# Питання

## 1. Методи статистичного відбору

1 Методи статистичного відбору

Способи відбору можна розділити на два види:

1. Відбір що не потребує розділення генеральної сукупності на частини:

а) простий випадковий відбір без повернення

б) -//- з поверненням

2. Відбір, при якому генеральна сукупність розбивається на частини:

а) типовий випадковий відбір;

б) механічний -//-

в) серійний -//-

Простий випадковий відбір - відбір, при якому елементи по одному вибирають із всієї генеральної сукупності. Якщо після обміряння повертаємо об'єкт - простий відбір з поверненням; не повертаємо - простий відбір без повернення.

Типовий випадковий відбір - відбір, при якому об'єкти обирають не з усієї генеральної сукупності, а з окремої її „типової" частини. Застосовується, коли групи дуже відрізняються (різні станки тощо).

Механічний випадковий відбір - відбір, при якому генеральну сукупність ділять на N груп, із кожної групи обирають одну деталь.

Серійний випадковий відбір – об'єкти обирають із генеральної сукупності не по одному, а „серіями", котрі підвергаються суцільному обстеженню.

## 2. Варіаційний ряд. Статисти[illegible]
[illegible]

# 2. Варіаційний ряд. Статистичний розподіл

## 2. Варіаційний ряд. Статистичний розподіл

**Вибірка** – сукупність випадково відібраних об'єктів

**Генеральна сукупність** – вся сукупність об'єктів, з яких проводиться вибірка.

**Об'єм** – к-ть елементів вибірки.

Нехай із генеральної сукупності вилучена вибірка, причому

$$x_1 \to n_1$$
$$x_2 \to n_2$$
$$\ldots$$
$$x_k \to n_k$$

$\sum n_i = n$ – об'єм вибірки

**Варіанта** – кожне значення вибірки, яке ми розглядаємо

**Варіаційний ряд** – послідовність варіант у зростаючому порядку.

**Частота** – к-ть спостережень, а їх відношення до об'єму вибірки $W_i = \frac{n_i}{n}$ – **відносна частота**

**Статистичний розподіл вибірки** – перелік варіант, частот або відносних частот, що їм відповідає.

| $x_i$ | 3 | 4 | 5 |
|---|---|---|---|
| $n_i$ | 1 | 2 | 3 |
| $W_i$ | $\frac{1}{6}$ | $\frac{2}{6}=\frac{1}{3}$ | $\frac{3}{6}=\frac{1}{2}$ |

приклад статистичного розподілу

Іноді задають у вигляді інтервалу.

## 3. Емпірична функція розподілу

### 3. Емпірична функція розподілу

Нехай відомий статистичний розподіл частот кількісної ознаки $\xi$.

$n_x$ – к-ть спостережень $x$

$n$ – об'єм вибірки.

$\omega_x = \frac{n_x}{n}$ – відносна частота події $x$.

Емпірична функція розподілу (функція розподілу вибірки) – ф-ція $F^*(x)$, яка визначає для кожного $x$ частоту (відносну) події $\xi < x$:

$$F^*(x) = \frac{n_x}{n}.$$

де $n_x$ – число варіант, менших за $x$.

Теоретична функція розподілу – функція розподілу генеральної сукупності.

Згідно з теоремою Бернуллі:

$$\lim_{n\to\infty} P\{|F(x)-F^*(x)| < x\} = 1.$$

Це дає можливість наближувати теоретичну функцію розподілу емпіричною.

### Властивості емпіричної функції розподілу

1) $F^*(x) \in [0,1]$

2) $F^*(x)$ – неспадна

3) $F^*(x) = 0$ при $x \le x_1$, де $x_1$ – найменша варіанта

4) $F^*(x) = 1$ при $x > x_k$, де $x_k$ – найбільша варіанта

# 4. Графічне представлення статистичного розподілу.

4 Графічне представлення статистичного розподілу

Полігон частот — ломана лінія, відрізки якої з'єднують точки $(x_1, n_1), (x_2, n_2), \dots, (x_k, n_k)$



Полігон відносних частот — ломана лінія, відрізки якої з'єднують точки $(x_1, w_1), (x_2, w_2), \dots, (x_k, w_k)$



Гістограма частот (для інтервалів)



Гістограма в-них частот

5. Генеральна і вибіркова середня. Властивості. Групова та загальна середні

5. Вибіркова середня, її властивості

Нехай для вивчення генеральної сукупності відносно кількісної ознаки $X$ вилучена вибірка об'ємом $n$.

Вибіркова середня $\bar{x}_в$ – середнє арифметичне значення вибіркової сукупності.

$x_1, \dots, x_k$

$n_1, \dots, n_k$ — об'єм вибірки: $n_1+n_2+\dots+n_k=n$

$$\bar{x}_в = \frac{1}{n}\sum_{i=1}^{k} x_i n_i$$

Вибіркова середня в загальному випадку є випадковою величиною, що залежить від вибірки.

Генеральна середня $\bar{x}_г$ – середнє арифметичне значення ознаки генеральної сукупності.

$$\bar{x}_г = \frac{1}{N}\sum_{i=1}^{k} x_i n_i$$

В якості оцінки генеральної ~~сукупності~~ середньої береться випадкова середня

$$M\bar{x}_в = \bar{x}_г \qquad Mx_i = a$$

$$M\bar{x}_в = M\left(\frac{1}{n}\sum_{i=1}^{k} x_i n_i\right) = \frac{1}{n}\, a \underbrace{\sum_{i=1}^{n} n_i}_{=n} = a.$$

Вибіркове середнє є незміщеною оцінкою генеральної середньої.

Стійкість вибіркових середніх:

Якщо є вибірка генеральної сукупності великого об'єму, то: ~~що~~

чим більше обсяг вибірок тим менше будуть відрізнятися вибіркові середні

5. Генеральна і вибіркова середня. Властивості. Групова та загальна середня

Групова середня – середнє арифметичне значень ознаки, що належать групі.

Загальною середньою $\bar{X}$ називають середнє арифметичне значень ознаки, що належать всій сукупності.

6. Генеральна і вибіркова дисперсія. Властивості. Групова, внутрішньогрупова та міжгрупова дисперсія. Незміщена оцінка генеральної дисперсії

6. Вибіркова дисперсія, її властивості

Вибіркова дисперсія $D_В$ – середнє арифметичне квадратів відхилення спостережених значень ознаки від їх середнього значення:

$$D_В = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x}_В)^2 n_i$$

Окрім дисперсії для характеристики розсіювання значень ознаки ~~використовують~~ вибіркової сукупності навколо його середнього значення використовують середнє квадратичне відхилення.

Вибіркове середнє квадратичне відхилення – квадратний корінь із вибіркової дисперсії:

$$\sigma_В = \sqrt{D_В}.$$

Теорема

Дисперсія дорівнює різниці середнього в квадраті і квадрату середнього

$$D = \overline{x^2} - (\overline{x})^2$$

Доведення

$$D = \frac{\sum n_i (x_i - \overline{x})^2}{n} = \frac{\sum n_i \left(x_i^2 - 2x_i\overline{x} + (\overline{x})^2\right)}{n} =$$

$$= \frac{\sum n_i x_i^2}{n} - 2\overline{x}\,\frac{\sum n_i x_i}{n} + (\overline{x})^2\,\frac{\sum n_i}{n} = \overline{x^2} - 2(\overline{x})^2 + (\overline{x})^2 =$$

$$= \overline{x^2} - (\overline{x})^2. \blacksquare$$

6. Генеральна і вибіркова дисперсія. Властивості. Групова, внутрішньогрупова та міжгрупова дисперсія. Незміщена оцінка генеральної дисперсії

Генеральна дисперсія $D_\Gamma$ – називається середнє арифметичне квадратів відхилень значень ознаки від генеральної сукупності

$$D_\Gamma = \frac{1}{N}\sum_{i=1}^{k}(x_i - \overline{x}_\Gamma)^2 n_i$$

$$D_\Gamma = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x}_\Gamma)^2$$

Групова дисперсія $D_{\Gamma P}$ – називають дисперсію значень ознаки, що належать групі відносно групової середньої:

$$D_{j\Gamma P} = \frac{1}{N_j}\sum_i n_i^{(j)}(x_i - \overline{x}_j)^2, \quad N_j = \sum n_i^{(j)}$$

Внутрішньогрупова дисперсія $D_{B\Gamma}$ – зважене середнє групових дисперсій:

$$D_{B\Gamma} = \frac{1}{N}\sum N_j D_{j\Gamma P}$$

Зовнішньогрупова дисперсія $D_{M\Gamma}$ – дисперсія групових середніх відносно спільної середньої:

$$D_{M\Gamma} = \frac{1}{N}\sum N_j D_{j\Gamma P}$$

Незміщена оцінка ген. дисперсії виправлена дисп.

$$S^2 = \frac{n}{n-1}D_B = \frac{n}{n-1}\cdot\frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \overline{x}_B)^2 = \frac{1}{n-1}\sum n_i(x_i - \overline{x}_B)^2$$

## 7. Статистичні оцінки параметрів розподілу. Їх властивості.

### 7 Статистичні оцінки параметрів розподілу. Їх властивості.

Нехай маємо дані вибірки, наприклад, значення кількісної ознаки $x_1, x_2, \dots, x_n$, отримані у результаті n спостережень.

Розглядаємо $x_1, x_2, \dots, x_n$ як незалежні випадкові величини $\xi_1, \xi_2, \dots, \xi_n$.

Статистична оцінка невідомого параметра теоретичного розподілу — функція від спостережених випадкових величин (наприклад, для мат. очікування оцінкою служить функція

$$\bar{\xi} = \frac{\xi_1 + \xi_2 + \xi_3 + \dots + \xi_n}{n}$$
)

Властивості оцінок

Нехай $\Theta^*$ — статистична оцінка невідомого параметру $\Theta$.

1. Незміщена статистична оцінка — оцінка, мат. сподівання якої дорівнює оцінюваному параметру при б-якому обсязі вибірки:

$$M(\Theta^*) = \Theta$$

2. Зміщена статистична оцінка — оцінка, —//— не дорівнює —//—.

3. Ефективна статистична оцінка — оцінка, котра при заданому об'ємі вибірки n має найменшу можливу дисперсію.

4. Слушна (консистентна) оцінка — оцінка, що при $n \to \infty$ прямує за ймовірністю до оцінюваного параметру.

$$P\{\Theta^* - \Theta\} \to 0$$

8. Нерівність Крамера-Рао для абсолютно неперервних розподілів (лема і теорема).

8 Нерівність Крамера-Рао для абсолютно неперервних розподілів (лема і теорема).

**Лема**

Якщо щільність $f(x,\theta)$, $x\in R^n$ абсолютно неперервного вибіркового вектора є двічі диференційована по параметру $\theta$, а її похідні мажоровані інтегрованими функціями і виконується умова

$$\left|\frac{\partial f(x,\theta)}{\partial\theta}\right| < J_1(x), \quad \left|\frac{\partial^2 f(x,\theta)}{\partial\theta^2}\right| \le J_2(x)$$

$$\int_{R^n} J_i(x)\,dx<\infty, \quad i=1,2,\dots, \quad x\in R^n$$

Умова: $M\left(\frac{\partial}{\partial\theta}\ln f(\xi,\theta)\right)^2<\infty$

$$M\left(\frac{\partial^2}{\partial\theta^2}\ln f(\xi,\theta)\right)<\infty \quad \forall\theta\in\Theta$$ – замкн. інтервал

то $M\left(\frac{\partial}{\partial\theta}\ln f(\xi,\theta)\right)=0$

$$D\left(\frac{\partial}{\partial\theta}\ln f(\xi,\theta)\right)=-M\frac{\partial^2}{\partial\theta^2}\ln f(\xi,\theta)$$

**Доведення**

Підрахуємо мат. сподівання (за умовами леми вони існують):

$$M\frac{\partial}{\partial\theta}\ln f(\xi,\theta)=\int_{R^n}\frac{\partial}{\partial\theta}\ln f(x,\theta)\cdot f(x,\theta)\,dx = \int_{R^n} f'_\theta(x,\theta)\,dx.$$

$$M\frac{\partial^2}{\partial\theta^2}\ln f(\xi,\theta) = \int_{R^n}\frac{\partial^2}{\partial\theta^2}\ln f(x,\theta)\cdot f(x,\theta)\,dx = \int_{R^n}\left(\frac{\partial}{\partial\theta}\frac{f'(x,\theta)}{f(x,\theta)}\right)f(x,\theta)\,dx = \int_{R^n}\left(\frac{f''(x,\theta)f(x,\theta)-(f'(x,\theta))^2}{f^2(x,\theta)}\right)f(x,\theta)\,dx =$$

$$=\int_{R^n} f''(x,\theta)\,dx - \int_{R^n}\left(\frac{f'(x,\theta)}{f(x,\theta)}\right)^2 f(x,\theta)\,dx = \int_{R^n} f''(x,\theta)\,dx - M\left(\frac{\partial}{\partial\theta}\ln f(\xi,\theta)\right)^2 \quad (1)$$

## 9. Наслідки з нерівності Крамера-Рао для абсолютно неперервних розподілів

9. Наслідки з нерівності Крамера-Рао для абсолютно неперервних розподілів.

Наслідок 1

Якщо оцінка $\theta^*$ - незміщена, і для неї нерівність Крамера-Рао перетворюється на рівність, то $\theta^*$ є ефективною оцінкою Крамера-Рао:

$$\frac{1}{I(\theta)} = D(\theta^*)$$

**Доведення**

$$D(\theta^*) = \frac{1}{I(\theta)} \le D(\hat\theta)$$

$$D(\theta^*) \le \inf_{\hat\theta:\, M\hat\theta = \theta} D\hat\theta$$

Оскільки $\theta^*$ - незміщена оцінка, то

$$\inf_{\hat\theta:\, M\hat\theta = \theta} D\hat\theta \le D\theta^*$$

Отже, $\theta^*$ - ефективна оцінка пар. $\theta$.

