

Математичні методи в біології

1. Загальна х-ка застосування мат. методів у біологічних дослідженнях.
2. Основні статистичні показники для х-ки сукупності експериментальних даних. Визначення та зміст показників. 
3. Форма подання рез-тів вибіркового експерименту. Теор. очікувані діапазони мінливості інд. даних і вибірових середніх значень.

Література

- Гланц С. Медико-биологическая статистика. Пер. С англ..– М., Практика, 1998.– 459с.
- Атраментова Л.А., Утевская О.М. Статистические методы в биологии , Горловка, 2008
- Гумецький Р.Я., Паляниця Б.М., Чабан М.Є. Математичні методи в біології : Теоретичні відомості, програмований практикум, комп'ютерні тести / Навч. посібник.– Львів: ЛНУ, 2004. – 112 с. Деол, 2008. – 324 с.
- Деркач М.П, Гумецький Р.Я., Чабан М.Є. Курс варіаційної статистики. – К.: «Вища школа», 1977.– 208 с.
- Компьютерная биометрика / Под ред. В.Н.Носова. (Колл. авторов).– М.: Изд-во МГУ, 1990. – 232 с.
- Лакин Г.Ф. Биометрия : Учеб. пособие для биол. специальностей вузов / 4-е изд.– М.: Высш. школа, 1980. – 293 с.
- Литвин І.І., Конончук О.М., Дещинський Ю.Л. Інформатика : Теоретичні основи і практикум / Підр.– Львів: Нов.світ, 2004. – 304 с.
- Рокицкий П.Ф. Биологическая статистика : Учеб. пособие для биол. фак. ун-тов./ Изд. 4-е.– Минск: Высш. школа, 1978. – 320 с.
- Тарасова В.В. Екологічна статистика (з блочно-модульною формою контролю знань). Підручник. – К.: Центр учбової літератури, 2008.– 392 с.
-

Історія

- Біля витоків біометрії стояв Френсіс Гальтон (1822-1911). Спочатку Гальтон готувався стати лікарем. Однак, навчаючись в Кембриджському університеті, він захопився природознавством, метеорологією, антропологією, спадковістю і теорією еволюції. У його книзі, присвяченій природній спадковості, виданій в 1889, вперше було введено слово *biometry*; в цей же час він розробив основи кореляційного аналізу. Гальтон заклав основи нової науки і дав їй ім'я.

Історія

- Однак перетворив її в наукову дисципліну математик Карл Пірсон (1857-1936). В [1884](#) Пірсон отримав кафедру прикладної математики в [Лондонському](#) університеті, а в 1889 році познайомився з Гальтоном і його роботами. Велику роль у житті Пірсона зіграв зоолог Уелдон [\[En\]](#). Допомагаючи йому в аналізі реальних зоологічних даних, Пірсон запровадив у 1893 р. поняття середнього квадратичного відхилення і коефіцієнта варіації. Намагаючись математично оформити теорію спадковості Гальтона, Пірсон в 1898 р. розробив основи множинної регресії. У 1903 р. Пірсон розробив основи теорії спряженості ознак, а в 1905 р. опублікував основи нелінійної кореляції і регресії.

Історія

- Наступний етап розвитку біометрії пов'язаний з ім'ям великого англійського статистика [Рональда Фішера](#) (1890-1962). Під час навчання в Кембриджському університеті Фішер познайомився з працями [Менделя](#) і Пірсона. У 1913-1915 роках Фішер працював статистиком на одному з підприємств, а в 1915-1919 роках викладав фізику і математику в середній школі. З 1919 року Фішер працював статистиком на дослідній сільськогосподарській станції в Ротамстеді, де він пропрацював до 1933 року. Потім з 1933 по 1943 рік Фішер працював професором в Лондонському університеті, а з 1943 року по 1957 рік завідував кафедрою генетики в Кембриджі. За ці роки ним були розроблені теорія вибірових розподілів, методи [дисперсійного](#) і [дискримінантного](#) аналізу, теорії планування експериментів, [метод максимальної правдоподібності](#) та ін., що становить основу сучасної прикладної статистики та математичної генетики.

Розділи біометрії

- Описова статистика.
- Оцінювання гіпотез, порівняльний аналіз.
- Кореляційний , регресійний і дисперсійний аналіз залежностей.
- Методи аналізу структури та класифікації багатовимірних даних.
- Методи аналізу часових рядів (кроскореляційний, спектральний, автокореляційний).
- Методи планування експерименту та моделювання біологічних процесів і систем.

