законів /в хемії це вже робиться/ та їх застосування до соціяльних явищ може дати згодом не лише дуже цікавий науковий матеріял, давчи нам змогу заглянути нарешті в найглибші тайники структури людської громади, структури, що про неї ми тепер лише фантазуємо, але й очевидно може мати чимале практичне значення".

"В далекім майбутнім легко можемо уявити собі країну, що побіч з метеорологічною, сейсмічною та іншими сітками чисто фізикальних лябораторій матиме й сітку "соціяльних лябораторій", що будуть старанно вивчати, може експериментувати та в усякому разі дуже уважно слідкувати за змінами всіх родів соціяльних коєфіцієнтів "К", скеровувчи діяльність адміністративних та господарських органів в бажаному напрямі". Далі він вказує на те, що й закон Макевелля, який визначає розподіл швидкостей на групу свобідно рухомих молекул, незв'язаних між собою жадними фізичними силами, також згодом знайде собі застосування до галузі соціяльних явищ.[x]

Щодо масових фізичних явищ, то видатний фізик Смолуховський міг перед компетентними представниками науки твердити: "закони фізики стають законами статистичними, хто не розуміє статистики, не може розуміти модерньої фізики".

Проф. Д-р. В. Л а с к а [xx] зовсім певно висловлюється, щодо поняття статистики; на його думку, "сміливо можна сказати, що статистика є природничою наукою своїм предметом і своєю методою, експериментом її є сукупність певного аргументу /змінної/, здібна витворити якусь характеристичну властивість або, ясніше кажучи, певну криву сукупности. На жаль, статистика вважається ще сьогодня частіше за домен правників, а математична статистика за нешкідливу іграшку декількох учених, або найліпше цікавою аплікацією рахування імовірностей:. Таким чином статистична школа Георга Майра, так звана "описова школа", чим далі, тим більше не задовольняє представників статистичної наукової думки. Можна певно ствердити, що взагалі математичне думання і математичне конструювання та визнання багатьох проблем соціяльної економіки набуває в все більшого значення. Досить для цього звернутися хочби до сучасних досліджень господарської коньюнктури.

Це не значить одначе, що принципи правільного побудування статистичної сукупности та її опису мають другорядне значення; навпаки, на нашу думку, чим більше застосовуємо математичну методу до аналізи атипових колективних явищ, тим з більшою увагою та суворістю мусимо додержуватися засад правильного статистичного числення, поскільки воно, як ми вже зазначили на початку своїх викладів, є перш за все матеріяльним численням.

Можливо, що занадто велике захоплення математичною методою одночасно з деяким занедбанням суворости статистичного опису явищ і їх характеристики сприяло тому, що математична метода в статистиці знаходить ще й досі досить значну кількість скептиків.

x/ Червоний Шлях № 8-9 1924 р. стор. 191.

xx/ Vybrané kapitoly z matematické statictiky

Цим, між іншим, пояснюється й те, що значну /першу/ частину основ науки про статистичну методу ми присвятили тим головним формам і типам статистичного числення, завдяки яким одержуємо статистичні сукупності та з яких будуємо статистичні ряди.

Йдучи по шляху наукового дослідження статистичною методою явищ, ми мусимо твердо засвоїти, що чим складнішою є природа атипових колективних явищ, починаючи від явищ неорганічної природи через явища біологічної природи аж до явищ соціяльно-економічної природи - тим з більшою обережністю мусимо оперувати математичними формулюваннями для відшукання постійних відношень і причинових зв'язків в явищах, що досліджуємо. Емпірична дійсність висловлена в статистичних сукупностях і рядах у більшості не є строго математичною конструцією і принципи теорії математичних функцій до динамічних процесів соціяльної економіки можуть прикладатись поки що як аналогії /подібності/, а не як тотожності. Лише в цьому ми вбачаємо і трудність застосування математичної методи до соціяльно-економічних явищ. Отже статистика не є математика, але без останньої вона не має того лихтаря, який світить їй в темряві складних процесів, що відбуваються в економіці й суспільстві. Закономірності цих процесів можуть бути цілком науково формульовані лише за допомогою математики. Що це так, можна пересвідчитись і на прикладанні репрезентативної методи до статистичного дослідження складної економічної природи, як це ми бачили на прикладі болгарського сільсько-господарського перепису й його розроблення.

К О Р Е Л Я Ц І Я.

/ Елементарні способи відшукання кореляційної залежности /.

65. Поняття кореляції. Під к о р е л я ц і є ю в головному розуміємо причиновий зв'язок між двома або декількома змінними величинами. В одних випадках певна змінна величина є в прямому зв'язку з другою, в інших випадках певні змінні величини є в прямому /безпосередньому/ зв'язку лише з деякими величинами і в посередньому зв'язку з іншими; інколи певні змінні величини є в зв'язку з деякими величинами і в той же час не мають ніякого впливу на інші. Про яку б з цих можливостей зв'язку ми не говорили, мусимо знати, що корелятивний зв'язок завжди означає певний причиновий зв'язок між всіма або декількома величинами. Отже в тих випадках, коли зіставлення статистичних рядів і їх величин не припускають логічно причинового зв'язку між собою, маємо справу не з кореляцією, лише з коваріяцією, себто з певним ступенем паралелізму або антипаралелізму, збіжности чи розбіжности величин, d'une concomitance , за влучним виразом французького статистика Mareh'a. Нема сумніву, наприклад, що між числом днів з дощами в місяці та між числом проданих парасолів в якійсь крамниці існує причиновий зв'язок. Тут можемо говорити про кореляцію. Звичайно, що цей причиновий зв'язок не є таким простим, як це здається на перший погляд. Певна кількість парасолів продається і в такі місяці, в які не було жадного дощу; в місяці з подвійним числом днів з дощем, не продається обов'язково подвійне число парасолів. Поскільки, одначе, в наведеному випадку йде про безсумнівний причиновий зв'язок, користання кореляційною методою для визначення ступеня залежности явищ, буде цілком законним.

Навпаки, можемо спостерегти цілковиту згоду, збіжність, наприклад, між зміною ціни масла й полотна, але прямого причинового зв'язку між цими двома змінними величинами не мається, і через те кореляційна статистична аналіза тут не до діла, навіть і тоді, коли б з'ясували, що коефіцієнт кореляції між цими змінними є рівний 0,98, себто майже рівний одиниці /повний позитивний зв'язок/.

Відціль випливає, що для правильного користання методою кореляції тре-

ба мати логічно-справедливу гіпотезу, цілком розумне припущення того, що між певними змінними дійсно може існувати кореляційний зв'язок. Лише після того можемо вжити графічного або більш точного, математичного способу визначення ступеня кореляції.

Простішим для нас буде такий випадок, коли маємо лише два статистичних ряди, дві змінні величини, з яких одна є лише причиною, а друга наслідком. В цьому випадку можемо вважати статистичний ряд, що означає причину, незалежною змінною, себто таким статистичним рядом /змінною/, на який не має ніякого впливу статистичний ряд, що вважаємо за наслідок. Математично це визначається як зв'язок між змінною ознакою факторіяльною й функціональною. В попередніх лекціях ми вже зупинялися на тому, що таке є факторіяльні й функціональні ознаки. Зазначили також те, що просте констатування зв'язку між ознаками досягається зіставленням статистичних рядів в комбінаційних таблицях. Одначе комбінаційна таблиця, давчи можливість встановлення позитивного або негативного зв'язку певних статистичних рядів чи змінних, не дає можливости кількісно визначити як ступінь існуючого зв'язку, так і розмір інтенсивности цього зв'язку. Це завдання й розв'язується методою кореляції.

Ситуація скомплікованости зв'язку між двома рядами, або змінними мало ускладнюється в тих випадках, коли фактично маємо такі два ряди, з яких один є не єдиною виключно причиною, а другий не є єдиним виключним випадком, але перший можемо вважати за основу, рішавчу причину /вплив/, а другий за основний наслідок, бо важливим тут є те, що причиновий зв'язок між цими двома змінними на основі якогось ясного принципу, теорії чи розумної гіпотези безсумнівно існує.

Значно складнішими випадками будуть такі, коли обидва ряди є наслідком певної причини або ряду причин, з якими мусимо рахуватися при аналізі. Ця складність стає ще більшою й тяжчою до розв'язання, коли маємо до діла з рядами, в яких певна причина є у зв'язку лише з деяким наслідком /змінною/ і не має ніякого впливу на решту рядів або змінних, в той час як інша причина є в зв'язку знову таки лише з деякими наслідками і не впливає на решту.

Маємо, наприклад, два ряди A і B, що є наслідками. Логічно до них стосуються три групи причин: ті, які стосуються виключно до ряду A, ті, які стосуються виключно до ряду B, і ті, що стосуються до обох рядів A і B. В кожній групі причин можемо далі розрізняти два види, причини, що визначені квантитативно, які можемо міряти, і такі, яких міряти не можемо, себто які не можемо формулювати для аналізи статистичною методою. В кожнім виді причин вимірних чи невимірних можемо також розрізняти ті, про які знаємо, від тих, про які при всьому бажанні відомостей не маємо.

Коли зважимо при тому, що ряд A може бути в певному зв'язку з рядом B, себто може на нього впливати і зворотньо, що ряд B може мати певний вплив на ряд A - будемо мати дуже складну ситуацію зв'язків, певних взаємозалежностей, яка є характерною для явищ соціяльної економіки, та яку мусимо розв'язувати методою кореляції. Нема чому дивуватися, що в таких випадках навіть складна, строго математична аналіза не дасть безпомилкових результатів.

Таким чином не можемо від кореляційної методи чекати науково досконалих результатів в значній кількості тих випадків і явищ, для яких характерним є співіснування зв'язків багатьох причин і наслідків. Навіть з теоретичного боку метода кореляції не є остільки закінчена, виспіла, щоб надавати їй універсальне значення при розв'язанні проблеми кількісного визначення зв'язків у соціяльно-економічних явищах.

Мусимо це зазначити через те, що метода ця в сучасних економічних дослідженнях дуже розповсюджена, хоч можливість її корисного прикладання не є завжди оправданою. Для доцільного використання цієї методи доброї математичної підготовки не досить. Не дооцінювати процесу логічної аналізи, яка набуває тут особливого значення і мусить завжди йти раніше математичного формулювання ні в якому разі не можна, потрібна тут попередня досконала об'єктивна аналіза статистичного матеріялу.

А все ж таки практика американської статистики, починавчи з напряму,

даного Гарвардським комітетом, а потім Bradford[1] Bitbay Smit' в статистичній лабораторії Вашингтонського Bureau of Agricultural Ekonomic щодо аплікації методи кореляції на практичні проблеми в хліборобстві, а також конструювання системи господарського барометру, який щомісячно видає American Telegraph and Telephone Company до певної міри свідчить про те, що кореляційна аналіза економічних явищ не тільки цікава формально, але й обіцяє плідний розвиток надалі.

66. Критерії кореляційної аналізи. В зв'язку з вищенаведеним можемо поділити кореляцію на підставі таких основних критерій:

I/ Коли маємо на меті відшукати розмір й силу зв'язку між двома лише змінними /рядами/ - маємо просту кореляцію. У випадках зв'язку між декількома рядами маємо до діла або з складною кореляцією або багатоподільною. Досліджуємо певне явище яко комбінований наслідок декількох причин - будемо мати складну кореляцію. Наприклад, прибуток промислового підприємства аналізуємо в залежності від кількости виробленого товару, а ціни останнього від розміру усунення страт з виробничого процесу, від інтенсивности праці робітників. Коли одначе, студіюємо явище в той спосіб, що беремо на увагу завжди дві змінних /ряди/ і їх взаємний вплив однієї на другу, в той час як решту змінних, які також виконують певний вплив, приймаємо за сталі /або для них визначаємо також якусь кореляцію/ - тоді маємо кореляцію багатоподільну. Наприклад, хочемо знайти як змінюється прибуток в залежності від того, як змінюється кількість виробленого товару, але при решті рівних умов, себто решту чинників вважаємо за константні.

2/ Другим критеріяєм може нам служити н а п р я м о к з м і н и в з м і н н и х. Коли зріст або спад величин однієї змінної збігається зі зростом або спадом величин другої і то так, що високі вартості однієї змінної співіснують з високими вартостями другої, а невеликі однієї з невеликими вартостями другої - маємо к о р е л я ц і ю п о з и т и в н у. Коли ж зріст величин першої змінної супроводиться спадом величин другої змінної і навпаки - тоді маємо к о р е л я ц і ю н е ґ а т и в н у.

3/ Нарешті, критерієм мусить служити також відношення змін у поодиноких змінних величинах. Коли зміни в одній змінній /ряді/ будуть супроводитись змінами в другій, а то так, що приростання й спадання величин в обох змінних знаходяться в константному відношенні - маємо л і н і й н у к о р е л я ц і ю. Коли ж це відношення в ході величин рядів міняється - маємо до діла з к о р е л я ц і є ю к р и в о л і н і й н о ю.

67. Корелятивна й функціональна залежності. Само собою розуміється, що математичне формулювання й визначення таких залежностей може базуватися на принципах теорії функцій. Одначе, основним положенням функцій є, що за функціональної залежности разом з кількісною змінною однієї величини, обов'язково відбувається одна певна зміна другої величини, коли вони дійсно функціонально пов'язані.

Коли обсяг газу, на підставі закона Бойля-Маріота, в умовах однакової температури, є зворотньо-пропорційній величині тиснення, - себто, коли $v, v_1 \ldots$ є обсяги, а $p, p_1 \ldots$ тиснення, то $v:v_1 = p:p_1$, а відціль $pv = p_1 v_1 \ldots$ .... вираз цей є величина стала й щоб ця сталість збереглася, треба, щоб кожній зміні величини p відповідала одна певна зміна v.

Одначе далеко не завжди одно явище з такою строгою математичною послідовністю пов'язане з іншими. Себто принцип, що величини незалежної змінної або аргументу строго відповідають величинам залежної змінної або функції, який висловлюється формулою $y = f/x/$ не може бути строго додержаний в кореляційній методі. Цей основний розділ між функціональною залежністю й кореляційною залежністю в статистичних рядах залишається й стає важливим в тих випадках, коли переходимо від простих ондозначних функцій до дробових функцій, що порушують для певних вартостей аргументу /змінної/ однозначність і безперервність.

Таким чином між конструкцією функціональної залежности або зв'язку та кореляційним зв'язком в статистичних рядах є лише певна аналогія /подібність/, а не тотожність. За кореляційної залежности величині аргументу відповідає не одна лише певна величина функції, а пересічна аритметична величина функції, що виводиться з окремих величин, які можуть бути різноманітно розкидані біля своєї середньої, але індивідуально не збігаються з нею. Отже за кореляційної залежности окремій величині незалежної змінної може відповідати певний варіяційний ряд зв'язаних з нею величин, і через те в характері розташування окремих вартостей ряду навколо середньої та в послідовності цих величин відповідно хитанням величин незалежної змінної виявляється пошукувана залежність. Таким чином між двома квантитативними ознаками /змінними/ існує колективна[x] залежність тоді, коли при зміні числових вартостей однієї ознаки змінюється імовірності поодиноких вартостей другої ознаки. Користуючись відомим з попередньої лекції виразом "закон розподілу", що означає систему можливих вартостей ознаки та відповідних цим вартостям імовірностей, - можемо зазначити, що між двома квантитативними ознаками існує колективна залежність тоді, коли зі змінами вартостей однієї ознаки, змінюється закон розподілу другої ознаки, а зокрема найважливіша характеристика цього закону розподілу - математична надія. Такого роду залежність і називається кореляційною залежністю або кореляцією. Отже п о з и т и в н о ю кореляцією остання буде тоді, коли зі зростом величин однієї ознаки, стають імовірнішими високі вартості другої ознаки, або інакше кажучи, росте математична надія другої ознаки. При зворотньому стані будемо мати н е г а т и в н у кореляцію.

Нарешті відсутність кореляційної залежности маємо тоді, коли при зміні вартостей однієї ознаки залишається математична надія другої ознаки без зміни.

Раніш ми вказували на те. що статистика є матеріяльне числення, числа статистичні не є математичні символи, і статистика не є математика. Через те прикладання математичних конструкцій до статистиних даних мусить робитися з великою обережністю і на підставі строго логічної аналізи. Те що ми казали відносно прикладання конструкцій рахування імовірностей до статистичного матеріялу залишається не в меншій мірі й тут.

Одначе при складній структурі сучасного соціяльно-господарського життя не маємо можливости й вибору для розв'язання проблем господарських іти іншим шляхом, ніж залишаючи деталі, брати на увагу лише вартості середні, які певне господарське явище характеризують і репрезентують значно ліпше, ніж ряд деталей. Що середні вартості при рішенню проблеми господарських задовольняюче заступають ряд деталей, з'ясувалось і практично в багатьох випадках прикладення кореляційної методи, що дала в американських дослідженнях не мало важливих і цікавих результатів зокрема в дослідженнях господарської коньюнктури.

8. Кореляційна таблиця й кореляційний коефіцієнт. Для того, щоб відшукати кореляційний зв'язок між двома статистичними рядами, їх, звичайно, комбінують подібно комбінаційній таблиці в вигляді спеціяльної кореляційної таблиці; в цій таблиці поземим рядкам відповідають клясові інтервали, за якими розподілено величини однієї ознаки, а прямовісним рядкам клясові інтервали, за якими розподілено величини другої ознаки.

Схема цієї таблиці має наступний вигляд:/див. на слідуючій сторінці/.

Припустимо, що комбінуємо /шукаємо зв'язок/ вік нареченого й вік нареченої, обох у час шлюбу. Вік нареченого розподіляємо в ряд, наприклад, в п'ятилітніх інтервалах, або інших, менших, а кляси цієї змінної розміщуємо в таблиці, себто в напряму вісі ОХ , а розподіл віку нареченої по тих же самих інтервалах клясових в легенді, себто в напрямку вісі ОУ.

---

x/ Колективною залежністю, за справедливим завваженням С.Кона, вважаємо кореляційний зв'язок через те, що цю залежність чи зв'язок можна емпірично спостерегти лише на цілих сукупностях випадків, а не на окремих випадках.

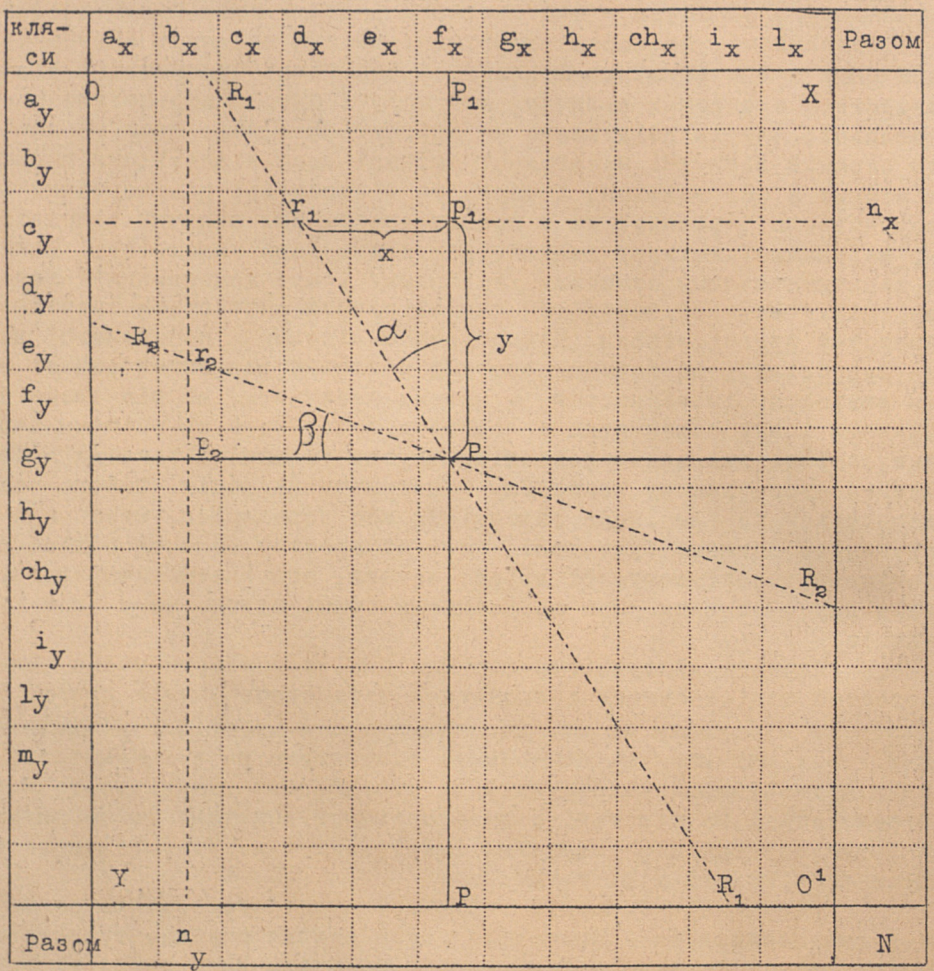
a_x, b_x, c_x - класові інтервали ряду в напрямку ОХ.

a_y, b_y, c_y - теж у напрямку ОУ.

n_x ... суми окремих рядків /в останньому прямовісному стовбці/.

n_y ...суми окр мих рядків /в останній поземій графі/.

