

## Розділ 1. Основи біометрії

### Тема 1. Мета і задачі дисципліни. Основні визначення. Історія розвитку біометрії (Б).

Б – наука про статичний аналіз масових явищ в біології, тобто таких явищ, в масі яких виявляються закономірності, що не виявляються на одиночних випадках спостережень.

Предметом Б служить будь-який біологічний об'єкт, якщо спостереження, що проводяться над ним, отримують кількісне вираження. Зазвичай спостереження проводяться не на одиночних, а на групових об'єктах, наприклад на особах одного і того ж виду, статі, віку, які розглядаються як складові елементи або члени групового об'єкту і називаються одинацями спостереження. Сукупність таких відносно однорідних, але індивідуально різних одиниць, що об'єднуються в відношенні деяких загальних умов для спільного (групового) вивчення, називається статистичною сукупністю (СС).

Поняття СС – одне з фундаментальних понять біометрії. Воно базується на принципі якісної однорідності її складу. Не можна об'єднувати в одну групу індивідуумів різної статі і віку, коли вирішуються такі, наприклад, задачі, як нормування харчування, стандартизації взуття і одягу, виготовленого для масового споживання, оскільки наперед відомо, що з віком і залежно від статі, індивідуумів змінюється їх потреба в харчуванні і не залишаються без зміни розміри і пропорції тіла. Не можна вивчати закономірності модифікації (спадкових змін фенотипа) на генетично неоднорідному матеріалі, об'єднуючи їх в одну сукупність чистопородних особин тощо.

СС може складатися не лише з аморфної маси однорідних одиниць, але і з різних за складом внутрішньо однорідних груп (особин, кліток тощо), що об'єднуються відносно прийнятих в досліді умов для спільної обробки. У таких випадках сукупність початкових даних називається статистичним комплексом (СК). Питання про структуру СС вирішується дослідником залежно від об'єкту і мети дослідження. Якої б форми і вміст не набувала СС, вона завжди представляє деяку систему, що не зводиться до арифметичної суми складових її одиниць або компонентів. У СС існує внутрішній зв'язок між частиною і цілим, одиничним і спільним, яка знаходить своє вираження в статистичних законах, що діють у сфері масових явищ. На ці закони спирається Б.

### Ознаки, їх властивості і класифікація

Спостереження над біологічними об'єктами проводяться за тими або іншими ознаками, тобто такими характерними особливостями в будові і функціях живого, за якими можна відрізнити одну одиницю спостереження від іншої, порівнювати між собою.

Всі біологічні ознаки варіюють, тобто змінюються від випадку до випадку в визначених межах. Варіювання – характерна властивість всього живого. Не треба мати виняткову спостережливість для того, щоб в масі однорідних особин бачити

більш менш помітні індивідуальні відмінності у величині, забарвленні, поведінці і інших ознаках і властивостях індивідуумів.

У всіх таких і подібних випадках величина кожної ознаки коливатиметься в деяких межах від однієї одиниці спостереження до іншої. Ці коливання величини однієї і тієї ж ознаки, спостережувані в спільній масі його числових значень, називається варіаціями (від лат. *variatio* – зміна, коливання), а окремі числові значення варійованої ознаки прийнято називати варіантами (від лат. *varians*, *variantis* – помітний, такий, що змінюється).

Всі біологічні ознаки варіюють, але не всі піддаються безпосередньому виміру. Звідси слідує їх поділ на: якісні або атрибутивні і кількісні. Якісні ознаки не піддаються безпосередньому виміру і враховуються по наявності їх у членів даної сукупності (наприклад, стать, масть, колір пелюсток).

Поділ ознак на кількісні і якісні умовне (хоча і необхідне з точки зору Б): у кожній якості можна виявити множину кількісних градацій (градації кольору пелюсток квітки).

Біологічні ознаки можна класифікувати по-різному в залежності від того, що береться за основу класифікації. Якщо основу класифікації складає той або інший спосіб угруповання біометричних даних, то ознаки ділять на альтернативних, порядкових, рангових і так далі. Спільною ж основою для класифікації ознак служить в одних випадках міра, а інших – рахунок. Це відноситься не лише до кількісних, але і до якісних ознак. На цій підставі біологічні ознаки поділяються на: мірні або метричні і рахункові (меристичні). Мірні ознаки варіюють безперервно: їх величина може в визначених межах (від – до) приймати будь-які числові значення, тоді як рахункові ознаки варіюють переривчасто або дискретно – їх значення завжди виражаються тільки цілими числами.

На мові математики величина будь-якої варійованої ознаки називається змінною випадковою величиною. На відміну від постійних величин, що позначаються початковими буквами латинського алфавіту, змінні величини прийнято позначати останніми в латинському алфавіті прописними буквами  $X$ ,  $Y$ ,  $Z$  ..., а їх числові значення, тобто варіанти – відповідними рядковими буквами того ж алфавіту  $x_1, x_2, x_3, \dots, x_n$  або  $y_1, y_2, y_3, \dots, y_n$  і так далі. Загальне позначення будь-якої варіанти позначається символом  $x_i, y_i, z_i$  і т. д., де індекс  $i$  вказує на порядковий номер варіанти.

#### Джерела варіювання ознак

Ознаки варіюють під впливом різних, у тому числі і чисельних випадкових величин. На результатах же спостережень позначаються ще і похибки, що допускаються при вимірах. Досвід показав, що виміри, як би точно вони не проводилися, завжди супроводжуються більш менш помітними похибками. Ці похибки або помилки виникають через недостатню точність вимірювальних приладів і інструментів (технічні похибки), від особистих якостей дослідника, його навиків і майстерності в роботі (особисті помилки) і ряду інших, що не піддаються регулюванню і усуненню, причин (випадкові похибки).

Технічні і особисті похибки, що об'єднуються в категорію систематичних, тобто не випадкових, можна значною мірою зменшити, удосконалюючи технічні

засоби, умови спостережень і особистий досвід. Ці заходи дозволяють звести розміри таких похибок до мінімуму, якими можна нехтувати. Випадкові ж похибки залишаються і, разом з природним варіюванням, впливають на результати спостережень. Проте в порівнянні з природним варіюванням ознак випадкові похибки вимірів, як правило, незначні. Тому варіювання результатів спостережень розглядається зазвичай як варіювання ознак, що вивчаються.

### Точність вимірів

Практично кожна ознака має свою міру. Як показує досвід, немає необхідності гнатися за уявною точністю вимірів, коли вона практично не потрібна. Це положення відноситься і до вимірюваних об'єктів і до обчислень узагальнених статистичних характеристик. "Обчислення, – писав А.Н. Крилов в "Лекції про наближені обчислення", – можна проводити як завгодно точно, але результат обчислення не може бути точніше за тих даних ..., на яких воно засноване".

Значення Б в дослідницькій роботі і в професійній підготовці біологів і фахівців суміжних областей знань стало очевидним вже тоді, коли були відкриті статистичні закони, що діють у сфері масових явищ. Але біологи не відразу оцінили важливість цих відкриттів. Потужним стимулом до розвитку експериментальних досліджень і використання кількісних методів стало еволюційне вчення Ч. Дарвіна. Дарвінізм дозволив глибше заглянути в суть біологічної мінливості і зробити перехід від типологічної, жорстко детерміністської концепції виду до ймовірностно-статистичного механізму еволюції.

З розвитком експериментальної біології застосування біометрії до вивчення варіюваних об'єктів стало насущною необхідністю. Утвердився і правильний підхід до Б, в якій сталі бачити не просто повчання до використання різних технічних прийомів, вживаних при обробці результатів вимірів, а науку, що займається статистичним аналізом масових явищ в біології.

Сучасна Б – обширна галузь знань, в ній нерозривно пов'язані питання планування біологічних експериментів і методика статистичного аналізу їх результатів.

Науково технічний прогрес і що з року в рік зростаючий потік наукової інформації призвели до перегляду багатьох традиційних форм і методів досліджень. У пошуках принципово нових вирішень біологічних проблем виникла ідея цілісності: окремі факти почали розглядатися не ізольовано один від одного, а як невід'ємні частки деякої системи, як наслідок спільної закономірності, що властива даній системі в цілому. Системний підхід призвів до заперечення колишнього уявлення про структуру живого як сума взаємодіючих часток. Спільне почало розглядатися ні як сума складових його компонентів, а щось більше, внутрішньо пов'язана множина складним чином взаємодіючих елементів, що не зводиться до їх сумарних властивостей і тим більше до властивостей кожного окремо узятим елементу структури.