Наслідок 2

Нехай щільність розподілу вибіркового вектора $f(x,\theta)$, $x \in R^n$, $h(x)$ - статистика, для якої викон. умови теореми.

Якщо $\frac{\partial}{\partial\theta} \ln f(x,\theta) = C(\theta)\left(h(\theta) - \theta\right)$, $x \in R^n$ (1)

то $\hat\theta = h(\xi)$ - ефективна оцінка параметру $\theta$.

**Доведення**

За умовами теореми, (1) перетворює нерівність Крамера-Рао на рівність, а, отже, $\hat\theta = h(\xi)$ - еф. оц.

Наслідок 3

Нехай $\xi = (\xi_1, \dots, \xi_n)$ - вибірка із розподілу з щільністю $g(x,\theta)$, $x \in R^n$. Для $f(x,\theta)$ виконані умови теореми:

$$f(x,\theta) = \prod_{i=1}^{n} g(x_i,\theta)$$

Тоді нерівність Крамера-Рао матиме вигляд:

10. Нерівність Крамера-Рао для дискретних розподілів (лема, теорема, наслідок)

## 10. Нерівність Крамера – Рао для дискретних розподілів (лема, теорема, наслідок)

Нерівність Крамера – Рао має майже той самий вигляд для дискретного розподілу

Якщо в нас є вибірковий вектор $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ і його розподіл $P(\cdot, \theta)$, який називається дискретним, якщо існує не більше ніж зліченна множина $X \subset R^n$, така, що

$$P_\xi(x) = P(x, \theta) \qquad x \in X$$

**Лема**

Якщо розподіл $P(x, \theta)$ дискретного вибіркового вектора $\xi = (\xi_1, \ldots, \xi_n)$ двічі диф. по $\theta$, тобто існують

$$\frac{\partial}{\partial \theta} P(x, \theta), \quad \frac{\partial^2}{\partial \theta^2} P(x, \theta)$$

і ряди абсолютно збіжні

$$\sum_{x \in X} \left| \frac{\partial}{\partial \theta} P(x, \theta) \right| < \infty \qquad \sum_{x \in X} \left| \frac{\partial^2}{\partial \theta^2} P(x, \theta) \right| < \infty$$

рівномірно в-но $\theta$ і виконується умова

$$M \left( \frac{\partial}{\partial \theta} \ln P(\xi, \theta) \right)^2 < \infty \qquad M \left( \frac{\partial^2}{\partial \theta^2} \ln P(\xi, \theta) \right) < \infty$$

то

$$M \left( \frac{\partial}{\partial \theta} \ln P(\xi, \theta) \right) = 0$$

$$D \left( \frac{\partial}{\partial \theta} \ln P(\xi, \theta) \right) = -M \frac{\partial^2}{\partial \theta^2} \ln P(\xi, \theta)$$

Доведення аналогічно неперервному випадку.

**Теорема** Нехай дискретний вектор $\xi = (\xi_1, \dots, \xi_n)$ з розподілом $P(x, \theta)$ і нехай $\hat{\theta} = h(\xi)$ — незміщена оцінка параметру $\theta$, причому ряд

$$\sum_{x \in X} \frac{\partial}{\partial \theta} \left( h(x) P(x, \theta) \right)$$ збіжний абсолютно і рівномірно в-но $\theta$.

Тоді

$$\frac{1}{I(\theta)} \leq D(\hat{\theta})$$

причому знак рівності досягається, коли

$$\frac{\partial}{\partial \theta} \ln P(x, \theta) = C(\theta) (h(x) - \theta)$$

Доведення аналогічно неперервному випадку

**Наслідок**

Якщо $\xi_1, \dots, \xi_n$ — вибірка дискретного розподілу і для сумісного розподілу $P(x, \theta)$ виконується умови теореми

$$P(x_1, \dots, x_n, \theta) = \prod_{i=1}^{n} P(x_i, \theta)$$

Тоді нерівність Крамера – Рао приймає вигляд

$$\frac{1}{n M \left( \frac{\partial}{\partial \theta} \ln P(\xi_1, \theta) \right)^2} \leq D\hat{\theta}$$

# 11. Метод моментів отримання оцінок.

11 Метод моментів отримання оцінок

Нехай $\xi_1, \xi_2, \dots, \xi_n$ – вибірка з розподілом $F(\cdot; \theta_1, \theta_2, \dots, \theta_s)$

Метод полягає в тому, що ми прирівнюємо теоретичні моменти емпіричним того ж порядку:

$$m_k(\theta_1, \dots, \theta_s) = \int_R x^k F(dx; \theta_1, \dots, \theta_s)$$

$\hat{m}_k$ – вибіркові моменти $k$-го порядку, які підраховані при значенні оцінок $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_s$.

Оскільки у нас уже є $S$ оцінок, то пишемо $S$ рівнянь:

$$\hat{m}_1 = \int_{R^1} x F(dx, \hat{\theta}_1, \dots, \hat{\theta}_s)$$

$$\hat{m}_2 = \int_{R^1} x^2 F(dx, \hat{\theta}_1, \dots, \hat{\theta}_s)$$

$$\dots$$

$$\hat{m}_s = \int_{R^1} x^s F(dx, \hat{\theta}_1, \dots, \hat{\theta}_s)$$

Кожне з рівнянь прирівнюємо до емпіричного моменту необхідного порядку.

Для того, щоб методом можна було користуватись, необхідно, щоб були відомі теорет. моменти.

# 12. Метод максимальної правдоподібності отримання оцінок.

12. Метод максимальної правдоподібності отримання оцінок

Нехай є вибірковий вектор $\xi = (\xi_1, \dots, \xi_n)$ і його реалізація $\xi(\omega)$ — значення в точці

$$\xi(\omega) = (\xi_1(\omega), \dots, \xi_n(\omega)),$$

а також розподіл $F(\cdot, \theta)$, де $\theta$ — параметр.

Задача полягає в тому, щоб оцінити значення параметру, тобто з $\theta \in \Theta$ вибрати таку оцінку $\hat{\theta}$, яку ми будемо використовувати в якості невідомого значення параметру.

Щільність реалізації $\xi(\omega)$ набуває max значення, тобто

$$f(\xi(\omega)) = \max_{\theta \in \Theta} f(\xi(\omega), \theta)$$

В якості оцінки $\hat{\theta}$ параметра $\theta$ беремо точку $\xi(\omega)$, в якій досягається найбільше значення.

Функція максимальної правдоподібності вибірки $(\xi_1, \dots, \xi_n)$ розподілу $F(\cdot, \theta)$, що залежить від параметру $\theta$ — це функція $L(\theta) = L(\theta_1, \theta_2, \dots, \theta_3)$, яка визначається рівністю:

$$L(\theta) = f(\xi, \theta), \ \theta \in \Theta$$ — коли вибірковий вектор абсолютно неперервний

$$L(\theta) = P(\xi, \theta), \ \theta \in \Theta$$ — коли вибірковий вектор дискретний.

# 13. Оцінка максимальної правдоподібності, її властивості

13. Оцінка максимальної правдоподібності, її властивості.

Функція максимальної правдоподібності вибірки $\xi=(\xi_1,\dots,\xi_n)$ з розподілом $F(\cdot,\theta)$, що залежить від параметру $\theta$ – це функція $L(\theta)=L(\theta_1,\dots,\theta_s)$, яка визначається рівністю:

$$L(\theta) = f(\xi,\theta),\ \theta\in\Theta$$ – коли вибірковий вектор абсолютно неперервний

$$L(\theta) = P(\xi,\theta),\ \theta\in\Theta$$ – коли вибірковий вектор дискретний

Оцінкою максимальної вірогідності параметра $\theta$ називається оцінка $\hat\theta$, в якій функція максимальної правдоподібності досягає найбільшого значення

$$L(\hat\theta) = \max_{\theta\in\Theta} L(\theta)$$

або

$$\ln L(\hat\theta) = \max_{\theta\in\Theta} \ln L(\theta)$$

Якщо функція максимальної правдоподібності $L(\theta_1,\dots,\theta_s)$ неперервно диф. по всіх параметрах $\theta_1,\dots,\theta_s$, то для вираження р-ня

$$\ln L(\hat\theta_1,\dots,\hat\theta_s) = \max_{(\theta_1,\dots,\theta_s)\in\Theta} \ln L(\theta_1,\dots,\theta_s)$$

достатньо ~~знайти~~ розв. рівняння

$$\frac{\partial}{\partial\theta_i} \ln L(\hat\theta_1,\dots,\hat\theta_s) = 0, \quad i=\overline{1,s}$$

Система р-нь називається р-ння максимальної вірогідності

Оцінки, знайдені цим методом, будуть консистентними, асимптотично-нормальними і мають найменшу дисперсію порівняно з іншими асимптотичними нормальними оцінками.

19. Довірчі інтервали, довірча ймовірність

Інтервальна оцінка — оцінка, яка визначається двома числами — кінцями інтервала

Нехай $\theta^*$ служить оцінкою невідомого параметру $\theta$.

Якщо $\delta > 0$ і $|\theta - \theta^*| < \delta$, то чим менше $\delta$, тим оцінка точніше. Таким чином, $\delta$ характеризує точність оцінки.

Якщо візьмемо число $\gamma$ і будемо вимагати, щоб

$$P\{|\theta^* - \theta^*| < \delta\} = \gamma$$

то $\gamma$ — надійність (довірча ймовірність)

$\gamma \sim 1$ : 0,95 ; 0,99

$$P\{\theta^* - \delta < \theta < \theta^* + \delta\} = \gamma$$

$\gamma$ — ймовірність того, що інтервал $(\theta^* - \delta; \theta^* + \delta)$ покриває невідомий параметр $\theta$.

Довірчий інтервал — інтервал $(\theta^* - \delta; \theta^* + \delta)$, який покриває невідомий параметр з заданою ~~ймовірністю~~ надійністю $\gamma$.

## 14. Довірчі інтервали для оцінки математичного сподівання нормального розподілу при відомій дисперсії.

20. Довірчі інтервали для оцінки мат. сподів. нормального розподілу при відомій дисперсії

Нехай у нас є вибірка $\xi_1, \dots, \xi_n$ з нормальним розподілом $N(a, \sigma)$

Задача — побудувати довірчий інтервал, який покриває параметр з заданою надійністю $\gamma$

$$\bar{\xi} = \frac{1}{n}\sum_{i=1}^{n} \xi_i$$

$N\left(a, \frac{\sigma}{\sqrt{n}}\right)$ — випливає із мат. сподівання та дисперсії

$$P\{|\bar{\xi} - a| < \delta\} = \gamma$$

$$P\{|\xi - a| < \delta\} = 2\Phi\left(\frac{\delta}{\sigma}\right)$$

замінивши $\xi$ на $\bar{\xi}$ та $\sigma(\bar{\xi}) = \frac{\sigma}{\sqrt{n}}$, отримаємо

$$P\{|\bar{\xi} - a| < \delta\} = 2\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right) = 2\Phi(t)$$

$$\delta = \frac{t\sigma}{\sqrt{n}}$$

$$P\left\{\bar{\xi} - \frac{t\sigma}{\sqrt{n}} < \delta < \bar{\xi} + \frac{t\sigma}{\sqrt{n}}\right\} = 2\Phi(t) = \gamma.$$

Довірчий інтервал $\left(\bar{\xi} - \frac{t\sigma}{\sqrt{n}}; \bar{\xi} + \frac{t\sigma}{\sqrt{n}}\right)$ покриває параметр $a$ із надійністю $\gamma$. Дану оцінку називають класичною

## 15. Довірчі інтервали для оцінки математичного сподівання нормального розподілу при невідомій дисперсії.

21. Довірчі інтервали для оцінки мат. сподівання нормального розподілу при невідомій дисперсії

Нехай маємо вибірку $\xi_1, \ldots, \xi_n$ з норм. розподілом $N(a, \sigma)$

Побудуємо по даним вибірки випадкову величину:

$$t = \frac{\bar{\xi} - a}{S/\sqrt{n}}$$

S – виправлене середньоквадр. відхилення,

$$S = \frac{n}{n-1}\sigma$$

Випадкова величина має розподіл Стьюдента. Щільність розподілу Стьюдента:

$$S(t,n) = B_n \left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}},$$

де $B_n = \dfrac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)}\,\Gamma\left(\frac{n-1}{2}\right)}$

$$P\left\{ \left| \frac{\bar{\xi} - a}{S/\sqrt{n}} \right| < t_\gamma \right\} = 2\int_0^{t_\gamma} S(t,n)\,dt = \gamma$$

$$P\left\{ \bar{\xi} - \frac{t_\gamma S}{\sqrt{n}} < a < \bar{\xi} + \frac{t_\gamma S}{\sqrt{n}} \right\} = \gamma$$

Отже, довірчий інтервал $\left(\bar{\xi} - \frac{t_\gamma S}{\sqrt{n}}, \bar{\xi} + \frac{t_\gamma S}{\sqrt{n}}\right)$ покриває параметр $a$ із надійністю $\gamma$.