N - сума цілого розподілу /сума величин останнього прямовісного стовбця й останньої поземої графи/.



В середині таблиці, у відповідних стовбцях і рядках що взаємно перехрещуються, містимо відповідні частоти /фреквенції/, себто число шлюбів з певним віком нареченого й нареченої. В останньому правому стовбці /разом/ даємо суми поземих рядків, себто суми шлюбів наречених відповідного вікового інтервалу, а в останній нижній графі таблиці даємо суми прямовісних стовбців /граф/, себто кількість шлюбів наречених відповідного вікового інтервалу. В правому нижньому розі таблиці даємо N , себто число всіх шлюбів взагалі, суму цілої фреквенції таблиці. Вік нареченого й нареченої є ті дві змінні величини /ознаки/, зв'язок яких спостерігаємо, математично з'ясовуємо й визначаємо коефіцієнтом кореляції /чи вірніше коваріації/. Повна кореляція в цій таблиці між двома змінними виявилася б в тому, що аритметична середня як змінної ряду поземого, так і змінної ряду прямовісного збіглися б в місці /точці/, де обидва ряди перетинаються, а лінія, що з'єднує місце перетину середніх з точками перетину ординат була б простою діягоналлю таблиці від лівого верхнього кута до правого долішнього. Ця кореляція означала б, що шлюб береться як правило /як вказує пересічна/ особами завжди певного віку. Коли ж аритметичні середні кож-

ного рядка, а подібно й кожного стовбця /графи/ збіглися б до того ж місця /інтервалу/ як і аритметичні середні відповідного цілого ряду /останього стовбця, або долішнього рядка/, тоді б ці середні місцево напросто розійшлися б, а прості, що їх з'єднують стали б одна до одної прямовісно.

Це б вказувало на те, що обидві спостережені змінні не мають нічого спільного, є між собою незалежні, а коваріяція їх /кореляція/ рівняється нулю.

Вищенаведена схема кореляційної таблиці графічно показує ясно випадок повної незалежности змінних X і У. Дійсно, у випадку повної незалежности будуть аритметичні середні кожного ряду знаходитися в тих же місцях /точках/ які аритметична середня в підсумковому рядку останньому і будуть лежати на простій $\overline{P_1P_1}$, подібно й перехічні прямовісних стовбців будуть лежати на простій $\overline{P_2P_2}$. Прості $\overline{P_1P_1}$ і $\overline{P_2P_2}$ стоять прямовісно одна до одної та перетинаються в точці P, що є спільною для поземого й прямовісного рядів та уявляє з себе перехічну цілого розподілу, цілої таблиці. Ця позиція є звичайно теоретично ідеальною.

В дійсності одначе розподіл частоти у поземому і в прямовісному рядах /змінних/ не є однаковим і пересічні роодиноких рядів не впадуть на прості $\overline{P_1P_1}$ та $\overline{P_2P_2}$, але більш-менш будуть відхилюватися від цих простих, склад давчі самостійні криві, які одначе можемо вирівняти в прості лінії. На поданій таблиці пересічні д і й с н и х р я д і в /поземих і прямовісних/ дають прості $\overline{R_1R_1}$ і $\overline{R_2R_2}$, а відхилення від теоретичних середніх P_1 і P_2 міряють нам кути $\measuredangle P_1PR_1 = \alpha$ і $\measuredangle P_2PR_2 = \beta$. Ці відхилення простих між позицією теоретичною /P_1 і P_2/ і дійсною /R_1 і R_2/ й є те, що визначає нам ступінь кореляції, що мусимо обчислити.

Прості $\overline{R_1R_1}$ та $\overline{R_2R_2}$ згідно розмірам зазначених відхилень /рядів/ можуть у межах прямокутників OP_1PP_2 і $PP_2O'P_1$ посідати безмежно велике число позицій. Чим більшою є сума відхилень від $\overline{P_1P_1}$ і $\overline{P_2P_2}$, себто сума кутів $\alpha+\beta$, тим більше $\overline{R_1R_1}$ та $\overline{R_2R_2}$ до себе наближуються, а ступінь кореляції між обома явищами, що заступають прості /або пересічні/ $\overline{R_1R_1}$ і $\overline{R_2R_2}$ є вищим. Це може йти так далеко, аж прості $\overline{R_1R_1}$ і $\overline{R_2R_2}$ одна з одною будуть збігатися. В цьому випадку сума відхилень буде максімум, себто $\alpha+\beta = 90^0$. З цього випливає, що коли сума відхилень /від теоретичних середніх/ рівняється нулю, є між обома явищами безперечна незалежність / $\overline{R_1R_1}$ впаде на $\overline{P_1P_1}$, а $\overline{R_2R_2}$ на $\overline{P_2P_2}$ /. Всі позиції, що можуть давати кутам α і β той або інший певний розмір визначають і більшу або меншу кореляцію.

Таким чином нахил лінії $\overline{R_1R_1}$ і $\overline{R_2R_2}$ міряється кутами α і β, або їх $tg\,\alpha$ і $tg\,\beta$ /тангенсами/. Ці лінії звуться лініями регресії.

Як же обчислити відхилення, що утворилися від розходження простих $\overline{R_1R_1}$ від $\overline{P_1P_1}$ і $\overline{R_2R_2}$ від $\overline{P_2P_2}$?

Завважимо, що P є геометричне місце /точка/ для середньої цілого розподілу обох змінних, воно є там, де перехрещуються середня P_1/середня з останнього поземого підсумкового рядку/ з середньою P_2 /середня з останнього підсумкового стовбця таблиці/.

Припустимо далі простіший випадок, коли середні поземих ліній падають точно на просту $\overline{R_1R_1}$. Хай буде P_2 середня вартість для всіх змінних величин у прямовісних стовбцях /середня з останнього підсумкового стовбця/, що визначена теоретично простою $\overline{P_2P_2}$. Проста $\overline{R_1R_1}$ перетинає просту $\overline{P_2P_2}$ в точці P. Тоді прямовісна проста в точці P мусить перетинати просту OX або УО в точці p_1, що означає середню для всіх змінних у поземих рядках /середня з останнього підсумкового стовбця/.

Відхилення простої $\overline{R_2R_2}$ від прямовісної простої P_1P_1, себто

$\frac{r_1p_1}{Pp_1} = \frac{x}{y} = tg\,\alpha$ означимо через b_1, а всі відхилення, себто віддаленість

любої точки на простій $\overline{R_1R_1}$ від простої $\overline{P_1P_1}$ - загально літерою x, а всі відхилення, себто віддаленість любої точки на простій $\overline{R_2R_2}$ від простої $\overline{P_2P_2}$- літерою y. Тоді для любого поземого рядка, в яких число спостере-

жень є m - сума відхилень всіх членів від P_1P_1 буде рівною $\sum/x/ = n.b_1.y$ /бо $x:y = tg\alpha = b_1$, відціль $x = b_1.y$/. Для всіх поземих рядків /для цілої таблиці/ буде сума відхилень від $\overline{P_1P_1} = \sum\sum/x/ = \sum n.b_1.y$ або $\sum\sum/x/ = b_1\sum/n.y$ /, понеже b_1 є константним. Через те, що $\sum/n.y/$ є сума всіх відхилень од середньої P_2 і є рівною, за відомим правилом про помилки від середньої - нулю і $\sum\sum/x/$ також є рівною нулю. Також P_1 - мусить бути середньою для всіх змінних величин у поземих рядках, і таким чином P, яко перетин задовольняючий обом напрямкам, буде середньою цілого розподілу /цілої дістрібуції/.

Коли знаємо, що $\overline{R_1R_1}$ дійсно проходить через точку P, залишається нам визначити b_1. Для цього візьмемо вираз $\sum/xy/ : N$, себто середню суми всіх пар з'єднаних відхилень x та y, при чому N є обсяг цілої сукупности, і цей вираз означимо через p. Вище ми встановили рівняння $\sum/x/ = nb_1y$, а коли помножемо це рівняння на y /на відхилення від P_2 /, дістанемо для кожного поземого рядка $y\sum/x/ = \sum/xy/ = y.n\,b_1.y = b_1ny^2$, а для цілої таблиці буде $\sum/xy/ = b_1\sum/ny^2/$. Ми одначе знаємо, що сума квадратів відхилень, поділена числом членів сукупности /тут N / є рівною квадрату середньої похибки, себто $\sum/ny^2/ : N = б^2$, або $\sum/ny^2/ = Nб_y^2$, а також $\sum/xy/ = b_1Nб_y^2$. Відціль $b_1б_y^2 = \sum/xy/ : N = p$; нарешті $b_1 = p : б_y^2$, себто b_1 визначимо, поділивши середню суми всіх попарно взятих відхилень x та y на квадрат штандартної помилки, обчисленої з підсумкового останнього ряду.

Ті ж самі міркування справедливі й для того випадку коли $\overline{k_2R_2}$ є проста, на яку в'яжуться середні стовбців прямовісних рядків і подібним способом можемо визначити також b_2, яко вираз, що міряє відхилення від поземої вісі $\overline{P_2P_2}$, себто $b_2 = p : б_x^2$, інакше кажучи, середня сум всіх попарно взятих y та x, поділена на квадрат штандартної похибки, обчисленої з підсумкової останньої графи таблиці /$б_x^2$/.

Геометрична середня величин b_1 та b_2 і є коефіцієнт кореляції, що означається літерою r, себто

$$\sqrt{b_1.b_2} = \sqrt{\frac{p}{б_x^2}\cdot\frac{p}{б_y^2}} = \frac{p}{б_x.б_y} = r$$

З цього випливає, що, підставивши замість p або $b_2б_y^2$, або $b_2б_x^2$ буде

$$r = \frac{b_1б_y}{б_x} = \frac{b_2б_x}{б_y}$$, або $b_1 = r.\frac{б_x}{б_y}$, та $b_2 = r.\frac{б_y}{б_x}$

Таким чином вираз для r є функцією b_1 та b_2 в тому розумінні, що коли ці дві змінні ознаки є абсолютно незалежні й рівняються нулю, тоді і r є рівним нулю. Знак позитивний чи негативний для r визначається відповідно середній добутків p, так що r є позитивним /+r/, коли великим вартостям x відповідають великі вартості y, коли ж великим вартостям x відповідають малі вартості y, тоді r є негативним. Само собою розуміється, що вартість r має границі $\pm$ I. Коли лінії регресій $\overline{R_1R_1}$ і $\overline{R_2R_2}$ зіллються, тоді $tg\beta$ дорівнюватиме $ctg\alpha$ і $tg\alpha$ дорівнюватиме $ctg\beta$, за тригонометрічною формулою $tg\alpha.ctg\beta = I$, а через те, беручи на увагу, що в випадку злиття ліній регресії $ctg\alpha = tg\beta$,- можемо підставленням одержати, що $tg\alpha.tg\beta = I$. Відціль випливає, що в ідеальному випадку коефіцієнт кореляції

$$r = \sqrt{tg\alpha.tg\beta} = I.$$

Коли $r = \pm 1$, це означає, що всі попарні варіяції знаходяться між собою в відношенні $x:y = б_x:б_y$. Ці властивості оправдуються в формулі $r = p:б_xб_y$ і це рівняння є виразом для ступеня кореляції між двома змінними, а r є дійсно коефіцієнтом коваріяції /кореляції/ двох змінних ознак. Коли замінимо величини

$$p = \frac{\sum/xy/}{N}, \quad б_x = \sqrt{\frac{\sum/x^2/}{N}}, \quad \text{а} \quad б_y = \sqrt{\frac{\sum/y^2/}{N}}$$

себто відхиленнями членів од середніх порівняних рядів /змінних/, одержимо для r ще таку формулу:

$$r = \frac{\frac{\sum|xy|}{N}}{\sqrt{\frac{\sum|x^2|}{N} \cdot \frac{\sum|y^2|}{N}}} = \frac{\sum|xy|}{\sqrt{\sum|x^2| \cdot \sum|y^2|}}$$

Отже, коли маємо два ряди, дві змінні й хочемо знайти, яку вартість має для них коефіцієнт кореляції, чинимо так: I/ Обчислимо аритметичні пересічні з обох рядів P_1 і P_2 ; 2/ Обчислимо відхилення всіх членів од відповідних середніх, себто x та y /завжди з відповідними знаками $\pm$/. 3/ Обчислимо квадрати відхилень / x^2 , y^2 /; 4/ Знайдемо добутки цих відхилень / x.y - з оглядом на знаки/; 5/ Квадрати відхилень і добутки підсумуємо $\left[\sum|x^2|, \sum|y^2|, \sum|xy|\right]$; 6/ Результати обчислень дамо до формули r .

Коли раніше ми відшукали дисперсію обох рядів згідно з стандартною похибкою, себто вже маємо обчисленими величини $б_x$ і $б_y$, тоді можемо скористатися для коефіцієнта кореляції з формули $r = \sum|xy| : N б_x б_y$, маючи на увазі, що N є число членів однакове в обох рядах.

69. Коефіцієнти регресії. Коефіцієнт кореляції визначає числово ступінь зв'язку або паралелізму, що може існувати між ходом двох або більше явищ. Встановити одначе відношення залежности ходу різних явищ між собою може рівняння регресії.

За допомогою них можемо розв'язати наступне питання: коли зміниться вартість A , яка їй відповідає зміна вартости B ? На це питання дадуть нам відповідь величини b_1 та b_2 , які ми вище визначили при встановленні коефіцієнта кореляції. Ці величини назвав Ґальтон коефіцієнтами регресії в своїому досліді дідичности фізичних розмірів людського тіла, в якому встановив, що в середньому сини по розмірах тіла відхиляються від батьків, а потім знов повертаються /відціль " regresse"/ до загального типу людности. Ми вже зазначили, що в наведеній вище кореляційній таблиці прості $\overline{R_1R_1}$ та $\overline{R_2R_2}$ є лініями регресії. Для аналітичних рівнянь цих простих, йдучих через середню кореляції P , яко через початок

$$x = r.\frac{б_x}{б_y}.y, \quad \text{а} \quad y = r.\frac{б_y}{б_x}.x,$$

дав Юл назву - х а р а к т е р и с т и ч н і р і в н я н н я. Головна функція цих рівнянь є в тому, щоб була можність при зміні величин одного ряду / x / відгадати п е р е с і ч н у зміну другого ряду / y / x/

70. Обчислення коефіцієнта кореляції. Для з'ясування способу визначення коефіцієнта кореляції можемо скористатися наступним конкретним матеріялом, поданим у Чубера / Die statistischen Forschungsmetoden /; відшукується коефіцієнт кореляції між двома ознаками рослини Trientalis europeas - товщиною стебла / x / та довжиною найдовшої пелюстки / y /, з метою поділу цих рослин в залежності від згаданих двох ознак. Дані зведемо в наступній кореляційній таблиці:/див. на слідуючій сторінці/.

Наведена таблиця скомбінована з додержанням важливого правила рівности інтервалів, що складають для товщини стебла 0,I мм., а для довжини пелюстки 6,0 мм., при чому для ряду довжини пелюстки взято не клясові інтервали з границями "від - до", а пересічні вартості інтервалів, що звуть часто типами відповідної будови ознаки.Вона дає уяву до якої міри розподіл вартостей однієї ознаки змінюється зі зміною вартостей другої ознаки, а також до якої міри і в якому напрямку йдуть найчастіші вартості ознаки y при збільшенню вартостей ознаки x. і зворотньо. Коли доповнимо таку таблицю ще обчисленням

---

x/ Для тих, хто не ознайомлений з основами вищого аналізу, рекомендую простудіювати невеликий, але добре складений курс доц.О.Коваленка: Основи вищої математики. Видання УТГІ.

Довжина пелюстка в міліметр.	Товщина стебла в мм.											Число випадків
	0,425	0,525	0,625	0,725	0,825	0,925	1,025	1,125	1,225	1,325	1,425	
10,5	I XVI		A									I
16,5	I XII	4 IX	I VI	I III								7
22,5	I VIII	9 VI	16 IV	3 II	I 0			B				30
28,5		2 III	9 II	22 I	9 0	2 I	I II					45
34,5			8 0	19 0	20 0	4 0	I 0					52
40,5	I IV			7 I	18 0	12 I	6 II	4 III				48
46,5				I II	8 0	9 II	3 IV	2 VI	I VIII			24
52,5						3 III	6 VI	4 IX	I XII	D		14
58,5							2 VIII	2 XII	I XVI	2 XX		7
64,5			C						I XX	3 XXV		4
70,5									I XXIV		I XXXV	2
Число всіх випадків	4	15	34	53	56	30	19	12	5	5	I	234

аритметичних середніх з вартостей ознаки У для кожного інтервалу ознаки Х та аритметичних середніх з вартостей ознаки Х для кожного інтервалу ознаки У, тоді в таблиці знайде свій аритметичний вираз і важливіша характеристика розподілу однієї ознаки для кожного інтервалу другої ознаки, себто математична надія У. Зміна цієї величини при змінах вартостей ознаки Х є для нас основним критерієм характеру чи ступеня кореляційної залежности.

Для того, щоб могли ясніше уявити собі кореляцію в даній таблиці можемо зобразити таблицю графічно та накреслити й лінії регресії /див! на слідуючій сторінці/.

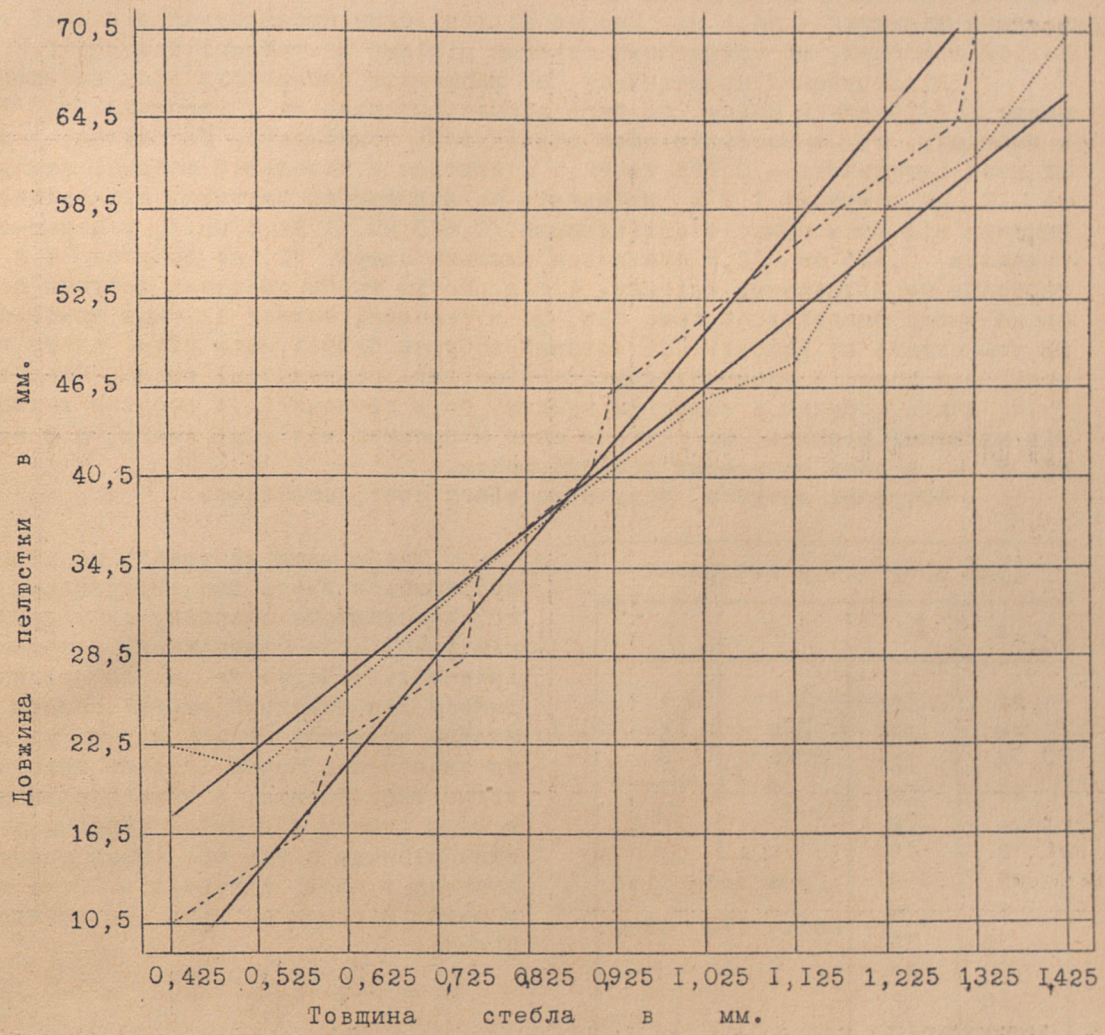
На подібному рисунку наносяться на вісі абсциси вартості або інтервали однієї ознаки,/в залежності від того чи є змінна тягла чи переривна/, а на вісі ординати вартості або інтервали другої. Серединами інтервалів ведемо поземі й прямовісні рівнобіжні, а на перетинках їх означаємо вартості аритметичних середніх з вартостей Х для поодиноких клясів по У та аритметичні середні з вартостей У для поодиноких клясів по Х. Далі ведемо "теоретичну лінію регресії" У по Х так, щоб наближче дотикалася пунктів тів, що означають аритметичні середні з вартостей У , а також і лінії регресії Х по У так, щоб найближче досягала пунктів аритметичних середніх з вартостей Х.