Системний підхід до аналізу складних явищ – одна з провідних ідей сучасного природознавства, у тому числі і біології. Ця ідея і пов'язані з нею

способи моделювання складних явищ зачіпають і сам спосіб мислення. Б має безпосереднє відношення до всіх цих фактів. Спираючись на закони, що діють в статистичних сукупностях, Б озброює дослідника не лише потрібними знаннями в області статистичного аналізу масових явищ, розширюючи тим самим його науковий кругозір. Б розкриває перед ними діалектику зв'язку між часткою і цілим, між одиничними фактами і їх сукупністю, між причиною і наслідком, випадковим і необхідним в явищах живої природи. Вона показує, що в хаосі випадковості, який кається, виявляються закономірності, доступні опису точними математичними методами. Б служить об'єктивною основою порівняльного аналізу, без якого неможливе пізнання реальної дійсності.

Б вносить вагомий вклад до скарбниці наших знань про природу і про способи її пізнання. При цьому вона не робить замах на історично складенні і до цього часу оправдані описові методи дослідження, не заперечує і не підмінює їх, а лише озброює дослідника ідеями і методами, які змінюються при вивченні варіюючих об'єктів.

Б – наука формальна і використовувати її потрібно уміло, з врахуванням специфіки явищ, що вивчаються. Перефразовуючи відомі слова Гекслі, можна сказати: Б – це свого роду жорнова, ”... які всяку засипку змеле, але цінність помелу визначається виключно цінністю засипаного“.

### Відмінні риси Б, її місце в системі біологічних наук

Б склалася в прикордонних областях (на стику) між біологією і математикою. Її розвиток пов'язаний з перетворенням біології з науки описової в науку точну, що заснована на вимірах, на використанні кількісних оцінок при вирішенні біологічних задач. З формальної точки зору Б представляє ”сукупність математичних методів, вживаних в біології“. Ці методи вона запозичує головним чином з області математичної статистики і теорії вірогідностей.

Це не ототожнює Б з математичною статистикою (МС) і тісно пов'язаною з нею теорією вірогідностей (ТВ). Тут треба зважати на специфіку Б і її відмінні риси.

МС і ТВ – розділи математики: це науки суто теоретичні, що розглядають масові явища в абстрактній формі, незалежно від їх конкретного змісту.

Б – наука прикладна, така, що має справу з конкретними фактами, які вона аналізує за допомогою методів МС і ТВ, що складають в сукупності те, що називають статистичним аналізом. Б має на увазі не математичні а суто біологічні цілі, використовуючи математико-статистичні методи, але власним кутом зору, пристосовувавши їх до задач і специфіки біологічних досліджень. Б – це розділ біології, а не математики; вона має свій предмет і займає визначене місце в системі біологічних наук.

Зв'язки сучасної біології з математикою багатосторонні. На їх стику виникли різні напрями математичній біології. Кожен напрям має свої задачі і відносно до них використовує відповідні математичні методи.

Б хоча і тісно пов'язана з математичною біологією, але ототожнюватися з нею не повинна. Математична біологія підходить до вирішення біологічних задач дедуктивно, висуваючи на перший план математичні моделі з подальшою

перевіркою їх дослідом. Б ж спирається на індуктивний метод, відправляючись від конкретних фактів, які вона аналізує методами математичної статистики і ТВ.

Сучасна Б – це розділ біології, вмістом якого є планування спостережень і статичний аналіз їх результатів. Причому під спостереженням в широкому сенсі мають на увазі процес планомірного отримання і накопичення фактичних даних незалежно від того, як він здійснюється – в експерименті або безпосереднім описом явищ, що вивчаються. Планування спостережень і обробка їх результатів – безперервно пов'язані задачі статистичного аналізу.

### Історія становлення біометрії як науки

Біометрія як відносно самостійна наукова дисципліна склалася в другій половині XIX ст. Проте її джерела витікають до більш раннього періоду в історії природознавства: на той час, коли виміри біологічних об'єктів почали розглядати як метод наукового пізнання. Буржуазне суспільство, що прийшло на зміну феодалізму, потребувало розвитку точних знань про природу; актуальним для цього часу став афоризм Галілея (1564–1642): „Вимірною все вимірюване і зроби не вимірюване вимірюваним”.

У 1614 р. з'явилася книга Сантаріо (1561–1636) „Про статичну медицину”. В 1680 р. вийшла в світ книга Бореллі (1608–1679) „Про рух тварин”. У 1768 р. французький гіпполог Буржеля видав свою працю „Екстер'єр коня”. У цій книзі наведено набір вимірів, необхідних для визначення придатності коней до тієї або іншої служби. Характерно, що в цей же час, тобто в XVIII столітті, розвивається військова антропологія, що спирається на результати виміру тіла чоловіків призовного віку з метою відбору придатних до несення військової служби. Підставою для кількісної оцінки будови тіла тварин і людини служив, очевидно, той факт, що зовнішні параметри тіла тварин, а також і будова тіла людини знаходяться в певному зв'язку з їх фізичними і психічними властивостями. Щоб точніше виразити цей зв'язок; візуальну оцінку властивостей тіла тварин і людей по їх зовнішньому вигляду (екстер'єру) почали доповнювати його вимірами. А оскільки результати вимірів варіювали, потрібно було досліджувати цю мінливість. У 1718 р. в Лондоні вийшла в світ книга французького математика А. де Муавра (1667–1754) „Вчення про випадки”. Вимірявши зріст у 1375 дорослих жінок і розташувавши результати вимірів в ряд, він виявив закономірність, що відповідає відомому в теорії вірогідності закону нормального розподілу. Виникла необхідність інтеграції методів біології з методами теорії вірогідності і математичної статистики.

Теорія вірогідності і математична статистика виникли в середині XVII ст. незалежно одна від одної. Стимулом до обґрунтування теорії вірогідності послужив розвиток грошових стосунків в буржуазному суспільстві. Відому роль при цьому зіграли азартні ігри – кидання монет, гральних костей, ігри картярства, – які виявилися простими моделями, що дозволили відмітити закономірність в поведінці випадкових подій масового характеру. Біля витоків теорії вірогідності стояли французькі учені П. Ферма (1601–1665) і Б. Паскаль (1623–1662), а також голландський математик і природодослідник Х. Гюйгенс (1629–1696).

Вагомий вклад до становлення теорії вірогідності внесли Я. Бернуллі (1654–1705) і А. де Муавр. Проте найбільш істотний розвиток отримала ця теорія в працях таких видатних математиків, як П. Лаплас (1749–1827), К. Гаус (1777–1855), С. Пуассон (1781–1840), а також П.Л. Чебишев (1821–1894) і його петербурзька школа.

Розвиток математичної статистики пов'язаний з проблемами державознавства. До середини XVII століття в економічно розвинених країнах Європи накопичилася така кількість відомостей про демографію, страхову справу, а також в області торгівлі, охорони здоров'я і інших галузях господарства, що знатися на них за допомогою способів описової статистики стало майже неможливим. Назріла гостра необхідність пошуку нових методів аналізу статистичних даних, їх теоретичного обґрунтування. Задача зводилася до того, щоб по частці судити про стан цілого, тобто по вибірці робити висновок про всю сукупність суспільних явищ в цілому, повний опис яких ставав справою дуже трудомісткою і дорогою. Розробка теорії вибіркового методу зближувала математичну статистику з висновками теорії вірогідності, що з'явилося важливою віхою на шляху до виникнення біометрії.

Першим, хто вдало об'єднав методи антропології і соціальної статистики з висновками теорії вірогідності і математичної статистики, був бельгійський антрополог і статистик А. Кетле (1796–1874). У 1835 р. в Брюсселі вийшла книга Кетле „Про людину і розвиток його здібностей або досвід соціальної фізики”. Друге видання цієї книги з'явилося в 1869 р. під заголовком „Соціальна фізика або досвід дослідження про розвиток людських здібностей”. На великому фактичному матеріалі Кетле вперше показав, що досить різні фізичні особливості людини і навіть його поведінка підкоряються загалом, закону розподілу вірогідності, що описується формулою Гауса-Лапласа. В іншій праці, „Про соціальну систему і закони, що керують нею” (1848), Кетле описав людське суспільство не як суму індивідів або співтовариство людей, що проживають на певній території, а як деяку систему, що підкоряється суворим законам природи, що не залежні від волі людей. Нарешті, в праці „Антропологія” (1871) Кетле показав, що відкриті ним статистичні закономірності поширюються не лише на людське суспільство, але і на всі інші живі істоти.

