

МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ
Київський національний університет будівництва і архітектури

ТЕОРІЯ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНА СТАТИСТИКА

Модуль II. Математична статистика

Конспект лекцій

для студентів, які навчаються за напрямом підготовки
6.030510 «Товарознавство і торговельне підприємництво»
(усі форми навчання)

Київ – 2012

УДК 517(07)

ББК 22.171

Д64

Укладачі: О.В. Доля, канд. фіз.-мат. наук, доцент

С.А. Теренчук, канд. фіз.-мат. наук, доцент

Рецензент: С.В. Білощицька, канд. техн. наук, доцент

Відповідальний за випуск В.В. Демченко, канд. техн. наук,
доцент

*Затверджено на засіданні кафедри прикладної математики,
протокол №17 від 17 травня 2012 року.*

Доля О.В.

Д64 Теорія ймовірностей та математична статистика. Модуль II.
Математична статистика: конспект лекцій. / О.В. Доля,
С.А. Теренчук. – К. КНУБА, 2012 р. – 44 с.

Розглянуто теоретичний матеріал до Модуля II, що вивчається за програмою дисципліни «Теорія ймовірностей та математична статистика» для підготовки студентів, які навчаються за напрямом підготовки 6.030510 «Товарознавство та торговельне підприємництво» (усі форми навчання).

УДК 517(07)

ББК 22.171

© О.В. Доля, С.А. Теренчук, 2012

©КНУБА, 2012

Зміст

Передмова.....	4
Лекція №1. Вибірка та її характеристики. Емпірична функція розподілу, гістограма.....	4
Лекція №2. Оцінювання невідомих параметрів розподілів. Класифікація оцінок. Вибіркове середнє та дисперсія, мода, медіана.....	7
Лекція №3. Ефективні оцінки. Достатні статистики.....	10
Лекція №4. Метод моментів та метод максимальної правдоподібності.....	15
Лекція №5. Надійні інтервали.....	19
Лекція №6. Перевірка статистичних гіпотез. Критерії згоди.....	21
Лекція №7. Перевірка гіпотез про рівність математичних сподівань та дисперсій для нормальних сукупностей.....	28
Лекція №8. Лінійна регресія.....	30
Лекція №9. Елементи дисперсійного аналізу.....	37
Список літератури.....	41
Додатки.....	42

Передмова

Конспект лекцій складається з дев'яти лекцій, що відповідають навчальній робочій програмі з курсу «Теорія ймовірностей та математична статистика». Модуль II «Математична статистика».

В ньому викладені теоретичні відомості та підібрано задачі з теорії оцінювання невідомих параметрів, перевірки статистичних гіпотез, регресійного та дисперсійного аналізів. У додатку вміщено основні статистичні таблиці.

Конспект лекцій може бути корисним для студентів всіх форм навчання.

Лекція №1

ВИБІРКА ТА ЇЇ ХАРАКТЕРИСТИКИ. ЕМПІРИЧНА ФУНКЦІЯ РОЗПОДІЛУ, ГІСТОГРАМА

Випадковою **вибіркою** об'єму n (чи вибіркою) називається випадковий вектор $\xi_1, \xi_2, \dots, \xi_n$, де ξ_i – незалежні і однаково розподілені $F_{\xi}(x) = P\{\xi_i < x\}, i = 1, 2, \dots, n$. Іноді кажуть, що вибірка $\xi_1, \xi_2, \dots, \xi_n$ добута з **генеральної сукупності** випадкової величини ξ з **функцією розподілу** $F_{\xi}(x)$. Реалізацію вибірки будемо позначати відповідно $X = (x_1, x_2, \dots, x_n)$.

Розташуємо величини x_1, x_2, \dots, x_n у порядку зростання: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$, де $x_{(1)} = \min\{x_1, x_2, \dots, x_n\}$, $x_{(2)}$ – друга за величиною серед x_1, x_2, \dots, x_n , $x_{(n)} = \max\{x_1, x_2, \dots, x_n\}$. Позначимо через $\xi_{(k)}$ випадкову величину, яка для кожної реалізації X вибірки ξ набуває значення $x_{(k)}$, $k = 1, 2, \dots, n$. Отримали нову послідовність випадкових величин $\xi_{(1)}, \xi_{(2)}, \dots, \xi_{(n)}$, які називаються **порядковими статистиками**. Причому вони задовольняють нерівність:

$$\xi_{(1)} \leq \xi_{(2)} \leq \dots \leq \xi_{(n)}.$$

Ця послідовність називається **варіаційним рядом вибірки**.

Позначимо через $v_n(x)$ випадкову величину, рівну числу елементів вибірки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$, значення яких менші x , і позначимо

$F_n^*(x) = \frac{v_n(x)}{n}$. Функція $F_n^*(x)$ називається **емпіричною функцією розподілу**. Її можна використовувати як оцінку $F_\xi(x)$. Легко бачити, що $F_n^*(x)$ – випадкова величина, яка набуває значення $\left\{1; 2; \dots; \frac{n-1}{n}; 1\right\}$ з ймовірністю:

$$P\left\{F_n^*(x) = \frac{k}{n}\right\} = C_n^k (F_\xi(x))^k (1 - F_\xi(x))^{n-k}.$$

Для кожної реалізації X вибірки ξ функція $F_n^*(x)$ задовольняє всі властивості функції розподілу: змінюється від 0 до 1, неперервна зліва, неспадна.

Якщо всі компоненти вектора X різні, то:

$$F_n^*(x) = \begin{cases} 0, & \text{якщо } x \leq x_{(1)} \\ \frac{k}{n}, & \text{якщо } x_{(k)} < x \leq x_{(k+1)}, \\ 1, & \text{якщо } x > x_{(n)}. \end{cases}$$

Якщо деякі компоненти вектора X повторюються, то реалізацію вибірки зручніше задавати таблицею (**статистичний ряд**):

Значення	y_1	y_2	...	y_s
Частота	m_1	m_2	...	m_s

де y_1, y_2, \dots, y_s – різні значення даних варіаційного ряду $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$, а m_i – кількість повторів значення y_i у цьому ряді, $i = 1, 2, \dots, s$. Легко бачити, що $m_1 + m_2 + \dots + m_s = n$. Тоді:

$$F_n^*(x) = \begin{cases} 0, & \text{якщо } x \leq y_{(1)} \\ \frac{m_1 + m_2 + \dots + m_k}{n}, & \text{якщо } y_k < x \leq y_{k+1}, \\ 1, & \text{якщо } x > y_s. \end{cases}$$

Теорема Гливенка. $P\left\{\lim_{n \rightarrow \infty} \sup_x |F_n^*(x) - F_\xi(x)| = 0\right\} = 1$.

Теорема Колмогорова. Якщо $F_\xi(x)$ – неперервна, тоді для будь-якого $t > 0$

$$\lim_{n \rightarrow \infty} \left\{ \sqrt{n} \sup_x |F_n^*(x) - F_\xi(x)| < t \right\} = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 t^2}.$$

Теорема Колмогорова показує, що $F_n^*(x)$ дає рівномірну оцінку з точністю до величини порядку $\frac{1}{\sqrt{n}}$.

Нехай тепер ξ – неперервна випадкова величина зі щільністю $p(x)$. Для оцінки $p(x)$ з реалізації $X = (x_1, x_2, \dots, x_n)$ вибірки ξ розіб'ємо множину значень ξ на s інтервалів довжини h_i , $i = 1, 2, \dots, s$. Нехай x_i^* – середина i -го інтервалу, v_i – кількість елементів x_i , $j = 1, 2, \dots, n$, які потрапили в i -й інтервал. Тоді $\frac{v_i}{nh_i}$ – оцінка щільності в точці x_i^* .

Прямокутники з основами h_i і висотами $\frac{v_i}{nh_i}$, $i = 1, 2, \dots, s$ у прямокутній системі координат називаються **гістограмою вибірки**. Якщо на гістограмі ординати відповідні x_i^* послідовно з'єднати відрізками прямих, то здобута ламана буде полігоном **частот**. Полігон частот є також статистичним аналогом теоретичної щільності. Для інтервальних оцінок

$$F_n(x_i^*) = \frac{1}{n} \left[\sum_{k=1}^{i-1} m_k + \frac{m_i}{2} \right].$$

Приклад розв'язання задачі

Задача. Побудувати полігон частот та емпіричну функцію розподілу для вибірки, представлені статистичним рядом:

y_i	1	4	6
m_i	10	25	15

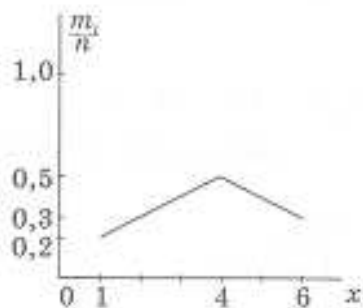


Рис.1.1

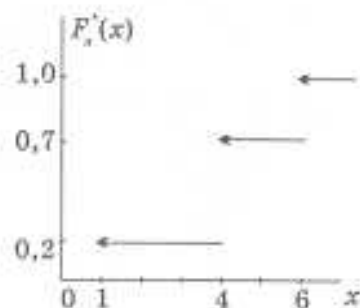


Рис.1.2

Розв'язання. На рис. 1.1 зображено полігон частот, а на рис. 1.2 – емпіричну функцію розподілу.

ОЦІНЮВАННЯ НЕВІДОМИХ ПАРАМЕТРІВ РОЗПОДІЛІВ. КЛАСИФІКАЦІЯ ОЦІНОК. ВИБІРКОВЕ СЕРЕДНЄ ТА ДИСПЕРСІЯ, МОДА, МЕДІАНА

Нехай $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ – вибірка з генеральної сукупності з розподілом $F_\xi(x, \theta)$. Причому функція розподілу $F_\xi(x, \theta)$ спостережуваної випадкової величини ξ має відому функціональну форму, але залежить від невідомого параметра θ . Цей параметр може бути будь-якою точкою заданої параметричної множини Θ .

Завдання оцінювання таке: використовуючи статистичну інформацію, яка міститься у вибірці ζ , зробити статистичні висновки про справжнє значення θ_0 невідомого параметра. Далі будь-яку функцію від вибірки будемо називати **статистикою**. При цьому множиною визначення статистики є вибірковий простір R^n , а множиною значень – Θ .

При точковому оцінюванні параметра необхідно знайти таку статистику $h = h(\xi)$, значення якої при заданій реалізації $X = (x_1, x_2, \dots, x_n)$ вибірки ζ приймають за наближене значення параметра θ_0 . У цьому випадку функцію $h(\zeta)$ називають **оцінкою параметра θ_0** . Для того, щоб порівнювати різні оцінки і вибрати кращу, існує така їх класифікація.