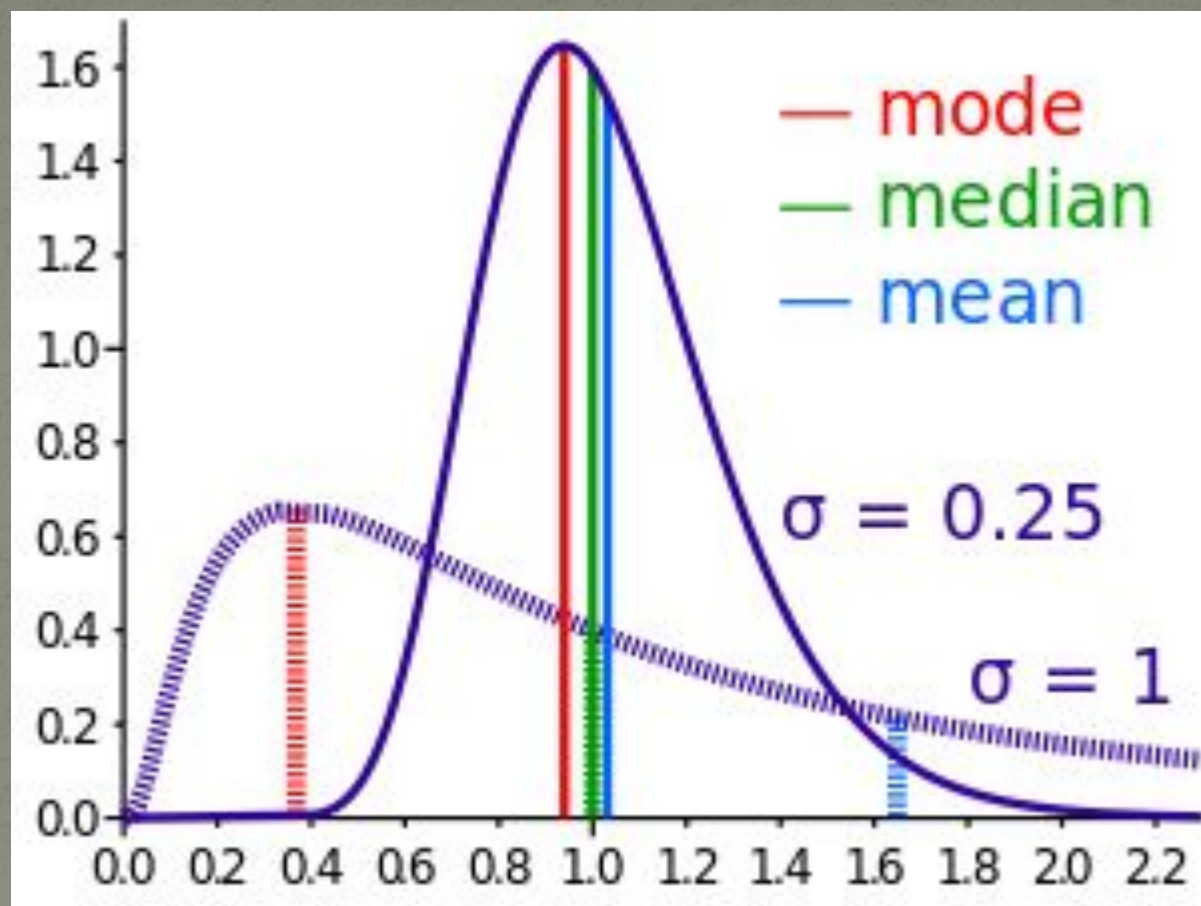
Якісні та кількісні ознаки

- Біологічним об'єктам властива велика різноманітність морфологічних, фізіологічних та інших ознак. Їх можна розділити на кількісні і якісні.
- До *якісних ознак* відноситься стать, забарвлення шерсті, тип тіло- будови та ін. Більшість з якісних ознак мають тільки два можливих альтернативних стани, наприклад стать жіноча або чоловіча, альбінізм – пігментованість, здорові – хворі тварини.
- *Кількісні ознаки* відрізняються тим, що можуть бути виміряні (в кілограмах, сантиметрах, процентах і т .п.) До кількісних ознак відносять живу вагу, вміст цукру в крові та ін.

- В біометрії масовий матеріал називають *генеральною сукупністю*, що становить мету дослідження. Теоретично це безмежно велика або наближається до безмежності сукупність, тобто популяція.
- *Вибіркою* називають множину об'єктів, що випадково відібрані з генеральної сукупності. Кількість об'єктів вибірки називають її об'ємом і позначають n .

Показники , що характеризують середній рівень

- *Середнє арифметичне (вибіркове)*
- **Як математична величина, вибіркове середнє арифметичне має наступні властивості:**
- Сума відхилень від середнього дорівнює нулю.
- Якщо всі значення вибірки збільшити або зменшити, помножити або поділити на одне й те ж число, середнє значення зміниться аналогічно.
- Із збільшенням кількості вимірювань точність оцінки зростає і середнє наближається до математичного очікування (лише у випадку нормального розподілу, $n \rightarrow \infty$).
- Середнє суми або різниці двох вибірок дорівнює сумі або різниці середніх, якщо вибірки однакового розміру.