З робіт Кетле виходило, що задача статистики полягає не в одному лише збиранні і класифікації статистичних даних, а в їх аналізі, метою якого має бути відкриття закономірностей, що діють у сфері масових явищ. Знання цих закономірностей і повинно було перетворити статистику на джерело наукового пізнання соціальних і біологічних явищ.

Дослідження Кетле стало поворотним пунктом в історії статистичної науки. Кетле одним з перших переконливо показав, що випадковості, що спостерігаються в живій природі, внаслідок їх повторюваності виявляють внутрішню тенденцію, яку можна досліджувати і описати точними математичними методами.

А. Кетле заклав основи біометрії. Створення ж математичного апарату цієї науки належить англійській школі біометриків XIX ст., на чолі якої стояли Ф. Гальтон (1822–1911) і К. Пірсон (1857–1936). Ця школа виникла під

впливом геніальної праці Ч. Дарвіна (1809–1882), що зробив переворот в біологічній науці. Спростувавши пануюче тоді уявлення про незмінність біологічних видів, Дарвін протиставив йому еволюційне вчення, поклавши в основу принцип природного відбору. Цей принцип базується на статистичному характері причинно-наслідкових стосунків, що складаються в живій природі; він підтверджує гегелівську концепцію про внутрішній зв'язок між випадковістю і необхідністю, між причиною і наслідком, часткою і цілим.

Революція, що здійснена Дарвіном в біологічній науці, поставила перед ученими цілий ряд великих і невідкладних задач, серед яких на першому плані опинилася проблема мінливості і спадковості організмів. Вирішення цієї проблеми стало потужним стимулом до розвитку експериментальних методів і, як наслідок, до розвитку біометрії.

Одним з тих, хто випробував на собі вплив геніальної праці Дарвіна „Походження видів” (1859), був його двоюрідний брат Ф. Гальтон. Сильне враження справили на Гальтона і праці Кетле, особливо його „Соціальна фізика” і „Антропологія”. Тому не дивно, що саме Гальтону належить перша спроба застосувати статистичні методи до вирішення проблеми спадковості і мінливості організмів. Зачинаючи з 1865 р. Гальтон опублікував ряд оригінальних робіт з антропології і генетики. На великому фактичному матеріалі він підтвердив висновки Кетле про те, що не лише фізичні, але і розумові здібності людини розподіляються за законом вірогідності, що описаний формулою Гауса – Лапласа.

Гідним продовжувачем досліджень Гальтона став його учень К. Пірсон – професор Лондонського університету. Отримавши в 1884 р. кафедру прикладної математики і механіки, Пірсон зайнявся вивченням проблеми спадковості й мінливості організмів. Він створив математичний апарат біометрії; розвинув вчення про різні типи кривих розподілу, розробив метод моментів (1894) і критерій погодження „хі-квадрат” (1900). Пірсон ввів в біометрію такі показники, як середнє квадратичне відхилення (1894) і коефіцієнт варіації (1896). Йому належить удосконалення методів кореляції і регресії Гальтона (1896, 1898). Разом з Д. Гальтоном і Уельдоном Пірсон організував випуск журналу „Біометрика” (1901), редактором якого він залишався до кінця свого життя. Цей журнал зіграв важливу роль в пропаганді біометричних методів, в створенні англійської школи біометриків.

Розроблені Гальтоном і Пірсоном біометричні методи увійшли до золотого фонду математичної статистики. Проте спроби Гальтона застосувати ці методи до вирішення проблеми спадковості організмів виявилися невдалими. Гальтон і Пірсон вважали, що за зовнішньою схожістю між родичами можна судити про ступінь їх спорідненості. Це було помилкою, на яку вказав данський вчений В. Йогансен (1857–1927). У дослідах з квасолею Йогансен прийшов до важливого висновку про те, що біологічні проблеми повинні вирішуватися за допомогою математики, але не як математичні задачі. „Статистиці, – писав Йогансен, – завжди повинен передувати біологічний аналіз, інакше результати можуть бути „статистичною брехнею”. Математика повинна надавати допомогу, а не служити керівною ідеєю”. Це був новий, реалістичний підхід до оцінки ролі математичних методів в біологічних дослідженнях.

Значення біометрії в дослідницькій роботі біологів стало очевидним вже тоді, коли були відкриті статистичні закони, що діють у сфері масових явищ. Але біологи не відразу оцінили всю важливість цих відкриттів: по-перше, тому, що статистичні методи базувалися на великих кількостях спостережень, а по-друге, вони вимагали великої обчислювальної роботи, до чого у біологів, звиклих до роботи на нечисленних вибірках, не було навиків.

Положення розпочало змінюватися після того, як була обґрунтована теорія малої вибірки. Піонером в цій області став учень Х Пірсона В. Гессе (1876–1937), який опублікував в журналі „Біометрика” свою працю під псевдонімом „Стюдент”. Подальший розвиток теорія малої вибірки отримала в праці Пірсона і особливо Р. Фішера (1890–1962), який вніс величезний вклад до біометрії, збагативши її новими методами статистичного аналізу. Впродовж ряду років Фішер працював науковим співробітником Ротамстедської сільськогосподарської дослідної станції, а з 1933 р. – на посаді професора кафедри прикладної математики Лондонського університету. Потім, з 1943 по 1957 р., Фішер завідував кафедрою генетики в Кембриджі.

Вдало сполучаючи в своїй особі біолога-експериментатора і математика-статистика, Фішер привніс в біометрію не лише нові методи, але і нові ідеї. Він заклав основи планування експериментів – теорії, яка в даний час отримала подальший розвиток і стала відносно самостійним розділом біометрії. Фішер ввів в біометрію цілий ряд нових термінів і понять і переконливо показав, що планування експериментів і обробка їх результатів – це дві нерозривно пов'язані задачі статистичного аналізу. Класичні праці Фішера стали новою віхою в історії біометрії. Вони довели, що біометрія – не просто настанова до використання різних технічних прийомів, вживаних при обробці результатів спостережень, а щось більше – наука, що займається статистичним аналізом масових явищ в біології.

Розглядаючи історію біометрії, не можна не відзначити той величезний вклад в розвиток теорії вірогідності і математичної статистики, який внесли такі вчені нашої країни, як С.Н. Бернштейн (1880–1968), А.Я. Хінчин (1894–1958), Е.Е. Слуцький (1880–1948), А.І. Хотімський (1892–1939), Б.С. Ястремський (1877–1962), В.І. Романовський (1879–1954), В.С. Немчинов (1894–1964) і багато інших, особливо А.Н. Колмогоров і його школа, що отримали світове визнання.

Перший підручник з теорії вірогідності був виданий в Росії в 1846 р. акад. В.Я. Буняковським (1804–1889). А перше повне зведення біометричних методів було складене в 1909 р. акад. А.В. Леонтовічем (1869–1943). В 1910 р. з'явилися „Нариси з теорії статистики” А.І. Чупрова (1874–1926). В 1916 р. вийшло в світ третє видання настанов з статистичних методів А.А. Кауфмана (1864–1919).

Потік біометричної літератури помітно зріс в нашій країні в ХХ ст. Біометричні методи почали застосовувати в самих різних галузях біології і суміжних наук. В.В. Алпатов (1898–1979) і його послідовники з успіхом застосували біометрію в бджільництві. Н.А. Плохинський (1899–1987), П.Ф. Рокицький (1902–1977), А.С. Серебровський (1892–1948) і ін. – в області генетики і селекції тварин. М.В. Ігнат'єв (1894–1959) і Ю.П. Зибін встановили кількісні параметри для розкрою і стандартизації взуття і одягу, що



виготовляється для масового споживання. В.В. Бунак (1891–1979), І.І. Шмальгаузен (1884–1963) застосували біометрію до вивчення закономірностей росту і розвитку організмів тварин і людини. П.В. Терентьєв (1903–1970) обґрунтував метод кореляційних плеяд, зоогеографічне „правило оптимуму” А.А. Ляпунов (1911–1973) був одним із засновників математичної біології і кібернетики в нашій країні. У короткому розділі неможливо перерахувати імена всіх учених, що внесли свій внесок до розвитку біометрії.