Означення 1. Оцінка $\hat{\theta} = h(\xi)$ параметра θ називається **незміщеною**, якщо $M(\hat{\theta}) = \theta$. Якщо $M(\hat{\theta}) = \theta + b(\theta)$, то величину $b(\theta)$ будемо називати **зсувом оцінки $\hat{\theta}$** .

Означення 2. Послідовність оцінок $\hat{\theta}_n = h(\xi_1, \dots, \xi_n)$, $n = 1, 2, \dots$ параметра θ називається **спроможною**, якщо $\hat{\theta}_n \xrightarrow{P} \theta, n \rightarrow \infty$.

Означення 3. Послідовність оцінок $\hat{\theta}_n = h(\xi_1, \dots, \xi_n)$, $n = 1, 2, \dots$ параметра θ називається **сильно спроможною**, якщо $\hat{\theta} \rightarrow \theta$ з імовірністю одиниця при $n \rightarrow \infty$.

Означення 4. Послідовність оцінок $\hat{\theta}_n = h(\xi_1, \dots, \xi_n)$, $n = 1, 2, \dots$ параметра θ називається **асимптотично нормальною**, якщо $\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow N(0, \sigma^2), n \rightarrow \infty$, тобто $\hat{\theta}_n$ має нормальний розподіл $N\left(0, \frac{\sigma^2}{n}\right)$.

Символ \xrightarrow{P} означає збіжність за ймовірністю: $\lim_{n \rightarrow \infty} P\{|\hat{\theta}_n - \theta| > \varepsilon\} = 0$, для будь-якого $\varepsilon > 0$, а символ \Rightarrow означає слабу збіжність функцій розподілів.

Позначимо клас усіх незміщених оцінок параметра θ через M_0 . Додатково припустимо, що дисперсії всіх оцінок з класу M_0 скінченні: $D(\hat{\theta}) = (\hat{\theta} - \theta)^2 < \infty$ для будь-якого $\hat{\theta} \in M_0$.

Означення 5. Оцінка $\hat{\theta}^*$ параметра θ називається **оптимальною**, якщо $D(\hat{\theta}^*) = \inf_{M_0} D(\hat{\theta})$.

Незміщеною, сильно спроможною та асимптотично нормальною оцінкою математичного сподівання є вибіркове середнє $\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$.

Через $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ будемо позначати реалізацію цієї оцінки. **Вибіркове середнє** для дискретного статистичного ряду знаходиться за формулою $\bar{x} = \frac{1}{n} \sum_{i=1}^s y_i m_i$, $n = \sum_{i=1}^s m_i$. **Вибіркове середнє** для інтервального статистичного ряду знаходиться за формулою $\bar{x} = \frac{1}{n} \sum_{i=1}^s y_i^* m_i$, де

$$y_i^* = \frac{y_i + y_{i+1}}{2} - \text{середина інтервалу } [y_i, y_{i+1}).$$

$$\hat{S}^2 = \frac{1}{n-1} \sum_{i=1}^s (\xi_i - \bar{\xi})^2 - \text{незміщена оцінка дисперсії } \sigma^2.$$

$$\bar{S}^2 = \frac{1}{n-1} \sum_{i=1}^s (\xi_i - \bar{\xi})^2 - \text{зміщена оцінка } \sigma^2.$$

Якщо математичне сподівання a відоме, то незміщеною, сильно спроможною і асимптотично нормальною оцінкою дисперсії, є оцінка:

$$S^2 = \frac{1}{n} \sum_{i=1}^s (\xi_i - a)^2.$$

Розв'язок x_p рівняння $F_\xi(x) = p$, де $p \in (0, 1)$, називається **p -квантиллю** розподілу $F_\xi(x)$. При $p=0,5$ квантиль називають **медіаною** розподілу. **Вибірковою p -квантиллю** $Z_{n,p}$ називають порядкову статистику:

$$Z_{n,p} = \begin{cases} \xi_{([np]+1)}, & \text{якщо } np \text{ дробове,} \\ \xi_{(np)}, & \text{якщо } np \text{ ціле.} \end{cases},$$

де $Z_{n,p}$ – це елемент вибірки, зліва від якого знаходиться частка $\frac{[np]}{n} \leq p$ спостережень і $Z_{n,p}$ – порядкова статистика з максимальним номером, що задовольняє цю властивість. Величина $Z_{n,1/2}$ називається **вибірковою медіаною**.

Реалізацію величини $Z_{n,1/2}$ позначають Me . Для дискретних статистичних рядів:

$$Me = \begin{cases} x_{(m)}, & \text{при } n = 2m - 1, \\ \frac{x_{(m)} + x_{(m+1)}}{2}, & \text{при } n = 2m. \end{cases}$$

Для інтервальних статистичних рядів:

$$Me = y_i + h_i \frac{\frac{n}{2} - \sum_{k=l}^{i-1} m_k}{m_i},$$

де y_i – початок медіанного інтервалу, тобто такого, якому відповідає перша з нагромаджених частот, що перевищує половину всіх спостережень, h_i – довжина інтервалу, m_i – частота медіанного інтервалу.

Мода – це елемент, який найчастіше трапляється у вибірці. Для дискретного статистичного ряду:

$$Mo = y_k, \text{ якщо } m_k = \max_i m_i.$$

Для інтервального статистичного ряду:

$$Mo = y_i + h_i \frac{m_i - m_{i-1}}{2m_i - m_{i-1} - m_{i+1}},$$

де y_i – початок інтервалу з найбільшою частотою, m_i – частота i -го інтервалу.

Приклад розв'язання задачі

Задача. Обчислити вибіркове середнє, дисперсію, моду, медіану для вибірки:

Інтервал	$[-2;0)$	$[0;4)$	$[4;6)$	$[6;10]$
m_i	5	10	20	15

Розв'язання. Вибіркове середнє для інтервального статистичного ряду:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^s y_i^* m_i = \frac{1}{50} (-1 \cdot 5 + 2 \cdot 10 + 5 \cdot 20 + 8 \cdot 15) = 4,7$$

$$M_2 = \frac{1}{n} \sum_{i=1}^s (y_i^*)^2 m_i = \frac{1}{50} ((-1)^2 \cdot 5 + 2^2 \cdot 10 + 5^2 \cdot 20 + 8^2 \cdot 15) = 30,1.$$

$$\bar{S}^2 = M_2 - (\bar{x})^2 = 30,1 - (4,7)^2 = 8,01.$$

$$\hat{S}^2 = \frac{n}{n-1} \bar{S}^2 = \frac{50}{49} \cdot 8,01 \approx 8,17.$$

$$Me = 4 + 2 \cdot \frac{25 - 15}{20} = 5.$$

$$Mo = 4 + 2 \cdot \frac{20 - 10}{40 - 10 - 15} = 5,33.$$

Лекція №3

ЕФЕКТИВНІ ОЦІНКИ. ДОСТАТНІ СТАТИСТИКИ

Нехай $L(X, \theta)$ – щільність розподілу вибіркового вектора $\zeta = (\xi_1, \xi_2, \dots, \xi_n)$, $X \in R^n$, $\theta \in \Theta$ (або ймовірність у дискретному випадку). Якщо ζ – вибірка з генеральної сукупності з розподілом $F_\xi(x, \theta)$, то $L(X, \theta) = \prod_{k=1}^n p(x_k, \theta)$, де $p(x_k, \theta)$ – щільність випадкової величини ξ (або ймовірність у дискретному випадку). Функція $L(X, \theta)$ називається **функцією правдоподібності**.

Функція $I(\theta) = M \left(\frac{\partial}{\partial \theta} \ln L(\zeta, \theta) \right)^2$ називається **кількістю інформації за Фішером**. У разі, якщо $\xi_1, \xi_2, \dots, \xi_n$ – незалежні однаково розподілені випадкові величини з щільністю $p(x, \theta)$,

$$I(\theta) = n I_1(\theta) = n M \left(\frac{\partial}{\partial \theta} \ln p(\xi, \theta) \right)^2.$$

Нехай для деякої функції $h(\theta)$ існує незміщена оцінка $\hat{h}(\zeta)$.

Теорема (нерівність Крамера–Рао). Нехай функція $L(X, \theta)$ двічі диференційована за θ і:

$$M \left| \frac{\partial}{\partial \theta} \ln L(\zeta, \theta) \right| < \infty, \quad M \left| \frac{\partial^2}{\partial \theta^2} \ln L(\zeta, \theta) \right| < \infty, \quad M \left| \frac{\partial}{\partial \theta} \ln L(\zeta, \theta) \right|^2 < \infty,$$

функція $h(\theta)$ диференційована і:

$$D\hat{h}(\zeta) < \infty, \quad \int_{R^n} \left| \hat{h}(X) \frac{\partial}{\partial \theta} L(X, \theta) \right| dX < \infty,$$

тоді

$$D\hat{h}(\zeta) \geq \frac{(h'(\theta))^2}{I(\theta)}.$$

Нерівність перетворюється в рівність тоді і тільки тоді, коли:

$$\frac{\partial \ln L(\zeta, \theta)}{\partial \theta} = \pm A(\theta) (\hat{h}(\zeta) - h(\theta))$$

з імовірністю одиниця, де

$$A(\theta) = \frac{I(\theta)}{h'(\theta)}.$$

Якщо $h(\theta) = \theta$, то $M(\hat{h}(\zeta) - \theta)^2 \geq I^{-1}(\theta)$.

Нехай тепер $\hat{h}(\zeta)$ зміщена оцінка θ і $M\hat{h}(\zeta) = \theta + b(\theta)$. Тоді

$$D\hat{h}(\zeta) \geq \frac{(1 + b'(\theta))^2}{I(\theta)}.$$

Якщо нижня межа нерівності двох останніх нерівностей для незміщеної чи зміщеної оцінки відповідно досягається, то така оцінка називається **ефективною**.

Статистика $T(\zeta)$ називається **достатньою** для невідомого параметра θ , якщо умовна щільність (або ймовірність у дискретному випадку) $L(X/t; \theta)$ випадкового вектора ζ при умові $T(\zeta) = t$ не залежить від θ .

Це означає, що статистика T містить усю інформацію про параметр θ , що є у вибірці. Достатня статистика задає оптимальний спосіб зображення статистичних даних, що особливо важливо при обробці великих масивів статистичної інформації. Потрібно знайти достатню статистику мінімальної вимірності, що подає дані в найбільш стислому вигляді. При цьому йдеться про **мінімальну достатню статистику**. Очевидно, вибірка ζ є достатньою статистикою, але ця статистика тривіальна, оскільки не скорочує даних.