## 16. Довірчі інтервали для оцінки середньоквадратичного відхилення нормального розподілу. Оцінка точності вимірювань.

11. Довірчі інтервали для оцінки середньоквадратичного відхилення нормального розподілу. Оцінка точності вимірювань

Маємо вибірку $\xi_1, \ldots, \xi_n$. Потрібно побудувати із дов. інтервал, покриваючий параметр $\sigma$ із надійністю $\gamma$.

$$P\{|\sigma - s| < \delta\} = \gamma$$

$$s - \delta < \sigma < s + \delta$$

$$s\left(1 - \frac{\delta}{s}\right) < \sigma < s\left(1 + \frac{\delta}{s}\right) \qquad q = \frac{\delta}{s}$$

$$s(1 - q) < \sigma < s(1 + q)$$

Введемо випадкову величину $\chi = \frac{S}{\sigma}\sqrt{n-1}$, $S$ - випадкове середньокв. відхилення

$$\frac{S^2}{\sigma^2}(n-1) \sim \chi^2$$

Якщо величина розподілена із степ. вільності $k = n-1$, то щільність розподілу має вигляд:

$$f_{\chi^2}(x) = \frac{x^{\frac{k}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} = \frac{x^{\frac{n-3}{2}} e^{-\frac{x}{2}}}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}$$

$$g(y) = f(\psi(y)) \cdot |\psi'(y)|$$

$$\chi = \varphi(\xi) = \sqrt{\xi} \qquad \psi(\chi) = \chi^2 \qquad |\psi'(\chi)| = 2\chi \ (\text{оскіл. } \chi > 0)$$

$$g(\chi) = f(\psi(\chi)) \cdot |\psi'(\chi)| = \frac{(\chi^2)^{\frac{n-3}{2}} e^{-\frac{\chi^2}{2}}}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)} \cdot 2\chi = \frac{\chi^{n-2} e^{-\frac{\chi^2}{2}}}{2^{\frac{n-3}{2}}\Gamma\left(\frac{n-1}{2}\right)} = R(\chi, n)$$

$$P\{\chi_1 < \chi < \chi_2\} = \int_{\chi_1}^{\chi_2} R(\chi, n)\, d\chi = \gamma.$$

Розглянемо випадки:

1) $q < 1$

$$\frac{1}{S(1+q)} < \frac{1}{\sigma} < \frac{1}{S(1-q)} \quad \Big| * S\sqrt{n-1}$$

$$\frac{\sqrt{n-1}}{1+q} < \frac{S\sqrt{n-1}}{\sigma} < \frac{\sqrt{n-1}}{1-q}$$

$$\frac{\sqrt{n-1}}{1+q} < \chi < \frac{\sqrt{n-1}}{1-q}$$

$$\int_{\frac{\sqrt{n-1}}{1+q}}^{\frac{\sqrt{n-1}}{1-q}} R(\chi, n)\, d\chi = \gamma$$

Знаходимо $q = q(\gamma, n)$ з табл.

2) $q > 1$

$$0 < \sigma < S(1+q)$$

$$\frac{1}{S(1+q)} < \frac{1}{\sigma} < \infty \quad \Big| * S\sqrt{n-1}$$

$$\frac{\sqrt{n-1}}{1+q} < \frac{S\sqrt{n-1}}{\sigma} < \infty$$

$$\frac{\sqrt{n-1}}{1+q} < \chi < \infty$$

$$\int_{\frac{\sqrt{n-1}}{1+q}}^{\infty} R(\chi, n)\, d\chi = \gamma$$

Знаходимо $q = q(\gamma, n)$ з таблиці.

17. Оцінювання ймовірності біноміального розподілу за відомою частотою.

23. Довірчі інтервали для оцінки ймовірності біноміального розподілу за відомою частотою

Нехай проводиться невідома випробування з невідомою ймовірністю $p$ появи події $A$ в кожному випробуванні. Необхідно оцінити $p$ по відносній частоті.

А. Точкова оцінка: $\omega = \frac{m}{n}$; $\xi_m \Rightarrow m$

Випадкова величина, яка є частотою: $\xi_\omega = \frac{\xi_m}{n}$

Оскільки $M\xi_m = np$, то $\xi_\omega = p$ – оцінка є незміщеною.

Знайдемо її дисперсію

$$D\xi_\omega = D\left(\frac{\xi_m}{n}\right) = \frac{1}{n^2} D\xi_m = \frac{npq}{n^2} = \frac{pq}{n}$$

Середньоквадратичне відхилення: $\sigma_\omega = \sqrt{\frac{pq}{n}}$.

Б. Інтервальна оцінка: для норм. випадк. величини,

$$P\{|\xi - a| < \delta\} = 2\Phi\left(\frac{\delta}{\sigma}\right) \quad \xi \sim N(a, \sigma^2)$$

$$P\{|\xi_\omega - p| < \delta\} = 2\Phi\left(\frac{\delta}{\sigma_m}\right)$$

Задамо надійність $\gamma$, з якою будуватимемо дов. інтервали.

$$P\{|\xi_\omega - p| < \delta\} = 2\Phi\left(\frac{\delta \cdot \sqrt{n}}{\sqrt{pq}}\right); \quad t = \frac{\delta\sqrt{n}}{\sqrt{pq}}$$

Маємо р-ня: $2\Phi(t) = \gamma$, $t = \frac{\delta\sqrt{n}}{\sqrt{pq}} \Rightarrow \delta = t\sqrt{\frac{pq}{n}}$

Отже, $P\left\{|\xi_\omega - p| < t\sqrt{\frac{pq}{n}}\right\} = 2\Phi(t) = \gamma$.

З надійністю $\gamma$ виконується рівність

$$\left|\frac{m}{n} - p\right| < t\sqrt{\frac{pq}{n}} = t\sqrt{\frac{p(1-p)}{n}}$$

Розглянемо випадки:

а) $\frac{m}{n} > p$: $\left(\underbrace{\frac{m}{n}}_{=\omega} - p\right) < t\sqrt{\frac{p(1-p)}{n}}$ (привести в квадрат, поделить на $n$, получить кв. уравнение)

$$p_1 = \frac{n}{t^2+n}\left(\omega + \frac{t^2}{2n} \mp t\sqrt{\frac{\omega(1-\omega)}{n} + \left(\frac{t}{2n}\right)^2}\right)$$

$p_2$ ?+

$p_1 < p < p_2$

б) $\frac{m}{n} < p$

$$p_1 = \omega - t\sqrt{\frac{\omega(1-\omega)}{n}} \qquad p_2 = \omega + t\sqrt{\frac{\omega(1-\omega)}{n}}$$

$p_1 < p < p_2$.

# 18. Принцип практичної впевненості

## 10.1. Принцип практической уверенности

Прежде чем перейти к рассмотрению понятия статистической гипотезы, сформулируем так называемый **принцип практической уверенности**, лежащий в основе применения выводов и рекомендаций с помощью теории вероятностей и математической статистики.

*Если вероятность события A в данном испытании очень мала, то при однократном выполнении испытания можно быть уверенным в том, что событие A не произойдет, и в практической деятельности вести себя так, как будто событие A вообще невозможно.*

Этот принцип не может быть доказан математически; он подтверждается всем практическим опытом человеческой деятельности, и мы постоянно (хотя и бессознательно) им руководствуемся. Например, отправляясь самолетом в другой город, мы не рассчитываем на возможность погибнуть в авиационной катастрофе, хотя некоторая (весьма малая) вероятность такого события все же имеется.

Обратим внимание на то, что принцип практической уверенности о невозможности маловероятных событий сформулирован «при однократном выполнении испытания». Если же произведено много испытаний, в каждом из которых вероятность события $A$ даже очень мала, то существенно повышается вероятность того, что событие $A$ произойдет хотя бы один раз в массе испытаний. Действительно, пусть вероятность $P(A) = \alpha$, где $\alpha << 1$. Тогда вероятность события $B$, состоящего в том, что событие $A$ произойдет хотя бы один раз в $n$ независимых испытаниях, по формуле **(1.29)** равна (при $\alpha << 1$):

$$P(B)=1-(1-\alpha)^n \approx 1-(1-n\alpha)=n\alpha,$$

т.е. вероятность $P(B)$ увеличилась по сравнению с $P(A)$ в $n$ раз.

Таким образом, при многократном повторении испытаний мы уже не можем считать маловероятное событие $A$ практически невозможным.

Вопрос о том, насколько мала должна быть вероятность $\alpha$ события $A$, чтобы его можно было считать практически невозможным, выходит за рамки математической теории и решается в каждом отдельном случае с учетом важности последствий, вытекающих из наступления события $A$. В одних случаях считается возможным пренебрегать событиями, имеющими вероятность меньше 0,05, а в других, когда речь



идет, например, о разрушении сооружений, гибели судна и т.п., нельзя пренебрегать событиями, которые могут появиться с вероятностью, равной 0,001.

19. Поняття статистичної гіпотези та статистичного критерію. Види гіпотез. Види помилок. Значущість та потужність критерію.

24. Поняття статистичної гіпотези та статистичного критерію. Види гіпотез. Види помилок. Значущість та потужність критерію.

Статистична гіпотеза - б-яке твердження про вигляд розподілу або властивості розподілу

Основна (нульова) гіпотеза $H_0$ - гіпотеза, яку перевіряють

Альтернативна (конкуруюча) гіпотеза - $H_1$ - гіпотеза, яка протиречить нульовій.

Проста гіпотеза - не містить додаткових параметрів

Складна гіпотеза - гіпотеза, що склад з кінечної або нескінченної к-ті простих гіпотез.

Статист. Критерій $K$ - статистика, яку обирають для перевірки гіпот.

Ймовірність того, що допуститься помилка I роду, називається рівнем значущості

Види помилок:

1) Помилка I роду - відкинута гіпотеза $H_0$ за умови, що вона правильна

2) Помилка II роду - прийняти гіпотезу $H_0$, якщо вона неправильна

Потужність критерію - ймовірність того, що нульова гіпотеза буде відкинута, якщо вірна альтернативна гіпотеза.

## 20. Прості гіпотези. Лема Неймана-Пірсона.

### §1. Прості гіпотези. Лема Неймана-Пірсона

Нехай вибірка $x_1, \dots, x_n$ проведена з генеральної сукупності з ф-цією розподілу $F(x)$. Тоді маємо дві гіпотези: (прості):

$H_0: \theta = \theta_0$ $\quad$ $\theta_0, \theta_1$ – відомі

$H_1: \theta = \theta_1$ $\quad$ функції розподілу.

Побудуємо функцію правдоподібності критерію $H_i$:

$$\prod_{i=1}^{n} p(x_i, \theta_i(x))$$

$$X : \{x_1, \dots x_n\} : \prod_{i=1}^{n} p(x_i, \theta_1(x)) \ge c \prod_{i=1}^{n} p(x_i, \theta_0(x))\}, \quad c \ge 0.$$

c – довільна константа,

Відношення правдоподібності

$$l(x) = \frac{\prod_{i=1}^{n} p(x_i, \theta_1)}{\prod_{i=1}^{n} p(x_i, \theta_0)}$$

$P\{X|H_0\} = g(c)$, $\quad$ g – незрост.: $c_1 \le c_2$ $\quad g(c_1) \ge g(c_2)$

$$cg(c) = cP\{X|H_0\} = c \int_X \prod_{i=1}^{n} p(x_k; \theta_0)\, dx_1, \dots, dx_n \le$$

$$\le \int_X \prod_{i=1}^{n} p(x_k, \theta_1)\, dx_1, \dots dx_n \le 1.$$

$$g(c) \le \frac{1}{c} \xrightarrow[c \to 0]{} 0$$

$\exists c : g(c) = \alpha$ – умова не необхідна

$$P\{X|H_0\} = g(c) = \alpha.$$

$\frac{1}{10}$ 196

Лема Неймана-Пірсона

Серед всіх критеріїв $\alpha$ з заданим рівнем значущості які перевіряють гіпотези $H_0, H_1$, критерій відношення правдоподібності є найбільш потужним

$$l(x) = \frac{\prod_{i=1}^{n} p(x_i, \theta_1)}{\prod_{i=1}^{n} p(x_i, \theta_0)} \quad c$$

Доведення

Позначимо через $Y$ крит. множину б-якого іншого критерію, який має рівень значущості $\alpha$.

Доведемо, що

$$P\{X|H_1\} \geq P\{Y|H_1\}$$

$$P\{X \setminus (X \cap Y)|H_0\} = P\{X|H_0\} - P\{X \cap Y|H_0\} = \alpha - P\{X \cap Y|H_0\}$$

$$P\{Y \setminus (X \cap Y)|H_0\} = \alpha - P\{X \cap Y|H_0\}$$

$$\forall (x_1 \dots x_n) \in X \qquad c\prod_{i=1}^{n} p(x_i, \theta_0) > \prod_{i=1}^{n} p(x_i, \theta_1)$$

$$P\{X \setminus (X \cap Y)|H_1\} = \int_{X \setminus (X \cap Y)} \prod_{i=1}^{n} p(x_i, \theta_1)\, dx_1 \dots dx_n \geq$$

$$\geq c \int_{X \setminus (X \cap Y)} \prod_{i=1}^{n} p(x_i, \theta_0)\, dx_1 \dots dx_n = cP\{X \setminus X \cap Y|H_0\} =$$

$$= cP\{Y \setminus X \cap Y|H_0\} - c\int_{Y \setminus (X \cap Y)} \prod_{i=1}^{n} p(x_i, \theta_0)\, dx_1 \dots dx_n >$$

$$> \int_{Y \setminus (X \cap Y)} \prod_{i=1}^{n} p(x_i, \theta_1)\, dx_1 \dots dx_n = P\{Y \setminus X \cap Y|H_1\}$$

$+$ $P\{X \cap Y|H_1\}$ — додамо до лівої і правої частини нерівності.

$$P\{X|H_1\} \geq P\{Y|H_1\} \blacksquare$$

## 21. Порівняння двох дисперсій нормальних генеральних сукупностей.

26. Порівняння двох дисперсій нормальних генеральних сукупностей

Нехай в нас є дві вибірки генеральних сукупностей $\xi$ і $\eta$

обʼєми

$\xi$ $n_1$ $\xi_1, \dots, \xi_{n_1}$

$\eta$ $n_2$ $\eta_1, \dots, \eta_{n_2}$

Перевіримо гіпотезу $H_0: D\xi = D\eta$ при заданому рівні значущості $\alpha$.