Дальший спосіб обчислення ведеться так:

З даних в таблиці чисел середня товщина стебла /Мх'/ усіх 234 рос-

Кореляція між трубостю стебла та довжиною найбільшої пелюстки

Рисунок



лин буде рівної 0,814 мм., штандартне відхилення /середне квадратове/ $б_x$ = 1,894 інтервалів = 0,189 мм./середня величина інтервалу 0,1 мм./; середня довжина пелюстки /M_y/ = 36,244 мм.; $б_y$ = 1,825 інтервалу = 10,95 мм. /величина інтервалу 6 мм./. Необхідно ще обчислити $\sum xy$, що робиться так:

Середню величину кожного інтервалу зіставляється з аритметичною середньою цілої кількости випадків, - зіставляємо, наприклад, 0,425 мм. із середньою аритметичною 0,814 мм., тут знак відхилення буде від'ємний і величина відхилення буде рівною 0,425 - 0,814 = -0,389; так само відхилення по другій змінній /у/ буде рівним 10,5 - 36,244 = -25,744. Помноживши абсолютні величини відхилень однієї ознаки на абсолютні величини відхилень другої, одержані добутки треба помножити на кількість випадків, що припадають на кожну клітину. Повторюючи такі дії до кожної звповненої клітини та складаючи одержані добутки, одержимо пошукувану величину $\sum xy$.

Як бачимо, рахункова операція займає при такому обчисленні надзвичайно багато часу.

Для полегшення обчислення користуються іншим способом. Для визначення величин відхилення x та y береться за початок відлічування довільну основу /а не середню/, що сходиться з середньою якогось інтервалу і від неї вже обчислюють відхилення, а потім роблять відповідне виправлення в розмірах відхилення, що сталося в наслідок обліку від довільно взятої основи. Найліпше

за довільну основу брати середину такого інтервалу, який найбільше наближується до середньої /М/. Для даних наведеної таблиці /двох змінних/ треба взяти за довільну основу два пункти - для грубости стебла - 0,825 мм., а для довжини пелюстки - 34,5 мм. Процес відлічування провадиться від тієї хрестоподібної фігури, що обкреслена грубими лініями на таблиці кореляції.

Відлічування провадиться по напрямках інтервалів обох змінних, при цьому відлічується місце, що його займає інтервал як у напрямку у, так і в напрямку х, а наслідки обох відлічувань помножують. Наприклад, у перетині двох інтервалів - 0,725 та 22,5 ставиться в наведеній таблиці римська цифра II як добуток I x 2 інтервалів по відношенні клітини, яка вміщає це означення від обох пунктів одлічування /0,825 мм. і 34,5 мм./, у перетині інтервалів 0,525 та 22,5 ставиться римська цифра УІ, як добуток 2 х 3 інтервалів по відношенню клітини, і т.д. Потім треба одержані добутки помножити на число випадків, а саме так, що в середині кожної клітини арабські цифри помножують на римські. Ці останні добутки будуть мати різні знаки в залежності від того, в якому квадраті розташовані помножувані числа. Квадранти А і Д дають добутки з додатнім знаком, бо в квадранті А добутки виникають від множення різниць, що мають в обох напрямках від'ємні знаки, а в квадранті Д - в обох напрямках додатні знаки.

Одержані добутки, беручи на увагу знак, складають:

Сума добутків у квадратах			
A	D	B	C
22	I2	- 2	- 7
I8	I2	- 2	- 4
6	I2	4	- 2
64	I8		I3
54	I2		
8	I2		
3	8	Сума добутків:	
6	9	68I - I7= +664	
36	36		
I2	36		
I6	I2		
25I	I6		
	24		
	I6		
	40		
	20		
	75		
	24		
	36		
25I + 430			
68I			

Треба мати на увазі, що обчислення провадиться ввесь час інтервалами, через що, розкладаючи одержану суму добутків між усіма 234 випадками, одержимо 2,8376 інтервалів. Через те, що обчислення провадилося від довільно взятої основи, пошукувана величина $p = \Sigma xy/:N$ не може бути визначеною точно і треба внести відповідне виправлення, а саме порівняти довільну основу з дійсною середньою і одержана різниця з тим або іншим знаком виявляється в числі інтервалів. Отже помилка в межах міряння по грубости стебла буде рівною:

$$K = \frac{0,8I4 - 0,825}{0,I} = -0,II$$ /інтервалів 0,I/;

помилка для довжини пелюстки -

$$l = \frac{36,244 - 34,5}{6} = 0,2906$$ інтервалів по шість мм.

Помилка добутку: $lm = -0,0320$.

Відіймаючи знайдену поправку від раніш знайденої величини одержимо:

$p = 2,8376 + 0,0320 = 2,8696$ інтервалів.

Отже, коли штандартні похибки /середні квадратові відхилення/ визначається у величинах інтервалів, коефіцієнт кореляції буде рівним:

$$r = \frac{2,8696}{I,894 \cdot I,825} = 0,830.$$ Кореляція позитивна й значна.

Одержана нами величина коефіцієнта кореляції, як величина емпірична, подібно іншим важливим характеристикам статистичних рядів має свою теоретичну похибку. Приблизне визначення середньої похибки коефіцієнту кореляції, як показав Пірсон, формулюється так:

$$б_r = \frac{1 - r^2}{\sqrt{N}}$$

а для середньої похибки коефіцієнту регресії формула буде такою:

$$б_{by/x/} = \frac{б_y}{б_x}\sqrt{\frac{1 - r^2}{N}}$$

$$б_{bx/y/} = \frac{б_x}{б_y}\sqrt{\frac{1 - r^2}{N}}$$

Ми ці формули лише зазначаємо, бо математичне обґрунтування їх базується на дуже тяжкій аналізі складних функцій.

формулу середньої похибки коефіцієнта кореляції до визначеної нами вище емпіричної величини цього коефіцієнта, знайдемо, що

$$б_r^{x/} = \frac{1 - /0,830/^2}{\sqrt{234}} = \frac{0,311}{15,3} = 0,020$$

Отже, в конечному вигляді коефіцієнт кореляції із знайденою поправкою визначиться так:

$$r = 0,830 \pm 0,020 = 0,810 \text{ або } 0,850$$

Межі хитань коефіцієнту кореляції показують оскільки знайдена емпірично величина може відступати в той або інший бік в залежності від причин чисто випадкових.

Лекція двадцять п'ята.

70. Колективний характер кореляційної залежности. В тих випадках, коли мислимо функціональну залежність між двома якимись явищами, чи ознаками, завше маємо, що за припущенням певної вартости ознаки x є можливою лише одна відповідна вартість ознаки y і таким чином всі спостережені випадки з певною вартістю x будуть геометрично з'ображені одною точкою на площі. Інакше буває тоді, коли залежність між квантитативними ознаками не має певного, однозначного характеру, але стохастичний (колективний) характер, що завше властивий кореляції. Тут кожній вартості ознаки x відповідає не одна, а цілий ряд можливих вартостей ознаки y і навпаки - кожній даній вартості ознаки y відповідає також не одна, а ряд вартостей ознаки x. Тут залежність характеризується тільки тим, що зі зміною вартостей ознаки x змінюється певним способом конструкція імовірности різних можливих вартостей ознаки y , а з'окрема й математичне очікування цієї ознаки, і навпаки, зі зміною вартостей ознаки y змінюється конструкція імовірности різних вартостей ознаки x , а з'окрема й математичне очікування цієї ознаки.

Таким чином у випадку однозначного функціонального зв'язку всі можливі конкретні перебіги явища, характеризовані даною вартістю ознаки x , є також характеризовані однією певною вартістю ознаки y , у випадку ж кореляційного зв'язку можуть бути характеризовані цілим рядом вартостей y ,

x) Інколи означається, як Sr.

а також при збільшенню або зменшенню вартости ознаки x частоти можливих конкретних перебігів явища, характеризованих різними вартостями ознаки y , будуть змінюватись; а з'окрема буде змінюватись і пересічна вартостей ознаки y в цих можливих конкретних перебігах, себто математичне очікування.

Те, що математичне очікування є аритметичною середньою з вартостей ознаки для всіх конкретних перебігів явища, випливає з наступного. Коли імовірність даної вартости ознаки y є відношенням числа конкретних перебігів явища, тхарактеризованих даною вартістю y , до цілої кількости конкретних перебігів явища, скажемо $\frac{v_i}{v}$, а математичне очікування є пересічною з вартостей y , зваженою їх імовірностями, тоді:

$$E_{[y]} = y_1\frac{v_1}{v} + y_2\frac{v_2}{v} + y_3\frac{v_3}{v} + \ldots\ldots\ldots$$

це й є аритметична середня з вартостей y для всіх конкретних перебігів явища (тут $v = v_1 + v_2 + \ldots\ldots$]

Вищенаведене означає також, що специфікуючи можливі конкретні перебіги явища поступово після різних (наприклад зростаючих) вартостей ознаки x , одержимо в таких групах можливих перебігів явища все більші пересічні вартості ознаки y - для випадку прямої (позитивної) кореляції й все менші пересічні вартості y - для випадку непрямої (негативної) кореляції.

Таким чином корелятивний зв'язок завше носить стохастичний характер і теорія імовірностей набирає великої ваги для освітлення такого роду зв'язків.

71. Конструкція імовірности та коефіцієнти регресії і кореляції.

Ми вже зазначили вище (стор. 215), що відношення залежности ходу різних явищ між собою встановлюється через рівняння регресії та коефіцієнта регресії. Себто, за допомогою них можемо розв'язувати питання: коли змінюється вартість x , яка їй відповідає зміна вартости y ? Само в собі це питання вже носить характер імовірности, а відповідь на нього можемо знайти в понятті математичного очікування (надії).

Отже певна конструкція імовірности безпосередньо може й мусить стосуватись до визначення ходу залежности різних явищ між собою. Ми обмежуємось лише простішою формою такої залежности - лінійною, лінійною кореляцією.

Повторимо тут, що під корелятивним зв'язком (стохастичним, колективним зв'язком) ми розуміємо такий зв'язок, між явищами, коли зі зміною вартостей однієї ознаки, змінюється закон розподілу вартостей другої ознаки, а разом з тим важливіша характеристика цього закону розподілу - математичне очікування. На стор. 217 ми подали також графічно образ лінійної регресії у конкретному випадку відшукання кореляції між довжиною пелюстки та товщиною стебла рослини Trientalis europeas.

Отже уявім собі, що точки, які означають обумовлені математичні очікування +) ознаки y для різних вартостей ознаки x створюють просту лінію - тоді ця проста може бути аналітично визначена рівнянням:

$$E^{[i]}[y] = a + bx_i \qquad (1)$$

Для певного опису цієї простої досить відшукати коефіцієнти рівняння (1) "a" та "b".

Щоб знайти зручну альгебричну форму для цих коефіцієнтів, буде більш доцільним дати рівнянню регресної простої інший вираз, а саме замінити в лівій частині рівняння (1) обумовлене математичне очікування y - відхиленням цього математичного очікування від звичайного(необумовленого) математичного очікування. ознаки y , себто $E[y]$, а в правій частині рівняння замінити вартість x_i , відхиленням цієї ж вартости від звичайного математичного очіку-

+) Ці обумовлені математичні очікування означимо $E^{[i]}[y]$

вання x , себто E[x] , тоді рівняння прийме такий вигляд:

$$E^{[i]}[y] - E[y] = a + b\left[x_i - E[x]\right] \qquad (2)$$

Тут відхилення обумовленого математичного очікування ознаки y від звичайного математичного очікування ознаки y є визначено, яко функція відхилення вартости ознаки x від її математичного очікування. Щоб було можливим коефіцієнти a й b визначити за помічч вартостей ознак x та y , помножимо рівняння на звичайну імовірність вартости "i" ознаки x , себто P_i , тоді одержимо:

$$P_{i/}\left[E^{[i]}[y] - E[y]\right] = P_{i/}\,a + P_{i/}\,b\left[x_i - E[x]\right]$$

Підсумувавши дані після "i" будемо мати:

$$\sum_i P_{i/}\left[E^{[i]}[y] - E[y]\right] = a\sum_i P_{i/} + b\sum_i P_{i/}\left[x_i - E[x]\right] \qquad (3)$$

Розглянємо ліву частину рівняння. В ній

$$E^{[i]}[y] = \sum_j P^{[i]}_{/j} \cdot y_j$$

маємо:

$$\sum_i P_{i/}\left[E^{[i]}[y] - E[y]\right] = \sum_i P_{i/}\left[\sum_j P^{[i]}_{/j}.y_j - E[y]\right]$$

Понеже далі можемо цілком управлено писати E[y] яко $\sum_j P^{[i]}_{/j}.\,E[y]$ бо $\sum_j P^{[i]}_{/j} = 1$, то одержимо:

$$\sum_i P_{i/}\left[E^{[i]}[y] - E[y]\right] = \sum_i P_{i/}\left[\sum_j P^{[i]}_{j}.y_j - \sum_j P^{[i]}_{/j}.E[y]\right]$$

$$= \sum_i P_{i/}\sum_j P^{[i]}_{/j}\left[y_j - E[y]\right] = \sum_i \sum_j P_{i/}\,.\,P^{[i]}_{j}\left[y_j - E[y]\right] \quad x]$$

Тому, що $P_{i/}\,.\,P^{[i]}_{j} = P_{i/j}$, то ми можемо написати:

$$\sum_i P_i\left[E^{[i]}[y] - E[y]\right] = \sum_i \sum_j P_{i/j}\left[y - E[y]\right]$$

Коли підрахуємо в правій частині рівняння після "i" для кожного - одержимо, яко перший сочинник, суму імовірностей:

$P_{1/j} + P_{2/j} + \ldots\ldots P_{k/j}$, що рівняється $P_{/j}$.

Дальшим підсумованням після j одержимо:

$$\sum P_{/j}\left[y_j - E[y]\right]$$

що є рівним нулі, бо математичне очікування відхилення від математичного очікування = 0. Тоді

$$\sum_i P_{i/}\left[E^{[i]}[y] - E[y]\right] = 0$$

Через те саме в другому члені правої частини рівняння (3)

$\sum_i P_{i/}\left[x_i - E[x]\right]$ також = 0, а тому, що $\sum_i P_{i/} = 1$

то з рівняння (3) випливає, що a = 0.

Себто при обраному нами регресному рівнянню мусить коефіцієнт "a" в рівнянню регресної простої рівнятись нулю.

+) Для ясности завважимо, що помноження суми всіх P_i (себто $_iP_i$) на суму всіх $P^{[i]}_j$ (себто $\sum_j P^{[i]}_j$) означає помноження кожного P_i кожним $P^{[i]}_j$; а сума цих добутків $P_i\,.P^{[i]}_j$ для всіх комбінацій вартостей "i" та "j", це й визначено символом подвійної суми $\sum\sum$ у виразі $\sum_i \sum_j P_i.\,P^{[i]}_j$.

Помноживши далі рівняння друге на $P_{i/}\left[x_i - E[x]\right]$ та підсумувавши після "i", одержимо:

$$\sum_i P_{i/}\left[x_i - E[x]\right]\left[E^{[i]}[y] - E[y]\right] = a\sum_i P_{i/}\left[x_i - E[x]\right] + b\sum_i P_{i/}\left[x_i - E[x]\right]^2 \qquad (4)$$

Як вище можемо вираз $\left[E^{[i]}[y] - E[y]\right]$ в лівій частині написати у формі:

$$\sum_j P^{[i]}_{/j}\left[y_j - E[y]\right] \quad , \text{ тоді } (\text{бо } P_{i/}.P^{[i]}_{/j} = P_{i/j})$$

набуде ліва сторона рівняння (4) форми:

$$\sum_i \sum_j P_{i/j}\left[x_i - E[x]\right]\left[y - E[y]\right] = E\left\{\left[x_i - E[x]\right]\left[y_j - E[y]\right]\right\} \qquad (5)$$

Вираз $E\left\{\left[x_i - E[x]\right]\left[y_i - E[y]\right]\right\}$, себто математичне очікування добутку відхилень x та y від їх математичних очікувань, має важливе значіння в теорії кореляції, його можемо спеціяльно означити як $\mu_{x,y}$.

Перший член правої частини рівняння (4) рівняється нулю. В другому члені $\sum_i P_{i/}\left[x_i - E[x]\right]^2$ є квадрат середнього відхилення (похибки) x, себто σ^2_x

Отже $$b = \frac{\mu_{x,y}}{\sigma^2_x} \qquad (6)$$

Це й є коефіцієнт регресії, який визначає просту регресії ознаки y після ознаки x та часто означається як $b_{y[x]}$.

З виразу регресійного рівняння

$$E^{[i]}[y] - E[y] = b_{y[x]}\left[x_i - E[x]\right]$$

в який можемо перетворити рівняння (2), випливає, що

$$b_{y[x]} = \frac{E^{[i]}[y] - E[y]}{x_i - E[x]}$$

З цього бачимо, що коефіцієнт регресії $b_{y[x]}$ є відношенням відхилення обумовленого математичного очікування ознаки y від її цілого математичного очікування - до відповідного відхилення вартости ознаки x від її математичного очікування.

Він вказує таким чином, на яку величину відхилиться обумовлене математичне очікування y від звичайного (необумовленого) математичного очікування y, коли вартість x відхилиться від свого звичайного математичного очікування на певну величину. Це значіння його випливає ясно, як було наведено в кореляційній таблиці та її розгляді (на стор.212 - 214), також з того, що в $b_{y[x]}$ геометрично є тангенсом кута, що створює регресна проста з віссю координат.

Тим самим способом, виходячи з другого регресного рівняння (x після y):

$$E^{[i]}[x] - E[x] = a + b\left[y_j - E[y]\right]$$

знайдемо, що в ньому $a = 0$ й що коефіцієнт

$$b_{x[y]} = \frac{\mu_{x,y}}{\sigma^2_y} \qquad (7)$$

Цей коефіцієнт регресії ознаки x після ознаки y цілком визначає відповідну другу регресну просту, і, як бачимо, не є завжди рівний першому коефіцієнту $b_{y[x]}$. Чисельника має з ним спільного, але в знаменнику різ-

в ньому
ниться, бо має квадрат середнього відхилення ознаки y . З цього випливає, що коефіцієнти регресії x після y та y після x є різні величини, за винятком випадку, коли середні відхилення обох ознак є однакові.

Коефіцієнти регресії $b_{y[x]}$ міряє залежність поки вона виявляється в змінах вартостей ознаки y виникших змінами вартостей ознаки x , а коефіцієнт регресії $b_{x[y]}$ залежність, що виявляється в змінах вартостей ознаки x при змінах ознаки y . Є важливим, одначе знайти якусь однозначну міру ступеня залежности, об'єднуючу як регресію y за ознакою x так й x за ознакою y .

Такою однозначною мірою залежности й буде геометрична середня з обох коефіцієнтів регресії; а саме:

$$r = \sqrt{\frac{\mu^2_{x,y}}{\sigma^2_x . \sigma^2_y}} = \frac{\mu_{x,y}}{\sigma_x . \sigma_y} \qquad (8)$$

Це й буде відомий коефіцієнт кореляції рівний, як бачимо, математичному очікуванню добутку відхилень x та y від їх математичних очікувань, поділеному добутком середніх похибок (штандартних відхилень) ознак x та y.