Теорема (критерій факторизації). Для того, щоб статистика $T(\zeta)$ була достатньою, необхідно і достатньо, щоб функція правдоподібності мала вигляд:

$$L(X, \theta) = \varphi(T(X), \theta) h(X),$$

де функції $h(X)$ і $T(X)$ не залежать від θ .

Приклади розв'язання задачі

Задача 1. Знайти оцінку параметрів нормального розподілу.

Розв'язання. Нехай $\zeta = (\xi_1, \xi_2, \dots, \xi_n)$ – вибірка з генеральної сукупності з нормальним розподілом $N(a, \theta)$. Розглянемо два випадки:

1) a – невідомий параметр, а дисперсія θ відома. Тоді:

$$L(X, a) = (2\pi\theta)^{-\frac{n}{2}} e^{-\frac{\sum_{k=1}^n (x_k - a)^2}{2\theta}},$$

$$\ln L(X, a) = -\frac{n}{2} \ln(2\pi\theta) - \frac{\sum_{k=1}^n (x_k - a)^2}{2\theta},$$

$$\frac{\partial \ln L(X, a)}{\partial a} = \frac{1}{\theta} \sum_{k=1}^n (x - a) = \frac{n}{\theta} \left(\frac{1}{n} \sum_{k=1}^n x_k - a \right) = \frac{n}{\theta} (\bar{x} - a).$$

$$\text{Тобто } \frac{\partial \ln L(\zeta, a)}{\partial a} = \frac{n}{\theta} (\bar{\xi} - a).$$

Знайдемо тепер $I(a)$:

$$I(a) = -M \frac{\partial^2 \ln L(\zeta, a)}{\partial a^2} = \frac{n}{\theta}, \quad D(\bar{\xi}) = \frac{\theta}{n}.$$

Отже, $D(\bar{\xi}) = I^{-1}(a)$ і $\frac{\partial \ln L(\zeta, a)}{\partial a} = I(a)(\bar{\xi} - a)$. Таким чином, оцінка

$\bar{\xi} = \frac{1}{n} \sum_{k=1}^n \xi_k$ параметра a ефективна. Крім того, вона сильно спроможна і асимптотично нормальна;

2) a – відомий параметр, а дисперсія θ невідома. Аналогічно:

$$\ln L(\zeta, \theta) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \theta - \frac{\sum_{k=1}^n (\xi_k - a)^2}{2\theta},$$

$$\frac{\partial \ln L(\zeta, \theta)}{\partial \theta} = -\frac{n}{2\theta} + \frac{1}{2\theta^2} \sum_{k=1}^n (\xi_k - a)^2 = \frac{n}{2\theta^2} \left(\frac{1}{n} \sum_{k=1}^n (\xi_k - a)^2 - \theta \right) = \frac{n}{2\theta^2} (S^2 - \theta).$$

Оцінка S^2 є незміщеною, сильно спроможною і асимптотично нормальною оцінкою дисперсії θ :

$$I(\theta) = -M\left(\frac{n}{2\theta^2} - \frac{1}{\theta^3} \sum_{k=1}^n (\xi_k - a)^2\right) = -\frac{n}{2\theta^2} + \frac{n\theta}{\theta^3} = \frac{n}{2\theta^2}.$$

Отже, $\frac{\partial \ln L(\zeta, \theta)}{\partial \theta} = I(\theta)(S^2 - \theta)$ і S^2 – ефективна оцінка параметра θ .

Її дисперсія $DS^2 = \frac{2\theta^2}{n}$.

Задача 2. Знайти оцінку параметра показникового розподілу.

Розв'язання. Нехай $\xi_1, \xi_2, \dots, \xi_n$ – незалежні однаково розподілені випадкові величини з $p(x, \theta) = \theta e^{-\theta x}$, $x \geq 0$, $M(\xi_k) = \frac{1}{\theta}$. Нехай $\hat{\theta}_n = \frac{n}{\sum_{k=1}^n \xi_k}$.

Випадкова величина $\gamma_n = \sum_{k=1}^n \xi_k$ має розподіл Ерланга, щільність якого:

$$p_{\gamma_n}(x) = \frac{\theta^n x^{n-1}}{(n-1)!} e^{-\theta x}, \quad x \geq 0. \quad \text{Тоді:}$$

$$\begin{aligned} M(\hat{\theta}_n) &= \int_0^\infty \frac{n}{x} \frac{\theta^n x^{n-1}}{(n-1)!} e^{-\theta x} dx = \frac{n\theta}{(n-1)!} \int_0^\infty t^{n-2} e^{-t} dt = \frac{n\theta}{(n-1)!} \Gamma(n-1) = \frac{n\theta(n-2)!}{(n-1)!} \\ &= \frac{n\theta}{n-1} = \theta + \frac{\theta}{n-1}, \end{aligned}$$

$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ – гамма-функція,

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha), \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

Отже, $\hat{\theta}_n$ – зміщена оцінка параметра θ , $b(\theta) = \frac{\theta}{n-1}$.

Знайдемо $I(\theta)$:

$$\begin{aligned} L(\zeta, \theta) &= \theta^n e^{-\theta \sum_{k=1}^n \xi_k}, \\ \ln L(\zeta, \theta) &= n \ln \theta - \theta \sum_{k=1}^n \xi_k, \end{aligned}$$

$$\frac{\partial \ln L(\zeta, \theta)}{\partial \theta} = \frac{n}{\theta} - \sum_{k=1}^n \xi_k,$$

$$\frac{\partial^2 \ln L(\zeta, \theta)}{\partial \theta^2} = -\frac{n}{\theta^2},$$

$$I(\theta) = \frac{n}{\theta^2}.$$

Далі знайдемо $D(\hat{\theta}_n)$:

$$M(\hat{\theta}_n^2) = \int_0^\infty \frac{n^2}{x^2} \frac{\theta^n x^{n-1}}{(n-1)!} e^{-\theta x} dx = \frac{n^2 \theta^2}{(n-1)!} \int_0^\infty t^{n-3} e^{-t} dt = \frac{n^2 \theta^2}{(n-1)!} \Gamma(n-2) = \frac{n^2 \theta^2}{(n-1)(n-2)};$$

$$D(\hat{\theta}_n) = \frac{n^2 \theta^2}{(n-1)(n-2)} - \frac{n^2 \theta^2}{(n-1)^2} = \frac{n^2 \theta^2}{(n-1)^2(n-2)}.$$

Далі
$$\frac{(1+b'(\theta))^2}{I(\theta)} = \frac{\left(1 + \frac{1}{n-1}\right)^2 \theta^2}{n} = \frac{\theta^2 n}{(n-1)^2};$$

$$D(\hat{\theta}_n) = \frac{n^2 \theta^2}{(n-1)^2(n-2)} > \frac{\theta^2 n}{(n-1)^2} \text{ для } n = 3, 4, \dots$$

Отже, оцінка $\hat{\theta}_n = \frac{n}{\sum_{k=1}^n \xi_k}$ не є ефективною оцінкою параметра θ .

Задача 3. Знайти достатню статистику для нормального розподілу $N(a, \theta)$.

Розв'язання.
$$L(X, a, \theta) = (2\pi\theta)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\theta} \sum_{k=1}^n (x_k - a)^2\right\} =$$

$$= (2\pi\theta)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\theta} \sum_{k=1}^n (x_k - \bar{x})^2 - \frac{1}{2\theta} \sum_{k=1}^n (\bar{x} - a)^2\right\}.$$

Згідно з критерієм факторизації $T(\xi) = \left(\bar{\xi}, \sum_{k=1}^n (\xi_k - \bar{\xi})^2\right)$. Якщо

дисперсія θ відома, то достатньою статистикою для параметра θ буде $\bar{\xi}$. Якщо ж, навпаки, a – відомий параметр, то достатньою статистикою для θ буде $\sum_{k=1}^n (\xi_k - a)^2$.

Задача 4. Знайти достатню статистику для параметра розподілу Коші.

Розв'язання. Нехай:

$$p(x, \theta) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}; \quad L(X, \theta) = \frac{1}{\pi^2} \prod_{k=1}^n \frac{1}{1 + (x_k - \theta)^2}.$$

Для параметра θ існує лише тривіальна достатня статистика $\zeta = (\xi_1, \xi_2, \dots, \xi_n)$.

Задача 5. Знайти достатню статистику для параметра рівномірного розподілу на відрізку $[-\theta, \theta]$.

Розв'язання.

$$p(x, \theta) = \begin{cases} 0, & x \notin [-\theta, \theta], \\ \frac{1}{2\theta}, & x \in [-\theta, \theta], \theta > 0, \end{cases}$$

$$L(X, \theta) = (2\theta)^{-n} \prod_{k=1}^n I(\theta - x_k) \cdot I(\theta + x_k),$$

$$\text{де } I(x) = \begin{cases} 1, & \text{при } x \geq 0, \\ 0, & \text{при } x < 0. \end{cases}$$

Легко бачити, що функцію правдоподібності можна зобразити у вигляді $L(X, \theta) = (2\theta)^{-n} I(\theta - x_{(n)}) I(\theta + x_{(1)})$. Тоді достатньою статистикою для параметра θ буде вектор $(\xi_{(1)}, \xi_{(n)})$, де $\xi_{(1)} = \min_{1 \leq k \leq n} \xi_k, \xi_{(n)} = \max_{1 \leq k \leq n} \xi_k$.

Лекція №4

МЕТОД МОМЕНТІВ ТА МЕТОД МАКСИМАЛЬНОЇ ПРАВДОПОДІБНОСТІ

Нехай $\zeta = (\xi_1, \xi_2, \dots, \xi_n)$ – вибірка з генеральної сукупності з розподілом $F_\xi(x, \theta)$, де $\theta = (\theta_1, \theta_2, \dots, \theta_s), \theta \in \Theta \subseteq R^s$. Припустимо, що у випадковій величині ξ , що спостерігається, є перші s моментів $\alpha_k = M\xi^k, k = \overline{1, s}$. При цьому вони є функціями від невідомих параметрів θ : $\alpha_k = \alpha_k(\theta)$. Нехай $X = (x_1, x_2, \dots, x_n)$ – реалізація вибірки ζ . Значення

оцінок параметрів $\theta_1, \theta_2, \dots, \theta_s$ за методом моментів знаходиться в результаті розв'язку системи рівнянь

$$\alpha_k(\theta) = \frac{1}{n} \sum_{i=1}^n x_i^k, \quad k = \overline{1, s}.$$

Оцінки, знайдені методом моментів, як правило, спроможні, але часто неефективні.