$MS_\xi^2 = D\xi$ ; $MS_\eta^2 = D\eta$, де s - виправлена вибіркова дисперсія

$$H_0: MS_\xi^2 = MS_\eta^2$$

① $H_1: D\xi > D\eta$

В якості критерію перевірки нульової гіпотези беремо відношення більшої випр. вибір. дисперсії до меншої, тобто випадкову величину із розподілом Фішера.

$$F_{спост} = \frac{S_Б^2}{S_М^2}$$ Ступені вільності: $k_1 = n_Б - 1$, $k_2 = n_М - 1$

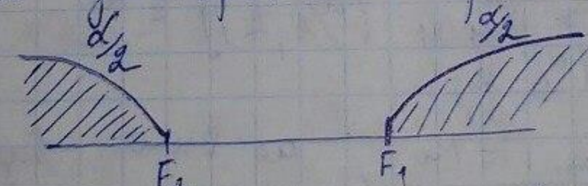
За таблицею крит. точок розподілу Фішера знаходимо $F_{кр}(\alpha, k_1, k_2)$.

$F_{спост} < F_{кр}$ - не можемо відхилити $H_0$

$F_{спост} > F_{кр}$ - відхиляємо $H_0$.

② $H_1: D\xi \neq D\eta$ $\qquad F_{спост} = \frac{S_Б^2}{S_М^2}$

Будуємо двосторонню крит. область




Найбільша потужн. критерію досягається тоді, коли:

$$P\{F < F_1\} = \frac{\alpha}{2}$$

$$P\{F > F_2\} = \frac{\alpha}{2}$$

$$F_2 = F_{кр}(\alpha/2; k_1, k_2)$$

$F_{сп} < F_{кр}$ - $H_0$

$F_{сп} > F_{кр}$ - $\cancel{H_0}$

22. Порівняння виправленої вибіркової дисперсії з гіпотетичною ген. Дисперсією нормальної сукупності.

26. Порівняння виправленої вибіркової дисперсії з гіпотетичною генеральною дисперсією норм. сукупності

Нехай генеральна сукупність розподілена нормально. ~~відому~~ значення гіпотетичної дисперсії $\sigma_0^2$.

$$H_0: MS^2 = \sigma_0^2$$

В якості крит. перевірки гіпотези використовуємо випадкову величину:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$ із ступ. вільності $k = n-1$

1) $H_1: \sigma^2 > \sigma_0^2$

Ймовірність потрапляння критерію в крит. обл. дорів. $\alpha$:

$$P\{\chi^2 > \chi^2_{кр}(\alpha, k)\} = \alpha$$

$\chi^2_{спост} = \frac{(n-1)S^2}{\sigma_0^2}$ ; $\chi^2_{кр}(\alpha, k)$ – з табл.

$\chi^2_{спост} > \chi^2_{кр}$ ; $\chi^2_{спост} < \chi^2_{кр}$ – ...

2) $H_1: \sigma^2 \neq \sigma_0^2$

$$\chi^2_{спост} = \frac{(n-1)S^2}{\sigma_0^2}$$

Будуємо двост. критичну область. Знаходимо праву та ліву крит. точки:

$$P\{\chi^2 < \chi^2_{лів.кр.}(\alpha/2, k)\} = \alpha/2$$

$$P\{\chi^2 > \chi^2_{пр.кр.}(\alpha/2, k)\} = \alpha/2$$

$\chi^2_{лів.кр}(1-\alpha/2, k)$ ; $\chi^2_{пр.кр.}(\alpha/2, k)$.

$\chi^2_{л.кр.} < \chi^2_{спост} < \chi^2_{пр.кр.}$ – $H_0$ ; не викон. – відх.

3) $H_1: \sigma^2 < \sigma_0^2$

$$P\{\chi^2 > \chi^2_{кр}(1-\alpha, k)\} = \alpha \qquad \chi^2_{кр}(1-\alpha; k)$$

$\chi^2_{набл} > \chi^2_{кр}$ – $H_0$ ; $\chi^2_{набл} < \chi^2_{кр}$ – відх.

При $k > 30$ крит. точку можна знайти по рівності Уілсона – Гілферті: $\chi^2_{кр}(\alpha, k) = k\left(1 - \frac{2}{9k} + z_\alpha\sqrt{\frac{2}{9k}}\right)$, $\Phi(z_\alpha) = \frac{1-2\alpha}{2}$

## 23. Порівняння двох середніх нормальних генеральних сукупностей дисперсії яких відомі (незалежні вибірки)

27 Порівняння двох середніх нормальних генер. сукупностей, дисперсії яких відомі (незалежні вибірки).

Нехай маємо дві генеральні сукупності $\xi$ та $\eta$

$\xi$ $\quad \xi_1, \dots, \xi_n \quad \bar{\xi}$

$\eta$ $\quad \eta_1, \dots, \eta_n \quad \bar{\eta}$ — вибіркові

За заданим рівнем значущості $\alpha$ перевірити $H_0$:

$$H_0: M\xi = M\eta$$

$$M\bar{\xi} = M\xi; \quad M\bar{\eta} = M\eta$$

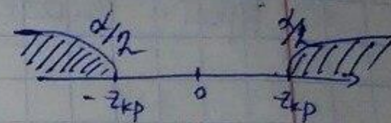
$$H_0: M\bar{\xi} = M\bar{\eta}$$

В якості критерію перевірки гіпотези викор. випад. величину:

$$z = \frac{\bar{\xi} - \bar{\eta}}{\sigma(\bar{\xi} - \bar{\eta})} = \frac{\bar{\xi} - \bar{\eta}}{\sqrt{\frac{D\xi}{n} + \frac{D\eta}{m}}} \sim N(0,1)$$

$$\sigma(\bar{\xi} - \bar{\eta}) = \sqrt{D(\bar{\xi} - \bar{\eta})} = \sqrt{D\bar{\xi} + D\bar{\eta}} = \sqrt{\frac{D\xi}{n} + \frac{D\eta}{m}}$$

1) $H_1: M\xi \neq M\eta$




$$P(z < -z_{kp}) = \frac{\alpha}{2} \qquad P(z > z_{kp}) = \frac{\alpha}{2}$$

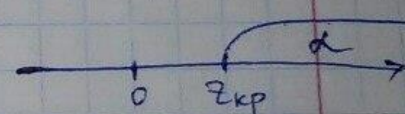
$$z_{спост} = \frac{\bar{\xi} - \bar{\eta}}{\sqrt{\frac{D\xi}{n} + \frac{D\eta}{m}}}$$

$$P\{0 < z < z_{kp}\} + P\{z > z_{kp}\} = \frac{1}{2}$$

$$\Phi(z_{kp}) + \frac{\alpha}{2} = \frac{1}{2} \Rightarrow \Phi(z_{kp}) = \frac{1-\alpha}{2}$$

$|z_{спост}| < z_{kp}$ — $H_0$ $\qquad$ $|z_{спост}| > z_{kp}$ — $\cancel{H_0}$

2) $H_1: M\xi > M\eta$

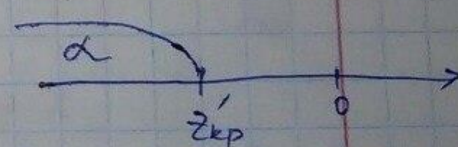



$$z_{спост} = \frac{\bar{\xi} - \bar{\eta}}{\sqrt{D\ldots}}$$

$$\Phi(z_{kp}) = \frac{1-2\alpha}{2}$$

$z_{спост} < z_{kp}$ — $H_0$ ; $\quad z_{спост} > z_{kp}$ — $\cancel{H_0}$

3) $H_1: M\xi < M\eta$




$$z_{спост} = \frac{\bar{\xi} - \bar{\eta}}{\sqrt{\ldots}}$$

$$z'_{kp} = -z_{kp}, \quad \Phi(z_{kp}) = \frac{1-2\alpha}{2}$$

$z_{спост} > -z_{kp}$ — $H_0$

$z_{спост} < -z_{kp}$ — $\cancel{H_0}$

## 24. Порівняння двох середніх нормальних генеральних сукупностей, дисперсії яких відомі (незалежні вибірки).

27 Порівняння двох середніх нормальних генер. сукупностей, дисперсії яких відомі (незалежні вибірки).

Нехай маємо дві генеральні сукупності $\xi$ та $\eta$

$\xi$: $\xi_1, \dots, \xi_n$ $\quad \bar\xi$

$\eta$: $\eta_1, \dots, \eta_n$ $\quad \bar\eta$ — вибіркові

За заданим рівнем значущості $\alpha$ перевірити $H_0$:

$$H_0: M\xi = M\eta$$

$M\bar\xi = M\xi$; $M\bar\eta = M\eta$

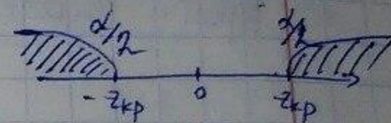
$$H_0: M\bar\xi = M\bar\eta$$

В якості критерію перевірки гіпотези викор. випадк. величину:

$$z = \frac{\bar\xi - \bar\eta}{\sigma(\bar\xi - \bar\eta)} = \frac{\bar\xi - \bar\eta}{\sqrt{\frac{D\xi}{n} + \frac{D\eta}{m}}} \sim N(0,1)$$

$$\sigma(\bar\xi - \bar\eta) = \sqrt{D(\bar\xi - \bar\eta)} = \sqrt{D\bar\xi + D\bar\eta} = \sqrt{\frac{D\xi}{n} + \frac{D\eta}{m}}$$

$$H_1: M\xi \neq M\eta$$




1. $P(z < -z_{kp}) = \frac{\alpha}{2}$ $\quad P(z > z_{kp}) = \frac{\alpha}{2}$

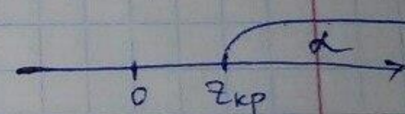
$z_{спост} = \frac{\bar\xi - \bar\eta}{\sqrt{\frac{D\xi}{n} + \frac{D\eta}{m}}}$

$$P\{0 < z < z_{kp}\} + P\{z > z_{kp}\} = \frac{1}{2}$$

$$\Phi(z_{kp}) + \frac{\alpha}{2} = \frac{1}{2} \Rightarrow \Phi(z_{kp}) = \frac{1-\alpha}{2}$$

$|z_{спост}| < z_{kp}$ — $H_0$ $\quad |z_{спост}| > z_{kp}$ — $\cancel{H_0}$

2. $H_1: M\xi > M\eta$

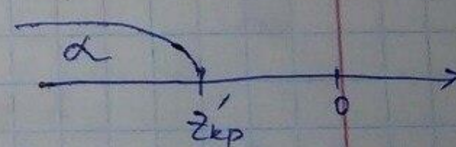



$z_{спост} = \frac{\bar\xi - \bar\eta}{\sqrt{\dots}}$

$$\Phi(z_{kp}) = \frac{1-2\alpha}{2}$$

$z_{спост} < z_{kp}$ — $H_0$ ; $\quad z_{спост} > z_{kp}$ — $\cancel{H_0}$

3. $H_1: M\xi < M\eta$




$z_{спост} = \frac{\bar\xi - \bar\eta}{\sqrt{\dots}}$

$z'_{kp} = -z_{kp}$, $\quad \Phi(z_{kp}) = \frac{1-2\alpha}{2}$

$z_{спост} > -z_{kp}$ — $H_0$

$z_{спост} < -z_{kp}$ — $\cancel{H_0}$

## 25. Порівняння двох середніх довільно розподілених генеральних сукупностей (великі незалежні вибірки).

28. Порівняння двох середніх довільно розподілених генеральних сукупностей (великі незалежні вибірки).

Нехай маємо дві довільно розподілені генеральні сукупності $\xi$ та $\eta$.

$\xi \quad \xi_1, \dots, \xi_n \quad n$ ($n, m$ – об'єми, $n \geq 30$, $m \geq 30$)

$\eta \quad \eta_1, \dots, \eta_m \quad m$

Тоді

$$Z'_{спост} = \frac{\bar{\xi} - \bar{\eta}}{\sqrt{\frac{D_B \eta}{m} + \frac{D_B \xi}{n}}}$$

Крит. значення $\rightarrow$ див. 27.

26. Порівняння двох середніх нормальних генеральних сукупностей, дисперсії яких невідомі й однакові (малі незалежні вибірки).