Зростаюча роль біометрії в дослідницькій роботі природно позначилася на підготовці фахівців біологічного профілю. Першим, хто ще в 1919 р. зачав читати студентам Московського університету курс біометрії з основами генетики був С.С. Четверіков (1880–1959). У 1924 р. він читав вже самостійний курс „Введення в біометрію”. Надалі курс біометрії в МГУ читали С.С. Алпатов, М.В. Ігнат'єв та ін.

Засновником Ленінградської школи біометриків був Ю.А. Філіпченко (1882–1930), який організував при Ленінградському університеті першу в нашій країні кафедру генетики. Філіпченко не лише уміло застосовував біометрію в дослідницькій роботі, але і пропагував її в нашій країні. Написана ним настанова з біометрії „Мінливість і методи її вивчення” ще за життя автора витримали чотири видання; (1923, 1925, 1927, 1929). Після смерті Філіпченко курс біометрії в Ленінградському державному університеті (ЛДУ) читав його учень А.І. Зуйтін, загиблий в 1942 г, в блокадному Ленінграді. Потім біометрію в ЛДУ почав викладати П.С. Терентьєв, талановитий учень С.С. Четверікова, що вніс помітний вклад до розвитку біометрії. Він перший ввів викладання біометрії стосовно великого практикуму з зоології хребетних тварин. П.В. Терентьєв був одним з перших організаторів курсів підвищення кваліфікації біологів з питань застосування математичних методів в біології і чотирьох Всесоюзних нарад (1958–1964), що присвячені цьому питанню.

Велика робота з навчання фахівців лісового і сільського господарства біометричними методами була проведена Ю.Л. Поморським (1893–1954). Багато зроблено в області пропаганди статистичних методів А.К. Мітропольським і іншими ученими нашої країни. На навчальних матеріалах Ю.А. Філіпченко, Ю.Л. Поморського, В.І. Романовського та інших учених виховалося ціле покоління вітчизняних біометриків.

### Етапи розвитку біометрії

Отже, біометрія в своєму історичному розвитку пройшла довгий і складний шлях – від чисто словесного опису біологічних об'єктів до їх вимірів, від статистичних зведень і таблиць до статистичного аналізу масових явищ. У історії біометрії можна відзначити декілька періодів, або етапів.

Перший період, описовий, бере свій початок в XVII столітті. В цей час відбувається перехід від словесного опису і елементарного кількісного обліку біологічних об'єктів до їх числових характеристик. Виміри розглядаються як метод наукового пізнання живої природи.

Другий період, що розпочався в першій половині XIX ст., ознаменований роботами А. Кетле. В цей час закладаються основи біометрії як науки, метою

якої є не опис явищ, а їх аналіз, направлений на відкриття статистичних закономірностей, які діють у сфері масових явищ. Біометрію розглядають одночасно і як науку, і як метод наукового пізнання.

Третій період, формалістичний, характеризується виникненням і розвитком англійської біометричної школи на чолі з Ф. Гальтоном і К. Пірсоном. В цей час створюють математичний апарат біометрії і роблять спроби застосувати його до вивчення проблеми спадковості і мінливості організмів.

Четвертий період, раціоналістичний, розпочинається з 1902 р. класичними дослідженнями Йогансена, який показав, що в області біологічних досліджень перше місце повинне належати біологічному експерименту, а не математиці. Математичні методи повинні застосовуватися як допоміжний апарат при обробці експериментальних даних.

П'ятий період в розвитку біометрії відкривають класичні роботи Стьюдента і Р. Фішера. В цей час створюються основи теорії малої вибірки, теорії планування експериментів, вводяться у вміст біометрії нові терміни і поняття. Всі ці новини пов'язані з революцією в біології, з ломкою застарілих принципів і понять в області дослідницької роботи, з посиленням процесу математизації біології. Відбувається все більш помітна спеціалізація біометрії, використання її методів в самих різних областях біології, медицини, антропології та інших суміжних науках.

## Тема 2. Вибірковий метод і групування первинних даних

### Генеральна сукупність і вибірка

Спостереження, що проводяться над біологічними об'єктами, можуть охоплювати всіх членів сукупності, що вивчається, без виключення і можуть обмежуватися обстеженням лише деякої частки членів даної сукупності. У першому випадку спостереження називатиметься суцільним або повним, а в другому – частковим або вибірковим. Суцільне спостереження дозволяє отримувати вичерпну інформацію про груповий об'єкт, в чому і полягає перевага цього способу перед способом вибіркового спостереження. Проте до суцільного спостереження прибігають не завжди. По-перше, тому, що ця робота пов'язана з великими витратами часу і праці, а по-друге, через практичну неможливість або недоцільність проведення такої роботи. Неможливо, наприклад, врахувати всіх мешканців зоо- або фітопланктону навіть невеликого водоймища, оскільки їх чисельність практично незора. Недоцільно висівати всю партію насіння для того, щоб визначити їх схожість або силу проростання. Тому в переважній більшості випадків замість суцільного спостереження вивченню піддають якусь частку обстежуваної сукупності, по якій і судять про її стан в цілому.

Сукупність, з якої відбирається деяка частка її членів для спільного вивчення, називається генеральною, а відібрана тим або іншим способом частка генеральної сукупності отримала назву вибіркової сукупності або вибірки. Об'єм генеральної сукупності, що позначається буквою  $n$ , теоретично нічим не обмежений, тобто генеральна сукупність мислиться як нескінченно велика множина щодо однорідних одиниць або членів, що складають її вміст. Практично

ж об'єм генеральної сукупності завжди обмежений і може бути різним, що залежить як від об'єкту спостереження, так і від задачі, що поставлена перед дослідником. Наприклад, при визначенні продуктивності тварин тієї або іншої породи або виду генеральну сукупність складуть всі особини даної породи або виду. Якщо ж питання про продуктивність тварин вирішується в межах даної області або району, то генеральну сукупність складе все поголів'я тварин породи, що вивчається, яка поширена в даній області або районі.

Об'єм вибірки, що позначається буквою  $n$ , може бути і порівняно великим, і малим, але він не може містити менше двох одиниць. Вибірковий метод є основним при вивченні статистичних сукупностей. Його перевага перед суцільним обліком всіх членів генеральної сукупності полягає в тому, що він скорочує час і витрати праці (за рахунок зменшення кількості спостережень), а головне – дозволяє отримувати інформацію про такі сукупності, суцільне обстеження яких практично неможливе або недоцільно.

### Репрезентативність вибірки

Основна задача, яка вирішується за допомогою вибіркового методу, зводиться до отримання такої інформації, що дозволяє більш менш точно судити про стан генеральної сукупності. Досвід показав, що вибірка досить добре відображає структуру генеральної сукупності. Проте, як правило, повного співпадання вибірових характеристик з характеристиками генеральної сукупності не буває. Щоб вибірка можливо повніше відображувала структуру генеральної сукупності, вона має бути досить представницькою, або репрезентативною (від лат. *represento* – представляю). Репрезентативність вибірки досягається способом рандомізації (від англ. *random* – випадок), або випадковим відбором варіант з генеральної сукупності, що забезпечує однакову можливість для всіх членів генеральної сукупності потрапити до складу вибірки.

Існує два основні способи відбору варіант з генеральної сукупності: повторний і безповторний. Повторний відбір проводиться за схемою „повернення” врахованих одиниць в генеральну сукупність, так що одна і та ж одиниця може потрапити у вибірку повторно. Цей відбір не впливає на склад генеральної сукупності і можливість кожної одиниці потрапити у вибірку не змінюється. При безповторному відборі враховані одиниці в генеральну сукупність не повертаються, кожна відібрана одиниця реєструється один раз. Можливість одиниць генеральної сукупності потрапити у вибірку змінюється при безповторному відборі, оскільки кожен попередній відбір впливає на результати подальшого і на склад генеральної сукупності, який теж змінюється. У практиці застосовується зазвичай безповторний випадковий відбір. Якщо піддаються виміру, наприклад, чоловіки призовного віку, то, змірявши одного з них, його повторно вже не вимірюють. Випадковий повторний відбір служить теоретичною моделлю, що дозволяє досліджувати процеси, які здійснюються в статистичних сукупностях, що має велике пізнавальне значення.