Нехай спостерігається випадковий вектор $\zeta = (\xi_1, \xi_2, \dots, \xi_n)$ зі щільністю $L(X, \theta)$, $X = (x_1, x_2, \dots, x_n) \in R^n$ і $\theta \in \Theta \subseteq R^n$. **Оцінкою максимальної правдоподібності** $\hat{\theta}_n$ називається така точка множини Θ , в якій функція правдоподібності $L(X, \theta)$ при заданому X набуває максимального значення. Тобто:

$$L(X, \hat{\theta}_n) = \max_{\theta \in \Theta} L(X, \theta).$$

У багатьох випадках знаходять $\max_{\theta \in \Theta} \ln L(X, \theta)$, причому максимум досягається в тих же точках, що і $\max_{\theta \in \Theta} L(X, \theta)$.

Якщо для кожного X з вибіркового простору R^n максимум $L(X, \theta)$ досягається у внутрішній точці Θ і функція $L(X, \theta)$ диференційована за θ , то оцінка $\hat{\theta}_n = (\hat{\theta}_n^{(1)}, \hat{\theta}_n^{(2)}, \dots, \hat{\theta}_n^{(s)})$ при заданій реалізації $X = (x_1, x_2, \dots, x_n)$ вектора ζ задовольняє систему рівнянь:

$$\frac{\partial L(X, \theta)}{\partial \theta_j} = 0, \text{ або } \frac{\partial \ln L(X, \theta)}{\partial \theta_j} = 0, \quad j = \overline{1, s}.$$

Останні рівняння називаються **рівняннями правдоподібності**.

Якщо для параметра θ існує достатня статистика $T(X)$, то розв'язок рівнянь правдоподібності є функцією від достатньої статистики.

Нехай θ – скалярний параметр. Якщо для параметра θ існує ефективна незміщена оцінка, то вона збігається з оцінкою максимальної правдоподібності.

Приклади розв'язання задач

Задача 1. Нехай $\xi_1, \xi_2, \dots, \xi_n$ – вибірка з генеральної сукупності зі щільністю:

$$p(x, \theta) = \frac{2x}{\theta^2} e^{-\frac{x^2}{\theta^2}}, \quad x \geq 0, \theta \geq 0.$$

Знайти методом моментів оцінку параметра θ . Чи буде ця оцінка незміщеною та спроможною?

Розв'язання. Спочатку знайдемо:

$$\begin{aligned} M(\xi_k) &= \int_0^\infty \frac{2x^2}{\theta^2} e^{-\frac{x^2}{\theta^2}} dx = 2 \int_0^\infty t e^{-t} \frac{\theta}{2\sqrt{t}} dt = \\ &= \theta \int_0^\infty \sqrt{t} e^{-t} dt = \theta \Gamma\left(\frac{3}{2}\right) = \frac{\theta}{2} \sqrt{\pi}. \end{aligned}$$

Тоді $\frac{\theta}{2} \sqrt{\pi} = \frac{1}{n} \sum_{i=1}^n x_i$ і оцінкою параметра θ буде:

$$\hat{\theta}_n = \frac{2\bar{\xi}}{\sqrt{\pi}} = \frac{2}{n\sqrt{\pi}} \sum_{k=1}^n \xi_k.$$

Таким чином, $\hat{\theta}_n$ є незміщеною і спроможною оцінкою (закон великих чисел).

Задача 2. Задана щільність рівномірного розподілу:

$$p(x, \theta) = \begin{cases} 0, & x \notin [-\theta, \theta], \\ \frac{1}{2\theta}, & x \in [-\theta, \theta], \theta > 0. \end{cases}$$

Знайти методом моментів оцінку параметра θ .

Розв'язання. Математичне сподівання $M(\xi_k) = \int_{-\theta}^{\theta} \frac{x}{2\theta} dx = 0$ і не залежить від θ . Знайдемо $M(\xi_k^2) = \int_{-\theta}^{\theta} \frac{x^2}{2\theta} dx = \frac{\theta^2}{3}$. Отже, рівняння має вигляд $\frac{\theta^2}{3} = \frac{1}{n} \sum_{i=1}^n x_i^2$ і оцінка параметра θ є $\hat{\theta}_n = \sqrt{\frac{3}{n} \sum_{k=1}^n \xi_i^2}$.

Задача 3. Знайти оцінку максимальної правдоподібності для параметрів нормального розподілу $N(a, \theta)$.

Розв'язання. $L(X, a, \theta) = (2\pi\theta)^{-\frac{n}{2}} e^{-\frac{1}{2\theta} \sum_{k=1}^n (x_k - a)^2}$.

Рівняння правдоподібності мають вигляд:

$$\begin{aligned} \frac{\partial \ln L(X, a, \theta)}{\partial a} &= \frac{1}{\theta} \sum_{k=1}^n (x_k - a) = 0, \\ \frac{\partial \ln L(X, a, \theta)}{\partial \theta} &= -\frac{n}{2\theta} + \frac{1}{2\theta^2} \sum_{k=1}^n (x_k - a)^2 = 0. \end{aligned}$$

Розв'язуючи ці рівняння відносно a і θ , дістанемо:

$$a_n = \frac{1}{n} \sum_{i=1}^n x_k = \bar{x}, \theta_n = \frac{1}{n} \sum_{i=1}^n (x_k - \bar{x})^2.$$

Знайдемо тепер другі частинні похідні:

$$\frac{\partial^2 \ln L(X, a, \theta)}{\partial a^2} = -\frac{n}{\theta} < 0,$$

$$\frac{\partial^2 \ln L(X, a, \theta)}{\partial a^2} = -\frac{n}{2\theta^2} - \frac{\sum_{i=1}^n (x_k - a)^2}{\theta^3} \Big|_{\substack{\theta=\hat{\theta}_n \\ a=\hat{a}_n}} = -\frac{n}{2\theta^2} < 0,$$

$$\frac{\partial^2 \ln L(X, a, \theta)}{\partial a \partial \theta} = -\frac{\sum_{i=1}^n (x_k - a)}{\theta^2} \Big|_{\substack{\theta=\hat{\theta}_n \\ a=\hat{a}_n}} = \frac{-\sum_{i=1}^n (x_k - \bar{x}) \cdot n^2}{\left(\sum_{i=1}^n (x_k - \bar{x})^2\right)^2} = 0,$$

оскільки $\sum_{i=1}^n (x_k - \bar{x}) = 0$.

Знайдемо визначник у точці $(a, \theta) = (\hat{a}_n, \hat{\theta}_n)$:

$$\begin{aligned} \Delta &= \begin{vmatrix} \frac{\partial^2 \ln L}{\partial a^2} & \frac{\partial^2 \ln L}{\partial a \partial \theta} \\ \frac{\partial^2 \ln L}{\partial \theta^2} & \frac{\partial^2 \ln L}{\partial \theta^2} \end{vmatrix} = \\ &= \frac{n^2}{2\theta^3} - \frac{\left(\sum_{i=1}^n (x_k - a)\right)^2}{\theta^4} \Big|_{\substack{\theta=\hat{\theta}_n \\ a=\hat{a}_n}} = \frac{n^5}{2\left(\sum_{i=1}^n (x_k - \bar{x})\right)^3} > 0. \end{aligned}$$

Тобто точка з координатами $(\hat{a}_n, \hat{\theta}_n)$ максимізує значення функції $\ln L(X, a, \theta)$, і для параметрів (a, θ) оцінкою максимальної правдоподібності є (\bar{x}, \bar{S}^2) .

Задача 4. Знайти оцінку максимальної правдоподібності для показникового розподілу зі щільністю $p(x, \theta) = \theta e^{-\theta x}$, $x \geq 0$.

Розв'язання. Функція правдоподібності має вигляд

$L(X, \theta) = \theta^n e^{-\theta \sum_{k=1}^n x_k}$, а рівняння правдоподібності:

$$\frac{\partial \ln L(X, \theta)}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^n x_k = 0.$$

Отже, $\theta_n = \frac{n}{\sum_{i=1}^n x_k}$. Оскільки $\frac{\partial^2 \ln L(X, \theta)}{\partial \theta^2} = -\frac{n}{\theta^2} < 0$, то θ_n – точка

максимуму функції $\ln L(X, \theta)$. Таким чином, оцінкою максимальної правдоподібності параметра θ буде $\frac{1}{\bar{\xi}}$. Ця оцінка є зміщеною і неефективною. Зауважимо, що коли показниковий розподіл задається у вигляді:

$$p(x, \theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, x \geq 0,$$

то оцінка максимальної правдоподібності $\theta_n = \bar{\xi}$ буде незміщеною і ефективною.

Лекція №5

НАДІЙНІ ІНТЕРВАЛИ

Нехай ζ – вибіркового вектор об'єму n , P_θ – розподіл вектора ζ , який залежить від невідомого параметра θ . При інтервальному оцінюванні невідомого параметра θ шукають такі дві статистики $h_1(\zeta)$ і $h_2(\zeta)$, в яких $h_1 < h_2$ і для яких при заданому $(1 - \alpha) \in (0, 1)$ виконується умова:

$$P_\theta \{h_1(\zeta) < \theta < h_2(\zeta)\} \geq 1 - \alpha.$$

Тоді інтервал $(h_1(\zeta), h_2(\zeta))$ називається **$(1 - \alpha)$ – надійним інтервалом**, ймовірність **$(1 - \alpha)$ – рівнем надійності**, а $h_1(\zeta)$ і $h_2(\zeta)$ – **нижньою та верхньою межами надійності**.

Інтервал $\left(\hat{\theta}_n - \frac{C_\alpha}{\sqrt{I(\hat{\theta}_n)}}, \hat{\theta}_n + \frac{C_\alpha}{\sqrt{I(\hat{\theta}_n)}} \right)$ – асимптотично найкоротший

$(1 - \alpha)$ – надійний інтервал. Значення величини C_α подано у табл. 5.1.

Таблиця 5.1

Значення C_α

$(1 - \alpha)$	0,8	0,9	0,95	0,99	0,999
C_α	1,28	1,64	1,96	2,57	3,29

Надійні інтервали для нормального розподілу

1. Надійний інтервал для математичного сподівання при відомій дисперсії:

$$\bar{\xi} - C_{\alpha} \frac{\sigma}{\sqrt{n}} < a < \bar{\xi} + C_{\alpha} \frac{\sigma}{\sqrt{n}}.$$

2. Надійний інтервал для дисперсії σ^2 , якщо математичне сподівання a відоме:

$$\frac{nS^2}{\chi_1^2} < \sigma^2 < \frac{nS^2}{\chi_2^2},$$

де $S^2 = \frac{1}{n} \sum_{k=1}^n (\xi_k - a)^2$, числа χ_1^2 і χ_2^2 такі, що:

$$P\{\chi^2(n) < \chi_1^2\} = \frac{\alpha}{2} \text{ і } P\{\chi^2(n) < \chi_2^2\} = \frac{\alpha}{2}$$

(табл. 1 дод. 1).