ожний коефіцієнт регресії може бути визначений через коефіцієнт кореляції, а саме:

$$b_{y[x]} = r \frac{\sigma_y}{\sigma_x} \quad ; \quad b_{x[y]} = r \frac{\sigma_x}{\sigma_y} \qquad (9)$$

Коли ознака y не корелюється з ознакою x , себто, коли всі обумовлені математичні очікування $E^{[i]}[y]$ суть рівні звичайному математичному очікуванню $E[y]$, тоді, і лише тоді, буде коефіцієнт лінійної кореляції $b_{y[x]}$ рівним нулю. Рівнож коефіцієнт $b_{x[y]}$ буде рівнятись нулю, коли ознака x не скорельована з ознакою y і коли всі обумовлені математичні очікування $E^{[i]}[x]$ рівняються $E[x]$.

Коли хоч один коефіцієнт регресії є рівним нулю, буде рівнятись нулю й коефіцієнт кореляції, що є їх геометричною середньою. Коли одначе коефіцієнт кореляції рівняється нулю, то, як випливає з виразу (9), буде й другий коефіцієнт регресії рівним нулю. Через те при лінійній кореляції може бути наявність або відсутність кореляційного зв'язку між ознаками лише взаємна.

Себто, коли ознака x корельована з ознакою y , також і ознака y обов'язково корелюється з ознакою x.

Одначе, коли не припускаємо лінійної кореляції, це правило не буде дійсним. Але тут ми обмежуємось лише характеристикою лінійної кореляції.

Що коефіцієнт кореляції у випадку відсутности корелятивного зв'язку є рівним нулю, випливає цілком із самого виразу (8).

Вираз $\mu_{x,y} = E\{[x - E[x]][y - E[y]]\}$
себто математичне очікування добутку здружених відхилень обох ознак від їх математичних очікувань мусить рівнятись нулю, коли певна тенденція до солідарности (здруження) додатніх відхилень x з додатніми відхиленнями y відсутня, або, навпаки, додатніх відхилень x з від'ємними (негативними) відхиленнями y . Коли ж така тенденція здруження додатніх відхилень x з додатніми відхиленнями y (себто високих вартостей x з високими вартостями y) існує, це вказує на позитивну кореляцію і тоді $\mu_{x,y}$ а разом з тим і коефіцієнт кореляції буде мати додатню величину й то тим більшу, чим частіше в можливих конкретних перебігах явища будуть зустрічатись відхилення з тим самим знаком та взаємно близьких величин. Для коефіцієнтів регресії не може бути даною якась верхня границя такого здруження або розходження, але коефіцієнт кореляції її має. Ця границя й буде від 0 до ±1, при тому завважимо, що вартості ±1 коефіцієнт кореляції може досягати лише у випадку однозначної

функціональної залежности та коли разом з тим кореляція є лінійною, в емпіричному ж матеріялі величина коефіцієнта кореляції може лише більш менш наближатись до своєї верхньої границі.

Завважимо ще, що коефіцієнт кореляції можна визначити також в наступній формі:

$$r = \frac{E[xy] - E[x].E[y]}{\sqrt{\left\{E[x^2] - \left[E[x]\right]^2\right\}\left\{E[y^2] - \left[E[y]\right]^2\right\}}} \qquad (10)$$

Дійсно: $\mu_{x,y} = E\left\{\left[x-E[x]\right]\left[y-E[y]\right]\right\} = E\left[xy-yE[x]-xE[y]+E[x].E[y]\right] =$

$= E[xy]- E[x]E[y]- E[x]E[y]+ E[x]E[y] = E[xy]- E[x]E[y].$

Що ж до знаменника в формулі(10), то квадрат штандартного відхилення можемо писати, яко математичне очікування квадрата ознаки мінус квадрат математичного очікування ознаки.

72. Кореляційне відношення. Для встановлення зв язку між квантитативними ознаками має велике значіння крім коефіцієнта кореляції ще й так звана величина кореляційного відношення, що означається $\eta_{y[x]}$.

Коли нам ходить про те, як великі зміни вартостей однієї ознаки в середньому спричинились через зміни вартостей ознаки другої, то можемо прямо зміряти середню великість таких змін, або відхилень обумовлених математичних очікувань у від звичайного (необумовленого) математичного очікування цієї ознаки. Інакше кажучи, можемо визначити середню похибку обумовлених математичних очікувань у біля звичайного математичного очікування у , себто:

$$\sqrt{\sum P_{i/}\left[E^{[i]}[y] - E[y]\right]^2}$$

Коли хочемо елімінувати тут загальну варіяцію ознаки у , тоді ділимо вищенаведену величину на загальну середню похибку ознаки у , одержуючи таким чином відносну міру змін ознаки у , а саме:

$$\frac{\sqrt{\sum P_{i/}\left[E^{[i]}[y] - E[y]\right]^2}}{б_y}$$

Ця величина, що часто вживається і в іншій формі:

$$\frac{\sum P_i\left[E^{[i]}[y] - E[y]\right]^2}{б_y^2} \qquad (11)$$

й зветься в теорії статистики кореляційним відношенням.

У випадкові лінійної регресії кореляційне відношення є рівним кореляційному коефіцієнту.

Візьмемо регресне рівняння:

$$\frac{E^{[i]}[y] - E[y]}{б_y} = r\,\frac{X_i - E[x]}{б_x}$$

піднісши його в квадрат та помноживши на $P_{i/}$, одержимо

$$P_{i/}\,\frac{\left[E^{[i]}[y] - E[y]\right]^2}{б^2_y} = r^2\,P_i\,\frac{\left[x_i - E[x]\right]^2}{б^2_x}$$

Підсумуємо після "i" і одержимо:

$$\frac{\sum_i P_i \left[E^{[i]}[y] - E[y]\right]^2}{\sigma^2 y} = r^2 \frac{\sum_i P_i \left[x_i - E[x]\right]^2}{\sigma^2 x}$$

В лівій частині рівняння маємо квадрат кореляційного відношення $\eta^2 y[x]$. Тому, що

$$\frac{\sum_i P_i \left[x_i - E[x]\right]^2}{\sigma^2 x} = \frac{\sigma^2 x}{\sigma^2 x} = 1$$

маємо, що $\eta^2 y[x] = r^2$. Треба мати на увазі, що такого зв'язку немає в випадку нелінійної (криволінійної) кореляції.

Одначе можуть бути доведені наступні два важливі правила:

1. Коли між y та x немає ніякого зв'язку, то кореляційне відношення ηy рівняється нулю, а коли y зв'язано з x точним зв язком, то ηy є рівним одиниці.

2. Коли $\eta y = 0$, то y ніяк не зв'язано з x та коли $\eta y = 1$, то y зв'язано з x точним зв язком.

Ці правила встановлюють значення величини ηy як міри зв'язку між y та x незалежно від того чи існує між ними зв'язок лінійний чи нелінійний.

Те ж саме можна довести і для ηx. себто вважавчи y незалежною величиною, а x залежною від y.

Таким чином кореляційні відношення ηx та ηy мають важливе значення для оцінки тісноти зв язку між x та y, коли такий зв'язок існує. Отже потрібно оцінити також і надійність (певність) величин кореляційних відношень одержаних для якогось конкретного випадку. Ця оцінка досягається через визначення середніх квадратових відхилень цих відношень, які формулюються так: (даємо їх без доведення).

$$\sigma[\eta y] = \frac{1 - \eta^2 y}{\sqrt{N}}, \qquad \sigma[\eta x] = \frac{1 - \eta^2 x}{\sqrt{N}} \qquad (12)$$

В своєму викладі ми обмежуємось лише поняттям елементарної кореляції.

Нелінійна регресія ознаки y після ознак x або навпаки, означає, що регресійна лінія не має форму простої, але кривої, або параболи, для якої рівняння регресії має форму

$$E^{[i]}[y] = a + bx_i + cx^2_i + dx^3_i + \ldots\ldots$$

Для повного опису таких регресій не досить, як бачимо, двох коефіцієнтів, а цілий ряд їх, в залежності від того, якого порядку є парабола.

З аналізом таких залежностей, і тим студентам нашого Інституту, які вважають себе досить підготовленими ми рекомендуємо ознайомитись по спеціяльних підручниках теорії статистики, що їх зазначаємо в кінці нашого курсу.

СТАТИСТИКА АТРИБУТІВ (ЯКІСНИХ АРҐУМЕНТІВ).

73. Клясифікація.

Означимо через A, B, C ....M якісь прикмети за якими розподілена певна статистична сукупність S, при тому ці прикмети є якісними, а протилежні цим прикметам ознаки назвемо a, b, c,.....m.

Наприклад населення може розподілитись нп чоловіків A і жінок- a на грамотних - B і неграмотних b на міщан - C і селян - c і т. под.

Символами AB, A b, Ba .....Abc ми означимо групи членів сукупности S , яким властиві одночасно декілька з прикмет A, B, a, b,..... і т.д.

Тоді можемо сказати, що S розподіляється на кляси : AB, Ab, Ba ..... Кляси, що складаються з додатніх прикмет, назвемо додатніми, а кляси до яких входять прикмети a, b, c....m - назвемо відємними.

Обсягом кляси будемо звати число членів сукупности, що до кляси належать, та означати його однаковими літерами, як і кляси, тільки в дужках (A), (AB) ..... (abc)..... Обсяги кляс уявляють з себе ніщо інше як частоти відповідних ознак в сукупності S.

Кількість прикмет кожного члена кляси одночасно, будемо називати порядком кляси, так що A, B, a, b, ... є кляса 1-го порядку, а AB, A b, Bc ..... - 2-го порядку, ABC, A bc ..... 3-го порядку.

Правило I. Обсяг любої кляси може бути визначений, яко сума обсягів кляс вищого порядку, на які розподіляється дана кляса.

Візьмемо наприклад клясу AB. Кожний член цієї кляси або має прикмету C, або її не має. З цього випливає, що (AB) = (ABC) = (ABc). Подібно до цього (ABC) = (ABC D) + (ABCd); (ABc) =[ABc D) + (ABc d).

Таким чином (AB) = (ABC D) + (ABC d) + (ABc D) + (ABc d) і правило цілком ясне.

Для повного перечислення обсягів всіх можливих клясів якої - будь сукупности досить знати обсяги її клясів найвищого порядку. Наприклад, маємо за переписом населення повіту з таким розподілом:

Населення	грамотних(B)		негрампотних (b)	
	чоловіків	жінок	чоловіків	жінок
	(A)	(a)	(A)	(a)
Тубольці (C)	3391	2637	3659	4631
Нетубольці (c)	7650	1095	169559	158939

В таблиці подано обсяги всіх кляс 3-го порядку, що можуть бути збудовані за трьома ознаками A, B, C та протилежними a, b, c. По них легко знайти обсяги всіх кляс нижчих порядків:

$$(AB) = (ABC) + (ABc) = 3391 + 7650 = 11041.$$

$$(A\,b) = (A\,bC) + (A\,bc) = 3659 + 169559 = 173218.$$

$$A = (AB) + (A\,b) = 11041 + 173218 = 184259.$$

$$N = (ABC) + (aBC) + (A\,bC) + (a\,bC) + (ABc) + (aBc) + (A\,bc) + (a\,bc) =$$

$$= 351561.$$

Правило II. Обсяг любої кляси кожної сукупности можна визначити через обсяги її додатніх кляс.

Мали вже очевидну рівність (AB) = (ABC) + (ABc), відціль (ABc) = = (AB) - (ABC); подібно (ABc d) = (ABc) - (ABc D) та (ABc D) = (AB D) - - (ABC D); через це (ABc d) = (AB) - (ABC) - (AB D) + (ABC D).

Для повного перечислення обсягів всіх кляс кожної сукупности досить знати обсяги всіх її додатніх кляс. Для сукупности , вказаної вище в таблиці ,обсяги додатніх кляс будуть такі:

$$N = 351561.$$

$$(A) = 184259 \; ; \; (B) = 14773 \; ; \; (C) = 14318$$

$$(AB) = 11041 \; ; \; (AC) = 7050 \; ; \; (BC) = 6028$$

$$(ABC) = 3391.$$

Не тяжко далі знайти й обсяги решти кляс, а саме:

$$(ABC) = (AB) - (ABC) = 11041 - 3391 = 7650$$

$$(A\,\bar{b}C) = (AC) - (ABC) = 7050 - 3391 = 3659$$

$$(A\,\bar{b}) = (A) - (AB) = 184259 - 11041 = 173218$$

$$(a\,\bar{b}) = (a) - (aB) = N - (A) - (B) + (AB) = 163570 \text{ і т.д.}$$

АСОЦІЯЦІЯ ЯКІСНИХ ОЗНАК.

74. Поняття асоціяції. Асоціяція або залежність, дисоціяція, або незалежність між якісними ознаками в статистичних сукупностях характеризується тим, що при наявности ознаки, або явища A змінюється імовірність наявности ознаки або явища B. Коли наявність ознаки A збільшує імовірність ознаки B, то таку залежність (зв'язок) між цими ознаками вважаємо за позитивний (прямий), коли ж імовірність ознаки B зменшується, то такий зв'язок вважаємо за негативний (непрямий). Нарешті, коли імовірність однієї ознаки не змінюється в зв'язку з співіснуванням другої вважаємо, що залежности або зв язку між ознаками не існує. Ця асоціяція так само носить стохастичний характер й у випадках залежностей між квантитативними (кількосними) ознаками суть кого ми вже з'ясували вище.

Для зручности й ясности в дальших висновках і твердженнях мусимо нашу загальну дефініцію асоціяції символізувати. Отже маємо дві ознаки - A і B . Імовірність якоїсь ознаки A в загальному явищі, одиниці якого складають статистичну сукупність, означимо через P_A, а імовірність ознаки B - через P_B; імовірність відсутности ознаки A в тій самій сукупности означимо через P_a та відсутність ознаки B через P_b .

Імовірність появлення ознаки B за припущенням існування ознаки A, себто обумовлену імовірність ознаки B, будемо означати $P_B^{(A)}$. Обумовлену імовірність ознаки B за відсутністю співіснування A, означимо $P_B^{(a)}$.

Формулюємо далі колективний зв язок або відсутність його між ознаками A і B на підставі нашої символізації, тоді будемо мати:

1. $P_B^{[A]} > P_B^{[a]}$ себто обумовлена імовірність ознаки B за припущенням наявности ознаки A є більшою, ніж при відсутності існування цієї ознаки - залежність пряма позитивна

2. $P_B^{[A]} < P_B^{[a]}$ залежність між ознаками негативна.

3. $P_B^{[A]} = P_B^{[a]}$ залежности не існує: A і B незалежні між собою.

Формулювати це можемо і в іншій формі, коли завважимо, що залежність між A і B існує тоді, коли імовірність B за припущенням існування A є більшою, або навпаки - меншою, ніж імовірність B, узгляднена без уваги на те, чи існує чи не існує ознака A. Себто ознака B є незалежною від A тоді, коли імовірність B при умові існування A є рівною своїй звичайній імовірности, встановленій без огляду на наявність A. Цю останню імовірність означаємо як P_B і тоді нове формулювання прийме наступні форми:

4. $P_B^{[A]} > P_B$ залежність позитивна.

5. $P_B^{[A]} < P_B$ залежність негативна.

6) $P_B^{[A]} = P_B$ незалежність.

що останні три форми зв'язку є лише іншою формою перших трьох, видно з наступного:

Звичайна імовірність появлення ознаки B буде P_B. Ознака B може з явитись або разом з ознакою A, або без неї. Імовірність першої можливости, за теоремою множення імовірностей залежних подій або явищ, буде рівною $P_A.P_B^{(A)}$; імовірність другої можливости $P_a.\ P_B^{(a)}$, а за теоремою додавання імовірностей:

$$P_B = P_A P_B^{[A]} + P_a.P_B^{[a]} \qquad (7)$$

У випадку $P_B^{(A)} = P_B^{(a)}$, формула (7) набере форми

$$P_B = P_B^{[A]}.[P_A + P_a]$$

Через те, що $P_A + P_a = 1$, бо можливості A і a вичерпують ціле явище, буде $P_B = P_B^{(A)}$.

Аналогічно можемо довести, що коли

$$P_B^{[A]} > P_B^{[a]} \quad \text{то й} \quad P_B^{[A]} > P_B$$

З рівняння (7) випливає також, що

$$P_B^{[a]} = \frac{P_B - P_A P_B^{[A]}}{P_a} \qquad (8)$$

А через те, коли маємо $P_B^{(A)} > P_B^{(a)}$, будемо мати, що

$$P_B^{[A]} > \frac{P_B - P_A P_B^{[A]}}{P_a}$$

Тому, що $P_A + P_a = 1$, одержимо, що $P_B^{(A)} > P_B$. Відповідно можемо переконатись, що, коли $P_B^{(A)} < P_B^{(a)}$ то й $P_B^{(A)} < P_B$.

Нарешті сама теорема множення імовірностей незалежних між собою явищ дає нам окрему форму зв'язку або його відсутности. Вона каже: імовірність співіснування (одночасної появи) двох незалежних явищ є рівною добутку їх імовірностей. Отже, коли імовірність співіснування A і B означаємо $P_{A,B}$, тоді випадок незалежности цих ознак формулюється як

$$P_{A,B} = P_A.\ P_B \qquad (9)$$

На підставі рівнянь (6) і (9) можемо стверднти, що

$$P_A.P_B^{[A]} = P_{A,B}$$

Так само від формули (4) перейдемо до взору:

$$P_{A,B} > P_A.\ P_B \qquad (10)$$

А від формули (5) до взору:

$$P_{A,B} < P_A.\ P_B \qquad (11)$$

З наведених формул можемо бачити, що залежність між ознаками може бути не тільки різного напрямку або знаку (позитивна, негативна) але й різного ступня, в залежності від того чи розділ між лівими і правими частинами нерівностей (1) і (2) або (4) і (5) або (10) і (11) є більшим чи меншим.

Щодо виявлення степеня або сили залежности заслуговують уваги формули (1) і (2), а також з формули (7) можна бачити, що коли імовірність P_A значно більша, ніж імовірність P_a - мусить імовірність P_B бути вельми близькою до імовірности $P_B^{(A)}$, незалежно від того, яка різниця буде між $P_B^{(A)}$ і $P_B^{(a)}$. Отже в такому випадку, коли імовірність здійснення явища A (себто P_A) є значно більшою, ніж P_a, або імовірність його нездійснення, тоді при вимірі сили залежности розділом між імовірністю явища B, за наявністю A та імовірністю явища B без огляду на наявність A, може вельми слабо виявитись і значна залежність між A і B.

Наприклад, шукаючи залежність між смертністю немовлят шлюбних і нешлюбних та порівнюючи імовірності смертности в першому році життя у тих і других, можемо знайти важливий розділ між цими імовірностями, що вказують на існування залежности (в даному випадку негативної).

Навпаки, порівнюючи імовірності смерти немовлят тільки шлюбних та всіх взагалі (себто обох категорій) не знайдемо замітної різниці, бо народження нешлюбної дитини є релятивно рідким випадком в цілій кількості народжень.

З цього погляду форми першого критерія заслуговують більшої уваги, ніж форми (10) і (11), що виведено з другого критерія простим помноженням P_A.

75. Емпіричні характеристики колективної залежности.

Вище, в своїх міркуваннях і характеристиках колективної залежности якісних подій чи ознак, ми виходили з апріорних імовірностей або характеристик, але в конкретній дійсности ці апріорні характеристики в більшості випадків не є можливими для безпосереднього спостереження дослідника. Не можемо, наприклад, спостерігати й порівнювати імовірности ознаки B за наявности або відсутности ознаки A. Ми можемо лише встановлювати відносну частоту ознаки B, в якійсь конкретній сукупності випадків (одиниць), коли присутня чи наявна й ознака A, та порівняти її (частоту) з відносною частотою ознаки B в другій сукупності випадків, коли ознака A була відсутньою. Не можемо спостереженням безпосередньо встановити імовірність смерти хорого на віспу, при умові ін'єкції ліку, та порівняти її з імовірністю смерти без ін'єкції. Можемо лише встановити відносну частоту смерти в певній сукупності ін'єктованих хорих та в певній сукупності неін'єктованих хорих і ці дві відносні частоти порівняти. Так само можемо для вияснення дідичного звязку короткозорости батьків і дітей, можемо встановити відносну частоту народження короткозорих дітей у певної сукупности короткозорих батьків та порівняти її з відносною частотою народження короткозорих дітей у іншій сукупності батьків з нормальним зором.