Ідеальний випадковий відбір, як повторний, так і безповторний, проводиться за способом жеребкування або лотереї, а також за допомогою випадкових чисел, що дозволяє повністю виключити суб'єктивний вплив на склад

вибірки. Суть цього способу полягає в наступному. На чисельно обмеженій, але досить великій штучній моделі генеральної сукупності методом повторного випадкового відбору утворюється ряд чисел, які заносяться в таблицю так, щоб вони мали однакову кількість цифр. Цим полегшується зручність використання такої таблиці в практичних цілях. Наприклад, при трьохзначності чисел цифра 8 заноситься в таблицю у вигляді 008, а число 69 – у вигляді 069 і так далі. Числа заносяться в таблицю без всякої послідовності, в випадковому порядку, тому вона і називається таблицею випадкових чисел.

Як користуватися таблицею випадкових чисел? Допустимо, що у віварії міститься 120 тварин. З них для досліду потрібно відібрати десять особин. Щоб відбір був дійсно випадковим, таким, що виключає суб'єктивний вплив на склад вибірки, потрібно поступити наступним чином: всім тваринам віварію надати номери від 1 до 120. Потім в таблиці знаходять десять таких чисел, які не перевищують 120. При використанні чотиризначної таблиці домовляються враховувати перші три цифри в кожному стовпці цієї таблиці (можна виходити і з іншої умови). У першому стовпці знаходимо числа 90 і 91 (див. 0905 і 0912), інших потрібних чисел в цьому стовпці тощо. У другому стовпці відбираємо числа 47 і 41. У третьому – 62, 84, 50 і 31. У четвертому стовпці опиняється решта чисел – 39 і 87. Всього вийшло десять чисел: 90, 91, 47, 41, 62, 84, 50, 31, 39 і 87. Особини з цими номерами включаємо до складу експериментальної групи.

Принцип рандомізації не виключає плановості відбору одиниць з генеральної сукупності і може здійснюватися по-різному залежно від задачі і організації експерименту. Розрізняють наступні види планового відбору:

- 1) типовий, або груповий;
- 2) серійний, або гніздовий;
- 3) механічний.

При типовому відборі генеральна сукупність заздалегідь ділиться на типові групи (ділянки, райони). Потім, в кожній групі випадковим способом відбирається однакове або пропорційне число одиниць, що об'єднуються потім в одну вибірку сукупність, яка і піддається статистичному аналізу.

У випадку серійного відбору генеральна сукупність, як і при типовому відборі, заздалегідь ділиться на групи, звані гніздами або серіями. Потім, по бажанню дослідника, із загальної кількості серій відбирається необхідне їх число для спільної обробки. При цьому серії можуть бути рівно- і різночисельними. Наприклад, з 50 груп підлітків у віці від 14 до 16 років намічено обстежувати вибірккові шість груп. Члени цих груп і об'єднуються в одну сукупність для спільної обробки. Таким чином, при серійному відборі на відміну від відбору типового з генеральної сукупності відбираються не окремі одиниці, а цілі серії або гнізда щодо однорідних одиниць.

Механічний відбір здійснюється по наступній схемі. Генеральна сукупність розбивається на декілька рівних частин або груп. Потім з кожної групи випадковим способом відбирають по одній одиниці. Наприклад, при обстеженні поля жита намічено відібрати 100 рослин (або колосків). Отже, поле має бути розбите на 100 однакових (квадратних) ділянок. Таким чином, при механічному відборі число відібраних одиниць дорівнює чисельності груп, на які розбита генеральна сукупність. Механічний відбір може проводитись і по іншій схемі,

коли в вибірку попадає кожна десята, сота і так далі одиниця генеральної сукупності. Наприклад, при проведенні ботанічних або зоологічних екскурсій у вибірку потрапляє кожен 10-й, 20-й екземпляр і так далі рослин або тварин даного виду, що зустрінуться.

Щоб вибірка була найбільш репрезентативною, необхідно наряду з правильно організованим відбором варіант звернути увагу на розмах варіювання ознаки і погоджувати з ним об'єм вибірки. Чим ширше розмах варіювання ознаки, тим більшим має бути і об'єм вибірки. Нечисленний склад вибірки при сильному варіюванні ознаки знижує її репрезентативність.

### Групування первинних даних

Спостереження над біологічними об'єктами проводяться одночасно по декількох ознаках, що дозволяє зібрати якнайповніші відомості. Результати спостережень фіксуються в щоденниках, протоколах, анкетах, бланках і в інших формулярах первинного обліку. Так, на маршрутних зоологічних і ботанічних екскурсіях, при проведенні експедицій і польових дослідів зручною формою обліку спостережень служить щоденник. В умовах лабораторного експерименту результати випробувань фіксують в протоколах, журналах, бланках і тому подібне. Форм і способи обліку результатів спостережень багато. Первинні документи обліку містять фактичний матеріал, що потребує обробки. Обробка розпочинається з впорядкування зібраних даних (дотримуючи правило однорідності складу вибірки), систематизації виражених числами фактів. Процес систематизації або впорядкування первинних біометричних даних з метою вибору вкладеної в них інформації, виявлення закономірності, якою слідує явище, що вивчається, то процес, називається угрупованням.

Групування — це не просто технічний прийом, а глибоко осмислена дія, що направлена на отримання правдоподібної і повноцінної інформації про об'єкт, що вивчається. Вибраний спосіб угруповання повинен відповідати вимозі поставленого завдання та добре узгоджуватися з вмістом явища, що вивчається. Тільки при одночасній наявності належного плану досліджень, вмілого проведення його і відповідних методів обробки дослідних даних дослідник може бути упевнений в надійності результатів, на яких він буде свої висновки.

Групування початкових даних може бути різним залежно від того, з якою метою і по яких ознаках вона проводиться. Найбільш прийнятною формою групування є статистичні таблиці.

Групування за однією ознакою називається простим, а за декількома ознаками — складним. Звідси і таблиці можуть бути простими і складними. Найбільш прості таблиці отримують при альтернативному групуванні, коли одна група членів сукупності протиставляється іншій, наприклад, хворі — здоровим, забарвлені — нефарбованим, високі — низьким, успішні — неуспішним тощо. Прикладом такого групування може служити таблиця з результатами обстеження стану гланд 265 учнів молодших класів (таблиця 1.1).

Складнішими є многопольні таблиці, що вживані при групуванні початкових даних для виявлення причинно-наслідкових стосунків між варіюючими об'єктами (таблиця 1.2). При цьому одні ознаки розглядаються як фактори, що впливають на інші ознаки, які називають результативними.

Дані, що приведені в таблиці 1.2, були отримані в науково-дослідному інституті ім. В.В. Докучаєва при випробуванні гречки сорту „Багатир” на врожайність (результативна ознака) залежно від попередників (впливаючі фактори).

Таблиця 1.1. Результати обстеження учнів молодших класів.

Класи, в яких проводився огляд учнів	Виявлено дітей		Всього
	здорових	хворих	
Треті і четверті	63	92	155
П'яті і шості	71	39	110
Всього	134	131	265

Зрозуміло, поняття факторних і результативних ознак застосовуються в кожному випадку окремо, бо те, що служить фактором в одному випадку, в іншому може виступати як результативна ознака.

Не менш складним виявляється групування вибірових даних при з'ясуванні зв'язку між варіюючими ознаками. У таких випадках числові значення ознак з врахуванням їх повторюваності в двовимірній сукупності групуються в кореляційну таблицю. Як приклад такого групування можна привести класичні дані Гальтона, що показують позитивний зв'язок між зростом батьків і зростом їх дорослих дітей (таблиця 1.3).

Таблиця 1.2. Результати врожайності гречки.

Попередники	Урожай гречки за повторностями, ц/га			Середній урожай
	1	2	3	
Горох ранозелений	23,7	20,1	20,5	21,4
Чечевиця	23,5	25,1	21,1	23,2
Чина степова №21	26,7	23,2	23,8	24,6
Ячмінь	26,0	24,9	25,3	25,4

Таблиця 1.3. Зв'язок зросту батьків та дорослих дітей.