3. Надійні інтервали для дисперсії a та σ^2 у випадку, якщо обидва параметри невідомі:

$$\bar{\xi} - t_{\alpha} \frac{\hat{S}}{\sqrt{n}} < a < \bar{\xi} + t_{\alpha} \frac{\hat{S}}{\sqrt{n}},$$

$$\frac{(n-1)\hat{S}^2}{\chi_2^2} < \sigma^2 < \frac{(n-1)\hat{S}^2}{\chi_1^2},$$

де t_{α} знаходиться з таблиці Стюдента (табл. дод. 2) так, щоб $P\{|t_{n-1}| < t_{\alpha}\} = 1 - \alpha$, де $k = n - 1$ – число степенів свободи,

$$\hat{S}^2 = \frac{1}{n-1} \sum_{k=1}^n (\xi_k - \bar{\xi})^2, \quad P\{\chi^2(n-1) < \chi_1^2\} = \frac{\alpha}{2} \text{ і } P\{\chi^2(n-1) < \chi_2^2\} = \frac{\alpha}{2}$$

(табл. дод. 1).

Приклад розв'язання задачі

Задача. Нехай $\zeta = (\xi_1, \xi_2, \dots, \xi_n)$ – вибіркового вектор генеральної сукупності з розподілом Пуассона з параметром θ . Побудувати $(1 - \alpha)$ – надійний інтервал.

Розв'язання. За умовою задачі:

$$P\{\xi_k = k\} = \frac{\theta^k}{k!} e^{-\theta}, \quad k = 0, 1, \dots \quad L(X, \theta) = e^{-n\theta} \frac{\theta^{\sum_{k=1}^n x_k}}{\prod_{k=1}^n x_k!}, \quad X = (x_1, x_2, \dots, x_n)$$

– реалізація вектора ζ

$$\frac{\partial \ln L(X, \theta)}{\partial \theta} = \frac{n}{\theta} (\bar{x} - \theta), \quad \frac{\partial^2}{\partial \theta^2} \ln L(X, \theta) = -\frac{n}{\theta^2} \bar{x}.$$

Оцінка максимальної правдоподібності знаходиться з рівнянь:

$$\frac{n}{\theta} (\bar{\xi} - \theta) = 0$$

i

$$\theta_n = \bar{\xi} = \frac{1}{n} \sum_{k=1}^n \xi_k, \quad I(\theta) = -M \frac{\partial^2}{\partial \theta^2} \ln L(\xi, \theta) = \frac{n\theta}{\theta^2} = \frac{n}{\theta}.$$

Тоді $(1 - \alpha)$ – надійний інтервал для параметра θ буде таким:

$$\bar{\xi} - C_\alpha \sqrt{\frac{\bar{\xi}}{n}} < \theta < \bar{\xi} + C_\alpha \sqrt{\frac{\bar{\xi}}{n}}.$$

Лекція №6

ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. КРИТЕРІЙ ЗГОДИ

Статистичною гіпотезою називається будь-яке твердження про вигляд або властивості розподілу випадкових величин, що спостерігаються в експерименті. Правило, згідно з яким гіпотеза H_0 , що перевіряється, приймається чи відхиляється, називається **статистичним критерієм** для перевірки гіпотези H_0 . Наведемо декілька прикладів статистичних гіпотез.

1. Гіпотеза про вигляд розподілу. Нехай $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ – вибірка з генеральної сукупності з невідомою функцією розподілу $F_\xi(x)$. Тоді $H_0: F_\xi(x) = F(x)$, де $F(x)$ повністю задана, або $H_0: F_\xi(x) \in M$, де $M = \{F(x, \theta), \theta \in \Theta\}$ – задане сімейство функцій розподілу.

2. Гіпотеза однорідності. Нехай маємо k вибірок $\xi = (\xi_1, \xi_2, \dots, \xi_n)$, $i = \overline{1, k}$ з генеральних сукупностей з функціями розподілу $F_i(x)$, $i = \overline{1, k}$. Потрібно перевірити гіпотезу про те, що це спостереження над однією і тією ж випадковою величиною, тобто $H_0: F_1(x) \equiv F_2(x) \equiv \dots \equiv F_k(x)$.

3. Гіпотеза незалежності. Одночасно спостерігаються дві випадкові величини ξ та η , $F_{(\xi, \eta)}(x, y)$ – невідома їхня сумісна функція розподілу. Потрібно перевірити гіпотезу про те, що ξ та η – незалежні випадкові величини, тобто $H_0: F_{(\xi, \eta)}(x, y) = F_\xi(x)F_\eta(y)$.

Розглянемо методи перевірки гіпотез описаних вище типів. У всіх випадках формулювалася лише одна гіпотеза H_0 , і необхідно перевірити, чи узгоджуються статистичні дані з цією гіпотезою, чи ні. Відповідні критерії називаються **критеріями згоди**. Якщо гіпотеза H_0 однозначно фіксує розподіл спостережень, то гіпотеза називається **простою**, у протилежному випадку – **складною**. У наведених вище прикладах лише гіпотеза $H_0: F_\xi(x) = F(x)$ є простою.

Опишемо основний процес, який використовується під час побудови критерію згоди. Вибирається деяка статистика $T = T(\zeta)$, яка є мірою розбіжності статистичного й теоретичного законів розподілу і називається **статистикою критеріїв** або **критерієм**. Далі знаходиться розподіл критерію T у припущенні, що розподіл спостережень збігається з гіпотетичним. Визначимо тепер таке число T_α , щоб:

$$P\{T \geq T_\alpha / H_0\} = \alpha,$$

де α – достатньо мале число. Число α називається рівнем значущості критерію. Якщо значення міри розбіжності T , обчислене за спостереженнями, більше або рівне T_α ($T \geq T_\alpha$), тоді відхилення від теоретичного закону вважається значущим, і гіпотеза відхиляється. Якщо ж $T < T_\alpha$, то відхилення не вважається значущим, тобто дані експерименту не суперечать гіпотезі.

Критерій Колмогорова про вигляд розподілу. Критерій Колмогорова застосовується в тих випадках, коли $F(x)$ – неперервна функція. Статистикою критерію вибирається величина:

$$D_n = D_n(\xi) = \sqrt{n} \sup_{-\infty < x < \infty} |F_n^*(x) - F(x)|,$$

де $F_n^*(x)$ – емпірична функція розподілу. За заданим рівнем значущості α знайдемо таке число λ_α , що:

$$P\{D_n \geq \lambda_\alpha / H_0\} \approx 1 - K(\lambda_\alpha) = \alpha.$$

У табл. дод. 4 наведено значення λ_α для різних α . Отже, одержимо таке правило перевірки гіпотези H_0 . Нехай $\lambda_n = D_n(X)$ – величина $D_n(\zeta)$, обчислена за реалізацією вибірки $X = (x_1, x_2, \dots, x_n)$. Якщо $\lambda_n \geq \lambda_\alpha$, то гіпотеза H_0 відхиляється, а при $\lambda_n < \lambda_\alpha$ – приймається.

Критерій згоди χ^2 про вигляд розподілу. Цей критерій можна використовувати для будь-яких розподілів. Розіб'ємо множину всіх

можливих значень спостережуваної випадкової величини ξ на r інтервалів $\Delta_1, \Delta_2, \dots, \Delta_r$, що не перетинаються. Якщо спостерігається дискретна випадкова величина, то Δ_i , $i = \overline{1, r}$ – різні значення цієї величини. Нехай v_i – кількість елементів вибірки, що потрапили в інтервал Δ_i , $P_i = P\{\xi \in \Delta_i / H_0\}$. Статистикою критерію χ^2 вибирається величина:

$$\chi_n^2 = \sum_{i=1}^r \frac{(v_i - np_i)^2}{np_i}.$$

Критерій перевірки гіпотези H_0 будується таким чином. Обчисливши значення χ_n^2 і вибравши рівень значущості α , за таблицею χ^2 – розподілу (табл. дод. 1) визначимо величину $\chi_{r-1, \alpha}^2 : P\{\chi^2(r-1) \geq \chi_{r-1, \alpha}^2\} = \alpha$. Якщо $\chi_n^2 \geq \chi_{r-1, \alpha}^2$, то гіпотеза H_0 відхиляється, якщо ж $\chi_n^2 < \chi_{r-1, \alpha}^2$, то гіпотеза приймається.

Критерій χ^2 можна використовувати для перевірки складних гіпотез. Нехай за вибіркою $\zeta = (\xi_1, \xi_2, \dots, \xi_n)$ потрібно перевірити гіпотезу $H_0: F_\xi(x) \in M$, де $M = \{F(x, \theta), \theta \in \Theta\}$ – задане сімейство розподілу. Невідомий параметр θ оцінюється за вибіркою і в статистику χ_n^2 підставляються ймовірності p_i , знайдені через $F(x, \hat{\theta}_n)$, де $\hat{\theta}_n$ – оцінка θ . Тоді граничний розподіл χ_n^2 буде χ^2 – розподіл з $(r-l-1)$ степенями свободи, де l – розмірність параметра θ . Далі критерії перевірки гіпотези будуються аналогічно наведеному критерію. Якщо $\chi_n^2 > \chi_{r-l-1, \alpha}^2$, то гіпотеза H_0 відхиляється, якщо ж $\chi_n^2 < \chi_{r-l-1, \alpha}^2$, то гіпотеза приймається.

Нехай $\zeta = (\xi_1, \xi_2, \dots, \xi_n)$ і $\gamma = (\eta_1, \eta_2, \dots, \eta_n)$ – дві вибірки з генеральних сукупностей з невідомими функціями розподілу $F_1(x)$ та $F_2(x)$ відповідно. Потрібно перевірити гіпотезу про те, що це спостереження над однією і тією ж випадковою величиною, тобто $H_0: F_1(x) \equiv F_2(x)$.

Критерій однорідності Смирнова – Колмогорова. Цей критерій використовується лише для неперервних розподілів. За міру розбіжності

приймається величина $\lambda_{nm} = \sqrt{\frac{nm}{n+m}} \sup_{-\infty < x < \infty} |F_{n1}^*(x) - F_{m2}^*(x)|$, де $F_{n1}^*(x)$,

$F_{m2}^*(x)$ – емпіричні функції розподілу, побудовані за першою і другою

вибірками. При $n \rightarrow \infty, m \rightarrow \infty$ розподіл випадкової величини λ_{nm} збігається до розподілу Колмогорова. Далі критерії будуються аналогічно критерію Колмогорова. За заданим рівнем значущості α знайдемо табличне число λ_α (табл. дод. 4). Якщо $\lambda_{nm} \geq \lambda_\alpha$, то гіпотеза H_0 відхиляється, а при $\lambda_{nm} < \lambda_\alpha$ – приймається.