Коли ми наведені вище апріорні імовірності в різних критеріях залежности замінимо емпіричними відносними частотами, тоді одержимо відповідні емпіричні критерії залежности. Означимо кількість спостережених випадків, в яких була наявною ознака A через символ (A), а в яких вона була відсутньою символом (a); теж саме для ознаки B через (b), а для сумісної наявности (або відсутности) обох ознак через (AB) чи (aḃ) та при умові наявности однієї і відсутности другої (Aḃ) чи (aB). Загальну кількість спостережень означимо через N. Тоді в цих символах відносні частоти, одповідаючі поодиноким імовірностям, про які ми говорили вище, будуть виглядати так:

$$P_A = \frac{(A)}{N} \quad ; \quad P_B = \frac{(B)}{N} \quad ; \quad P_a = \frac{(a)}{N} \quad ; \quad P_b = \frac{(b)}{N}$$

$$P_{AB} = \frac{(AB)}{N} \quad ; \quad P_{ab} = \frac{(ab)}{N} \quad ; \quad P\frac{(A)}{B} = \frac{(AB)}{(A)} \quad ;$$

$$P\frac{(B)}{A} = \frac{(AB)}{(B)} \quad \text{і т.д.}$$

Таким чином замість апріорних взорів критерія залежности (1-го), (2-го) й (3-го) - будемо мати:

$$\frac{(AB)}{A} > \frac{(aB)}{(a)} \tag{12}$$

$$\frac{(AB)}{(A)} < \frac{aB}{a} \tag{13}$$

$$\frac{(AB)}{(A)} = \frac{(aB)}{(a)} \tag{14}$$

Замість форми (4-ої), (5-ої) і (6-ої) будемо мати:

$$\frac{(AB)}{(A)} > \frac{(B)}{N} \tag{15}$$

$$\frac{(AB)}{(A)} < \frac{(B)}{N} \tag{16}$$

$$\frac{(AB)}{(A)} = \frac{(B)}{N} \tag{17}$$

Замість форми (9-ої), (10-ої) і (8-ої) будемо мати:

$$\frac{(AB)}{N} > \frac{(A)}{N} \cdot \frac{(B)}{N} \tag{18}$$

$$\frac{(AB)}{N} < \frac{(A)}{N} \cdot \frac{(B)}{N} \tag{19}$$

$$\frac{(AB)}{N} = \frac{(A)}{N} \cdot \frac{(B)}{N} \tag{20}$$

76. Можливість зворотности колективної залежности.

Характерною рисою колективної залежности явищ або ознак A і B є та, що існування A змінює імовірність явища B. В багатьох прикладах такий звязок означав би те, що явище A по суті речі слід вважати причиною B, і що останнє є наслідком, себто імовірності, які в таких випадках нам прийшлось би порівнювати, були б обумовленими імовірностями наслідку за припущенням певної причини.

Одначе, констатуючи колективну залежність порівнянням обумовленої імовірности одного явища за умовою існування або неіснування другого, дуже

часто не знаємо взагалі, що є причиною і що є наслідком, а констатована залежність нічого про це не говорить. Отже треба добре нам пам'ятати що дефініція колективної залежности, яко розділу імовірности однієї прикмети за умовою існування або неіснування другої прикмети (чи явища) об'єднує в собі не тільки випадки, в яких порівнювані імовірності є обумлвленими імовірностями наслідку, але й такі, в яких ці імовірності уявляють з себе обумовлені імовірності причин.

Отже тоді, коли не знаємо чи є прикмета A причиною чи наслідком прикмети B, виникає питання чи можемо, беручи за основу прикмету A, себто взявши її за причину і встановивши залежність від неї B, обернути зворотньо так встановлену залежність. Себто чи можемо взявши за основу прикмету A та порівнявши імовірності B за умов наявности і відсутности A, перейти до констатування однакової залежности, яку найшли раніш? Наприклад, порівняли ми імовірності глухонімоти у душевнохорих і здорових осіб та констатували, що перша імовірність більша другої, себто встановили пряму залежність. Коли, одначе, з дальших досліджень іншого характеру випливає, що не душевна хоріть є причиною глухонімоти, а навпаки, глухонімота є причиною душевної хороти, буде це означати, що ми порівняли обумовлені імовірності причини за припущенням певного наслідку і наслідку зворотнього. Виникає питання, чи сконстатуємо однакову колективну залежність, коли перейдемо у цьому випадку до більш звиклих для нас обумовлених імовірностей наслідків і порівняємо обумовлену імовірність наслідку за припущенням існування даної причини з обумовленою імовірністю його, за умовою відсутности даної причини. Отже у випадку порівняння імовірности душевної хоросты у глухонімих з імовірністю душевної хоросты у неглухоніміх чи виявиться й тут перша імовірність як більша, ніж друга?

В перший раз ми порівняли імовірності глухонімоти у душевнохорих і здорових осіб. Означимо прикмету причину символом A, а прикмету наслідок символом B, тоді глухонімота є A, а душевна хоріть B (навпаки, відсутність цих прикмет означимо через a і $\bar{b}$).

При порівнянню зазначених імовірностей користувались ми з наступного взору залежности:

$$P_A^{(B)} > P_A^{(\bar{b})} \qquad (21)$$

Чи дістанемо одначе нерівність такого самого знаку, себто залежність пряму, коли будемо порівнювати обумовлену імовірність душевної хоросты (B) за умовою наявности глухонімоти (A) та за умовою її відсутности (a). Себто чи є правильною нерівність (21), коли має підставу і нерівність відомої вже нам форми $P_B^{(A)} > P_B^{(a)}$, а коли $P_A^{(B)} < P_A^{(\bar{b})}$ чи можна твердити, що $P_B^{(A)} < P_B^{(a)}$? Нарешті, коли $P_A^{(B)} = P_A^{(\bar{b})}$, чи можна твердити, що $P_B^{(A)} = P_B^{(a)}$? На це питання мусимо відповісти позитивно. Дійсно, за теоремою множення імовірностей, залежних подій або явищ, імовірність сумісного відбуття явищ A і B (яку означаємо $P_{A,B}$), можемо написати або:

$$P_{A,B} = P_A P_B^{(A)} \quad \text{або} \quad P_{A,B} = P_B P_A^{(B)} .$$

Відціль $P_A P_B^{(A)} = P_B P_A^{(B)}$, або

$$P_A^{(B)} = \frac{P_A P_B^{(A)}}{P_B} \qquad (22)$$

Однаково одержимо:

$$P_A^{(b)} = \frac{P_A P_b^{(A)}}{P_b} \qquad (23)$$

Таким чином, коли $P_A^{(B)} = P_A^{(b)}$, то за рівняннями (22) та (23) можемо написати:

$$\frac{P_B^{(A)}}{P_B} = \frac{P_b^{(A)}}{P_b} = \frac{1 - P_B^{(A)}}{1 - P_B} ;$$

відціль випливає, що $P_B^{(A)} = P_B$. У випадку (як раніш було зазначено $P_B^{(A)} = P_B$, то й $P_B^{(A)} = P_B^{(a)}$, себто є доведеним, що з рівняння $P_A^{(B)} = P_A^{(b)}$ випливає рівняння $P_B^{(A)} = P_B^{(a)}$. Однаковим способом можемо довести, що з нерівности $P_A^{(B)} > P_A^{(b)}$ випливає нерівність $P_B^{(A)} > P_B^{(a)}$ й що з нерівности $P_A^{(B)} < P_A^{(b)}$ випливає, що $P_B^{(A)} < P_B^{(a)}$.

Отже, коли встановимо незалежність між прикметами A і B шляхом порівняння імовірности A за умов наявности і відсутности B, та коли порівняємо імовірність B за умов наявности і відсутности A, прийдемо до тотожньої незалежности між A і B.

Таким чином таке "обернення" критерія колективної залежности (що нас цікавить у випадку незнання, яка саме з прикмет, що знаходяться в колективному зв'язку є причиною, а яка наслідком) є цілком допустимим.

Одначе такий зворот критерія залежности є можливий щодо встановлення самої залежности та виявлення її знаку (позитивности чи негативности), але ні в якому разі її ступеня чи сили.

Дійсно, коли $P_B^{(A)} > P_B^{(a)}$ тоді й $P_A^{(B)} > P_A^{(b)}$, але розділ між $P_A^{(B)}$ і $P_A^{(b)}$, що є певним виразом ступеня колективної залежности не мусить рівнятися розділу між $P_B^{(A)}$ і $P_B^{(a)}$, бо може бути й більшим або меншим, ніж останній, і це залежить на тому, яким є взаємне відношення імовірностей P_A, P_a, P_B, P_b. Тільки за цілком певного взаємного відношення цих імовірностей, як то може бути доведеним, ці два розділи будуть взаємно рівними. З'окрема $P_B^{(A)} - P_B^{(a)}$ буде рівнятись $P_A^{(B)} - P_A^{(b)}$ лише тоді, коли $P_A P_a$ буде рівнятись $P_B P_b$. Коли напишемо цю умову у вигляді

$$P_A(1 - P_A) = P_B(1 - P_B) ,$$

то побачимо, що це означає ніщо інше, як рівність середніх помилок (штандартних) обох прикмет, бо середні похибки будуть:

$$\sqrt{P_A(1 - P_A)} \quad \text{та} \quad \sqrt{P_B(1 - P_B)} .$$

Таким чином, беручи за основу порівняння одну прикмету і не знавчи, певно, чи є вона причиною чи наслідком другої, можемо бути певні, що у випадку якби ми взяли за основу другу прикмету прийшлиб до встановлення залежности того самого знаку або тотожнього висновку про відсутність залежности але ступень або силу такої залежности могли б знайти іншу.

Само собою розуміється, що вищенаведені правила теоретичного взаємного звязку різних критеріїв про можливість "зворотности" залежности є дійсними і для емпіричних взорів цих критеріїв. Для зрозуміння одначе внутрішнього змісту емпіричних критеріїв треба пам'ятати, що у всіх критеріях колективної залежности між якісними прикметами, про які було говорено вище, йде справа про розділи, різниці між імовірностями і коли йдемо до їх пізнання через спостереження відповідних відносних частот, - мусимо в кожному випадку вирішити питання, до якої міри певна різниця між частотами є значною та невипадковою, себто, до якої міри вона характеризує різницю між відповідними імовірностями.

З попередніх викладів ми вже добре знаємо, що обчислення штандартного відхилення та модуля відносної частоти і дає нам ту міру, до якої кожна дана поєдинча відносна частота випадково+) відхилиться від відповідної імовірности.

Знаємо, що штандартна похибка або модуль різниці між відносними частотами формулюється між двома величинами x та y як

$$б_a = \sqrt{б^2x + б^2y}$$

так само й модуль ($c = б\sqrt{2}$) буде рівним

$$C_d = \sqrt{C^2x + C^2y}$$

В данному випадкові, в якому маємо до діла з похибкою різниці між двома відносними частотами, будемо мати середні похибки поєдинчих відносних частот, а за взором Лапласа (обернений)

$$\sqrt{\frac{\frac{m_1}{n_1}\left[1 - \frac{m_1}{n_1}\right]}{n_1}} \quad \text{та} \quad \sqrt{\frac{\frac{m_2}{n_2}\left[1 - \frac{m_2}{n_2}\right]}{n_2}} \quad \text{себто}$$

$$б_d = \sqrt{\frac{\frac{m_1}{n_1}\left[1 - \frac{m_1}{n_1}\right]}{n_1} + \frac{\frac{m_2}{n_2}\left[1 - \frac{m_2}{n_2}\right]}{n_2}}$$

а також модуль

$$C_d = \sqrt{\frac{2\frac{m_1}{n_1}\left[1 - \frac{m_1}{n_1}\right]}{n_1} + \frac{2\frac{m_2}{n_2}\left[1 - \frac{m_2}{n_2}\right]}{n_2}}$$

Наприклад, хочемо з'ясувати собі чи існує залежність між мертвонародженням та шлюбним чи нешлюбним походженням дитини. В 1923 році в Чехословаччині було всього 388650 народжених, з них 349279 було шлюбних і 39371 нешлюбних. З числа шлюбних дітей 341449 будо живих, а 7830 мертвих, з нешлюбних 37811 було живих, а 1560 мертвих.

Отже % мертвонароджених з числа шлюбних дітей дає 2.24%, а з числа нешлюбних дітей дає 3.96%, себто розділ між поданими відносними частотами складає 1.72%.

+) Під "випадково" тут розуміємо - "важливо", "засадничо".

$$\sigma_d = \sqrt{\frac{2.24 \times 97.76}{349.279} + \frac{3.96 \times 96.04}{39.371}} = 0.101\ \%$$

$C_d = 0.143\ \%$. Себто дійсний розділ є рівним 17 разів взятій своїй середній похибці і 12 разів взятому модулю. Таким чином, з досить великою імовірністю можемо припустити, що вказаний розділ (2.24 % та 3.96 %) не міг бути випадковим і що між мертвонародженням та шлюбним чи нешлюбним походженням дитини існує залежність.

З'ясуємо далі залежність між статтю народжених та живим або мертвим народженням. З 379.260 живонароджених в Чехословаччині було хлопців 195.835 або 51.64 %, а з 9.390 мертвонароджених було хлопців 5.194 або 55.31 %. Різниця складає 3.67 %.

$$\sigma_d = \sqrt{\frac{51.64 \times 48.36}{379.260} + \frac{55.31 \times 44.69}{9.370}} = 0.519\ \%$$

$C_d = 0.734\ \%$. Себто дійсна різниця 7.1 б і 5.0 С або залежність можемо вважати з великою імовірністю за існуючу. Мусимо, одначе, завше пам'ятати, що наведена вище формула модулю може прикладатись за певних умов (незалежність подій і т.д.), про які вже згадувалось у наших попередніх викладах.

77. Коефіцієнт асоціяції. Подібно тому, як для визначення напряму і сили залежности між квантитативними ознаками, ми знайшли й формулювали єдину міру виявлення такої залежности або незалежности через коефіцієнт кореляції так само і у випадках, коли маємо до діла з квалітативними прикметами, можна таким користуватись єдиною мірою, що характеризує звязок чи залежність між такого роду ознаками. Ця міра зветься коефіцієнтом асоціяції.

Означимо знайомий нам вираз

$$P_B^{(A)} - P_B^{(a)} = \beta_B$$

та відповідно $P_A^{(B)} - P_A^{(b)} = \beta_A$. Тоді на підставі наведених нами вище критеріїв залежности можемо встановити, що

$$P_A = P_B \cdot \frac{P_A P_a}{P_B P_b} \quad \text{або} \quad \beta_A = \beta_B \cdot \frac{\sigma_A^2}{\sigma_B^2}$$

так само $$\beta_B = P_A \cdot \frac{\sigma_B^2}{\sigma_A^2}$$

Геометрична середня з обох величин β_A і β_B дасть:

$$\sqrt{\beta_A \beta_B} = \sqrt{\beta_A^2 \cdot \frac{\sigma_B^2}{\sigma_A^2}} = \beta_A \cdot \frac{\sigma_B}{\sigma_A} = \beta_B \cdot \frac{\sigma_A}{\sigma_B} \qquad (24)$$

Отже кожна величина β_A і β_B поділена штандартною похибкою тої ж самої ознаки і помножена на штандартне відхилення (похибку) другої ознаки, дає тотожню величину, що й геометрична середня з β_A і β_B. Ця геометрична середня, що означається звичайно через " r", і є коефіцієнт асоціяції і може бути цілком зручною мірою залежности між кваліятивними ознаками.

З формули (24) випливає, що

$$r = P_B^{(A)} - P_B^{(a)} \cdot \sqrt{\frac{P_A P_a}{P_B P_b}} = P_A^{(B)} - P_A^{(b)} \cdot \sqrt{\frac{P_B P_b}{P_A P_a}} \qquad (25)$$

Через те при незалежності між A і B, коли $P_B^{(A)} = P_B^{(a)}$, а $P_A^{(B)} = P_A^{(b)}$, буде $r = 0$.

При повній залежності, коли $P_B^{(A)} = 1$; $P_B^{(a)} = 0$; $P_A = P_B$, а $P_a = P_b$ з виразу (25) виходить, що $r = 1$. Таким є теоретичне формулювання. Коли візьмемо замість імовірностей у взорі (25) відповідні емпиричні відносні частоти, тоді в кожній вартості такої емпиричної міри " r " буде матись певна частка випадковости і то тим більша, чим меншим є кількість спостережень, з яких r обчислено. Через те вартість $r = 1$ не буде в дійсности рішучо свідчити про цілком певну залежність між ознаками, так само як і вартість $r = 0$ не буде свідчить про цілковиту незалежність. Коефіцієнт асоціяції r може бути визначений і в іншій формі чим вирази (24) і (25). Понеже можна легко довести, що

$$P_B = \frac{P_B^{(A)} - P_B}{P_a}$$

можемо взір (25) писати у формі:

$$r = \frac{P_B^{(A)} - P_B}{P_a} \cdot \frac{\sqrt{P_A P_a}}{\sqrt{P_B P_b}} = \frac{(P_B^{(A)} - P_B) \cdot \sqrt{P_A}}{\sqrt{P_a P_B P_b}} = \frac{(P_B^{(A)} - P_B) \cdot P_A}{\sqrt{P_A P_a P_B P_b}}$$

а через те, що $P_A P_B^{(A)} = P_{A,B}$, то й

$$r = \frac{P_{A,B} - P_A P_B}{\sqrt{P_A P_a P_B P_b}} \qquad (26)$$

Нарешті від цього останнього взору можемо перейти й до третього взору для r, а саме:

Понеже $P_{A,B} + P_{A,b} + P_{a,B} + P_{a,b} = 1$; $P_{A,B} + P_{A,b} = P_A$, а $P_{A,B} + P_{a,B} = P_B$, можемо чисельник у взорі (26) писати у формі

$$P_{A,B} - P_A P_B = P_{A,B}(P_{A,B} + P_{A,b} + P_{a,B} + P_{a,b}) - (P_{A,B} + P_{A,b})(P_{A,B} - P_{a,B})$$

а це після множення дає:

$$P_{A,B} P_{a,b} - P_{A,b} P_{a,B}$$

Отже остаточно:

$$r = \frac{P_{A,B}P_{a,b} - P_{A,b}P_{a,B}}{\sqrt{P_A P_a P_B P_b}} \qquad (27)$$

78. Посередня залежність прикмет. В тому випадку, коли коефіцієнт асоціяції двох прикмет A і B є близьким до одиниці, себто між ними мається тісний звязок, може виникнути питання: чи цей звязок між прикметами не пояснюється тим, що ознаки A і B є звязані кожна з окрема з якоюсь третьою прикметою C ?

Для відповіді на це питання треба знайти: 1,- коефіцієнт асоціяції $r_{C(AB)}$ ознак A і B для тих членів сукупности, що мають прикмету C; 2,- коефіцієнт асоціяції $r_{c(AB)}$ прикмет A і B для тих членів сукупности, що не мають ознаки C. При чому під (AB) розуміємо кількість об'єктів, що мають одноразово дві ознаки A і B.

Якщо $r_{C(AB)}$ та $r_{c(AB)}$ є приблизно рівними, можемо припустити, що звязок поміж прикметами A і B є безпосередній; у протилежнім випадку треба дослідити асоціяції прикмет A і B з прикметою C і скласти коефіцієнти $r_{(AC)}$, $r_{(BC)}$.

Коефіцієнти $r_{C(AB)}$ та $r_{c(AB)}$ знайдемо з формул:

$$r_{C(AB)} = \frac{(ABC)(a\bar{b}C) - (aBC)(A\bar{b}C)}{\sqrt{(AC)(aC)(BC)(\bar{b}c)}} \qquad (28)$$

$$r_{c(AB)} = \frac{(ABc)(ab\bar{c}) - (aBc)(A\bar{b}c)}{\sqrt{(Ac)(ac)(Bc)(\bar{b}c)}} \qquad (29)$$

Ці формули й дають клефіцієнти асоціяції A і B в клясах C і c.

Коли обидві формули дають приблизно однакові величини для $r_{C(AB)}$ та для $r_{c(AB)}$, тоді можемо вважати, що звязок A і B є однаковим в клясах C і c, себто не залежать від C і уявляє в себе простий звязок. Коли б виявилось наприклад, що перше значення (r) є рівним нулю, а друге (r) не є рівним нулю, мусимо припустити, що звязок між A і B існує лише в клясі C і себто між ним є лише посередній звязок. Треба також завважити, що розходження величин r_C і r_c можемо вважати з великою імовірністю випадковим, коли величина цієї ріжниці не перевищує потроєної її середньої квадратової похибки, що є: рівною:

$$\sigma(r_C - r_c) = \sqrt{\sigma^2(r_C) + \sigma^2(r_c)} \qquad (30)$$

Візьмемо приклад наведений у Гальтона +) по дослідженню дідичности фарби очей в англійських родинах (78 родин). Получені ним дані дали наступне:

(ABC) = 1928	(a$\bar{b}$C) = 303
(ABc) = 596	(aBc) = 225
(A$\bar{b}$C) = 552	(a$\bar{b}$C) = 395
(A$\bar{b}$c) = 508	(a$\bar{b}$c) = 501

+) В. Романовський. стор. 319.