Зріст батьків, см	Зріст дітей, см								Всього го
	60,7	62,7	64,7	60,7	68,7	70,7	72,7	74,2	
74							4		4
72			1	7	11	17	20	6	62
70	1	2	21	48	83	66	22	8	251
68	1	15	56	130	148	69	11		430
66	1	15	19	56	41	11	1		144
64	2	7	10	14	4				37
Всього	5	39	107	255	287	163	58	14	928

Особливий інтерес для біолога представляє групування початкових даних в статистичні ряди – ряди числових значень ознаки, розташованих в певному порядку. Залежно від того, в якому стані (статистиці або динаміці) і за якими ознаками (якісними або кількісними) розглядаються явища або процеси, статистичні ряди поділяються на атрибутивні, варіаційні, динамічні або регресії. Прикладом атрибутивного ряду може служити наступний розподіл 870 солдатів племені Маорі за групами крові системи АВО:

Група крові	I (0)	II (A)	III (B)	IV (AB)
Число осіб, що мають дану групу	388	453	16	13

Як інший приклад атрибутивного ряду наведемо дані (таблиця 1.4), що показують закономірне збільшення гемоглобіну *Hb* в крові хребетних тварин (розпочинаючи з класу риб).

Таблиця 1.4. Вміст гемоглобіну в крові хребетних тварин.

Клас тварин	риби	амфібії	рептилії	птиці	ссавці
Кількість гемоглобіну на один кг маси тіла	1,6	2,9	3,8	11,2	11,7

Найбільше значення в курсі біометрії мають варіаційні ряди.

### Варіаційні ряди

Варіаційним рядом називається ряд чисел, що показує закономірність розподілу одиниць сукупності, що вивчається, за ранжируемими значеннями варіюючої ознаки (від франц. ranger – розташування в ряд за ранжиром, тобто за зростанням). Наприклад, з урожаю картоплі, що зібрана на одній з дослідних ділянок випадковим способом, тобто наздогад, було відібрано 25 бульб, в яких підраховувалося число очок. Результат опинився наступний: 6, 9, 5, 10, 7, 9, 8, 10, 9, 10, 8, 11, 9, 12, 9, 10, 8, 10, 11, 9, 10, 9, 8, 7, 11. Щоб знатися на цих даних,

розташуємо їх в порядку зростання числових значень ознаки, тобто ранжируємо так, щоб підрахувати, скільки разів кожна варіанта ( $x_i$ ) зустрічається в даній сукупності. Оскільки ознака варіює в межах від 5 до 12 одиниць, вибірка розподіляється таким чином:

Значення ознаки ( $x_i$ )	.....5	6	7	8	9	10	11	12
Число варіант ( $p_i$ )	.. 1	1	2	4	7	6	3	1

Цей подвійний ряд чисел, що показує, яким чином числові значення ознаки ( $x_i$ ) пов'язані з їх повторюваністю ( $p_i$ ) в даній сукупності, називається варіаційним рядом або рядом розподілу. Числа, що показують, скільки разів окремі варіанти зустрічаються в даній сукупності, називаються частотами або вагами варіант і позначаються  $p$  або  $f$ . Загальна сума частот завжди дорівнює об'єму даної сукупності, тобто  $n$ .

Частоти виражаються не лише абсолютними, але і відносними числами – в долях одиниці або у відсотках від загальної чисельності варіант, що складають дану сукупність. У таких випадках ваги називають відносними частотами або частотостями. Спільна сума частостей дорівнює одиниці.

Сукупність числових значень ознаки розташовується в безінтервальний або інтервальний варіаційний ряд залежно від того, як варіює ознака – в широкому або вузькому діапазоні. У першому випадку частоти розподіляються безпосередньо за ранжированими значеннями варіюючої ознаки, які набувають значення „класів” варіаційного ряду, а в другому – частоти розподіляються за окремими інтервалами, або проміжках (від – до), на які розбивається варіація ознаки в межах від мінімальної до максимальної варіанти сукупності. Ці проміжки, або класові інтервали, можуть бути рівними і нерівними за шириною. Звідси розрізняють варіаційні ряди рівно- і нерівноінтервальні. Величина рівних інтервалів визначається діленням розмаху варіювання ознаки на число груп або класів ( $k$ ), що намічаються при побудові варіаційного ряду:

$$i = \frac{(x_{\max} - x_{\min})}{k}, \quad (1.1)$$

де  $i$  — величина класового інтервалу,  $x_{\max}$  – максимальна, а  $x_{\min}$  – мінімальна варіанти сукупності.

Величина нерівних інтервалів встановлюється у кожному конкретному випадку з врахуванням особливостей об'єкту спостереження і задачі дослідження. Прикладом нерівноінтервального розподілу можуть служити дані А.Ф. Ковшар, що характеризують чисельність голубиних зграй в період гніздування залежно від числа особин в зграї (таблиця 1.5).

У нерівноінтервальних рядах зробити порівняння окремих груп або класів один з одним складно, оскільки по мірі зміни інтервалів змінюється і характер розподілу частот. В якості характеристики таких рядів зручніше використовувати щільність розподілу, тобто відношення частот або частностей до величини класових інтервалів.



Таблиця 1.5. Чисельність голубиних зграй в період гніздування.

Кількість особин в зграї (від – до)	Число зграй в період з 15 березня по 15 серпня	Щільність розподілу			
		абсолютна		відносна	
Одинаки	6 (18,2%)	6,0	1,0	18,2	1,7
2 – 5	19 (57,8%)	6,33	3,0	19,2	5,08
5 – 10	4 (12,1%)	0,8	0,8	2,42	1,36
10 – 20	2 (6,1%)	0,2	1,2	0,61	2,03
20 – 30	1 (3,0%)	0,1	1,3	0,3	2,2
30 – 50	1 (3,0%)	0,5	0,55	0,15	0,93

Нерівноінтервальне групування результатів спостереження частіш всього застосовується в області соціально-економічних досліджень і порівняно рідко в біології. Біометричні дані розподіляються зазвичай в рівноінтервальні варіаційні ряди, що полегшує зіставлення різних груп і дозволяє краще виражати закономірність варіювання, характеризувати варіюючі об'єкти середніми показниками.

#### Техніка побудови варіаційних рядів

Питання про розподіл вибірки в безінтервальний або інтервальний ряди вирішується залежно від розмаху і характеру варіювання ознаки: Якщо ознака варіює дискретно і слабо, тобто у вузьких межах, вибірка розташовується в безінтервальний ряд, якщо ж в широкому діапазоні, то незалежно від того, як варіює ознака, дискретно або безперервно вибірка розташовується в інтервальний ряд. При цьому важливо, щоб намічена ширина класового інтервалу задовольняла певним вимогам. Справа в тому, що побудову варіаційного ряду вимагає подвійна мета: виявлення закономірності варіювання ознаки як спосіб, що полегшує обчислення, та визначення статистичних характеристик варіюючого об'єкту. Грубе групування початкових даних, коли намічаються дуже широкі або надмірно вузькі класові інтервали, позначається як на характері розподілу частот за класами варіаційного ряду, так і на точності визначення нею числових характеристик. При виборі вузьких інтервалів точність числових характеристик ряду – середніх величин і показників варіації – підвищується, але ряд виходить дуже розтягнутим, погано осяжним, недостатньо таким, що чітко відображує закономірність варіювання. При широких інтервалах теж погано виявляються характерні риси варіювання, а головне – знижується точність статистичних характеристик ряду.

Звідси зрозумілими є спроби знайти такий спосіб або критерій, який дозволяв би у кожному конкретному випадку визначати те оптимальне число класів ( $k$ ), на яке треба розбити варіацію ознаки, щоб вийшла добре осяжним варіаційний ряд при достатній точності його числових характеристик. Приблизно це число ( $k$ ) визначають за допомогою таблиці 1.6.

Таблиця 1.6. Таблиця визначення числа класів.

Об'єм вибірки (від – до)	Число класів $k$
25—40	5—6
40—60	6—8
60—100	7—10
100—200	8—12
>200	10-15

Більш точно число класів  $k$ , що входить до складу формули (1), визначається за формулою Стерджеса

$$k = 1 + 3,32 * \lg(n),$$

або за формулою, що запропонували К. Брукс і Н. Карузерс :

$$k = 5 * \lg(n).$$

Техніка побудови варіаційного ряду зводиться до наступного. Склавши зведення початкових даних, відшуковують в масі числових значень ознаки мінімальну ( $x_{min}$ ) і максимальну ( $x_{max}$ ) варіанти. Потім за формулою (1.1) визначають величину класового інтервалу ( $i$ ). Якщо отримаємо, що  $i=1$ , то зібраний матеріал розподіляється в безінтервальний варіаційний ряд; якщо же  $i \neq 1$ , початкові дані треба розподілити в інтервальний ряд.