Критерій однорідності χ^2 . Цей критерій можна використовувати для перевірки даних, які мають дискретну структуру. Окрім того, за допомогою цього критерію можна перевіряти однорідність будь-якого скінченного числа вибірок (за критерієм Смирнова–Колмогорова можна порівнювати лише дві вибірки). Нехай проведено k послідовних серій незалежних спостережень, які складаються з n_1, n_2, \dots, n_k спостережень. При цьому в кожному експерименті може виникнути один з l наслідків, v_{ij} – число виникнень i -го наслідку в j -й серії, $n_j = \sum_{i=1}^l v_{ij}$, $n = n_1 + n_2 + \dots + n_k$ – загальна кількість спостережень. Потрібно перевірити гіпотезу H_0 про те, що всі спостереження проводилися над однією і тією ж величиною. Статистикою критерію є величина:

$$\chi_n^2 = n \sum_{i=1}^l \sum_{j=1}^k \frac{\left(v_{ij} - \frac{v_{i0} n_j}{n} \right)^2}{v_{i0} n_j}; v_{i0} = \sum_{j=1}^k v_{ij}.$$

У таблиці χ^2 – розподілу (табл. дод. 1) за заданим α і числом степенів свободи $m = (l - 1)(k - 1)$ знаходимо число $\chi_{m,\alpha}^2: P\{\chi^2(m) \geq \chi_{m,\alpha}^2\} = \alpha$. Якщо $\chi_n^2 > \chi_{m,\alpha}^2$, то гіпотеза H_0 відхиляється, якщо ж $\chi_n^2 < \chi_{m,\alpha}^2$, то гіпотеза приймається.

Критерій незалежності χ^2 . Критерій χ^2 дає змогу перевіряти також гіпотезу про незалежність двох випадкових величин ξ та η . Статистикою критерію є величина:

$$\chi_n^2 = n \sum_{i=1}^l \sum_{j=1}^k \frac{(v_{ij} - m_{ij})^2}{m_{ij}},$$

де v_{ij} – число випадків, коли одночасно спостерігалися $\xi = x_i$ та $\eta = y_j$ (для неперервних випадкових величин i та j – номери відповідних інтервалів),

$$m_{ij} = \frac{v_{i0}v_{0j}}{n}; v_{0j} = \sum_{i=1}^l v_{ij}; v_{i0} = \sum_{j=1}^k v_{ij};$$

l і k – число значень, яких набувають випадкові величини ξ та η ;
 $n \sum_{i=1}^l \sum_{j=1}^k v_{ij}$ – обсяг вибірки. Вибір табличного значення $\chi^2(l-1)(k-1), \alpha$ і прийняття рішення проводиться аналогічно описаній вище процедурі для критерію однорідності χ^2 .

Приклади розв'язання задачі

Задача 1. Нехай вибірку подано у вигляді таблиці частот:

Інтервал	[0;1)	[1;2)	[2;3)	[3;4)	[4;5)	[5;6)	[6;7)	[7;8)	[8;9)	[9;10]
m_i	35	16	15	17	17	19	11	16	30	24

Перевірити гіпотезу $H_0: F_{\xi}(x) = \begin{cases} 1, x > 10, \\ \frac{x}{10}, x \in [0;10], \\ 0, x < 0 \end{cases}$ рівномірний

розподіл на $[0;10], \alpha = 0.05$.

Розв'язання. Нехай x_i^* – середина i -го інтервалу. Емпіричну функцію розподілу $F_n^*(x)$ можна знайти за формулою:

$$F_n^*(x_i^*) = \frac{1}{n} \left(\sum_{k=1}^{i-1} m_k + 0.5m_i \right).$$

Усі результати обчислень наведено у таблиці ($n=200$):

Інтервал	[0;1)	[1;2)	[2;3)	[3;4)	[4;5)	[5;6)	[6;7)	[7;8)	[8;9)	[9;10]
m_i	35	16	15	17	17	19	11	16	30	24
x_i^*	0,5	1,5	2,5	3,5	4,5	5,5	6,5	7,5	8,5	9,5
F_n^*	0,087	0,215	0,292	0,372	0,457	0,547	0,622	0,69	0,805	0,94
$F(x)$	0,05	0,15	0,25	0,35	0,45	0,55	0,65	0,75	0,85	0,95
$ F_n^*(x) - F(x) $	0,037	0,065	0,042	0,022	0,007	0,003	0,028	0,06	0,045	0,01

$\lambda_n = \sqrt{200 \cdot 0.065} = 0.919$; $\lambda_{\alpha} = 1.358$. Оскільки $\lambda_n < \lambda_{\alpha}$, то гіпотеза H_0 приймається.

Задача 2. За спостереженнями, наведеними в таблиці, за допомогою критерію χ^2 перевірити гіпотезу про нормальний розподіл випадкової величини ($\alpha = 0,05$).

Інтервал	$[-4;0)$	$[0;2)$	$[2;4)$	$[4;6]$
m_i	20	40	30	10

Розв'язання. Обчислимо оцінки параметрів нормального розподілу за вибіркою:

$$\bar{x} = \frac{1}{100}(-40 + 40 + 90 + 50) = 1,4;$$

$$S^2 = \frac{1}{99} \cdot ((-2 - 1,4)^2 \cdot 20 + (1 - 1,4)^2 \cdot 40 + (3 - 1,4)^2 \cdot 30 + (5 - 1,4)^2 \cdot 10) = 4,48;$$

$$S = 2,1.$$

Тепер перейдемо до знаходження ймовірностей p_i , $i = \overline{1,4}$.

$$p_1 = P\{-4 \leq \xi < 0\} = P\left\{\frac{-4 - 1,4}{2,1} \leq \frac{\xi - a}{\sigma} \leq \frac{0 - 1,4}{2,1}\right\} = \\ = \Phi(2,57) - \Phi(0,66) = 0,4949 - 0,2455 = 0,2494.$$

$$p_2 = P\{0 \leq \xi < 2\} = P\left\{\frac{0 - 1,4}{2,1} \leq \frac{\xi - a}{\sigma} \leq \frac{2 - 1,4}{2,1}\right\} = \\ = \Phi(0,286) + \Phi(0,66) = 0,1140 + 0,2455 = 0,3595.$$

$$p_3 = P\{2 \leq \xi < 4\} = P\left\{\frac{2 - 1,4}{2,1} \leq \frac{\xi - a}{\sigma} \leq \frac{4 - 1,4}{2,1}\right\} = \\ = \Phi(1,23) - \Phi(0,286) = 0,3905 - 0,1140 = 0,2765.$$

$$p_4 = P\{4 \leq \xi < 6\} = P\left\{\frac{4 - 1,4}{2,1} \leq \frac{\xi - a}{\sigma} \leq \frac{6 - 1,4}{2,1}\right\} = \\ = \Phi(2,19) - \Phi(1,23) = 0,4855 - 0,3905 = 0,0950.$$

Далі результати запишемо у таблицю:

Інтервал	$[-4;0)$	$[0;2)$	$[2;4)$	$[4;6]$
m_i	20	40	30	10
p_i	0,2494	0,3595	0,2765	0,095
np_i	24,94	35,95	27,65	9,5
$(m - np_i)^2$	24,40	16,40	5,52	0,25
$\frac{(m_i - np_i)^2}{np_i}$	0,98	0,46	0,2	0,02

$\chi_n^2 = 1,66$. Кількість інтервалів $r = 4$, а кількість невідомих параметрів $l = 2$. Тоді число степенів свободи $k = r - l - 1 = 1$.

$\chi^2_{1;0.05} = 3.8$. Отже, $1,66 < 3.8$, гіпотеза приймається.

Задача 3. За допомогою критерію χ^2 перевірити гіпотезу однорідності двох вибірок ($\alpha = 0,05$).

x_i	1	2	3	4
v_{i1}	40	26	24	10
v_{i2}	30	20	30	20
v_{i0}	70	46	54	30

Розв'язання. За умовою задачі $n_1 = 100$, $n_2 = 100$, $n = 200$.

Тоді

$$\chi_n^2 = \frac{(40-35)^2}{35} + \frac{(26-23)^2}{23} + \frac{(24-27)^2}{27} + \frac{(10-15)^2}{15} + \frac{(30-35)^2}{35} + \\ + \frac{(20-23)^2}{23} + \frac{(30-27)^2}{27} + \frac{(20-15)^2}{15} = 6,2; \quad \chi^2_{3;0.05} = 7,8.$$

Отже, оскільки $6,2 < 7,8$, то гіпотеза однорідності приймається.

Задача 4. Проведено 300 спостережень одночасно над випадковими величинами ξ та η , які набувають значень 1; 2 і 1; 2; 3 відповідно. Кількість спостережуваних пар v_{ij} наведено в таблиці.

$\eta \backslash \xi$	1	2	3	v_{i0}
1	32	68	50	150
2	40	70	40	150
v_{oj}	72	138	90	300

Перевірити за допомогою критерію χ^2 , чи є незалежними випадкові величини ξ та η при рівні значущості 0,01.

Розв'язання. Знайдемо величини m_{ij} . Матриця (m_{ij}) , $i = \overline{1,2}$; $j = \overline{1,3}$ буде такою:

$$m_{ij} = \begin{pmatrix} 36 & 69 & 45 \\ 36 & 69 & 45 \end{pmatrix}, \text{ а матриця } ((v_{ij} - m_{ij})^2) = \begin{pmatrix} 16 & 1 & 25 \\ 16 & 1 & 25 \end{pmatrix}.$$

Далі знайдемо матрицю, елементами якої будуть величини $\frac{(v_{ij} - m_{ij})^2}{m_{ij}}$. Дістанемо $\begin{pmatrix} 0.44 & 0.014 & 0.55 \\ 0.44 & 0.014 & 0.55 \end{pmatrix}$. Підсумувавши елементи матриці, знаходимо, що $\chi_n^2 = 2,008$. Число степенів свободи $m = 2$, $\chi_{2;0.01}^2 = 9.2$. Оскільки $\chi_n^2 < \chi_{2;0.01}^2$, то гіпотеза незалежності приймається.