29. Порівняння двох середніх нормальних генеральних сукупностей, дисперсії яких невідомі та однакові (малі незалежні вибірки)

Нехай маємо дві генеральні сукупності, розподілені нормально і їх дисперсії невідомі. Вважаємо, що дисперсії однакові. Вибірки невеликі.

$$H_0 : M\xi = M\eta$$

В якості критерію перевірки гіпотези використовуємо випадкову величину із розподілом Стюдента

$$T = \frac{\bar{\xi} - \bar{\eta}}{\sqrt{(n-1)S_\xi^2 + (m-1)S_\eta^2}} \sqrt{\frac{nm(n+m-2)}{n+m}}$$

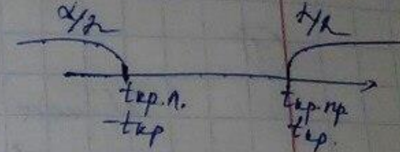
із $k = n+m-2$ ступенями вільності.

$$H_1 : M\xi \neq M\eta$$

(1) Будуємо двосторонню критичну область:

$$P\{T < t_{кр.лів.}\} = \frac{\alpha}{2}$$

$$P\{T > t_{кр.пр.}\} = \frac{\alpha}{2}$$




$$T_{спост} = \frac{\bar{\xi} - \bar{\eta}}{\sqrt{\ldots}} \cdot (\ldots)$$

Маючи рівень значущості $\alpha$ і $k = n+m-2$, знаход. $t_{кр}(\alpha, k)$

$|T_{спост}| < t_{кр}(\alpha, k)$ – не відхиляємо $H_0$

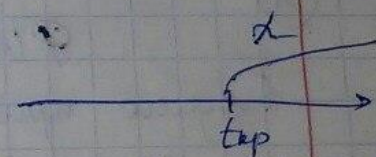
$|T_{спост}| > t_{кр}(\alpha, k)$ – $\cancel{H_0}$

(2)

$$H_1 : M\xi > M\eta$$

Будуємо критичну область

$$P\{T > t_{кр.пр.}\} = \alpha$$




Знаходимо $t_{кр.пр.}(\alpha, k)$ $\quad k = m+n-2$

$|T_{спост}| < t_{кр.пр.}(\alpha, k)$ – не відх. $H_0$

$|T_{спост}| > t_{кр.пр.}(\alpha, k)$ – $\cancel{H_0}$

(3)

$$H_1 : M\xi < M\eta$$

$$P\{T < t_{лів.кр}\} = \alpha \qquad t_{лів.кр.} = -t_{пр.кр.}$$

$T_{спост} > -t_{пр.кр.}$ – $H_0$

$T_{спост} < -t_{пр.кр.}$ – $\cancel{H_0}$

27. Порівняння вибіркової середньої з гіпотетичною генеральною середньою нормальної сукупності.

30. Порівняння вибіркової середньої з гіпотетичною генеральною середньою нормальної сукупності

Маємо генеральну сукупність $\xi$, яка розподілена нормально — $N(a_0, \sigma)$.

А. Дисперсія ген. сукупності відома: Маємо вибіркове середнє $\bar{x}$. З'ясуємо, чи істотна відмінність $\bar{x}$ від $a$:

$$H_0: a = a_0$$

$$M\bar{\xi} = a$$

$$H_0: M\bar{\xi} = a_0$$

В якості критерію перевірки гіпотези беремо випадкову величину, яка розподілена нормально:

$$U = \frac{\bar{\xi} - a_0}{\sigma(\bar{\xi})} = \frac{(\bar{\xi} - a_0)\sqrt{n}}{\sigma}$$

(1)

$$H_1: a \neq a_0$$

$$U_{спост} = \frac{(\bar{x} - a_0)\sqrt{n}}{\sigma}$$

$\Phi(u_{кр}) = \frac{1-\alpha}{2}$, $\alpha$ — рівень значущості критерію.

$|U_{спост}| < u_{кр}$ — $H_0$

$|U_{спост}| > u_{кр}$ — $\cancel{H_0}$

(2)

$$H_1: a > a_0$$

$$\Phi(u_{кр}) = \frac{1-2\alpha}{2}$$

Будуємо правостор. інтервал

$U_{спост} < u_{кр}$ — $H_0$

$U_{спост} > u_{кр}$ — $\cancel{H_0}$

(3)

$$H_1: a < a_0$$

Рахуємо критичну точку

$$\Phi(u_{кр}) = \frac{1-2\alpha}{2}$$

$$u'_{кр} = -u_{кр}$$

$U_{спост} > -U_{кр}$ — $H_0$
$U_{спост} < -U_{кр}$ — $\cancel{H_0}$

Б. Дисперсія ген. сукупності невідома

$H_0: a = a_0$

В якості критерію перевірки гіпотези беремо випадкову величину із розподілом Стьюдента:

$$T = \frac{(\bar{x} - a)\sqrt{n}}{S}$$ із $k = n-1$ ступ. вільності

1. $H_1: a \neq a_0$

$$T_{спост} = \frac{(\bar{x} - a)\sqrt{n}}{S}$$

Знаходимо критичну точку $t_{кр\ двост}(\alpha, k)$

$|T_{спост}| < t_{кр\ двост}$ — $H_0$
$|T_{спост}| > t_{кр\ двост}$ — $\cancel{H_0}$

2. $H_1: a > a_0$

$$T_{спост} = \frac{(\bar{x} - a)\sqrt{n}}{S}$$

$t_{кр\ пр}(\alpha, k)$

$T_{спост} < t_{кр\ пр}$ — $H_0$
$T_{спост} > t_{кр\ пр}$ — $\cancel{H_0}$

3. $H_1: a < a_0$

$$T_{спост} = \frac{(\bar{x} - a)\sqrt{n}}{S}$$

$t_{кр\ лів} = -t_{кр.\ прав.}$

$T_{спост} > -t_{кр.\ пр.}$ — $H_0$
$T_{спост} < -t_{кр.\ пр.}$ — $\cancel{H_0}$

## 28. Порівняння двох середніх нормальних генеральних сукупностей з невідомими дисперсіями (залежні вибірки)

31. Порівняння двох середніх нормальних генеральних сукупностей з невідомими дисперсіями (залежні вибірки)

Нехай маємо дві залежні генеральні сукупності, розподілені нормально:

$$\xi : \ x_1, x_2, \dots, x_n$$

$$\eta : \ y_1, y_2, \dots, y_n$$

$$H_0 : M\xi = M\eta$$

$$M\xi = M\eta \Leftrightarrow M\xi - M\eta = 0 \Rightarrow M(\xi - \eta) = 0 \Rightarrow M(\bar{D}) = 0$$

$$H_0 : M(\bar{D}) = 0$$

$$H_1 : M(\bar{D}) \neq 0.$$

В якості критерію перевірки гіпотези беремо випадкову величину із розподілом Стьюдента:

$$T = \frac{\bar{d}\sqrt{n}}{S_d} \quad \text{із } k = n - 1.$$

$$T_{спост} = \frac{\bar{d}\sqrt{n}}{S_d}$$

$$D : d_1, d_2, \dots, d_n \qquad d_i = x_i - y_i \qquad i = \overline{1,n}$$

$$\bar{D} = \frac{1}{n}\sum_{i=1}^{n} d_i = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i) = \bar{\xi} - \bar{\eta}$$

Знаходимо критичну точку $t_{кр}(\alpha, k)$:

$$|T_{спост}| < t_{двуст.кр} - H_0$$

$$|T_{спост}| > t_{двуст.кр} - \cancel{H_0}$$

# 29. Порівняння спостережуваної відносної частоти з гіпотетичною ймовірністю появи події.

32. Порівняння спостережуваної відносної частоти з гіпотетичною ймовірністю появи події

Нехай маємо достатньо велику к-ть $n$ незалежних випробувань; $p$ – ймовірність появи події в кожному з цих випробувань; $\frac{m}{n}$ – відносна частота появи події.

При задан. $\alpha$: $H_0: p = p_0$

В якості критерію перевірки гіпотези беремо випадкову величину

$$U = \frac{\left(\frac{m}{n} - p_0\right)\sqrt{n}}{\sqrt{p_0 q_0}}, \quad q_0 = 1 - p_0$$

1. $H_1: p \neq p_0$

$$U_{спост} = \frac{\left(\frac{m}{n} - p_0\right)\sqrt{n}}{\sqrt{p_0 q_0}}$$

Знаходимо критичну точку: $\Phi(u_{кр}) = \frac{1-\alpha}{2}$

$|U_{спост}| < u_{кр}$ – $H_0$

$|U_{спост}| > u_{кр}$ – $\cancel{H_0}$

2. $H_1: p > p_0$

$U_{спост}$ = –//–

Знаходимо критичну точку: $\Phi(u_{кр}) = \frac{1-2\alpha}{2}$

$U_{спост} < u_{кр}$ – $H_0$

$U_{спост} > u_{кр}$ – $\cancel{H_0}$

3. $H_1: p < p_0$

$U_{спост}$ = –//–

Знаходимо крит. точку: $\Phi(u_{кр}) = \frac{1-2\alpha}{2}$

$u'_{кр} = -u_{кр}$

$U_{спост} > -u_{кр}$ – $H_0$

$U_{спост} < -u_{кр}$ – $\cancel{H_0}$

## 30. Порівняння двох ймовірностей біноміальних розподілів.

### 33. Порівняння двох ймовірностей біноміальних розподілів

Нехай маємо дві генеральні сукупності:

$p_1$ – ймовірність появи події A для I сукуп.

$p_2$ – ймовірність появи події A для II сукуп.

$n_1$ – об'єм вибірки з I сукуп.

$n_2$ – об'єм вибірки з II сукуп.

$W_1(A)=\frac{m_1}{n_1}$ $W_2(A)=\frac{m_2}{n_2}$ – в-на частота появи події A

$$H_0: p_1=p_2=p$$

В якості критерію перевірки гіпотези візьмемо випадкову величину, розподілену наближено нормально

$$U=\frac{m_1/n_1-m_2/n_2}{\sqrt{p(1-p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$M(m_1)=n_1p_1$ $D(m_1)=n_1p_1(1-p_1)=n_1p(1-p)$

$M(m_2)=n_2p_2$ $D(m_2)=n_2p_2(1-p_2)=n_2p(1-p)$

$M\left(\frac{m_1}{n_1}-\frac{m_2}{n_2}\right)=0$ $D\left(\frac{m_1}{n_1}-\frac{m_2}{n_2}\right)=\frac{1}{n_1^2}D(m_1)+\frac{1}{n_2^2}D(m_2)=p(1-p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$

$$U_{спост}=\frac{m_1/n_1-m_2/n_2}{\sqrt{\frac{m_1+m_2}{n_1+n_2}\left(1-\frac{m_1+m_2}{n_1+n_2}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

1) $H_1: p_1\neq p_2$

$\Phi(u_{кр})=\frac{1-\alpha}{2}$

$|u_{спост}|<u_{кр}$ – $H_0$

$|u_{спост}|>u_{кр}$ – $\cancel{H_0}$

2) $H_1: p_1>p_2$

$\Phi(u_{кр})=\frac{1-2\alpha}{2}$

$u_{спост}<u_{кр}$ – $H_0$

$u_{спост}>u_{кр}$ – $\cancel{H_0}$

3) $H_1: p_1<p_2$

$\Phi(u_{кр})=\frac{1-2\alpha}{2}$

$u'_{кр}=-u_{кр}$

$u_{спост}>-u_{кр}$ – $H_0$

$u_{спост}<-u_{кр}$ – $\cancel{H_0}$

# 31. Критерій згоди Колмогорова.

34. Критерій згоди Колмогорова

Нехай у нас є вибірка $\xi = (\xi_1, \dots, \xi_n)$, $F_\xi(x)$ - ф-ція розподілу.

$H_0: F_\xi(x) = \mathcal{F}(x)$, де $\mathcal{F}(x)$ - деяка задана ф-ція розподілу

$H_1: \overline{H_0}$.

Такі критерії називаються критеріями згоди

Критерій згоди Колмогорова використовується, коли розподіл неперервний

$$* \quad D_n = D_n(\xi) = \sup_{-\infty < x < +\infty} |\hat{F}_n(x) - F(x)|,$$ $\hat{F}_n(x)$ - емпір. ф-ція розподілу

Якщо $H_0$ - істинна, то $\hat{F}(x)$ є консист. оцінкою.

Критична межа знаходиться за теор. Колмогорова

<u>Теорема Колмогорова</u>

Якщо $F(x)$ - неперервна, то $\forall t > 0$

$$\lim_{n \to \infty} P\{\sqrt{n} D_n \le t\} = K(t),$$ n визначається *

$K(t)$ - розподіл Колмогорова, $K(t) = \sum_{j=-\infty}^{+\infty} (-1)^j e^{-2j^2t^2}$

При $n \geq 20$ можемо користуватись даним критерієм

$$t_\alpha = \frac{\lambda_\alpha}{\sqrt{n}} \quad K(\lambda_\alpha) = 1 - \alpha$$

Будуватимемо критичну область

$$T_\alpha = \{t \geq t_\alpha\}$$

$$P\{D_n \in T_\alpha | H_0\} = P\{\frac{1}{\sqrt{n}} D_n \geq \lambda_\alpha | H_0\} \simeq 1 - K_\alpha = \alpha$$

$\sqrt{n} D_n \geq \lambda_\alpha$ - $H_0$ відкидаємо.

Критерій є асимптотичним.

$$D_n = \max\{D_n^+, D_n^-\}$$

$$D_n^+ = \max_{1 \le k \le n} \left(\frac{k}{n} - \mathcal{F}(\xi_{(k)})\right)$$

$$D_n^- = \max_{1 \le k \le n} \left(\mathcal{F}(\xi_{(k)}) - \frac{k-1}{n}\right),$$

де $\xi_{(k)}$ - варіанта.

## 32. Критерій згоди Xi^2 Пірсона

35. Критерій згоди $\chi^2$ Пірсона

Нехай у нас є вибірка $\xi = (\xi_1, \dots, \xi_n)$, $F_\xi(x)$ – функція розподілу.

$H_0: F_\xi(x) = F(x)$, де $F(x)$ – задана ф-ція розп.

$H_1: \overline{H_0}$

Такі критерії називаються критеріями згоди.

Критерій згоди $\chi^2$ Пірсона використовується для б-яких ф-цій розподілу, якщо $F(x)$ – неперервна.