Літера А означає світлу фарбу очей, "а" темну у дітей, літера В і b ті ж самі барви у батьків, а С і с і дідів. Пошукувалось дідичність ознак від дідів до внуків, поминувши батьків.

Пошукувані коефіцієнти знайдемо заміною в рівнянні

$$r = \frac{(AB)(ab) - (aB)(Ab)}{\sqrt{(A)(a)(B)(b)}}$$

В на С, а в рівняннях (28) і (29) перестановленням В й С. З поданих вище обсягів 3-го порядка (три ознаки) знайдемо необхідні нам обсяги агрегатів 1-го та 2-го порядку і тоді одержимо, що

$$r_{(AC)} = \frac{2480.726 - 698.1104}{\sqrt{3584.1424.3178.1830}} = 0,189$$

$$r_{B(AC)} = \frac{1928.225 - 303.596}{\sqrt{2524.528.2231.821}} = 0,162$$

$$r_{b(AC)} = \frac{552.501 - 395.508}{\sqrt{1060.896.947.1009}} = 0,086$$

Всі три коефіцієнти додатні, через те можна припустити, що передача ознак від дідів до внуків, через батьків, існує.

9. Коефіцієнт континґенції Пирсона як міра асоціяції многих ознак.

Наведені нами вище коефіцієнти асоціяції ознак не можуть задовольнити критеріям залежності ознак в тих випадках, коли маємо до діла з залежностями багатьох класифікованих прикмет. Пирсон і запропонував <u>таку міру звязку у вигляді коефіцієнта континґенції</u>.

Припустимо, що прикмета А має відмін A_1, A_2, ...A_k а взагалі A_i, а прикмета В та 1 відмін : B_1, B_2, B_3 .....B_l, а взагалі B_j . Імовірності появи відповідних відмін для прикмети А будуть P_1, P_2, ...P_k а взагалі P_i а для прикмети В суть P_1, P_2, ...P_l а взагалі P_j . Імовірність сукупної появи відміни "i" прикмети А з певною відміною " j " прикметй В означимо як $P_{i/j}$. На підставі вищенаведеного третього критерія залежності рівняння (9), (10), (11) буде для кожної такої пари у випадку незалежности справедливим:

$$P_{i/j} = P_i \cdot P_j$$

а колективна залежність буде тим більша, чим більшими в цілому будуть різниці $P_{i/j} - P_i \cdot P_j$. З цих різниць Пирсон визначає свій коефіцієнт континґенції (береться в іншому ніж Пирсон означенні А.Чупрова):

$$\varphi^2 = \sum_{i=1}^{k} \sum_{j=1}^{e} \frac{\left[P_{i/j} - P_{i/} \cdot P_{/j}\right]^2}{P_{i/} \cdot P_{/j}} \qquad (31)$$

Суми, які відносяться до вартостей "i" та до всіх вартостей " j " означають, що взяті будуть всі можливі комбінації (пари) відмін прикмет A і B і для кожної з них буде обчислений вираз

$$\frac{\left[P_{i/j} - P_i \cdot P_j \right]^2}{P_i \cdot P_j}$$

а сума таких виразів дасть пошукувану загальну міру залежности. Завважимо, що квадрат різниці $P_{i/j} - P_i \cdot P_j$ взятий через те, що сума перших степеней для всіх пар відмін дає обов'язково нуль. Понеже йде тут про всі можливі комбінації (пари) відмін прикмет A і B, то сума їх імовірностей

$\left[\sum_{i=1}^{k} \sum_{j=1}^{e} P_{i/j} \right]$ є рівною одиниці.

Таким чином, коли A і B є незалежні, тоді кожна різниця $P_{i/j} - P_i \cdot P_j = 0$, а $\varphi^2 = 0$ і навпаки φ^2 може рівнятись нулі тоді, коли прикмети є незалежні. При "цілковитій" залежності в точному розумінні причинового зв'язку, коли при появі даної відміни A може появитись лише одна певна відміна прикмети B де стало будь i = j, буде $P_{i/j}$ рівнятись нулю, та для всієї решти пар, де i = j , буде $P_{i/j} = P_i = P_j$. При таких умовах коефіцієнт буде рівним:

$$\varphi^2 = \sum_{i=1}^{k} \sum_{j=1}^{e} \frac{P_{i/j}^2}{P_i \cdot P_j} - 1 \qquad (32)$$

Такою є теоретична міра контингенції. Щодо емпіричної середньої міри контингенції, то вона буде випадати інакше, а саме:

$$\varphi^2 = \sum_i \sum_j \frac{\left[\frac{A_i B_j}{N} - \frac{A_i}{N} \cdot \frac{B_j}{N} \right]^2}{\frac{(A_i)}{N} \cdot \frac{B_j}{N}} = \sum_i \sum_j \frac{\left[A_i B_j - \frac{1}{N} (A_i)(B_j) \right]^2}{(A_i)(B_j)} \qquad (33)$$

Тут $(A_i B_j)$, (A_i) та (B_j) означають абсолютні кількості спостережених випадків з відповідною відміною прикмети, а

$$\frac{(A_i B_j)}{N} \qquad \frac{(A_i)}{N} \qquad \frac{(B_j)}{N}$$

відповідні емпіричні відносні частоти.
Можемо вжити коротших означень: $(A_i) = n_i$; $(B_j) = n_j$; $(A_i B_j) = n_{i/j}$ то середня міра контингенції набере виразу:

$$\varphi^2 = \sum_i \sum_j \frac{\left[n_{i/j} - \frac{1}{N} n_i \cdot n_j \right]^2}{N_i \cdot n_j} \qquad (34)$$

Розуміється, у вартостях цієї міри, яко вартостях емпіричних, будуть певні елементи випадковости, що не дають можливости трактувати цю міру, як міру повної залежности чи незалежности.

У випадку подвійно клясифікованих прикмет (P_{AB}), буде мати 4 різниці:

$$P_{AB} - P_A P_B \; ; \; P_{a\beta} - P_a P_\beta \; ; \; P_{A\beta} - P_A P_\beta \; ; \; P_{aB} - P_a P_B \; .$$

Всі вони по своїй абсолютній кількості рівні між собою. Тоді коефіцієнт контингенції для такого випадку буде рівним:

$$\varphi = \frac{P_{A,B} - P_A P_B}{\sqrt{P_A P_B P_a P_\beta}}$$

себто в цьому випадку він є ніщо інше як величина " r " яку ми подали вище.

Закінчуючи цим другу частину основ про статистичну методу мусимо ще раз підкреслити, що прикладення всіх наведених вище формульовань, тісно звязаних з теорією імовірностей, до статистичної дійсности, до емпіричних статистичних даних та рядів мусить робитись з надзвичайною обережністю, оскільки, як вище було многократно підкреслено, статистика не є математика і кореляційні звязки і залежності не є функціональні звязки і залежності.

Нарешті мусимо зазначити, що виклад про відшукання так званих нелінійних кореляційних залежностей і звязків, а також статистичну методу аналізи так званих часових статистичних рядів, ми мусимо залишити, не передбачаючи відповідної підготовки у слухачів нашого Інституту до такого рода аналізи, а почасти і через те, що вважаємо більш доцільним поповнити в певних межах цей пробіл в своєму пізнішому для наших студентів курсі "Дослідження господарської коньюнктури", що входить в програм навчання.

----пОп----

Запитання до лекцій 21 - 25.

1. Що таке з статистичного погляду "маса" ?

2. В чому зв'язок між інтегралом Лапласа та ймовірністю відхилень ?

3. На підставі чого визначається та частина якоїсь цілої сукупности одиниць, яка може задовольнити дослідника при вибірковому способі спостереження або дослідження ?

4. Через що спосіб інтерполірування криє в собі небезпеку щодо висновку ?

5. Що таке коефіцієнт асоціяції і з яким коефіцієнтом він має аналогію ?

----пОп----

Л І Т Е Р А Т У Р А.

Із згаданих підручників теорії статистики і праць, що торкаються основ статистики і статистичної методи й можуть бути дуже корисними для дальших студій та поглиблення знання з царини теоретичної статистики - рекомендуємо наступні праці:

А.

В мові англійській:

1. G.U.Yule: Introduction to the Theory of Statistics. 8 вид.; London. 1927. (Маються чеський й польський переклади).

2. F.C.Mills: Statistical Methods applied to Economics and Business. New York. 1924.

3. A.L.Bowley: Elements of Statistics. 5 вид. P.S.Kirey. London. 1926.

4. H.L.Rietz: Mathematical Statistics. Yowa. 1928.

Друга частина підручника Бовлі і особливо праця Рітца вимагають доброї математичної підготовки.

В мові німецькій.:

1. G.Th.Fechner: Kollektivmasslehre - Leipzig. 1897.

2. E.Blaschke: Vorlesungen über matematische Statistik. Leipzig. 1906.

3. E.Czuber: Die statistischen Forschungsmethoden. Wien. 1921.

4. W.Winkler: Statistik. Leipzig. 1925.

5. A.Tchuprow: Grundbegriffe und Grundprobleme der Korrelationstheorie. Berlin. 1925.

Книжка А.Чупрова є дуже цінною для наукового розуміння кореляції і тих можливостей, що дає ця спеціяльна метода для відшукання причинових зв'язків і взаємозалежностей в колективних атипових явищах. Підручники Е.Бляшке і Е.Чубера потрібують математичного підготовлення.

В мові італійській:

1. L.Benini: Principii di statistica metodologica. Torino. 1906.

2. G.Mortara: Elementi di statistica. Roma. 1917.

3. G.Mortara: Lezioni di statistica metodologica. Bologna. 1922.

4. A.Niceforo: Il metodo statistico, troria laplicationi alle scienze naturali, alle scienze sociali e all'arte. Messina. 1923. (Перекладена на французьку мову Paris.1925).

Дуже цінні праці Мортара вимагають доброго математичного підготовлення читача, книжка ж Нічефоро більш приступна та дає цілком ясне освітлення статистичної методи.

В мові французькій:

1. A.Julin: Principes de statistique théorique et appliquée. Tome I. Bruxelles. 1921.

2. J.Darmois: Statistique mathematique. Doin, Paris. 1928.

Остання дуже цінна книжка з теоретичного боку, але вимагає від читача доброго знання математики. Праця Жюлена, дуже добрий загальний підручник, цілком доступний для читача з середньошкільною підготовкою.

В мові російській:

1. А.А.Чупров: Очерки по теоріи статистики. Москва. 1910.

2. В.И.Романовский. Елементарний курс математической статистики. Ленинград. 1924.

3. А.А.Кауфман: Введение в теоретическую статистику. Ленинград. 1923.

4. А.А.Кауфман: Теория и методи статистики.

5. А.К.Митропольский: Основи статистики т.I. Ленинград. 1925.

6. М.В.Птуха: Очерки по теоріи статистики населенія и моральной. Петроград. 1916.

Щодо з'ясования основних проблем статистичної методи праці А.Чупрова і А.Кауфмана (Введение) заслуговують особливої уваги; яко підручник основ науки про статистичну методу також рекомендуємо працю В.Романовского.

В мові українській:

1. П.А.Вихляєв: Нариси теоретичної статистики (переклад з російської). 2 вид. 1927.

2. С.М.Романовський: Математична статистика (конспект-літогр.). Подєбради. 1928.

3. Сапігин: Варіяційна статистика. Вид.Книгоспілки. 1926.

В мові чеській:

1. St.Kohn: Základy teorie statistické metody. Praha. 1929.

2. J.Beneš: O statistice a její teorii a vědach a zajmech s nimiž souvisí. Praha. 1920.

Підручник небіжчика Станіслава Кона, бувшого слухача проф. А.Чупрова, є одним з кращих сучасних підручників основ статистичної методи, надзвичайно цінний по своїй методологічній послідовності та критичному відношенню і освітленню головних проблем статистичної методи. Більш-менш доступний також і з математичного боку.

В мові польській:

1. J.Czekanowski: Zarys metod statystycznych w zastosowaniu do antropologii. Prace Tow-twa Nauk. Warszawskiego - III. №5. 1913.

B.

Підручники по теорії імовірностей:

1. В.Дяконенко: Числення імовірностей.(літогр.)"Сіяч". Прага. 1928.

2. A.Cournot: Exposition de la théorie des chauces et des probabilités. Paris. 1848.

3. H.Poincaré: Calcul des probabilités. 2 вид. Paris. 1922.

4. E.Borel: Eléments de la théorie des probabilités. 3 вид. Paris. 1924.

5. Czuber E.: Warscheinlichkeitsrechnung und ihre Anwendung auf Fehlerausgleichung. Statistik und Lebensversicherung. т.1. 4 вид.. Leipzig. 1924. т.2. 3 вид. 1921.

6. Л.К.Лахтин: Курс теории вероятностей. Москва. 1924.

7. А.А.Марков: Исчисление вероятностей. 4 вид. Москва. 1924.

8. С.Бернштейн: Теория вероятностей. Москва. 1927.

C.

Заслуговують великої уваги студіюючих статистику також наступні спеціяльні праці:

1. Chr.Sigwart : Logik. 3 вид. Tübingen. 1904.

2. R.Meerwarth: Einleitung in die Wirtschaftstatistik. 1920. (мається в перекладі на російську мову).

3. Збірник: Статистический метод в научном исследовании. Под ред. М.Смит и А.Тимирязева. Москва. 1925.

4. Mitchell W.C.: 1) Index Numbers в збірнику "Business Statistics" Copeland. 1921.
2) The Making and Using of Index Numbers, Bulletin of the U.S.A. Bureau of Labor Statistics № 173. Washington. 1921.

Повний перегляд літератури можна знайти в підручнику А.А.Чупрова: "Очерки по теоріи статистики"(до 1910 р.), а також в брошурі А.Митропольского: "Предмет статистики". Ленинград. 1927, та в підручнику S.Kohn: "Základy teorie statistické metody. Praha. 1929.

----ппОпп----

О Г Л А В.

(В дужках подані підрозділи 1-го видання курсу).

Стор.

Передмова . . . . . I.

В С Т У П.

ПРЕДМЕТ ТА ЗАВДАННЯ СТАТИСТИКИ.

1. Поняття статистики. 2. Природа колективних атипових явищ. 3. Поняття випадку і випадкової причини. 4. Об єкт статистики. 5. Статистичне спостереження. 6. Суб єкт статистичного спостереження. 7. Масовість статистичного спостереження. 8. Числове спостереження і матеріяльна природа статистичного числа. 9. Органи статистичного спостереження. — 1 - 10

ЧАСТИНА ПЕРША

ОПИСОВЕ ЗАВДАННЯ СТАТИСТИКИ (ІДІОГРАФІЯ).

Гл.I. Статистична сукупність. — 11 - 14

10. Властивості статистичної сукупности. 11. Ознаки неподільних (одиниць) сукупности.

Гл.II. Статистичне числення, його основні форми і правила переведення. — 14 - 59

12. Одиниця статистичного числення. 13. Моменти часу й простору в статистичному численні. 14. Експедиційний і бланковий способи числення. 15. Типи статистичного числення: а) переписи населення, б) переписи професійні, в) переписи сільсько-господарські, г) переписи подвірні земські на Україні і в Росії, д) переписи промислові, е) інші типи статистичного числення. 16. Поточна реґістрація.

Гл.III. Габличне зведення матеріялу статистичного числення і його публікація. — 59 - 74

17.(18) Контроля матеріялу. 18(19). Зведення первісних даних статистичного числення: а) таблиці прості й комбінаційні, б) статистичний підмет і статистичний присудок, в) факторіяльні й функціональні ознаки в групуванні матеріялу. 19(19 літ.г). Плян зведення матеріялу. 20(19 літ. г) Централізоване й децентралізоване зведення матеріялу. 21(19 літ. д)Техніка рахування в статистичному зведенню матеріяла. 22(19 літ.е) Публікація зведеного матеріялу.

Гл.IУ. Похідні статистичні величини і статистичні ряди. — 75 - 94

23(20). Похідні величини. 24(21). Величини відносної частоти 25(22). Таблиці розподілу частоти (статистичні ряди): а) кількість клясових інтервалів та їх границі, в) рівність перемежнів клясових. 26(23). Техніка обліку частот по клясах. 27(24). Графічне зображення розподілу частоти. 28(25). Ідеальна крива частоти. 29.(26). Типи розподілу частоти: а) симетричний, б) мало-асиметричний, в) вельми-асиметричний, г) тип кривої зворотньої асиметричній або мало-асиметричній формам розподілу частоти.

Стор.

Гл.У. Середні величини. 95 - 118

30(27-27.1). Середні величини. 31(27.2). Аритметична середня. а) властивості аритметичної середньої та в) способи її обчислення. 32(28-29). Середня аритметична зважена (media ponderata). 33(30). Типові й атипові пересічні. 34(30). Медіана. 35(31). Модальна середня (Moda). 36(31). Геометрична середня. 37(31). Гармонійна середня.

Гл.УІ. Міра дисперсійної ознаки. 118 - 138

38(34-а). Загальні зауваження. 39(34-б). Штандартне відхилення (похибка). 40 - 41(34-д). Обчислення штандартного відхилення. 42(35). Правило "шости sigma ". 43(36). Коефіцієнт дисперсії. 44(37). Квартільне відхилення (quartil deviation) 45(38). Міра асиметрії або скошеності кривої. 46(38). Індексні числа.

ЧАСТИНА ДРУГА.

ДОСЛІДЖЕННЯ ПРИЧИНОВИХ ЗВ'ЯЗКІВ (НОМОГРАФІЯ).

Гл.УІІ. Статистична метода й теорія імовірностей. 139 - 210

47(39). Номографічні завдання статистичної методи. 48(40). Статистична метода і теорія імовірностей. 49(41 - 42). Основні правила рахування імовірностей. а)теорема додавання імовірностей, в) теорема множення імовірностей, 50 (43). Імовірності у випадку численности причин . 51 (41). Теорема Байєса. 52 (44). Закон великих чисел. 53 (44). Найімовірніша вартість відносної частоти і математичний доказ закона великих чисел. 54(46). Біноміальний розподіл частоти. 55(47). Рівнання Лапласа і теорема Бернуллі. 56(48). Модуль. 57(49). Міра точности відносної частоти. 58(50). Обернена теорема Бернуллі і взору Лапласа. 59(51). Нормальний розподіл частоти або крива Гавса. 60(52). Математичне очікування (надія). 61(53). Математичне очікування і теорема Чебишева. 62(54). Теореми Пуассона і Бернуллі. 63(55). Закон великих чисел і поняття "маси". 64(56). Апріорна середня похибка. 65(57). Інтеграл Лапласа й імовірність відхилень при означеному числі спостережень. 66(58). Емпірична перевірка закона великих чисел. 67(59). Коефіцієнт дівергенції. 68(60). Стабільність статистичних рядів і міра дисперсії В.Лексіса. 69(61). Репрезентативне статистичне спостереження (Вибіркова метода). 70(62). Інтерполяція. 71(63). Статистика й математика.

Гл.УІІІ. К о р е л я ц і я. 210 - 227
(Елементарні способи відшукання кореляційної залежности).

72(65). Поняття кореляції. 73(66). Критерії кореляційної аналізи. 74(67). Кореляційна і функціональна залежності. 75(68). Кореляційна таблиця і кореляційний коефіцієнт. 76(69). Коефіцієнти регресії. 77(70). Обчислення коефіцієнта кореляції. 78(70-а). Колективний характер кореляційної залежности. 79(71). Конструкція імовірностей та коефіцієнти регресії і кореляції. 80(72). Кореляційне відношення.

Стор.

Гл.IX. Статистика атрибутів (якісних аргументів). 227 - 241

81(73). Класифікація. 82(74). Поняття асоціації. 83(75). Емпіричні характеристики колективної залежности. 84(76). Можливість зворотности колективної залежности. 85(77). Коефіцієнт асоціяції. 86(78). Посередня залежність прикмет. 87(79). Коефіцієнт контінґенції Пірсона як міра асоціяції многих ознак.

Література . . . . . 242 - 244

ДОДАТКИ:

1. Формуляр світового сільсько-господарського перепису 1929-1930 р.р., організованого міжнароднім Сільсько-господарським Інститутом у Римі. . . . . 1 - 10

2. Основний і спеціяльний формуляр американського (З.Ш.П.А.) індустріяльного цензу 1925 р. . . . . 11 - 16

3. Інтерполяційна таблиця для знайдення границь абсолютної похибки (відхилення) релятивних чисел для вибірки переведеної у відношенню один до семи (N = 7n) 17

-----ппОпп-----

Д О Д А Т К И

до курсу доц. Л. Шрамченка "Основи теорії статистики"

Д О Д А Т О К I-2

/ до лекції четвертої /.