- Розглянемо числовий *приклад 1*. Визначали вміст вітаміну С у крові хворих людей (в мг%). Виконано 13 аналізів і одержано такі дані: 36,9; 45,4; 51,1; 48,3; 48,3; 39,7; 46,8; 42,6; 43,1; 39,7; 42,9; 35,2; 34,4. Обчислимо середнє арифметичне як найбільш типове для проведеної серії визначень.

$$M = \frac{\sum_{i=1}^{13} x_i}{n} = \frac{554,4}{13} = 42,6$$

Середнє гармонічне M_h застосовується тоді, коли результати спостережень виявляють обернену залежність, задані оберненими значеннями варіант. Обчислення проводять за формулою:

$$M_h = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

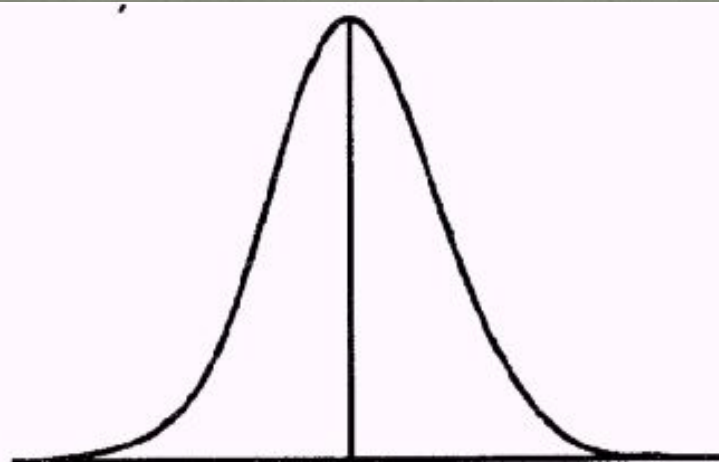
- Обчислення *середнього геометричного*
- необхідне для визначення темпу зміни ознаки, якщо вона вимірюється в часі чи в просторі. Наприклад, якщо досліджувати темп росту рослини залежно від часу доби – вночі ріст сповільнюється, в денну пору рослина росте швидше. Тут варто використовувати середнє геометричне, а не стандартне середнє арифметичне експериментальної вибірки.

$$M_g = \sqrt[n]{(x_1 \cdot x_2 \cdots x_n)}$$

- Середнє квадратичне M_q . Для більш точної числової характеристики мір площі застосовують середнє квадратичне. Його обчислюють за формулою

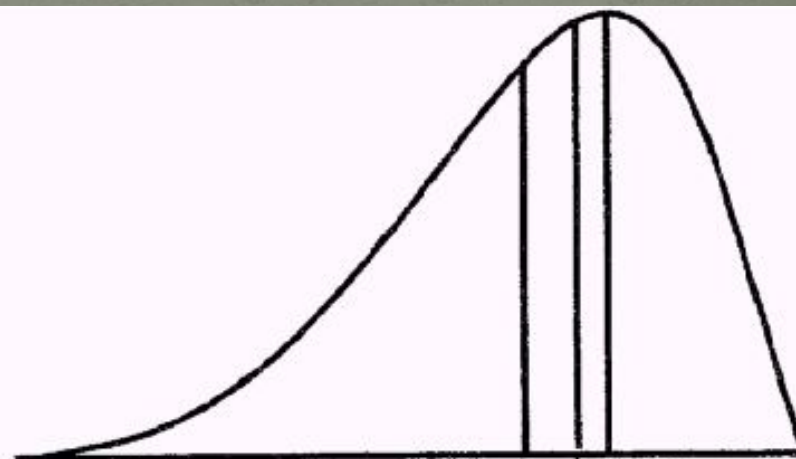
$$M_q = \sqrt{\frac{\sum x_i^2}{n}}$$

- У випадку великої сукупності експериментальних даних для визначення моди дані формують у варіаційний ряд. Спочатку їх сортують за спаданням значень вимірюваної ознаки. Потім дані рангують, для чого розбивають на k (>5) класів з класовим інтервалом, що дорівнює $h = (x_{max} - x_{min})/k$ і обчислюють частоту потрапляння даних сукупності у кожен з класів. Клас, який містить максимальну частоту, називають *модальним*, а його центральне значення називається *модою*. Варто зауважити, що моду не використовують тоді, коли розподіл даних мультимодальний, т.б. багатoverшинний.
- Отже, *варіаційний ряд* – це подвійний ряд чисел що складається з позначення класів і відповідних частот. Розглянемо приклад формування варіаційного ряду.