Нехай маємо вибірку $\xi = (\xi_1, \dots, \xi_n)$ об'єму $n$

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | $\xi$ | 1 | 2 | 3 | ... | $N$ | |
| ймов. – | $p$ | $p_1$ | $p_2$ | $p_3$ | ... | $p_N$ | $\sum_{i=1}^{N} p_i = 1$ |
| частот. – | $\omega$ | $\omega_1$ | $\omega_2$ | $\omega_3$ | ... | $\omega_N$ | $\sum_{i=1}^{N} \omega_i = n$ |

$\overline{\omega} = (\omega_1, \dots, \omega_N)$

Вектор $\overline{\omega}$ має поліноміальний розподіл, який має вигляд

$$p(\overline{\omega} = \overline{x}) = \frac{n!}{x_1!\, x_2! \dots x_N!}\, p_1^{x_1} \dots p_N^{x_N}; \quad \sum_{i=1}^{N} x_i = n$$

$$\overline{p}^0 = (p_1^{(0)}, \dots, p_N^{(0)}) \qquad \sum_{i=1}^{N} p_i^{(0)} = 1.$$

Статистика, на якій базується критерій Пірсона

$$\tilde{\chi}_n^2 = \tilde{\chi}_n^2(\overline{\omega}) = \sum_{i=1}^{N} \frac{(\omega_j - np_j^{(0)})^2}{np_j^{(0)}} = \sum_{i=1}^{N} \frac{\omega_i^2}{np_j^{(0)}} - n$$

Критична область:

$$T_\alpha = \{\tilde{\chi}_n^2 > t_\alpha\}, \quad t_\alpha \text{ – критична межа}$$

$$P\{\tilde{\chi}_n^2 > t_\alpha \mid H_0\} = \alpha$$

Теорема (одна з граничних)

$$F_{\tilde{\chi}_n \mid H_0} \longrightarrow \chi^2(N-1), \text{ де } (N-1) \text{ – ступені вільності}$$

Цей розподіл можемо використовув. при $n \geq 50$, $n_j \geq 5$ за таких обмежень:

$$P\{\tilde{\chi}_n^2 > t_\alpha \mid H_0\} \approx 1 - \chi_{N-1}(t_\alpha)$$

$\chi_{N-1}(t_\alpha)$ – ф-ція розподілу з $(N-1)$ ступ. вільності

$$t_\alpha = \chi^2(1-\alpha, N)$$

$H_0$ відкидаємо $\Leftrightarrow \{\tilde{\chi}_n^2 > \chi^2(1-\alpha, N-1)\}$.

# 33. Критерій однорідності Смірнова.

36. Критерій однорідності Смірнова

Нехай ми маємо дві вибірки $\xi = (\xi_1, \dots, \xi_n)$ і $\eta = (\eta_1, \dots, \eta_m)$ із однієї ген. сукупності із ф-ціями розподілу $F_\xi(x)$ і $F_\eta(x)$ відповідно.

$$H_0 : F_\xi(x) = F_\eta(x)$$

$$H_1 : F_\xi(x) \neq F_\eta(x)$$

Такі критерії назив. критеріями однорідності

$$D_{n,m} = \sup_{-\infty < x < +\infty} |\hat{F}_{1n}(x) - \hat{F}_{2m}(x)|$$

$\hat{F}_{1n}(x), \hat{F}_{2m}(x)$ – емпіричні функції розподілу.

Критична область:

$$T_\alpha = \{D_{n,m} > t_\alpha(n,m)\}$$

Теорема Смірнова

Якщо $F(x)$ – неперервна, то

$$\lim_{n,m \to \infty} P\left\{\sqrt{\frac{nm}{n+m}} D_{n,m} < t_\alpha \Big| H_0\right\} = K(t),$$ де $K(t)$ – ф-ція Колмогорова

$$K(t) = \sum_{j=-\infty}^{+\infty} (-1)^j e^{-2j^2t^2}$$

Дов.

$$t_\alpha(n,m) = \sqrt{\frac{1}{n} + \frac{1}{m}}\, \lambda_\alpha$$

$K(\lambda_\alpha) = 1 - \alpha$, $\alpha$ – квантиль Колмогорова

$$P\left\{D_{n,m} > \sqrt{\frac{1}{n} + \frac{1}{m}}\, \lambda_\alpha \Big| H_0\right\} = P\left\{\sqrt{\frac{nm}{n+m}} D_{n,m} > \lambda_\alpha \Big| H_0\right\} \approx$$

$$\approx 1 - K(\lambda_\alpha) = \alpha.$$

$H_0$ відкидаємо $\Leftrightarrow$ $D_{n,m} > \lambda_\alpha \sqrt{\frac{1}{n} + \frac{1}{m}}$

$$D_{n,m} = \max(D^+_{n,m}, D^-_{n,m})$$

$$D^+_{n,m} = \max_{1 \le j \le m}\left(\frac{j}{m} - \hat{F}_{1n}(\eta_{(j)})\right) = \max_{1 \le i \le n}\left(\hat{F}_{2m}(\xi_{(i)}) - \frac{i-1}{n}\right)$$

$$D^-_{n,m} = \max_{1 \le j \le m}\left(\hat{F}_{1n}(\eta_{(j)}) - \frac{j-1}{m}\right) = \max_{1 \le i \le n}\left(\frac{i}{n} - \hat{F}_{2m}(\xi_{(i)})\right)$$

де $\xi_{(i)}$ – варіанта.

# 34. Критерій однорідності Xi^2

## 37. Критерій однорідності $\chi^2$.

Нехай ми маємо дві вибірки $\xi(\xi_1,\dots,\xi_n)$ і $\eta=(\eta_1,\dots,\eta_m)$ із однієї ген. сукупності із ф-цією розподілу $F_\xi(x)$ і $F_\eta(x)$

$$H_0: F_\xi(x)=F_\eta(x)$$

$$H_1: F_\xi(x)\ne F_\eta(x)$$

Такі критерії називаються критерії однорідності

Маємо $k$ серій, кожна з серій склад. з $n_1 \dots n_k$.

$\bar\nu_i=(\nu_{i1},\dots,\nu_{iN})$ – частота $i$-серії

$\bar p_i=(p_{i1},\dots,p_{iN})$

Гіпотеза однорідності: $H_0: \bar p_1=\bar p_2=\dots=\bar p_k=\bar p(p_1,\dots,p_N)$

$\sum p_i=1$

$$\nu_i \sim M(n_i;\bar p)$$

$$M(\nu_{ij}|H_0)=n_i p_j \qquad j=\overline{1,N}$$

Обираємо статистику:

$$\chi^2_{n_1,\dots,n_k}(\bar p)=\sum_{i=1}^{k}\sum_{j=1}^{N}\frac{(\nu_{ij}-n_ip_j)^2}{n_ip_j}$$

$$\hat{\bar p}=\arg\max_{\bar p}\prod_{i,j}p_j^{\nu_{ij}}=\arg\max_{\bar p}\prod_{j}p_j^{\nu_{\cdot j}} \qquad \nu_{\cdot j}=\sum_{i=1}^{k}\nu_{ij}$$

$$\hat p_j=\frac{\nu_{\cdot j}}{n}, \quad j\in\overline{1,N}$$

$$\hat\chi_{n_1,\dots,n_k}=n\sum_{i=1}^{k}\sum_{j=1}^{N}\frac{1}{n_i\nu_{\cdot j}}\cdot\left(\nu_{ij}-\frac{n_i\nu_{\cdot j}}{n}\right)^2$$

$P\{\hat\chi_{n_1,\dots,n_k}>t_\alpha\}=\alpha$, $t_\alpha$ – межа за рівнем значущості $\alpha$

Для того, щоб підрахувати $t_\alpha$, використ. результат

$$P\{\hat\chi_{n_1,\dots,n_k}|H_0\}\to\chi^2((k-1)(N-1))$$

$$t_\alpha=\chi^2(1-\alpha;(k-1)(N-1))$$

$H_0$ відкидаємо $\iff \chi^2_{n_1\dots n_k}>\chi^2(1-\alpha;(k-1)(N-1))$

# 35. Рангові критерії однорідності. Критерій Вілкоксона.

38. Рангові критерії однорідності. Критерій Вілкоксона

Нехай в нас є дві незалежні вибірки $\xi=(\xi_1,\dots,\xi_n)$ і $\eta=(\eta_1,\dots,\eta_m)$ із ф-ціями розподілу $F_\xi(x)$ і $F_\eta(x)$ відповідно.

$H_0: F_\xi(x)=F_\eta(x)$

Отримаємо вибірку $(\xi_1,\dots,\xi_n,\eta_1,\dots,\eta_m)$, який відповідає $x^{(1)}, x^{(2)},\dots,x^{(n+m)}$, де $r=n+m$ – ранг.

$R(\xi_1)=r_1,\dots,R(\xi_n)=r_n$, $n\le m$.

$T=\sum_{i=1}^{n} r_i$ – критерій Вілкоксона

A) $n\le 25$, $m\le 25$

1. $x^{(1)},\dots,x^{(n+m)}$

$W_{спост}=\sum_{i=1}^{n} r_i$

2. Шукаємо нижню крит. точку

$\omega_{н.кр}(\frac{\alpha}{2}, n, m)$

3. Шукаємо верхню критичну точку

$\omega_{в.кр}=(n+m+1)n-\omega_{н.кр}$

а) $H_1: F_\xi(x)\ne F_\eta(x)$

$W_{спост}<\omega_{н.кр}$ або $W_{спост}>\omega_{в.кр}$ – $H_0$ відкидаємо

$\omega_{н.кр}<W_{спост}<\omega_{в.кр}$ – $H_0$ не відк.

б) $H_1: F_\xi(x)>F_\eta(x)$

$\omega_{н.кр}(\alpha, n, m)$

$W_{спост}<\omega_{н.кр}$ – $\cancel{H_0}$  $W_{спост}>\omega_{н.кр}$ – $H_0$

в) $H_1: F_\xi(x)<F_\eta(x)$

$\omega_{в.кр}=(n+m+1)n-\omega_{н.кр}(\alpha, n, m)$

$W_{спост}>\omega_{в.кр}$ – $\cancel{H_0}$  $W_{спост}<\omega_{в.кр}$ – $H_0$

Б) $n$ або $m>25$

а) $H_1: F_\xi(x)\ne F_\eta(x)$

$$\omega_{н.кр}\left(\frac{\alpha}{2}, n, m\right)=\left[\frac{(n+m+1)n-1}{2}-z_{кр}\sqrt{\frac{nm(n+m+1)}{12}}\right] \quad (*)$$

$\Phi(z_{кр})=\frac{1-\alpha}{2}$  [ ] – ціла частина

б) і в) зберігаються по формулі (*)  $\Phi(z_{кр})=\frac{1-2\alpha}{2}$

# 36. Однофакторний дисперсійний аналіз.

42. Однофакторний дисперсійний аналіз

$$x_{ij} = \mu + F_i + \varepsilon_{ij}$$

$x_{ij}$ – значення змінної, яку досліджуємо
$i$ – рівень фактору, $i=\overline{1,n}$
$F_i$ – можлив. значення рівнів фактору
$j$ – послідовність, $j=\overline{1,m}$
$\varepsilon_{ij}$ – випадкова компонента (збурення)
$\mu$ – константа

Основні передумови дисперсійного аналізу
1) $M\varepsilon_{ij}=0$
2) Збурення $\varepsilon_{ij}$ – взаємно незалежні
3) $D\varepsilon_{ij}=\sigma^2$
4) Збурення $\varepsilon_{ij}$ має нормальний закон розподілення $N(0;\sigma^2)$.

Нехай маємо $m$ партій виробів. З кожної партії обираємо $n_1=n_2=\ldots=n_m=n$ виробів.

$$\begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ x_{m1} & x_{m2} & \ldots & x_{mn} \end{pmatrix} = (x_{ij}), \quad i=\overline{1,m}, \; j=\overline{1,n}$$

$H_0$: $a_1=a_2=\ldots=a_m$

$$\overline{x}_{i\cdot} = \frac{\sum_{j=1}^{n} x_{ij}}{n}, \quad i=\overline{1,m}$$

$$\overline{x}_{\cdot\cdot} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} = \frac{1}{m}\sum_{i=1}^{m}\overline{x}_{i\cdot}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\overline{x}_{\cdot\cdot})^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\left((\overline{x}_{i\cdot}-\overline{x}_{\cdot\cdot})+(x_{ij}-\overline{x}_{i\cdot})\right)^2 =$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n}(\overline{x}_{ij}-\overline{x}_{\cdot\cdot})^2 + \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\overline{x}_{i\cdot})^2 + 2\sum_{i=1}^{m}\sum_{j=1}^{n}(\overline{x}_{i\cdot}-\overline{x}_{\cdot\cdot})(x_{ij}-\overline{x}_{i\cdot});$$

$= Q_1 + Q_2 + Q_3$

$$Q_3 = 2\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{i\cdot})(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot}) = 2\sum_{i=1}^{m}(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})\sum_{j=1}^{n}(x_{ij} - \bar{x}_{i\cdot}) = 0$$

$$Q_1 = n\sum_{i=1}^{m}(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2$$

$$Q = Q_1 + Q_2$$

$$Q_2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{i\cdot})$$

$Q_1$ – міжгрупова (факторна) сума квадратів відхилення

$Q_2$ – внутрішньогрупова (остаточна) сума квадратів відхилення

$$D_{заг} = D_{вн.гр.} + D_{міжгр}$$

$$S_1^2 = \frac{Q_1}{k_1} \qquad k_1 = m-1$$

$$S_2^2 = \frac{Q_2}{k_2} \qquad k_2 = m(n-1)$$

$$MS_1^2 = \frac{n}{m-1} M\left(\sum_{i=1}^{m}(\mu + F_i + \varepsilon_{ij} - \mu - F_\cdot - \varepsilon_{\cdot\cdot})^2\right) = \frac{n}{m-1} M\left(\sum_{i=1}^{m}(F_i - F_\cdot) + (\varepsilon_{i\cdot} - \varepsilon_{\cdot\cdot})\right)^2 = (\text{открыть скобки}) = \frac{n}{m-1} M\left\{\sum_{i=1}^{m}(F_i - F_\cdot)^2\right\} + \sigma^2$$

$$2(F_i - F_\cdot)(\varepsilon_{i\cdot} - \varepsilon_{\cdot\cdot}) = 0 \qquad \frac{n}{m-1} M(\varepsilon_{i\cdot} - \varepsilon_{\cdot\cdot})^2 = nM\sum\frac{(\varepsilon_i - \varepsilon_{\cdot\cdot})^2}{m-1} = n\cdot\sigma^2_{\bar{\varepsilon}_{i\cdot}} = n\frac{\sigma^2_{\varepsilon_{ij}}}{n} = \sigma^2$$

$$MS_2^2 = \frac{1}{m(n-1)} M\sum_{i=1}^{m}\sum_{j=1}^{n}(\mu + F_i + \varepsilon_{ij} - \mu - F_i - \varepsilon_{i\cdot})^2 = \frac{1}{m} M\left(\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(\varepsilon_{ij} - \varepsilon_{i\cdot})^2}{n-1}\right) =$$

$$= \frac{1}{m}\sum_{i=1}^{m}\sigma^2_{\varepsilon_{ij}} = \frac{1}{m}\cdot m\sigma^2 = \sigma^2$$