Формуляр світового сільсько-господарського перепису 1929 - 1930 рр. організованого Міжнародім Сільсько-Господарським Інститутом у Римі.

Пепередні завваження. Формуляр перепису має на меті одержати відомості про площі культур та стан скотарства, які мають значіння завдяки великому своєму простору чи розміру продукції або через свою вагу в міжнародній торговлі. Формуляр складений для обох півкуль світу разом з країнами тропічними, через те торкається також рослин і видів тварин, які в одних країнах існують, а в других ні. Розуміється, що рослини і види тварин, що не плекаються в даній країні, не мусять входити до формуляру перепису такої країни. Крім того, кожна країна може приєднати інші рослини або предмети перепису та поширити обсяг питань, щоб можна було одержати всі бажані додаткові інформації. / з цього приводу в кінці формуляру подано можливі додаткові питання/.

Мета перепису. Метою перепису є зібрати матеріял щодо розміру сільсько-господарського підприємства, врожаю рослин, стану скотарства лише для цілей статистичних як основу для здогадів та щорічної статистики, що переводиться в роках між переписами. Мусить також сприяти поступу сільського господарства та загального добробуту.

Перепис не має цілей фіскальних. Бажані інформації не мають ніякого зв'язку з накладанням податків, є суворо довірочні, їх можуть знати лише службовці статистичних урядів, а дані, що стосуються до кожного окремого підприємства не будуть публіковані.

Дефініція сільсько-господарського підприємства. По думці перепису сільсько-господарське підприємство складається із цілої площі землі, яка призначена цілком або частково для плекання сільсько-господарських рослин або тварин та яку обробляє, або користується одна особа як само - стійно, так і за поміччю інших осіб, незалежно від відносин власности, найму, розміру або місця розположення, незалежно від того, чи творить вона єдину територіяльну цілість чи більше число парцель, але яка вважається завше й уявляє з себе єдине господарське підприємство. Особа, яка господарить або керує підприємством, може бути власником, або наємцем, управителем, або особою, що обробляє, або управляє маєтком та господарить на ньому на яких будь інших, особливих умовах. Коли в господарстві господарять разом дві або більше осіб, мусить одна з них бути визначена за представника спільників та подати відомості для перепису. Підприємство може бути означено яко земельна парцеля, городництво, садівництво, маєток, двір, плянтація, господарство, заклад сільсько-господарський, власність або й інакше; в кожному, одначе, випадку мусить господарити або управляти ним одна особа. Участки заліснені, справу яких веде господар, записуються як частина єдиного підприємства, так само як шматки землі з будовами, на яких живуть та якими користуються робітники, що працюють в одному підприємстві.

Предметом числення є всі сільсько-господарські підприємства розміром не менше I гектара та результати виробництва з поля або городнього господарства та від скотарства, яких доріжнювались би в 1926 році продажній ціні не менше 100 долярів або відповідній ціні в валюті даної країни.

Про землі громадські подає відомості: I/ Орендар, коли він виключно сам господарює на них. 2/ В решті випадків сільський староста.

Завваження. Розуміється, що кожна держава може включити до перепису всі господарства, себто і менші I гектара, але в такому випадку є дуже бажаним, щоб при обробленню матеріялу ці менші господарства групувалися так, щоб можна було господарства розміром I гектар і вище порівнювати спільно для всіх держав.

Критична доба, до якої мусять відноситись питання перепису.

I. Північна півкуля: площі засіву і наслідки врожаю за календарний рік I929, а для скотарства і інших предметів на день I929 року, який буде вказаний пізніше.

2. Південна півкуля: площі засіву і наслідки врожаю в господарському році від I/УII I929 до 30/УI I930 р.; господарські тварини та інші предмети на день року, що кінчається 30/УI I930 року, який буде зазначений пізніше.

I. Особа господаря / власник, наємець, управитель іт.д./
I. Ім'я і прізвище.....
2. Поштова адреса .....

II. Розмір господарства.

I. Ціла площа сільсько-господарського підприємства ..... гект.
/Всі землі, на яких господарить вказана в розділі I особа. Сюди не належить земля винайнята іншим особам, або земля, врожай з якої збирає хтось інший; в такому випадку земля ця записується у формуляр, що відноситься до підприємства, до якого ті землі господарськи належать/.

2. Орна земля.....гект.
/Сюди належать також земля під постійне плекання лучних трав, незалежно від того чи знятий з них врожай чи ні, а також пар. Луки штучні /конюшина/, які були зорані року I928 або в якомусь з 4-х попередніх років, заносяться сюди також/.

Завваження. Коли плекається на одній площі трави і дерева, то слід приблизно визначити площу під тими і другими окремо та площу під травами віднести до орної землі, а решту до п.4-го - землі засаджені чагарниками і деревами. Коли трави і лози та чагарники так помішані, що не можна їх точно розподілити, то в залежності від більшости тої або іншої маси рослин, віднести всю таку площу або до орної землі або до п.4-го.

3. Луки тривалі і пасовиська.....гект.
/Сюди заносяться землі, що стало або принаймні шість останніх років використовуються під сінокоси і пасовиська. Сталі луки і пасовиська, на яких рослини деревисті, як також луки і пасовиська лісові, заносяться сюди тоді, коли лучні рослини переважають./

4. Культури кущаві та деревисті..... гект.
/Сюди не відноситься земля під лісом/.

Завваження. Коли на цих землях плекаються також лучні трави, то при розподілі площі треба керуватись вказівкою, наведеною в п.2-му.

5. Ліси..... гект.
/Сюди відносяться тільки землі зарослі лісом, який розробляється на дерево та інші предмети лісового виробництва/.

6. Болота, що дають вигоди, верескові або інші необроблені землі, що дають пожиток .....гект.
/Сюди віднести землю, яка дає якийсь рослинний продукт, але не належить до якої будь категорії земель, що вище визначені/.

7. Землі неродючі..... гект.
/Сюди відносяться землі, які не дають якого будь рослинного

продукту та не належать до категорії вище наведених/.

Завваження. Площа землі, зазначеної в п. 2-7му, мусить рівнятись площі, зазначеній в пакті І-му цього розділу.

III. Робітники.

/Рахується всіх осіб, що працюють в даному підприємстві в день перепису, з розподілом за статю і віком, зараховуючи також господаря, його дружину і членів родини, дійсно затруднених працею сільсько-господарською. Не рахується осіб, працюючих деінде, а також працюючих в домашньому господарстві./.

І. Особи працюючі стало в день перепису:

	Члени родини		Інші особи	
	мужчини	жінки	мужчини	жінки
а/ молодші 15 років.....	....	...	....	...
б/ 15 років і старші.....	....	...	....	...
2. Особи працюючі переходово в день перепису:				
а/ молодші 15 років.....	....	...	....	...
б/ 15 років і старші.....	....	...	....	...

Р о с л и н и т а в р о ж а й 1 9 2 9 р о к у.

Рослини другорядні. Коли на одній площі засіяні дві або декілька рослин і врожай кожної збирається окремо, треба їх подати окремо у відповідних розділах, зазначивши лише приблизну їх площу. Коли рослини так помішані, що не можна подати приблизної площі кожної, треба зазначати їх окремим пактом, при чому в цьому пакті треба показати площу і врожай другорядних рослин серед рослини важнішої; у решти рослин, мішано плеканих, подати також дані, але в дужках.

Рослини, що збираються одна по одній. Коли на тому ж участку, в тому ж році засіяна та зібрана після одної рослини друга, заноситься перша рослина на відповідне місце до вільного рядку, а рослина друга теж у відповідний рядок, але поруч з пактом у дужках. Цим буде усунено подвійне урахування площі, інакше була б почислена більша площа під рослинами, яку має господарство в дійсності.

Рослини, що дають подвійні вигоди. Коли якась рослина дає різні вигоди, як наприклад, льон і коноплі /насіння і волокно/ визначається лише площа у головної вигоди, а в другої повторюється лише в дужках.

IУ. Зернові культури.

І. Пшениця озима.....	гект.	6. Просо.....	гект.
2. Пшениця ярова.....	гект.	7. Кукурудза.....	гект.
3. Овес.....	гект.	8. Риж /нелупаний/.....	гект.
4. Ячмінь.....	гект.	9. Сарго.....	гект.
5. Жито.....	гект.	10. Суржик /сумішка жита з	
Інші зернові	гект.	пшеницею/.....	гект.

У. Стручкові / на насіння/.

І. Боби.....	гект.	4. Сочевиця.....	гект.
2. Соя.....	гект.	5. Інші стручкові на зерно	
3. Горох.....	гект.	/насіння/.....	гект.

УІ. Коренеплоди та бульбоплоди на їжу і на корм.

І. Бараболя.....	гект.	4. Арорут /Arow - root/ /....	гект.
2. Батата і солодка ба-		5. Інші коренеплоди і буль-	
раболя.....	гект.	боплоди.....	гект.

УІІ. Трави і стручкові на корм.

І. Конюшина, флеол, райґрас та подібні інші рослини, чисті або помішані, засіяні протягом останніх 5років та призначені знову до оранки:
а/ зібрано на сіно..... гектарів.
б/ не зібрано на сіно..... гектарів.
2. Люцерна:
а/ зібрані на сіно..... гектарів
б/ не зібрано на сіно..... гектарів
3. Інші трави:
а/ зібрано на сіно..... гектарів
б/ не зібрано на сіно..... гектарів

УІІІ. Рослини промислові.

а/ Рослини цукрові:
І. Троща.....гект. 2. Буряк цукровий.....гект.
3. Інші цукрові.... гект.
б/ Рослини текстильні:
І. Бавовна /без зерна/....гект. 2. Льон на волокно.....гект.
3. Коноплі на волокно.....гект. 4. Льон новозеляндський.....гект.
5. Абака /коноплі манільські/.....гект. 6. Юта і под.....гект.
7. Китайська крапива /Ramie, rhea/..... гект.
8. Сейсал /мехіканські коноплі/.....гект. 9. Інші рослини текстильні..... гект.
в/ Рослини олійні:
І. Льняне насіння.....гект. 2. Конопляне насіння..... гект.
3. Бавовняне насіння.....гект. 4. Бурські горішки /Arachides/ ..... гект.
5. Рицина /Райдерево/....гект. 6. Ріпак і рапс..... гект.
7. Сезам .....гект. 9. Інші олійні рослини.... гект.
8. Соняшник.....гект.
г/ Інші промислові рослини:
І. Тютюн.....гект. 2. Хміль..... гект.
3. Горчиця .....гект. 4. Індіго..... гект.
5. Інші промислові рослини.....гект.

ІХ. Столові рослини / городина/

І. Шпарги..... гект. 2. Целер..... гект.
3. Капуста..... гект. 4. Зелений горох..... гект.
5. Цибуля..... гект. 6. Горох лущений..... гект.
7. Томати..... гект. 8. Інші столові /городина/..гект.

Х. Рослини плекані на насіння.

/Сюди не належать рослини плекані на насіння та наведені в інших розділах/.
І. Буряк цукровий на насіння..... гект.
2. Конюшина на насіння..... гект.
3. Люцерна на насіння ..... гект.
4. Інші рослини плекані виключно на насіння.. гект.

ХІ. Засіяні площі, на якій урожай був знищений.

Ціла площа.....гект.

ХІІ. Цілорічний пар.

Ціла площа ..... гект.

Завваження. Розмір площі, показаної в розд. IУ - XII мусить рівнятись площі, показаній в розд. II до питання 2-го.

XIII. Луки сталі і пасовиська.

I. Землі призначені до сталого плекання трав або за добу останніх 6 років /рахуючи в тому і участки, що поросли деревиною, коли вони призначені переважно на сінокоси або пасовиська/.

а/ призначені на сіно..... гект.

б/ не призначені на сіно..... гект.

XIУ. В и н н и ц і.

	Число кущів /лоз/. родючих	неродючих
I. Винниці на виробу вина.....гект.	.....	.....
2. Призначені на сушку ягід.....гект.	.....	.....
3. Призначені на столові ягоди..гект.	.....	.....

XУ. Оливкові сади.

	Число дерев родючих	неродючих
I. Оливи для виробу олія.....гект.	.....	.....
2. Оливи для овочів до їжі.....гект.	.....	.....

XУI. Овочеві сади.

/Лише засаджені дерева, а не дико ростучі/.

	Число дерев родючих	неродючих
I. Марелі і броскви..... гект.	.....	.....
2. Мигдал ..... гект.	.....	.....
3. Ананаси..... гект.	.....	.....
4. Банани..... гект.	.....	.....
5. Черешні..... гект.	.....	.....
6. Цитровники /інші ніж цитрони, помаранці і мандарини/..... гект.	.....	.....
7. Даклі /фініки/..... гект.	.....	.....
8. Фики,..... гект.	.....	.....
9. Цитрони..... гект.	.....	.....
I0. Оріхи..... гект.	.....	.....
II. Кокосові горіхи..... гект.	.....	.....
I2. Помаранці та мандарини..... гект.	.....	.....
I3. Груші..... гект.	.....	.....
I4. Яблоні..... гект.	.....	.....
I5. Сливи..... гект.	.....	.....
I6. Інші овочеві дерева..... гект.	.....	.....

XУII. Дрібні плекані овочі.

I. Ціла площа / сюди не належать дико ростучі/.....гект.

XУIII. Культури кущові і деревисті.

		родючі	неродючі
I. Плантації	кави..... гект.	....	....
2. "	чаю..... гект.	....	....
3. "	какао..... гект.	....	....
4. "	перцю звичайного..... гект.	....	....

		родючих	неродючих
5. Плантації хінхіни	гект.	.....	.....
6. Рослини каучукові, гумові, що дають живицю			
а/ плантації каучуку Рага / Havea/	гект.	.....	.....
б/ " " Мексиканського /Castillia /	гект.	.....	.....
в/ " " Ceara /Manihot/	гект.	.....	.....
г/ " " Ассан / Ficus Elastika/	гект.	.....	.....
ґ/ " дерев гутаперчових	гект.	.....	.....
д/ " інших дерев і кущів, що дають ґуму, живицю або каучук	гект.	.....	.....
7. Маньока	гект.	.....	.....
8. Шовковиця	гект.	.....	.....
9. Бамбук	гект.	.....	.....
IO. Ione de rottin	гект.		
II. Саго	гект.	.....	.....
I2. Інші дерева і кущі незазначені раніш	гект.	.....	.....

XIX. Господарські тварини.

/Дані подаються відносно господарських тварин, що дійсно маються в день перепису. Записати треба всіх тварин, незалежно від того чи вони є власністю господаря чи ні, а також і тих, що знаходяться під доглядом чабанів /пастухів/ в отарах, чередах на землях громадських, на луках, болотах, горах або у лісах - взагалі на дібрах:, що не є охоплені переписом сільсько-господарських підприємств.

		Число.
а/ Коні.	I. Лошачки і лошички до I року	.....
	2. Коні від I - 3 років	.....
	3. Коні більше 3-х років	.....
	4. Племенні кобили і жеребці більш 3 р.	.....
б/ Мули:	I. Мули I до 2-х літ.	.....
	2. " більше 2 літ.	.....
в/ Осли:	I. Загальне число	.....
г/ Худоба:	I Телята до I року	.....
	2. Яловки і корови більше року	.....
	3. Бугаї більш року	.....
	4. Воли більше року	.....
	5. Корови, що дояться в день перепису	.....
	6. Корови, що доїлись цілий або частину року I929. /Не рахуються продані при кінці року/.	.....
ґ/ Вівці:	I. Ягнята до року	.....
	2. Барани більше року	.....
	3. Овечки більше року	.....
	4. Решта більше року	.....
д/ Кози:	I. Загальне число	.....
е/ Свині:	I. Кнурята до 6 місяців	.....
	2. Племенні кнурі більш 6 місяців	.....
	3. Свиней племенних більш 6 місяців.	.....
	4. Решта свиней більш 6 місяців	.....
є/ Птиця /дріб/	I Загальне число курей, когутів і курчат.	.....
	2. Загальне число качок	.....
	3. " " гусей	.....
	4. " " індиків	.....
ж/ Бджоли:	I. Число вуликів в день перепису	.....

Д О Д А Т О К.

I. Виробництво рослинне.

Одним із найголовніших завдань перепису є одержати якнайточніші дані про розмір врожаю окремих рослин. Цієї мети можна досягти подвійним способом:

I. Тим, що кожного господаря запитується, яку кількість врожаю кожної рослини одержав на заявленій площі засіву. Теоретично ця метода здається задовольняючою. Але на практиці не є так, через те, що дуже часто кількість врожаю і не міряється і не занотовується. Досвід показав, що ця метода досить задовольняюча для рослин, врожай яких продається цілий зразу, як наприклад, бавовна, льон, деякі види зернових, овочі та городина в деяких країнах. Дійсно, продана кількість важиться та взагалі кількість врожаю відносно певного краю або округи можна контролювати статистикою руху та прибуття товарів на головний ринок. Метода є менш задовольняюча, коли ходить про рослини, які частково або цілком вжиті в господарстві, як наприклад, сіно, городина, овочі, молоко, дріб і яйця.

2. Тим, що спеціялісти відгадують пересічний врожай кожної рослини на одиницю площі або пересічний продукт від господарських тварин про кожну територіяльну одиницю, до якої відносяться дані перепису. Добутком площі посіву на пересічний врожай, або число тварин на пересічний продукт однієї тварини можна одержати цілу кількість за рік

Цінність результатів одержаних цією методою залежить від ступня досконалости, з якою можливо одержати під час перепису дані про площі засівів та про кількість господарських тварин, а також від уважности та точности, з якою здогадом обчислюється врожай пересічний. Коли обидва складники відшукані точно, є ця метода однаково безпечна як і інформації індивідуального господаря; крім того вимагає менш часу і видатків та не є складною. Через те, що Міжнародній Сільсько-Господарський Інститут має великий інтерес на тому, щоб встановити цілий врожай для кожної рослини та продукцію кожного виду господарських тварин, дуже просить всі держави, щоб забезпечили відомості про виробництво тою або іншою методою, або комбінуванням обох метод. Без цих даних був би перепис неповний.

II. Продукти тваринництва.

Наступні питання приєднуються для того, щоб можна було одержати прямо від господарів дані про деякі продукти від плекання тварин. Колиб спроба одержати інформації безпосередньо від господарів не дала добрих результатів, Інститут жадає від влад, щоб це виробництво було встановлено здогадом спеціялістів у формі окремих анкет.

А/ Здобич молочарства:

I. Ціла кількість молока одержаного в р. 1929 .....літр.

а/ коровячого..... .....літр.

б/ інших тварин..... .....літр.

/Записується молоко, що давалось іншим тваринам, спожите в господарстві, продане або зужите іншим способом. Не записується молоко виссане молодняком./.

2. Масло вироблене в господарстві .....клгрм.

3 Сир вироблений в господарстві..... .....клгрм.

Б/ Здобич від дробу:

I. Число яєць зібраних у 1929 році..... .....клгрм.

В/ Вовна:

I. Число овечок в 1929 р. у господарстві стрижених.....

2. Загальна вага зисканої вовни..... .....клгрм.

/ треба вказати чи вовна тучна чи виправна/

Г/ Шерсть козяча :
I. Загальна вага зисканої шерсти ..... клгрм.
Ґ/ Мед: I. Меду, одержаного в 1929 р. ..... клгрм.
Д/ Шовкопрядка: Кількість насажених яєць..... грам
Кількість свіжої опряді зисканої в 1929 р....пкпрм.

Питання евентуальні.

Адаємо від всіх держав аби прийняли до перепису р.1930 всі питання подані в цьому формулярі на попередніх сторінках та підготовили все необхідне для з'ясування вислідів виробництва згідно з пропонованими методами в додатку. Далі подаються інші питання, на які бажано було б одержати відомості по можливості від більшого числа країн.

Багато цікавих річей було з формуляру виключено, щоб перепис не був утруднений питаннями менш важливими, щоб були зменшені видатки на перепис ,або через те, що відповіді на них не були б задовольняючими. Сюди відносяться питання про ціни та вартість живої та мертвої ваги худоби, зужитковання продуктів в господарстві та інші подібні питання.

I. Відносини в користуванні землею.

I. Кілька гектарів є в Вашій власності або у співвласності одного або декількох членів Вашої родини, коли Ви не є наємцем господарства, або не одержуєте платні за управління ним ..... гект.

2. Кільки гект. із цього господарства Ви найняли іншим особам...гект.

3. Кількома гект.цього господарства управляєте для інших осіб, яко оплачений керовник ..... гект.

4. Кількома гект. цього господарства управляєте на інших умовах, ніж плачений управитель ..... гект.

З а в в а ж е н н я. Суму гектарів цих 4-х пактів слід порівняти з площею вказаною в пакті I. Розд. II цього формуляру.

2. Дренаж.