Средняя
Медиана
Мода

Симметричное распределение

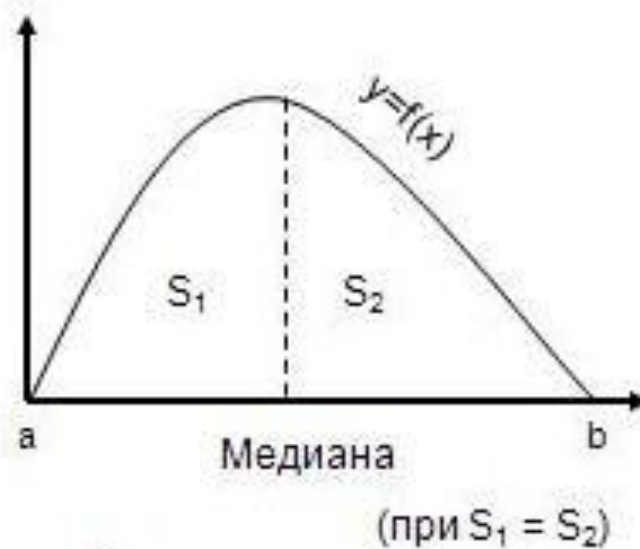


Мода
Медиана
Средняя

Смещенное распределение

- *Медіана* – це значення, яке ділить варіаційний експериментальний ряд на дві рівні за об'ємом групи. У випадку, коли вибірка має парну кількість елементів, значення медіани дорівнюватиме середньому значенню двох центральних членів ряду. На значення медіани не впливає зміна крайніх значень варіаційного ряду, якщо тільки менше від медіани залишається меншим, а більше від медіани залишається більшим. Через цю властивість медіану використовують замість середнього значення у тому випадку, якщо крайні значення варіаційного ряду різко відрізняються від решти.

Медиана (Median)



$$S = \int_a^b f(x) dx$$

Показники мінливості ознаки

- середнє квадратичне відхилення (σ),
- дисперсія,
- коефіцієнт варіації (C_v),
- нормоване відхилення (t),
- розкид даних ($x_{\max} - x_{\min}$).

- Основним критерієм мінливості є *середнє квадратичне відхилення*, який показує на скільки в середньому відхиляється за досліджуваною ознакою кожен член вибірки від середнього арифметичного даної сукупності.
- .

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - M)^2}{(n - 1)}}$$

Дисперсія

- Ще один показник, який служить для оцінки мінливості (варіації) сукупності даних – це емпірична *дисперсія*, яка обчислюється за формулою

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - M)^2}{n - 1}$$

Коефіцієнт варіації.

- Величина σ завжди іменована (кг, см, % і т.п.). Якщо потрібно порівняти ступінь мінливості ознак, що вимірюються в різних одиницях, то визначають коефіцієнт варіації і позначають його C_v .

$$C_v = \frac{\sigma}{M} \cdot 100\%$$

Стандартна похибка середнього арифметичного.

- Бажано, щоб статистичні похибки були якнайменші, тоді вибіркові параметри більш правильно характеризують генеральну сукупність. Величина статистичної похибки залежить від ступеня мінливості ознаки і від кількості даних.

$$m = \frac{\sigma}{\sqrt{n}}$$

$$m = \sqrt{\frac{\sum (x - M)^2}{n(n - 1)}}$$

- Статистично правильний запис результатів експерименту буде тоді, коли ми вкажемо інтервал змінювання середнього значення $M \pm m$ (n)
- (із зазначенням кількості варіант у вибірці).

- Довірчий інтервал для середнього визначається за формулою $[M - t_{n,p} m, M + t_{n,p} m]$, де M – середнє значення вибірки, m – стандартна похибка середнього, $t_{n,p}$ – коефіцієнт Стюдента, що відповідає рівню ймовірності ($p = 0,95$; $p = 0,99$ або $p = 0,999$), знайдений в таблиці значень нормального інтеграла ймовірностей залежно від кількості даних у вибірці n . Якщо n мале, то, звичайно, користуються таблицею Стюдента []. Як правило, вважається, що статистично достовірними є такі твердження, для яких $p \geq 0,95$

Довірчий інтервал індивідуальних значень

- Довірчий інтервал для окремих значень вибірки
- $[M - t_{n,p} \sigma, M + t_{n,p} \sigma]$,
- де σ – середньоквадратичне відхилення, M і $t_{n,p}$ – ті ж показники, які описані вище для довірчого інтервалу для середнього.
- Нагадаємо, що формули для довірчих інтервалів отримані саме для нормального розподілу даних, про який докладніше буде йти мова у наступному розділі.