I. Модель з фіксованими р-нями фактора:

$$MS_1^2 = \frac{n}{m-1}\sum_{i=1}^{n}(F_i - F_\cdot)^2 + \sigma^2 \qquad H_0: F_i = F_\cdot \quad i = \overline{1,m}$$

II Випадкова модель

$$\sigma_F^2 = M\left(\sum_{i=1}^{m}(F_i - F_\cdot)^2/(m-1)\right)$$

$$MS_1^2 = n\sigma_F^2 + \sigma^2 \qquad H_0: \sigma_F^2 = 0. \qquad S_1^2, S_2^2 \text{ – незміщені оцінки}$$

$$F = \frac{S_1^2}{S_2^2} > F_{\alpha,k_1,k_2} \text{ – } \cancel{H_0}$$

$$F = \frac{S_1^2}{S_2^2} \le F_{\alpha,k_1,k_2} \text{ – } H_0.$$

## 37. Функціональна, статистична та кореляційна залежності.

### 12.1. Функциональная, статистическая и корреляционная зависимости

В естественных науках часто речь идет о ***функциональной*** зависимости (связи), когда каждому значению одной переменной соответствует вполне *определенное значение другой*. Функциональная зависимость может иметь место как между детерминированными (неслучайными) переменными (например, зависимость скорости падения в вакууме от времени и т.п.), так и между случайными величинами (например, зависимость стоимости проданных изделий от их числа и т.п.).

В экономике в большинстве случаев между переменными величинами существуют зависимости, когда каждому значению одной переменной соответствует не какое-то определенное, а множество возможных значений другой переменной. Иначе говоря, каждому значению одной переменной соответствует *определенное (условное) распределение другой переменной*. Такая зависимость (связь) получила название ***статистической*** (или ***стохастической***, ***вероятностной***). (О ней уже шла речь в § 5.5.)

Возникновение понятия статистической связи обусловливается тем, что зависимая переменная подвержена влиянию ряда неконтролируемых или неучтенных факторов, а также тем, что измерение значений переменных неизбежно сопровождается некоторыми случайными ошибками. Примером статистической связи является зависимость урожайности от количества внесенных удобрений, производительности труда на предприятии от его энерговооруженности и т.п.

В силу неоднозначности статистической зависимости между $Y$ и $X$ для исследователя, в частности, представляет интерес усредненная по $x$ схема зависимости, т.е. закономерность в изменении среднего значения — условного математического ожидания[1] $M_x(Y)$ (математического ожидания случайной переменной $Y$, найденного при условии, что переменная $X$ приняла значение $x$) в зависимости от $x$.



Определение. *Статистическая зависимость между двумя переменными, при которой каждому значению одной переменной соответствует определенное среднее значение, т.е. условное математическое ожидание другой, называется* ***корреляционной***. Иначе, ***корреляционной зависимостью*** *между двумя переменными величинами называется функциональная зависимость между значениями одной из них и условным математическим ожиданием другой.*

Корреляционная зависимость может быть представлена в виде:

$$M_x(Y) = \varphi(x) \qquad (12.1)$$

или

$$M_y(X) = \psi(y). \qquad (12.2)$$

Предполагается, что $\varphi(x) \neq \text{const}$ и $\psi(y) \neq \text{const}$, т.е. если при изменении $x$ или $y$ условные математические ожидания $M_x(Y)$ и $M_y(X)$ не изменяются, то говорят, что корреляционная зависимость между переменными $X$ и $Y$ отсутствует.

Сравнивая различные виды зависимости между $X$ и $Y$, можно сказать, что с изменением значений переменной $X$ при ф у н к ц и о н а л ь н о й зависимости однозначно изменяется определенное *значение* переменной $Y$, при к о р р е л я ц и о н н о й — определенное *среднее значение* (условное математическое ожидание) $Y$, а при с т а т и с т и ч е с к о й — определенное (условное) *распределение* переменной $Y$ (рис. 12.1).

Таким образом, из рассмотренных зависимостей *наиболее общей выступает статистическая зависимость*[2]. Каждая корреляционная зависимость является статистической, но не каждая статистическая зависимость является корреляционной. Функциональная зависимость представляет частный случай корреляционной (об этом речь еще пойдет ниже, в § 12.3).

---

[1] Для условного математического ожидания в литературе используется также обозначение $M(Y \mid X = x)$.

[2] Хотя статистическая зависимость и является наиболее общей из рассмотренных, она не отражает любую возможную зависимость между переменными в условиях неопределенности. Например, можно предполагать, что существует некоторая зависимость между числом (продолжительностью) военных конфликтов и числом изобретений за определенный период времени. Эта зависимость хотя и сводится к зависимости между событиями с неопределенным исходом (могут произойти или не произойти), но не является статистической, ибо каждому значению одной переменной нельзя поставить в соответствие распределение другой, так как к таким единичным и неповторяемым в одинаковых условиях событиям, какими являются соответственно военные конфликты и изобретения, неприменимо само понятие вероятности (см. § 1.3).




Рис. 12.1

Уравнения (12.1) и (12.2) называются *модельными уравнениями*

## 38. Побудова рівнянь лінійної парної регресії методом найменших квадратів.

Уравнения (12.1) и (12.2) называются *модельными уравнениями регрессии* (или просто *уравнениями регрессии*) соответственно $Y$ по $X$ и $X$ по $Y$[1], функции $\varphi(x)$ и $\psi(y)$ — *модельными функциями регрессии* (или *функциями регрессии*), а их графики — *модельными линиями регрессии* (или *линиями регрессии*).

Для отыскания модельных уравнений регрессии, вообще говоря, необходимо знать *закон распределения двумерной случайной величины* $(X, Y)$. На практике исследователь, как правило, располагает лишь *выборкой* пар значений $(x_i, y_i)$ ограниченного объема. В этом случае речь может идти об оценке (приближенном выражении) по выборке функции регрессии. Такой наилучшей (в смысле метода наименьших квадратов) оценкой является *выборочная линия (кривая) регрессии* $Y$ по $X$

$$\bar{y}_x = \hat{\varphi}(x, b_0, b_1, \dots, b_p), \qquad (12.3)$$

где $\bar{y}_x$ — *условная (групповая) средняя* переменной $Y$ при фиксированном значении переменной $X = x$; $b_0, b_1, \dots, b_p$ — параметры кривой.

Аналогично определяется *выборочная линия (кривая) регрессии* $X$ по $Y$:

$$\bar{x}_y = \hat{\psi}(y, c_0, c_1, \dots, c_p), \qquad (12.4)$$

где $\bar{x}_y$ — *условная (групповая) средняя* переменной $X$ при фиксированном значении переменной $Y = y$; $c_0, c_1, \dots, c_p$ — параметры кривой.

Уравнения (12.3), (12.4) называют также *выборочными уравнениями регрессии* соответственно $Y$ по $X$ и $X$ по $Y$[2].

---

[1] Или $Y$ на $X$ и $X$ на $Y$.

[2] В дальнейшем для краткости там, где это очевидно по смыслу, мы часто и выборочные уравнения (линии) регрессии будем называть просто уравнениями (линиями) регрессии.



При правильно определенных аппроксимирующих функциях $\hat{\varphi}(x, b_0, b_1, \dots, b_p)$ и $\hat{\psi}(y, c_0, c_1, \dots, c_p)$ с увеличением объема выборки ($n \to \infty$) они будут сходиться по вероятности соответственно к функциям регрессии $\varphi(x)$ и $\psi(y)$.

Статистические связи между переменными можно изучать методами корреляционного и регрессионного анализа. *Основной задачей регрессионного анализа является установление формы и изучение зависимости между переменными. Основной задачей корреляционного анализа — выявление связи между случайными переменными и оценка ее тесноты.*

Вначале (§ 12.2, 12.3) познакомимся с основными понятиями корреляционного и регрессионного анализа, а затем (§ 12.4−12.7, 13.1−13.8) перейдем к более детальному изучению этих методов.

39. Оцінка тісноти кореляційної залежності. Коефіцієнт кореляції та його властивості.

44. Оцінка тісноти кореляційної залежності коеф. кореляції та його властивості

$$y_x - \bar{y} = b_{yx}(x - \bar{x})$$

$$\frac{y_x - \bar{y}}{S_y} = b_{yx}\left(\frac{S_x}{S_y}\right)\frac{x - \bar{x}}{S_x}$$

$$r = b_{yx}\cdot\frac{S_x}{S_y}$$

$r$ є показником тісноти кореляційного зв'язку і називається вибірковим коефіцієнтом кореляції

$$r = \frac{\overline{xy} - \bar{x}\bar{y}}{S_x S_y}$$

$$r = b_{xy}\frac{S_y}{S_x}$$

$$r^2 = b_{yx}\cdot b_{xy} \Rightarrow r = \pm\sqrt{b_{yx}\cdot b_{xy}}$$

$$r = \frac{\sum_{i=1}^{l}\sum_{j=1}^{m}(x_i - \bar{x})(y_j - \bar{x})n_{ij}}{nS_xS_y}$$

Властивості коеф. кореляції:

1. $-1 \le r \le 1$.

1. Якщо всі значення змінних змінити на одне й те саме число, або в одну й ту саму к-ть разів, то величина $r$ не зміниться.

3. При $r = \pm 1$ кореляційний зв'язок представляє лінійну функціональну залежність

4. При $r = 0$ лінійний кореляційний зв'язок відсутній. При цьому групові середні змінних $\bar{y}_x$ співпадають з їх загальними середніми, а лінії регресії $X$ та $Y$ паралельні осям координат.

## 40. Основні положення кореляційного аналізу. Двомірна модель.

### 45. Основні положення кореляційного аналізу. Двомірна модель

**Кореляційний аналіз** - метод, що застосовується коли дані спостережень або експерименту можна вважати випадковими і обраними із сукупності, розподіленої по багатомірному нормальному закону.

**Основна задача кореляційного аналізу** виявлення зв'язку між випадковими змінними шляхом точкової і інтервальної оцінок різних коефіцієнтів кореляції.

Розглянемо двомірну модель кореляційного аналізу: Щільність нормального розподілу двох змінних X та Y має вигляд:

$$\varphi_N(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-Z(x,y)},$$

$$Z(x,y) = -\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-a_x}{\sigma_x}\right)^2 - 2\rho\frac{x-a_x}{\sigma_x}\cdot\frac{y-a_y}{\sigma_y} + \left(\frac{y-a_y}{\sigma_y}\right)^2\right]$$

$a_x, a_y$ – мат. сподівання змінних X та Y

$\sigma_x^2, \sigma_y^2$ – дисперсії змінних X та Y

$\rho$ – коеф. кореляції

$$\rho = \frac{K_{xy}}{\sigma_x\sigma_y} = \frac{M(X-a_x)(Y-a_y)}{\sigma_x\sigma_y} = \frac{M(XY) - a_x a_y}{\sigma_x\sigma_y}$$

$K_{xy}$ – коваріація

$$M_x(Y) = a_y + \rho\frac{\sigma_y}{\sigma_x}(x-a_x)$$

$$M_y(X) = a_x + \rho\frac{\sigma_x}{\sigma_y}(y-a_y)$$

$$\sigma_x^2(Y) = \sigma_y^2(1-\rho^2)$$

$$\sigma_y^2(X) = \sigma_x^2(1-\rho^2)$$

Для оцінки генерального коеф. кореляції $\rho$ необхідно замінити параметри наступним чином:

$a_x \to \bar{x}$ $\quad \sigma_x^2 \to S_x^2$ $\quad K_{xy} \to \mu$

$a_y \to \bar{y}$ $\quad \sigma_y^2 \to S_y^2$

## 41. Перевірка значущості коефіцієнта кореляції.

46. Перевірка значущості коеф. кореляції

$H_0: \hat{r} \neq 0$

Нехай статистика має розподіл Стьюдента

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

$k = n-2$ - степені ~~свободи~~ вільності

Якщо $|t| = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} > t_{1-\alpha,k}$, то $\cancel{H_0}$

де $t_{1-\alpha,k}$ - табличне значення критерію Стьюдента, визначене на рівні значущості $\alpha$.

## 42. Кореляційне відношення та індекс кореляції.

### 47 Кореляційне відношення та індекс кореляції

$$S_y^2 = S'^2_{iy} + \delta^2_{iy}$$

$S_y^2 = \frac{1}{n}\sum_{j=1}^{m}(y_j - \bar{y})^2 n_i$ – спільна дисперсія змінної

$S'^2_{iy} = \frac{1}{n}\sum_{i=1}^{l} S^2_{iy} n_i$ – середня групова дисперсія $S^2_{iy}$

$S^2_{iy} = \frac{1}{n}\sum_{j=1}^{m}(y_j - \bar{y}_i)^2$

$\delta^2_{iy} = \frac{1}{n}\sum_{i=1}^{l}(\bar{y}_i - \bar{y})^2 n_i$ – міжгрупова дисперсія

$\eta_{yx} = \sqrt{\frac{\delta^2_{iy}}{S^2_y}}$, де $\eta^2_{yx}$ – емпіричний коеф детермінації

Основні властивості кореляційних відношень

1) $0 \le \eta_{yx} \le 1$

2) при $\eta_{yx} = 0$ кореляц. зв'язок відсут.