Ця рекомендація гарантує найменші втрати інформації, немінучі при обчисленні статистичних характеристик інтервального варіаційного ряду. Треба також дотримуватись правила, згідно якому величина класового інтервалу повинна відповідати точності, прийнятій при вимірі ознаки, що враховується. Наприклад, відсоток жиру в молоці корів ( $n=60$ ) вимірювався з точністю до 0,01% і варіював від 3,21 до 4,55%. У такому разі класовий інтервал повинен визначатися з точністю до 0,01 одиниці:

$$i = (4,55 - 3,21) / (1 + 3,32 \lg n) = 1,34 / (1 + 5,90) = 0,194 \approx 0,19.$$

Якщо точність виміру даної ознаки обмежити десятими долями відсотка, то величина класового інтервалу повинна відповідати їй:

$$i = (4,6 - 3,2) / 6,9 = 0,2.$$

У обох випадках вибірка повинна розташовуватись в інтервальний варіаційний ряд.

При побудові інтервального варіаційного ряду треба добиватися того, щоб мінімальна варіанта ( $x_{min}$ ) потрапляла приблизно в середину першого класового інтервалу. Ця умова гарантує побудову варіаційного ряду, який найбільш повно відповідає природі вивчаемого явища, а відповідно, найменші втрати інформації

про точність обчислених статистичних характеристик ряду. Даній умові задовольняє формула:

$$l = x_{min} - i/2 \quad (1.2)$$

де  $l$  – нижня границя першого класового інтервалу.

Так, при  $x_{min} = 3,21\%$  і  $i = 0,19$  знаходимо  $l = 3.21 - 0.19/2 = 3.115 \approx 3.12$ .

Намітивши класові інтервали, залишається розподілити за ними всі варіанти вибірки, тобто визначити частоти кожного класу. Але тут виникає питання, в які класи відносити варіанти, величина яких збігається з верхньою границею одного і з нижньою границею іншого, сусіднього класу. Наприклад, в який клас треба віднести варіанту 3,31 – в перший або в другий? Вирішується це питання по-різному. Допустимо в один і той же клас розмістити варіанти, які більше нижньої, але менше

8	10	6	10	8	5	11	7
10	6	9	7	8	7	9	11
8	9	10	8	7	8	8	11
7	10	8	8	5	11	8	10

дорівнюють його верхній границі. Частіше верхні границі класів зменшують на величину, що дорівнює точності, прийнятій при вимірі ознаки, наприклад, на 0,01, 0,1 і т. д., чим і досягається необхідне розмежування класів.

Оскільки числові характеристики тих об'єктів, що варіюють, – середня величина і показники варіації – обчислюються на безінтервальних рядах, інтервальний ряд доводиться перетворювати на ряд безінтервальний. Це досягається заміною класових інтервалів центральними або серединними значеннями ( $x_i$ ), які відстоять від їх нижніх границь ( $x_{ll}$ ) на величину, що дорівнює половині класового інтервалу ( $i$ ), тобто  $x_i = x_{ll} + i/2$ ; вони визначаються за напівсумою нижньої і верхньої границь інтервалів, а при їх розмежуванні – за напівсумою нижніх границь попереднього і наступного класів.

Розглянемо техніку побудови варіаційних рядів на прикладах.

Приклад 1. На свинофермі зареєстровано 64 опороси свиноматок. При цьому від кожної свиноматки отримана наступна кількість поросят:

12	7	5	7	9	7	5	10
8	9	7	12	8	9	6	7
8	7	11	8	6	7	9	10
6	7	6	12	8	10	6	11

У цій сукупності  $x_{min}=5$  і  $x_{max}=12$ , звідси

$$i=(12-5)/(1+3,32lg(64))=7/(1+6)=1.$$

Оскільки ознака варіює дискретно і величина класового інтервалу дорівнює одиниці, вибірку треба розподілити в безінтервальний ряд, тобто безпосередньо за ранжируемими значеннями ознаки, які і будуть класами даного ряду.

Ранжируємо вибірку в межах від 5 до 12 варіант і знаходимо їх частоти. Щоб при визначенні частот кожного класу не збитися з рахунку, треба побудувати допоміжну розрахункову таблицю, в якій перша графа заповнюється класами (в даному випадку ранжируемими значеннями ознаки), а друга служить для обліку частот ( $p$ ), що розподіляються за цими класами.

Приклад 2. На підставі багатолітніх клінічних спостережень, що проводилися в Сухумському розпліднику мавп, складена вибірка, яка включає 100 аналізів на вміст кальцію (мг %) в сироватці крові клінічно здорових мавп (павіанів-гамадрилів):

13,6	12,9	12,3	9,9	12,7	11,7	10,8	10,4	10,9	10,2
14,7	10,4	11,6	11,7	12,1	10,9	12,1	9,2	10,7	11,5
13,1	10,9	12,0	11,1	13,5	11,2	13,5	10,1	14,0	10,0
11,6	12,4	11,9	11,4	12,8	11,4	10,9	12,7	13,8	13,2
11,9	10,8	11,0	12,6	10,0	10,3	12,7	11,7	12,1	13,8
12,2	11,9	11,6	10,6	11,1	10,7	12,3	11,5	11,2	11,5
12,7	10,5	11,2	11,9	9,7	13,0	9,6	12,5	11,6	9,0
11,5	12,3	12,8	12,6	12,8	12,5	12,8	11,4	12,5	12,3
14,5	12,3	12,6	11,7	12,2	12,3	11,6	12,0	13,5	12,5
11,6	11,9	12,0	11,4	14,7	11,3	13,2	14,3	13,2	14,2

Потрібно згрупувати ці дані у варіаційний ряд. В даному випадку ознака варіює безперервно в межах від 9,0 до 14,7 мг %. Встановлюємо величину класового інтервалу:

$$i=(14,7-9,0)/(1+3,32lg100)=5,7/7,6=0,75\approx 0,8.$$

Величина  $i=0,75$  заокруглена до 0,8, що відповідає точності, що прийнята при вимірюванні ознаки.

Оскільки вибірку необхідно групувати в інтервальний варіаційний ряд, визначаємо нижній кордон першого класу:

$$x_{min} = 9,0 - 0,8/2 = 8,6.$$

Потім намічаємо наступні класові інтервали: 8,6—9,4—10,2—11,0—11,8—12,6—13,4—14,2—15,0. Вийшло вісім інтервалів. Розмежовуємо їх на величину, рівну точності виміру ознаки, тобто зменшуємо верхні кордони інтервалів на 0,1 мг %. Будуємо допоміжну таблицю і розносимо все 100 варіант по класових інтервалах.

Переводимо шифр частот в числа, сума яких повинна дорівнювати об'єму вибірки, тобто в даному випадку  $\sum p_i = 100$ .

Серединні значення класів розраховані за формулою  $x_i = x_{min} + i/2$ , тобто збільшенням до нижніх кордонів класів  $1/2$  класового інтервалу – величини, що дорівнює  $0,8/2 = 0,4$ . Тоді:  $8,6 + 0,4 = 9,0$ ;  $9,4 + 0,4 = 9,8$  і так далі. Такий же результат виходить, якщо середні значення класів розраховуються по напівсумах їх нижніх кордонів:  $(8,6 + 9,4)/2 = 9,0$ ;  $(9,4 + 10,2)/2 = 9,8$  і так далі. Таким чином, інтервальний варіаційний ряд перетворено на ряд безінтервальний.

Приклад 3. Яйценосність 80 курей, що містяться на птахофермі, варіювала від 208 до 250 яєць від однієї несучки в рік. З врахуванням цих даних визначено класовий інтервал:

$$i = (250 - 208) / (1 + 3,321 \sqrt{80}) = 42 / (1 + 1,63) = 5,75.$$

Оскільки класовий інтервал не дорівнює одиниці, результати спостережень мають бути згруповані в інтервальний варіаційний ряд, не зважаючи на те, що ознака варіює дискретно. Встановлюємо нижній кордон першого класу:  $i = 208 - 6/2 = 205$ . Встановлюємо такі класові інтервали: 205—211—217—223—229—235—241—247—253. Зменшуємо верхні кордони інтервалів на одиницю: 205—210; 211—216; 217—222; 223—228; 229—234; 235—240; 241—246; 247—252. Подальші дії, що відносяться до побудови варіаційного ряду, зрозумілі з попередніх прикладів.