Лекція №7

ПЕРЕВІРКА ГІПОТЕЗ ПРО РІВНІСТЬ МАТЕМАТИЧНИХ СПОДІВАНЬ ТА ДИСПЕРСІЙ ДЛЯ НОРМАЛЬНИХ СУКУПНОСТЕЙ

Нехай ξ та η – дві незалежні випадкові величини, кожна з яких має нормальний розподіл $N(a, \sigma_1^2)$ та $N(a, \sigma_2^2)$. У результаті спостережень цих випадкових величин отримано дві вибірки $\zeta = (\xi_1, \xi_2, \dots, \xi_{n1})$ та $\gamma = (\eta_1, \eta_2, \dots, \eta_{n2})$.

Гіпотеза про рівність математичних сподівань при відомих дисперсіях. Необхідно перевірити гіпотезу $H_0: a_1 = a_2$ проти альтернативної гіпотези $H_1: |a_1 - a_2| > 0$, σ_1^2 та σ_2^2 – відомі. Критична множина задається нерівністю:

$$R_{n1} = \left\{ (X, Y) : \frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \geq C_\alpha \right\},$$

$X = (x_1, x_2, \dots, x_{n1})$, $Y = (y_1, y_2, \dots, y_{n2})$ – реалізація вибірок ζ та γ , $\bar{x} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i$, $\bar{y} = \frac{1}{n_2} \sum_{i=1}^{n_2} y_i$, α – похибка першого роду, тобто ймовірність прийняти гіпотезу H_1 , коли правильна H_0 , $2\Phi(C_\alpha) = 1 - \alpha$.

Якщо $(X, Y) \in R_{n1}$, то приймається гіпотеза H_1 , якщо ж $(X, Y) \notin R_{n1}$, то приймається гіпотеза H_0 .

Гіпотеза про рівність математичних сподівань при невідомих дисперсіях. Нехай потрібно перевірити такі ж гіпотези, як і в

попередньому випадку, але $\sigma_1^2 = \sigma_2^2 = \sigma^2$ і величина σ^2 невідома. Критична множина задається нерівністю:

$$R_{n1} = \left\{ (X, Y) : \frac{|\bar{x} - \bar{y}|}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \frac{(n_1 - 1)\hat{S}_1^2 + (n_2 - 1)\hat{S}_2^2}{n_1 + n_2 - 2}}} \geq t_{n1+n2-2; \alpha} \right\},$$

$$\text{де } \hat{S}_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2, \quad \hat{S}_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2.$$

Число $t_{n1+n2-2; \alpha}$ знаходиться за таблицею Стюдента (табл. дод. 2) при числі степенів свободи $n_1 + n_2 - 2$ і:

$$P\{|t| < t_{n1+n2-2; \alpha}\} = 1 - \alpha.$$

Гіпотеза про рівність дисперсій при невідомих математичних сподіваннях. Нехай тепер потрібно перевірити гіпотезу $H_0: \sigma_1^2 = \sigma_2^2$ проти альтернативної гіпотези $H_1: \sigma_1^2 > \sigma_2^2$. При справедливості гіпотези H_0 випадкова величина $F = \frac{\hat{S}_1^2}{\hat{S}_2^2}$ має розподіл Фішера–Снедекора з $(n_1 - 1, n_2 - 1)$ степенями свободи. Тоді критична множина задається нерівністю:

$$R_{n1} = \left\{ (X, Y) : \frac{\hat{S}_1^2}{\hat{S}_2^2} \geq F_{(\alpha; k_1; k_2)} \right\},$$

$\hat{S}_1^2 > \hat{S}_2^2$, чого завжди можна досягти, змінивши індекси; $k_1 = n_1 - 1$, $k_2 = n_2 - 1$. Величина $F_{(\alpha; k_1; k_2)}$ знаходиться за таблицею розподілу Фішера–Снедекора (табл. дод. 3).

Гіпотеза про рівність дисперсій при відомих математичних сподіваннях. Ця гіпотеза перевіряється аналогічно попередній, але в даному випадку $F = \frac{S_1^2}{S_2^2}$, $S_1^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (\xi_i - a_1)^2$, $S_2^2 = \frac{1}{n_2} \sum_{i=1}^{n_2} (\eta_i - a_2)^2$. Якщо правильна гіпотеза $H_0: \sigma_1^2 = \sigma_2^2$, то випадкова величина F має розподіл Фішера–Снедекора зі (n_1, n_2) степенями свободи. Критична множина задається нерівністю:

$$R_{n1} = \left\{ (X, Y) : \frac{S_1^2}{S_2^2} \geq F_{(\alpha; n_1; n_2)} \right\}.$$

Приклади розв'язання задач

Задача 1. За двома вибірками об'єму $n_1 = 25, n_2 = 50$ з генеральних сукупностей випадкових величин ξ та η , які мають нормальний розподіл $N(a, \sigma_1^2)$ та $N(a, \sigma_2^2)$, визначено $\bar{x} = 9,79, \bar{y} = 9,60$. Перевірити гіпотезу $H_0: a_1 = a_2$ при $\sigma_1 = \sigma_2 = 0,3, \alpha = 0,01$.

Розв'язання. Маємо:

$$C = \frac{|\bar{x} - \bar{y}|}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{9,79 - 9,60}{0,3 \sqrt{\frac{1}{25} + \frac{1}{50}}} = 2,59.$$

Порівняємо значення C з $C_\alpha = 2,57$. Оскільки $2,59 > 2,57$, то гіпотезу про рівність математичних сподівань відхиляємо.

Задача 2. За двома вибірками об'єму $n_1 = 10, n_2 = 15$ з генеральних сукупностей випадкових величин ξ та η , які мають нормальний розподіл, визначено вибіркові дисперсії $\hat{S}_1^2 = 9,6$ і $\hat{S}_2^2 = 5,7$. Перевірити гіпотезу про рівність дисперсій випадкових величин ξ та $\eta, \alpha = 0,05$.

Розв'язання. Обчислимо:

$$F = \frac{\hat{S}_1^2}{\hat{S}_2^2} = \frac{9,6}{5,7} = 1,68, k_1 = 9, k_2 = 14.$$

Порівняємо F з $F_{(0,05; 9; 14)} = 2,65$. Маємо $1,68 < 2,65$. Тоді гіпотеза про рівність дисперсій ξ та η приймається.

Лекція №8

ЛІНІЙНА РЕГРЕСІЯ

Нехай ми маємо теоретичну залежність між величинами x та y у вигляді:

$$y = \varphi(x),$$

де $\varphi(x)$ – деяка функція. Проте в експерименті внаслідок можливих похибок вимірювань або невизначеностей у самому об'єкті в точці x_i спостерігається величина:

$$y_i = \varphi(x_i) + \delta_i,$$

де δ_i – деякі випадкові величини.

Треба за спостереженнями пар (x_i, y_i) , $i = \overline{1, n}$ зробити статистичні висновки щодо функції $\varphi(x)$.

Регресія. Нехай ξ та η – дві випадкові величини (залежні у загальному випадку), і ми хочемо знайти найкраще в деякому розумінні наближення величини η деякою функцією $g(\xi)$ від величини ξ .

Під найкращим наближенням будемо розуміти наближення в середньому квадратичному.

Означення. Величина $g(\xi)$ називається **найліпшим наближенням** величини η в середньому квадратичному, якщо:

$$M(\eta - g(\xi))^2 = \min M(\eta - \psi(\xi))^2.$$

Функція $g(\xi)$ є середньоквадратичною регресією величини η на величину ξ . Часто функцію $g(x) = M(\eta / \xi = x)$ називають **кореляційною залежністю** випадкових величин ξ та η , або **функцією регресії** випадкової величини η відносно ξ .

Лінійна регресія. Розглянемо регресію в класі лінійних функцій, тобто припустимо, що:

$$g(\xi) = \alpha\xi + \beta,$$

де α і β – невідомі параметри.

Введемо такі позначення:

$$m_1 = M(\xi), m_2 = M(\eta), \sigma_1^2 = D(\xi), \sigma_2^2 = D(\eta),$$

$$\mu = M(\xi - m_1)(\eta - m_2), \rho = \frac{\mu}{\sigma_1 \sigma_2},$$

тобто ρ – коефіцієнт кореляції.

Лінійна середньоквадратична регресія $g(\xi)$ величини η на величину ξ має вигляд:

$$g(\xi) = m_2 + \rho \frac{\sigma_2}{\sigma_1} (\xi - m_1).$$

Повернемося тепер до сформульованої на початку задачі про найліпше визначення функції $\varphi(x)$. Будемо вважати, що функція $\varphi(x)$ належить деякій параметричній сукупності функцій $\varphi(x, \alpha_1, \alpha_2, \dots, \alpha_k)$, і ми маємо спостереження (x_i, y_i) , $i = \overline{1, n}$.

Означення. Оцінкою невідомих параметрів $\alpha_1, \alpha_2, \dots, \alpha_k$ за методом найменших квадратів буде вектор $\alpha_1, \alpha_2, \dots, \alpha_k$ при якому досягається мінімум функції:

$$\Phi(\alpha_1, \alpha_2, \dots, \alpha_k) = \sum_{i=1}^n \left(y_i - \phi(\alpha_1, \alpha_2, \dots, \alpha_k) \right)^2,$$

а функція $\phi(\alpha_1, \alpha_2, \dots, \alpha_k)$ буде найліпшим середньоквадратичним наближенням, що відновлює залежність між x та y за результатами наших спостережень.

Якщо функція $\phi(x, \alpha_1, \alpha_2, \dots, \alpha_k)$ диференційована за аргументами $\alpha_1, \alpha_2, \dots, \alpha_k$, для знаходження величин $\alpha_1, \alpha_2, \dots, \alpha_k$ одержимо систему рівнянь:

$$\sum_{i=1}^n \frac{\partial \phi(x_i, \alpha_1, \alpha_2, \dots, \alpha_k)}{\partial \alpha_j} (y_i - \phi(x_i, \alpha_1, \alpha_2, \dots, \alpha_k)) = 0, j = \overline{1, k}.$$

Розглянемо важливий випадок, коли функція $\phi(x)$ має вигляд $\phi(x) = kx + b$, де k і b – невідомі параметри. У цьому випадку оцінками параметрів лінійної регресії будуть числа k і b , при яких функція $\Phi(k, b) = \sum_{i=1}^n (y_i - kx_i - b)^2$ досягається мінімуму.