(3) при $\eta_{yx} = 1$ між змінними $\exists$ функц. залежність

4) $\eta_{xy} \ne \eta_{yx}$

Емпіричне кореляційне відношення є показником розсіювання точок корел. поля відносно емпіричної лінії регресії

$R_{yx}$ – індекс кореляції (теор. кореляційне відношення)

$$R_{yx} = \sqrt{\frac{\delta^2_y}{S^2_y}} = \sqrt{1 - \frac{S'^2_y}{S^2_y}}$$

$0 \le |r| \le R \le \eta \le 1.$

$(\bar{y}_x - \bar{y}) = b_{yx}(x - \bar{x})$

У випадку лінійної моделі індекс кореляції $R_{yx}$ дорівнює коеф. кореляції $r$: $R_{yx} = |r|$ (або $R_{xy} = |r|$).

Статистика $F$ має розподіл Фішера:

$$F = \frac{\eta^2 (n-m)}{(1-\eta^2)(m-1)} \qquad k_1 = m-1, \quad k_2 = n-m.$$

Якщо $F > F_{\alpha, k_1, k_2}$, то $\eta$ суттєво відрізн. від 0.

Якщо $F = \frac{R^2(n-[illegible])}{1-R^2} > F_{\alpha, k_1, k_2}$, то $R$ – істотний,

де $k_1 = 1$, $k_2 = n-2$

## 43. Багатомірний кореляційний аналіз. Множинний і частковий коефіцієнт кореляції.

48. Багатомірний кореляційний аналіз. Множинний і частковий коеф. кореляції

Нехай маємо сукупність випадкових змінних $X_1, X_2, \dots, X_i, \dots, X_j, \dots, X_p$, що має сумісний норм. розподіл.

Кореляційна матриця

$$Q_p = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{21} & 1 & \dots & \rho_{2p} \\ \dots & \dots & \dots & \dots \\ \rho_{p1} & \rho_{p2} & \dots & 1 \end{bmatrix}$$

– складається з парних коеф. кореляції $\rho_{ij}$ $(i,j=1,2,\dots,p)$

Основна задача багатовимірного кореляційного аналізу – оцінка кореляційної матриці $Q_p$ за вибіркою. Ця задача вирішується визначенням матриці вибіркових коеф. кореляції:

$$q_p = \begin{bmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \dots & \dots & \dots & \dots \\ r_{p1} & r_{p2} & \dots & 1 \end{bmatrix}, \text{ де } r = \frac{\overline{xy} - \bar{x}\bar{y}}{s_x s_y}$$

У багатомірному кореляційному аналізі визначають 2 задачі:

а) визначення частоти зв'язку однієї змінної із сукупністю інших $(p-1)$ змінних;

б) визначення —//— між змінними при фіксуванні або виключенні впливу інших $q$ змінних, де $q \le (p-2)$.

Множинний коеф. кореляції: а) задача вирішується за допомогою

$$R_{i,1\dots p} = \sqrt{1 - \frac{\det q_p}{q_{ii}}}, \text{ де } q_{ii} \text{ – алгебр. доповнення } r_{ii}$$

Запровадимо статистику:

$$F = \frac{R^2(n-p)}{(1-R^2)(p-1)}, \text{ де } R^2 \text{ – вибірковий множинний коеф. детерм.}$$

Якщо $F > F_{\alpha, k_1, k_2}$, то $R$ суттєво відрізняється від 0.

$k_1 = p-1$

$k_2 = n-p$

Частковий коеф. кореляції між $x_i$ та $x_j$ при фіксованих значеннях (p-2) змінних змінних є:

$$r_{ij,1\dots p} = \frac{-Q_{ij}}{\sqrt{Q_{ii}Q_{jj}}}$$

$Q_{ii}$, $Q_{jj}$ - алгебраїчні доповнення $r_{ii}$ та $r_{jj}$ відпов.

$-1 \le r_{ij,1\dots p} \le 1$.

## 44. Коефіцієнт рангової кореляції Спірмена та перевірка його значущості.

19/10 Рангова кореляція. Коефіцієнти рангової кореляції Спірмена і Кендалла, перевірка їх значущості.

Якщо об'єкти ранжовані по двум ознакам, то існує можливість оцінити частоту зв'язку між ознаками, основуючись на рангах, тобто частоту рангової кореляції.

Коеф. рангової кореляції Спірмена

$$\rho = 1 - \frac{6\sum_{i=1}^{n}(r_i - s_i)^2}{n^3 - n}$$

Формула діє у випадку, коли немає різних об'єктів із зв'язаними номерами.

Якщо вони є, то діє формула:

$$\rho = 1 - \frac{\sum_{i=1}^{n}(r_i - s_i)^2}{\frac{1}{6}(n^3 - n) - (T_r + T_s)}$$

$$T_r = \frac{1}{12}\sum_{i=1}^{m_r}(t_r^3 - t_r) \qquad T_s = \frac{1}{12}\sum_{i=1}^{m_s}(t_s^3 - t_s)$$

$m_r, m_s$ – к-ть рангів, що повторюються

$t_r, t_s$ – к-ть рангів, що не відрізняються

При $n > 10$ статистика має розподіл Стьюдента:

$$t = \frac{\rho\sqrt{n-2}}{\sqrt{1-\rho^2}}, \quad k = n-2 \text{ – ступені вільності}$$

Якщо $|t| > t_{1-\alpha, n-2}$, то $\rho$ буде істотним на рівні $\alpha$.

коеф. рангової кореляції Кендалла

$$\tau = 1 - \frac{4K}{n(n-1)}$$

K – статистика Кендалла

При $n > 10$ статистика:

$$t = \tau\sqrt{\frac{9n(n-1)}{2(2n+5)}}$$

Якщо $t > t_{кр}$, де $\Phi(t_{кр}) = 1 - \alpha$, то $\tau$ істотний на рівні $\alpha$.

## 45. Коефіцієнт рангової кореляції Кендалла та перевірка його значущості.

19/10 Рангова кореляція. Коефіцієнти рангової кореляції Спірмена і Кендалла, перевірка їх істотності.

Якщо об'єкти ранжовані по двум ознакам, то існує можливість оцінити частоту зв'язку між ознаками, основуючись на рангах, тобто частоту рангової кореляції.

Коеф. рангової кореляції Спірмена

$$\rho = 1 - \frac{6\sum_{i=1}^{n}(r_i - s_i)^2}{n^3 - n}$$

Формула діє у випадку коли немає різних об'єктів із зв'язаними номерами.

Якщо вони є, то діє формула:

$$\rho = 1 - \frac{\sum_{i=1}^{n}(r_i - s_i)^2}{\frac{1}{6}(n^3 - n) - (T_r + T_s)}$$

$$T_r = \frac{1}{12}\sum_{i=1}^{m_r}(t_r^3 - t_r) \qquad T_s = \frac{1}{12}\sum_{i=1}^{m_s}(t_s^3 - t_s)$$

$m_r, m_s$ – к-ть рангів, що повторюються

$t_r, t_s$ – к-ть рангів, що не відрізняються

При $n>10$ статистика має розподіл Стьюдента:

$$t = \frac{\rho\sqrt{n-2}}{\sqrt{1-\rho^2}}, \quad k = n-2 \text{ – ступені вільності}$$

Якщо $|t| > t_{1-\alpha, n-2}$, то $\rho$ буде істотним на рівні $\alpha$.

коеф. рангової кореляції Кендалла

$$\tau = 1 - \frac{4K}{n(n-1)}$$

K – статистика Кендалла

При $n>10$ статистика:

$$t = \tau\sqrt{\frac{9n(n-1)}{2(2n+5)}}$$

Якщо $t > t_{кр}$, де $\Phi(t_{кр}) = 1-\alpha$, то $\tau$ істотний на рівні $\alpha$.

## 46. Основні положення парного регресійного аналізу.

2/ ч

51. Модель лінійної регресії

$\varphi(x) = \beta_0 + \beta_1(x)$ – ф-ція лінійна в-но $\beta_0$ і $\beta_1$

Лінійна парна регресійна модель:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$y_i, x_i$ – вибіркові пари

$\beta_0, \beta_1$ – оцінювані параметри

$\varepsilon_i$ – збурення

Основні передумови регресійного аналізу

1. $\varepsilon_i$ – випадкова величина, $x_i$ – невипадкова велич.
2. $M(\varepsilon_i) = 0$
3. $D(\varepsilon_i) = \sigma^2$ – постійна (умова рівнозмінності)
4. Збурення $\varepsilon_i, \varepsilon_j$ – некорельовані: $M(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$.
5. Збурення $\varepsilon_i$ – нормально розподілена випад. велич.

5 передумова необхідна для оцінки точності рівняння регресії та її параметрів.

Дія неврахованих випадкових факторів визначається за допомогою остаточної дисперсії. Незміщеною оцінкою цієї дисперсії є вибіркова остаточна дисперсія:

$$S^2 = \frac{\sum_{i=1}^{n} (y_{x_i} - y_i)^2}{n-2} = \frac{\sum_{i=1}^{n} e_i^2}{n-2}$$

$y_{x_i}$ – групова середня

$e_i = y_{x_i} - y_i$ – залишок регресії

## 47. Перевірка значущості рівняння регресії.

### 52. Перевірка значущості рівняння регресії

Перевірити значущість р-ня регресії — встановити, чи відповідає мат. модель, що виражає залежність між змінними, експериментальним даним і чи достатньо включених в рівняння пояснюючих змінних для опису залежної змінної

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(y_{x_i}-\bar{y})^2+\sum_{i=1}^{n}(y_i-y_{x_i})^2$$

$$Q = Q_R + Q_e$$

$$Q_3=2\sum_{i=1}^{n}(y_{x_i}-\bar{y})(y_i-y_{x_i})=0.$$

Випадкові величини $S_R^2$ і $S_e^2$ мають розподіл $\chi^2$ відповідно із $k_1=m-1$ і $k_2=n-m$ ступенями вільності відповідно:

$$S_R^2=\frac{Q_R}{m-1} \qquad S_e^2=\frac{Q_e}{n-m}$$

$$F=\frac{S_R^2}{S_e^2}=\frac{Q_R(n-m)}{Q_e(m-1)}$$

$F > F_{\alpha,k_1,k_2}$, то рівняння регресії є значущим на рівні $\alpha$

де $F_{\alpha,k_1,k_2}$ — табличне значення критерію Фішера

# 48. Множинний регресійний аналіз.

## 53. Множинний регресійний аналіз

Використовується при дослідженні залежності змінної $y$ від декількох змінних $X_1, X_2, \dots, X_n$.

Модель множинної лінійної регресії:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i$$

$Y = (y_1\ y_2\ \dots\ y_n)'$ – вектор значень залежної змінної

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$$ –

матриця значень пояснюючих змінних

$\beta = (\beta_0\ \beta_1\ \dots\ \beta_p)'$ – вектор параметрів розміру $(p+1)$

$\varepsilon = (\varepsilon_1\ \varepsilon_2\ \dots\ \varepsilon_n)'$ – вектор збурень розміру $n$.

В матричній формі:

$$Y = X\beta + \varepsilon.$$

Оцінкою цієї моделі за вибіркою є р-ня

$$Y = XB + e, \quad B = (B_0\ B_1\ \dots\ B_p)', \quad e = (e_1\ e_2\ \dots\ e_n)'$$

$$\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} B_0 \\ B_1 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum y_i x_i \end{pmatrix}$$

Розв'язком р-ня є вектор $B = (X^T X)^{-1} X^T Y$.

Вибіркове рівняння множинної регресії:

$$y_{x_0} = X_0^T b.$$

параметрів лінійної регресії

Для оцінки вектора невідомих $\beta$ викорис. МНК.

$$e^T e = (e_1\ e_2 \dots e_n)\begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix} = e_1^2 + e_2^2 + \dots + e_n^2 = \sum_{i=1}^{n} e_i^2$$

Умова мінімізації остаточної суми квадратів:

$$S = \sum_{i=1}^{n} (y_{x_i} - y_i)^2 = \sum_{i=1}^{n} e_i^2 = e^T e = (y - Xb)^T (y - Xb) \to min$$

$$(Xb)^T = b^T X^T$$

$$S = y^T y - b^T X^T y - y^T X b + b^T X^T X b =$$

$$= y^T y - 2b^T X^T y + b^T X^T X b \to min$$

$$\frac{\partial}{\partial b}(b^T c) = c \qquad \frac{\partial}{\partial b}(b^T A b) = 2Ab$$

$$\frac{\partial S}{\partial b} = -2X^T y^T + 2X^T X b = 0 \quad \Rightarrow \quad X^T X b = X^T y^T$$

$$X^T X = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & \dots & \dots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix} =$$

$$= \begin{pmatrix} n & \sum x_{i1} & \dots & \sum x_{ip} \\ \sum x_{i1} & \sum x_{i1}^2 & \dots & \sum x_{i1} x_{ip} \\ \dots & \dots & \dots & \dots \\ \sum x_{ip} & \sum x_{i1} x_{ip} & \dots & \sum x_{ip}^2 \end{pmatrix} \quad \det X^T X \neq 0.$$

$$X^T y = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & \dots & \dots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum y_i x_{i1} \\ \dots \\ \sum y_i x_{ip} \end{pmatrix}$$