### Графічне зображення варіаційних рядів

Для наочного представлення закономірності варіювання тієї або іншої кількісних ознак варіаційні ряди зображають у вигляді геометричних фігур в системі прямокутних координат. Так, якщо з'єднати прямими лініями геометричні точки, що зв'язують значення класів (відкладаються по осі абсцис) з їх частотами, що відкладаються по осі ординат, вийде лінійний графік, названий варіаційною кривою або кривою розподілу.

При побудові графіка безінтервального варіаційного ряду по осі абсцис відкладають значення класів, а по осі ординат – частоти. Висота перпендикулярів, що проставлені по осі абсцис, відповідає частотам класів. Сполучаючи вершини перпендикулярів прямими лініями, отримують геометричну фігуру у вигляді багатокутника, що називається полігоном розподілу частот (рис. 1.1).

При побудові графіка інтервального варіаційного ряду по осі абсцис відкладають межі класових інтервалів. В результаті виходить стовпчикова геометрична фігура, що називається гістограмою розподілу частот (рис. 1.2). Якщо з серединних точок вершин прямокутників гістограми опустити

перпендикуляри на вісь абсцис, гістограма перетворюється на полігон розподілу. Сполучаючи точки вершин прямокутників гістограми прямими лініями, отримуємо варіаційну криву.

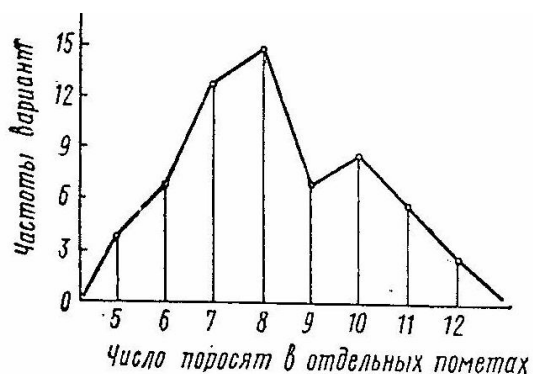


Рис. 1.1. Полігон розподілу чисельності поросят в 64 опоросах свиноматок.

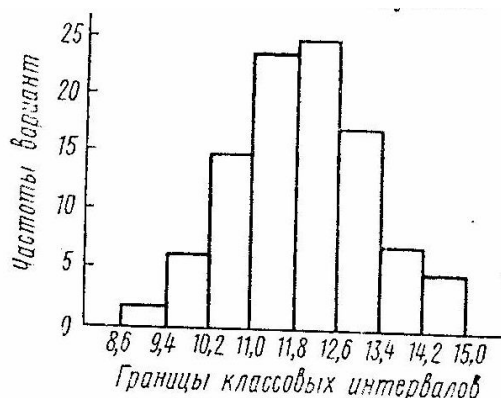


Рис. 1.2. Гістограма розподілу кальцію (мг%) в сировотці крові павіанів-гамадрилів.

Отже, будь-який варіаційний ряд в загальному вигляді описується графічно варіаційною кривою, яка має не лише ілюстративне значення але і використовується для статистичного аналізу варіюючих ознак.

При побудові графіків варіаційних рядів масштаби на осях координат вибираються довільно, але з таким розрахунком, щоб висота варіаційної кривої (тобто максимальна ордината) відносилася до її основи приблизно, як 5:8. Відкладаючи по осі абсцис класи варіаційної лави, треба доводити їх зліва і справа до нульових класів, які не містять жодної варіанти. Недотримання цих правил приводить до небажаних результатів: графік виходить із загостреною вершиною і вузькою основою або у вигляді надмірно розтягнутої по ширині фігури сплюснення. У обох випадках варіаційна крива виявляється такою, що недостатньо чітко відображує характерні риси варіювання ознаки, яка вивчається.

Окрім варіаційної кривої емпіричний ряд розподілу можна представити у вигляді кумуляти та огиви. Кумуляту отримують якщо по осі абсцис відкладати значення класів, а по осі ординат – накопичення частоти з подальшим з'єднанням геометричних точок прямими лініями. На рис. 1.3 зображена кумулята розподілу кальцію в сироватці крові павіанів-гамадрилів. На відміну від куполоподібної варіаційної кривої кумулята має S-образну форму. Накопичені частоти отримують послідовним підсумовуванням або кумуляцією (від лат. *sumulato* – збільшення, накопичення) частот в напрямку від мінімальної класової варіанти або від першого класу до кінця варіаційного ряду.

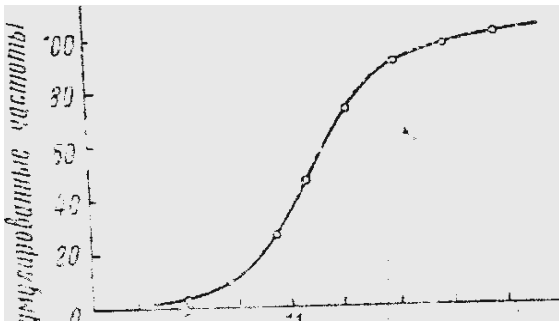


Рис. 1.3. Кумулята розподілу кальцію (мг<sup>о</sup>%) в сироватці крові павіанів-гамадрилів.

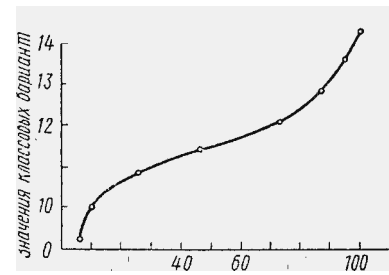


Рис. 1.4. Оги́ва розподілу кальцію (мг<sup>о</sup>%) в сироватці крові павіанів-гамадрилів.

В даному випадку частоти розподілу кальцію в сироватці крові павіанів-гамадрилів кумульовані таким чином:

Частоти ( $p_i$ ) ....	2	6	15	23	25	17	7	5
Кумуляти частот ( $Sp_i$ )	2	8	23	46	71	88	95	100.

Якщо накопичені частоти відкладати по осі абсцис, а значення класів – по осі ординат з подальшим з'єднанням геометричних точок прямими лініями, вийде графік, що називається *огивою* (рис. 1.4).

Для огиви і кумуляти властива більш обтічна форма, ніж це спостерігається відносно емпіричних варіаційних кривих, які зазвичай виглядають у вигляді ламаних ліній. Ця властивість дозволяє інколи віддавати перевагу цим графікам перед емпіричною варіаційною кривою. Зокрема, огива дозволяє кращим чином порівнювати одночасно декілька емпіричних розподілів неоднакового об'єму, особливо в тих випадках, коли частоти рядів виражаються у відсотках від загальної кількості спостережень.

Для кумуляти сукупності, що мають нормальний закону розподілу, характерно те, що її центральна точка ділить ряд накопичених частот навпіл і збігається з центром розподілу. Це властивість кумуляти дозволяє використовувати її при визначенні середніх доз випробовуваних речовин, що викликають ефект у 50% піддослідних тварин. Оскільки кумулята виражає залежність між дозами впливаючого аргента (наприклад, лікарських або токсичних речовин) і у відповідь реакцію організму, її називають „кривою ефекту” (Ван-дер-варден, 1960) або „характеристичною кривою” (Треван, 1927).

Огива дозволяє порівнювати один з одним одночасно декілька емпіричних розподілів неоднакового розміру.

Треба відзначити, що невміла побудова графіків призводить до того, що вони получаются або в вигляді гостроверхих геометричних фігур з вузькою основою, або плосковершинними, занадто розтягнутими по осі абсцис. В обох випадках графіки будуть погано наочними, що нечітко відобрадають закономірності варіювання.

Уникнути цих недоліків дозволяє правило „золотого перетину”, згідно з яким основа геометричної фігури відноситься до її висоти, як  $1:0,62$ . Відносно до побудови варіаційної кривої масштаби на осях прямокутних координат треба вибирати з таким розрахунком, щоб основа кривої було в  $1,5 - 2$  рази більше її висоти (тобто максимальної ординати). Відкладаючи по осі абсцис класи варіаційного ряду, треба також доводити крайні із них до нульових класів, в яких не міститься ні однієї варіанти. В результаті цього варіаційній кривій надається закінчений добре оглядовий (наочний) вигляд.