I. Площа землі цього господарства, яка забезпечена сталими штучними осушуючими канавами, каналами із цегли і т.п. .....гект.
/ Сюди відносяться земля дренажована сталими каналами, або рурами назовні або під землею/

3. Обводнення.

I. Подайте площу господарства, що є обводненою.....гект.
/Дані відносно лише штучних каналів, канав, рур, помп, що накачують воду. Ручне заливання сюди не стосується/

4. Робітники в господарстві.

Яко доповнення до даних про число робітників, що подаються в розділі III формуляру, пропонується поставити питання нижченаведені в тих країнах, які хочуть одержати інформації про добу праці, заробітну платню та умови житлові робітництва.

I. Денна доба праці, на протязі якої працюють робітники в даному господарстві за платню:
а/ Підчас жнив.....годин
б/ в інші періоди.....годин

2. Платня сільсько-господарських робітників:

	влітку		зимою	
	мужчин	жінок	мужчин	жінок
а/ у поденних платня за годину	...	...	....	...
в/ у місячних:				
I/ пересічна платня за місяць на особу в грошах	...	...	....	...
2/ пересічна платня натурою /страва, житло, паливо і п./ в грошовій вартості за місяць і особу	...	...	....	...

З а в в а ж е н н я. Сума I/ і 2/ мусить дати цілу пересічну місячну платню робітника.

3. Житло робітників, працюючих за плату та осіб зайнятих у господарстві:

Із загального числа осіб, зайнятих у цьому господарстві:

	Мужчин	Жінок
I. Скільки користується окремим помешканням, що належить господарству ?		
а/ за рахунок части своєї платні	.....	.....
б/ без якого будь зв'язку з платнею	.....	.....
2. Скільки їх живе в підприємстві		
а/ в помешканні господаря	.....	.....
б/ в спеціяльних робітничих помешканнях	.....	.....
в/ в інших будовах або якось інше	.....	.....
3. Скільки їх живе поза господарством або незалежно від від нього	.....	.....

5. Угноєння.

Угноєння є важливим чинником сільсько-господарського робітництва і в багатьох країнах бажано одержати відповідні дані про його види і кількість. Пропонуються дві методи:

I. Безпосередня анкета у сільських господарів в зв'язку з сільсько-господарським переписом.

2. Анкетою через промисловців та торговців штучними угноєннями, а також через контрольні уряди або регулюючі довіз уряди / фінансові/ крайові або провінціяльні.

6. Запаси в руках сільських господарів.

Обчислення готових запасів сільсько-господарського виробництва на певний момент має велике значіння для кожної країни і для народньої та міжнародньої торговлі. Видимі запаси, себто запаси, що знаходяться у транспорті або в громадських елеваторах, можна почислити задовольняюче, бо тут іде про велику кількість, яка порівнюючи легко встановлюється. Невидимі запаси, себто запаси, які ще знаходяться у господарів, або в першого покупця чи в млинах, або у приватних складах, які, одначе, є частиною всіх запасів, дуже тяжко обчислити і навіть встановити здогадом. Через те пропонується, щоб уряди підчас перепису встановили запаси продуктів сільського господарства певного значіння,як,наприклад, кукурудзи, текстильних рослин, тютюну, кави, чаю, вовни і т.п., які малися на продаж у маєтках, або в млинах, складах або в сільських господарів.

Такі відомості мали б велику ціну, яко основа для обчислення здогадів щорічних запасів.

7. Сільсько-господарські машини.

Питання, про яке треба зібрати інформації, дотикається перепису сільсько-господарських машин та знарядля. Такий перепис є можливий включенням відповідних питань до формуляру. Окремі види машин можуть бути додані у формулярі згідно з потребою кожної країни. Є бажаним, одначе, щоб Інститут одержав відомості про види машин, які він пропонує для кожної країни, щоб забезпечити порівняння даних між різними країнами.

Треба з'ясувати кількість таких машин у господарстві в день перепису:

а/ плугів
б/ борон
в/ сіялок
г/ косарок
ґ/ молотилок
д/ ґенераторів, погонних сил.

8. Школки, оранжерії та парники.

I. Кількість кв, метр. під шклом I/I 1930 р. .....гект.
2. " гектарів відкритого простору .....гект.
3. Сума одержана р. 1929-го за продаж молодих овочевих дерев, кущів або саженців винограду .....франків
4. Сума, одержана за продаж молодих лісових дерев та оздобних кущів.....франків
5. Сума одержана за квітки та за розсаду їх.....франків
6. Сума одержана за продаж городини та за розсаду.....франків

9. Квіти.

Плекані головним чином на продаж.....гект.
/Записати гектари та десяти частини гектарів. Занотувати площу засаджену головно квітами на продаж. Не заносити квіти плекані головно на насіння, яке належить занести до пакту 4 розд.IX формуляру, або квіти плекані в оранжеріях вже занесені до п.5 розд.8 цих питань евентуальних./

IO. Особливі домашні тварини в деяких країнах.

/Подати число згідно із станом в день перепису/.

	Кількість.
I. Верблюди: молодші року.....	.....
старші року.....	.....
2. Лами, гуанаки[x]/ і вікуни[x]/.....	.....
3. Олені	.....
4. Слони: молодші року	.....
старші року	.....
5. Буйволи: а/ телята до року	.....
б/ молоді від I-2-х років	.....
в/ решта старші 2-х років.....	.....
6. Струсі	.....
7. Звірята плекані в неволі на футри і шкіри:	
а/ лисиці	.....
б/ скунці	.....
в/ інша звірина	.....
г/ зайці та кролики	.....

x/ Від лами, що водяться в Болівії, Еквадорі та в Андах Південної Америки.

До курсу "Основи теорії статистики"

Д О Д А Т О К 2.

/ до лекції шостої /

Основний формуляр американського індустріяльного цензу 1925 р.

Д о в і р о ч н о. ряд № .....

Ф о р м а 100.

ДЕПАРТАМЕНТ ТОРГІВЛІ.

ЦЕНЗОВЕ БЮРО.

Вашинґтон.

Законом встановлено обов'язок для кожного промисловця подати цензові відомості. Строга дискретність буде дотримана для всіх відповідів.

І Н Д У С Т Р І Я Л Ь Н И Й Ц Е Н З, 1 9 2 5

пожаданий актом конґресу, ствердженим 3 березня 1919 року.

Г Е Н Е Р А Л Ь Н А С К Е Д У Л А /

Відомості жадаються від всіх підприємств, які були в чинности хоч яку-будь частину року, закінченого 31 грудня 1925 р., але на бажання, дані можуть відноситись до ділового року, що найбільше відповідає календарному року. Цензове Бюро не робитиме жадних публікацій, які б відкривали операції поодиноких підприємств; жадному незаприсяженому урядовцеві не буде дозволено продивлятись звідомлення.

---

1. ОПИС ПІДПРИЄМСТВА.- Окремі звіти жадаються від підприємств, що знаходяться в різних штатах і для тих в окремих містах, що мають 10.000 мешканців або більше.

a. Назва підприємства.....

b. Прізвище власника або керівника.....

c. Місце осідку закладу або фабрики { Штат..... місто..... вулиця та Ч.....графство.....

d. Чи заклад знаходиться в межах території, що належить місту? /Так чи ні/

e. Поштова адреса головної контори.....

/Коли цей звіт охоплює більше ніж одну фабрику, подайте назву і місце кожної під "Примітками"/

---

2. ХАРАКТЕР ІНДУСТРІЇ.

a. Продукти.....

/перечислити головні роди вироблених продуктів або зробленої праці/.

b. Вживані матеріяли.....

/перечислити головні вживані матеріяли в порядку важливости/

3. ДОБА ЧИННОСТИ:

a. Число днів, в які підприємство було в чинності на протязі року....

b. Число нормальних годин праці найманих робітників для зміни..... на тиждень.....

---

4. ЗАНЯТІ ОСОБИ: - /Не повідомляйте за кожну особу більш ніж під одною з літер - a, b, c /.....

a. Власник або члени фірми /не прикладайте це до сполучених товариств/.....

b. Службовці за платню - Подайте число платних урядовців, корпорацій дозорчих, директорів, контрольорів, бухгальтерів, зборщиків і інших надзірчих і адміністративних службовців на виплатному реєстрі 15 грудня, або на найближчий репрезентаційний день .....

c. Робітники, включаючи і поштучних робітників. Подайте число всіх робітників кваліфікованих і некваліфікованих, таких як механіки, машинові оператори, зборщики, упаковщики та інші, включаючи і поштучних робітників, занятих у операціях виробництва на 15-е кожного місяця, або на найближчій репрезентаційний день:

Січень..... Лютий..... Березень..... Квітень..... Травень.....

Червень..... Липень..... Серпень..... Вересень..... Жовтень.....

Листопад..... Грудень.....

---

5. УТРИМАННЯ, ПЛАТНЯ І МАТЕРІЯЛИ. /Подайте лише найточніше визначені видатки; капітал і різні мішаного характеру видатки не мусять сюди включатись. Не подавайте жадного роду видатків більше ніж визначених під одною з літер a, b, c і d/.

a. Утримання - Подайте цілу квоту, виплачену платним урядовцям, показаним вище /Питання 4.літ.c / на протязі періоду, що охоплюється цим звітом..... .....Долярів.....

b. Платня - Подайте цілу квоту, виплачену платним робітникам по азаним вище /Питання 4 літ.c./ на протязі періоду, охопленого цим звітом..... ..... Долярів.....

c. Вартість вжитих матеріялів.- Подайте цілу вартість всіх матеріялів, або сирівців, фабричних засобів, палива, сили і упаковочних матеріялів, проданих з товарами, дійсно вжитих у виробництві дібр, яке провадилось на протязі періоду, охопленого цим звітом та охоплює продукти під п.6.....Долярів.....

d. Контрактова праця.- Подайте цілу квоту, виплачену за працю, зроблену іншими особами ніж вищі службовці, з матеріялами, доданими вашим підприємством, у виробництві предметів, вказаних під запитом 6. Не включайте коштів будівельної і ремонтної праці іншого підприємства.....Долярів.....

Р а з о м Долярів.....

---

6. ПРОДУКТИ /ФАБРИКАТИ/: Вартість продуктів і праці виконаної на протязі року. - Подайте продажну вартість loco фабрика всіх без виключення продуктів, вироблених на протязі року /як продані, так і ні/ та з побічними продуктами включно. Головні продукти мусять бути показані окремо і загальна кількість і вартість подана для кожного. До вартости ураховується і кошти

упаковки, коли остання продана разом з товарами. Сумарні розписки контрактової праці і податкової праці, зроблених для других на протязі року, мусять бути подані в рядках f і g

П р о д у к т и /Подайте головні продукти окремо/.	Кількість	Вартість
a.....	.....	Доляр.....
b.....	.....	Доляр.....
c.....	.....	Доляр.....
d. Всі інші /означити/.....		Доляр.....
e.....		Доляр.....
f. Контрактова праця і матеріяли додані іншими /сумарні розписки/...		Доляр.....
g. Податкова праця і ремонтування /сумарні розписки/...		Доляр.....
Ціла вартість продуктів...		Доляр.....

h. Який відсоток складає ціла вартість, вказана вище, до можливої вартости при умові праці підприємства на повну здібність?
Оцінено в.....%

7. СИЛА.- Подайте всі силові заряджения в підприємстві активні і запасові /Подайте оцінку моторів у кінських силах і ґенераторів або динам у кіловатах.

К л я с а	Число	Число кінських сил
Первісні /головні/ двигуни:		
a. Парові машини.....	.....	.....
Парові турбіни.....	.....	.....
Мотори внутрішнього горіння /олійні, газові, газоолійні і т.ін./.....	.....	.....
Водяні кола і водяні турбіни.....	.....	.....
Електричні мотори, що працюють купленим током.....	.....	.....
b. Другорядна сила: Електричні отори, що працюють власним током.....	.....	.....
		Сумарна оцінка кіловатів
c. Електричні ґенератори або динами.....	.....	.....

	Погонна сила	Для інших індустр.цілей: Індустріяльне нагрівання	Електролітичні і інші електрохемічні цілі.
d. Електричний ток загалом, спожитий на протязі року: Куплений /кіловат-години/...	.....	.....	.....
Вироблений в підприємстві /кіловат-години/.....	.....	.....	.....

e. Середнє електричне індустріяльне навантаження в I925 р. кіловати: Загалом...../I/ Для погонної сили...../2/ для індустр.нагрівання..... /3/ для електролітичних і інших електрохемічних цілей.....

Цим свідчу, що інформації подані в скедулі є правильні і вичерпувчі та охоплюють період від.....1925 р. до ..... 1925 р.

.....
Підпис і офіційне звання особи, що подала інформації.

.....
Підпис спеціяльного агента

.....
адреса

Примітки: .....

.....

x - x - x - x - x

Спеціяльний формуляр американського індустріяльного цензу 1925 р.

Довірочно.

Форма 169.

ряд.№.....

Так само як і в формі Ч.100.

ДЕПАРТАМЕНТ ТОРГІВЛІ.

ЦЕНЗОВЕ БЮРО.
Вашинґтон.

ІНДУСТРІЯЛЬНИЙ ЦЕНЗ, 1925.

пожаданий актом конґресу, ствердженим 3 березня 1919 року.

БУРЯКОВИЙ ЦУКОР.

Звіти жадаються від всіх підприємств, які були в чинності хоч яку будь частину року закінченого 31 грудня 1925, але на бажання, дані можуть відноситись до ділового року, що найбільше відповідає календарному року. Цензове Бюро не робитиме жадних публікацій, які б відкривали операції поодиноких підприємств; жадному незаприсяженому урядовцеві не буде дозволено продивлятись звідомлення.

1. ОПИС ПІДПРИЄМСТВА.
2. ХАРАКТЕР ІНДУСТРІЇ.
3. ДОБА ЧИННОСТИ.
4. ЗАНЯТІ ОСОБИ.
5. ПЛАТНЯ І МАТЕРІЯЛИ

тотожні як у генеральній скедулі форма 100.

6. ПРОДУКТИ - Подайте кількість ізагальну продажну вартість фабрика всіх вироблених продуктів на протязі сезону 1924-1925 р.р. /як проданих так і ні/, а також побічні продукти, включаючи оцінену вартість тих, що зроблені по контракту, або на основі поділу, зглядно клясифікації що наведена вище. Тех мусять бути включені і вартости проданої разом з продуктами упаковки.

Р і д	Одиниця міри	Кількість	Вартість
Ц У К О Р:			
Дроблений	фунти.....	.....	Дол.....
Недроблений /Всі матеріяли вжиті в процесі виробки обчислюється як еквівалент дробленого цукру/	тони,2000 ф.....	.....	Дол.....
Мелясси - відкинуті від процесу сепарації цукру, але не включаючи ті, що вживаються для мелясової бурякової маси	тони,2000 ф.....	.....	Дол.....
Меляси /інші ніж відпадки/, продані для цілей інших ніж сепарація цукру	тони, 2000 ф.....	.....	Дол.....
Сироп./проданий для людського споживання/.	галони.....	.....	Дол.....
Сушена бурякова маса /Виключаючи мелясову бурякову масу/	тони, 2000 ф.....	.....	Дол.....
Вогка бурякова маса.	тони, 2000 ф.....	.....	Дол.....
Мелясова бурякова маса	тони, 2000 ф.....	.....	Дол.....
Всі інші продукти /означити/:.....			Дол.....
.....			Дол.....
Р а з о м.....			Дол.....

Оцінка навантаження підприємства в %% як у формулярі I00 під п.6.

7. БУРЯКИ ЗУЖИТІ І ЗАСАДЖЕНА ПЛОЩА НА РІК, що охоплений цим звітом:

С т а т т я	Зужито буряків /тони в 2000 ф/	Число засаджених акрів	Вартість
Культивовані прямо цукроваром	.....	.....	Дол.....
" арендаторами цукровару.....	.....	.....	Дол.....
" по контракту іншими особами ніж арендарами.....	.....	.....	Дол.....
Куплені зужиті буряки.....	.....	---------	Дол.....
Зужиті на підставі контракту, або на основі поділу.....	.....	---------	Дол.....
Р а з о м			Дол.....

8. ЯКІСТЬ ЗУЖИТИХ БУРЯКІВ:

Середній %% сахарози..... Середній коефіцієнт чистоти.....

9. ПОДАЙТЕ ПРОЦЕС СЕПАРАЦІЇ ЦУКРУ З МЕЛЯСИ, як що це робиться.

.....

.....

.....

IO. СИЛА і підтвердження щодо правдивости поданих відомостей, а також підписи осіб і примітки тотожні з генеральною скедулою форма IOO.

x - x - x - x - x - x

Для того, щоб не заплишити поза промисловим численням /цензом/ чисто адміністративних управлінь, підприємств, або генеральних контор їх, що найчастіше знаходяться у великих промислових містах і безпосередньо територіяльно та технічно майже не зв'язані з технічним виробництвом або фабриками певного підприємства, Бюро цензу вживає також окрему скедулу - форми IOI, під назвою: "Формуляр /скедула/ адміністративний і генеральної контори".

В ньому мусить бути показано: назва підприємства, прізвище власника або керівника, точне місце находження генеральної контори. Число службовців генеральної контори, себто такої, яка контролює дві або більше фабрик та які не можуть бути приділені до яких-будь інших закладів підприємства.

Таким чином, ця група службовців не може бути показана на генеральній скедулі-форма IOO, а також у спеціяльних формулярах; число урядовців, їх платня за рік охоплений цензом також показується в цьому формулярі адміністративному - форма IOI.

Нарешті в цьому формулярі вимагається подати назву і місце находження всіх закладів виробництва, які контролюються також генеральною конторою є цензове Бюро через це має матеріял про те, які технічно-виробничі одиниці входять в те або інше ціле підприємство.

Д о д а т о к 3.

/ до лекції 23./

Інтерполяційна таблиця для знайдення границь абсолютної похибки релятивних чисел /відсотків/ для вибору переведеному в відношенні 1:7 / себто N =7n /.

Для відносної частоти 100 m/n =	Для абсолютного числа одиниць /випадків/ у виборі n =								
	100	200	300	400	500	600	700	800	900
	Границі похибки відповідного відсотка /гр.1/ складають плюс мінус %:								
2%	±2,75	1,94	1,59	1,38	1,23	1,12	1,04	0,97	0,92
4%	3,85	2,74	2,22	1,92	1,72	1,57	1,45	1,36	1,28
6%	4,67	3,30	2,70	2,33	2,09	1,91	1,77	1,65	1,56
8%	5,33	3,77	3,08	2,66	2,38	2,18	2,01	1,88	1,78
10%	5,89	4,17	3,40	2,94	2,64	2,40	2,23	2,08	1,96
12%	6,39	4,51	3,69	3,19	2,89	2,61	2,41	2,26	2,13
14%	6,82	4,82	3,93	3,41	3,04	2,78	2,58	2,41	2,27
16%	7,20	5,09	4,16	3,60	3,22	2,94	2,72	2,55	2,40
18%	7,55	5,34	4,36	3,77	3,37	3,08	2,87	2,66	2,52
20%	7,86	5,55	4,54	3,93	3,51	3,21	2,97	2,78	2,62
22%	8,14	5,76	4,70	4,07	3,64	3,32	3,08	2,88	2,71
24%	8,39	5,93	4,84	4,19	3,75	3,42	3,17	2,97	2,80
26%	8,62	6,09	4,97	4,31	3,85	3,52	3,26	3,05	2,87
28%	8,82	6,24	5,09	4,41	3,94	3,60	3,33	3,12	2,94
30%	9,01	6,37	5,20	4,50	4,03	3,67	3,40	3,18	3,00
32%	9,17	6,48	5,29	4,58	4,10	3,74	3,46	3,24	3,05
34%	9,31	6,58	5,37	4,65	4,16	3,80	3,52	3,29	3,10
36%	9,43	6,67	5,44	4,71	4,22	3,85	3,56	3,33	3,14
38%	9,54	6,75	5,50	4,77	4,26	3,89	3,60	3,37	3,17
40%	9,63	6,80	5,56	4,81	4,30	3,93	3,64	3,40	3,21
42%	9,70%	6,86	5,60	4,85	4,34	3,96	3,66	3,43	3,23
44%	9,76	6,89	5,63	4,87	4,36	3,98	3,69	3,45	3,25
46%	9,79	6,92	5,65	4,89	4,38	4,00	3,70	3,46	3,26
48%	9,82	6,94	5,67	4,91	4,39	4,01	3,71	3,47	3,27
50%	9,82	6,95	5,67	4,91	4,39	4,01	3,71	3,47	3,27

SLAVISCHE BIBLIOTHEK IN PRAG
SLOVANSKÁ KNIHOVNA V PRAZE
22. II. 1938
1122.