Тоді:

$$k = \frac{\sum_{i=1}^n y_i x_i - n \bar{y} \bar{x}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}, b = \bar{y} - k \bar{x},$$

$$\text{де } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

У випадку поліноміальної регресії:

$$\phi(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \dots + \alpha_m x^m,$$

оцінки невідомих параметрів $\alpha_0, \alpha_1, \dots, \alpha_m$ знаходяться з системи лінійних алгебраїчних рівнянь:

$$\sum_{j=0}^m S_{k+j} \alpha_j = Z_k, k = \overline{0, m},$$

$$\text{де } S_i = \sum_{i=1}^n x_i^j, Z_k = \sum_{i=1}^n y_i x_i^k, k = \overline{0, m}, j = \overline{0, 2m}.$$

Якщо значення x_i відомі без похибок, а значення y_i незалежні та рівноточні, то оцінка дисперсії (похибка вимірювань) величини y_i визначається формулою:

$$\sigma^2 = \frac{\Phi_{min}}{n-m-1}, \text{ де } \Phi_{min} = \sum_{i=1}^n \left(y_i - \sum_{j=0}^m \alpha_j x_j \right)^2.$$

Оцінки σ_k^2 дисперсій коефіцієнтів α_k визначаються за формулами:

$$\sigma_{\alpha_k}^2 = \frac{\Delta_{kk} \sigma^2}{\Delta},$$

де Δ – визначник системи, а Δ_{kk} – алгебраїчне доповнення до елемента, який стоїть на діагоналі й має індекс k у визначнику Δ . У випадку лінійної регресії

$$\sigma_{\alpha_0}^2 = \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \frac{\Phi_{min}}{n-2},$$

$$\sigma_{\alpha_1}^2 = \frac{n}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \frac{\Phi_{min}}{n-2}.$$

Якщо величини y_i мають нормальний розподіл, то для коефіцієнтів α_k справджуються такі надійні інтервали:

$$\alpha_k - t_{n-m-1; \alpha} \sigma_{\alpha_k} < \alpha_k < \alpha_k + t_{n-m-1; \alpha} \sigma_{\alpha_k},$$

де α_k – оцінки, отримані методом найменших квадратів, а число $t_{n-m-1; \alpha}$ знаходиться за таблицею розподілу Стюдента (табл. дод. 2) при числі степенів свободи $k = n - m - 1$ і $P\{|t| < t_{n-m-1; \alpha}\} = 1 - \alpha$.

Вибірковий коефіцієнт кореляції визначається за формулою:

$$\hat{r} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}.$$

У деяких випадках функція $\varphi(x)$, яка не є многочленом, може бути зведена до нього заміною змінних. Приклади такої заміни наведені в таблиці.

№	Початкова	До якого вигляду	Замінна змінної
---	-----------	------------------	-----------------

	функція	приводиться	
1	$y = Ae^{kx}$	$Z = \alpha_0 + \alpha_1 x$	$Z = \ln y, \alpha_0 = \ln A, \alpha_1 = k$
2	$y = Bx^\beta$	$Z = \alpha_0 + \alpha_1 u$	$Z = \ln y, u = \ln x, \alpha_0 = \ln B, \alpha_1 = \beta$
3	$y = \alpha_0 + \frac{\alpha_1}{x}$	$y = \alpha_0 + \alpha_1 u$	$u = \frac{1}{x}$
4	$y = \alpha_0 + \frac{\alpha_1}{x^\beta}$	$y = \alpha_0 + \alpha_1 u$	$u = \frac{1}{x^\beta}$

Коефіцієнт кореляції рангів. У деяких випадках натрапляємо на ознаки, які не піддаються кількісним оцінкам. Тоді кожній оцінці можна поставити у відповідність порядковий номер, який назвемо **рангом**. Нехай n осіб за якістю A мають ранги X_1, X_2, \dots, X_n , а за якістю B – Y_1, Y_2, \dots, Y_n , де всі X та Y є перестановками n перших чисел натурального ряду. $d_k = X_k - Y_k$ – різниця рангів.

Тоді **коефіцієнт кореляції рангів Спірмена**, або **коефіцієнт щільності зв'язку** між A та B , визначається за формулою:

$$\rho = 1 - \frac{6 \sum_{k=1}^n d_k^2}{n^3 - n}.$$

Є й інші показники щільності зв'язку між рангами. Якщо не можна визначити рангову відмінність декількох осіб, то беруть середній ранг. У цьому випадку використовують **коефіцієнт кореляції рангів Кендела**:

$$\rho = \frac{\frac{n^3 - n}{6} - (T_x + T_y) - \sum_{k=1}^n d_k^2}{\sqrt{\left(\frac{n^3 - n}{6} - 2T_x\right)\left(\frac{n^3 - n}{6} - 2T_y\right)}},$$

де $T_x = T_y = \frac{\sum_{i=1}^l (t_i^3 - t_i)}{12}$, t_i – число об'єднаних рангів для X та Y .

Приклади розв'язання задач

Задача 1. Припускаючи, що кореляційна залежність величин η та ξ лінійна ($\varphi(x) = kx + b$), за результатами спостережень, наведених у таблиці, знайти коефіцієнти k, b , похибку вимірювань, надійні інтервали для коефіцієнтів та вибірковий коефіцієнт кореляції, $1 - \alpha = 0,95$.

x_i	2	4	6	8	10
y_i	4	6	8	7	10

Розв'язання. Усі результати обчислень занесемо в таблицю ($n = 5$).

№	x_i	y_i	x_i^2	$x_i y_i$	y_i^*	$(y_i - y_i^*)^2$
1	2	4	4	8	4.4	0.16
2	4	6	16	24	5.7	0.09
3	6	8	36	48	7	1
4	8	7	64	56	8.3	1.69
5	10	10	100	100	9.6	0.16
Σ	30	35	220	236		3.1

$$\Delta = n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 = 5 \cdot 200 - 30^2 = 200,$$

$$\bar{x} = 6, \bar{y} = 7, \hat{k} = \frac{236 - 5 \cdot 42}{40} = 0,65,$$

$$\hat{b} = 7 - 0,65 \cdot 6 = 3,1, \hat{y}_i = 0,65x + 3,1, \Phi_{\min} = 3,1.$$

Тоді $\hat{\sigma}^2 = \frac{3,1}{5-2} = 1,03, \quad \hat{\sigma}_b^2 = \frac{220}{200} \cdot 1,03 = 1,133, \quad \hat{\sigma}_b = 1,064,$

$$\hat{\sigma}_k^2 = \frac{5}{200} \cdot 1,03 = 0,02575, \quad \hat{\sigma}_k = 0,16.$$

З таблиці розподілу Стюдента (табл. дод. 2) знаходимо $t_{3;0,05} = 3,18$.
Тоді з імовірністю 0,95:

$$3,1 - 3,18 \cdot 1,064 < b < 3,1 + 3,18 \cdot 1,064,$$

тобто $-0,283 < b < 6,483$ і:

$$0,65 - 3,18 \cdot 0,16 < k < 0,65 + 3,18 \cdot 0,16,$$

тобто $0,141 < k < 1,159$.

Для визначення вибіркового коефіцієнта кореляції знайдемо суми:

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \frac{\Delta}{n} = 40;$$

$$\sum_{i=1}^n (y_i - \bar{y})^2 = (4-7)^2 + (6-7)^2 + (10-7)^2 = 20;$$

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} = 236 - 5 \cdot 42 = 26.$$

Тоді $\rho = \frac{26}{\sqrt{20 \cdot 40}} = \frac{13}{10\sqrt{2}} = 0,919.$

Задача 2. Припускаючи, що кореляційна залежність величин η та ξ експоненціальна, тобто $\varphi(x) = Ae^{-kx}$, за результатами спостережень, наведених у таблиці, знайти коефіцієнти A та k .

x_i	0	1	2	3	4	5	6	7	8	9	10
y_i	100	75	55	40	30	20	15	10	10	5	5

Розв'язання. Усі результати обчислень занесемо в таблицю ($z_i = \ln y_i$):

№	x_i	y_i	z_i	x_i^2	$x_i z_i$	\hat{z}_i
1	0	100	4.6	0	0	4.59
2	1	75	4.3	1	4.3	4.28
3	2	55	4	4	8	3.97
4	3	40	3.7	9	11.1	3.66
5	4	30	3.4	16	13.6	3.35
6	5	20	3	25	15	3.04
7	6	15	2.7	36	16.2	2.73
8	7	10	2.3	49	16.1	2.42
9	8	10	2.3	64	18.4	2.11
10	9	5	1.6	81	14.4	1.80
11	10	5	1.6	100	16	1.49
Σ	55		33.5	385	133.1	

$$n = 11, \bar{x} = 5, \bar{z} = 3.04, \alpha_1 = -\hat{k} = \frac{133.1 - 11 \cdot 5 \cdot 3.04}{385 - 11 \cdot 25} = -0.31,$$

$$\alpha_0 = \ln \hat{A} = 3.04 + 0.31 \cdot 5 = 4.59.$$

Отже, $\hat{k} = 0.31$ і $\hat{A} = e^{\alpha_0} = 98.5$.

Тоді $\varphi(x) = 98.5 e^{-0.31x}$.

Задача 3. На змаганнях з фігурного катання судді розподілили місця між учасниками змагань таким чином:

Учасники	A	B	C	D	E	F	G	H	I	J
1 суддя	1.5	1.5	3	4	6	6	6	8	9.5	9.5
2 суддя	1	2	4	4	4	6	7	8	9	10

Встановити, наскільки об'єктивні оцінки суддів, тобто наскільки тісний зв'язок між оцінками.

Розв'язок. Перший суддя поділив перше місце між учасниками А та В. Їхній об'єднаний ранг $\frac{1+2}{2} = 1.5$. Учасники D, E, G поділили 5, 6, 7 місця. Їхній об'єднаний ранг дорівнює 6 і т.д. Обчислимо величини T_x, T_y . При обчисленні T_x маємо: А та В – два об'єднаних ранги, D, E, G – три об'єднаних ранги й I, J – два об'єднаних ранги. Таким чином:

$$T_x = \frac{(2^3 - 2) + (3^3 - 3) + (2^3 - 2)}{12} = 3, \quad T_y = \frac{(3^3 - 3)}{12} = 2;$$

$$\hat{r} = \frac{\frac{(1000 - 10)}{6} - (3 + 2) - 7}{\sqrt{\left(\frac{(1000 - 10)}{6} - 6\right)\left(\frac{(1000 - 10)}{6} - 4\right)}} = 0,956.$$

Звідси можна зробити висновок, що оцінки, поставлені суддями учасникам змагань, об'єктивні.

Лекція №9

ЕЛЕМЕНТИ ДИСПЕРСІЙНОГО АНАЛІЗУ

Дисперсійний аналіз – це статистичний метод аналізу результатів спостережень, які залежать від різних, одночасно діючих факторів, вибір найбільш важливих із них і оцінка їхнього впливу.

Однофакторний дисперсійний аналіз. В основі однофакторного аналізу лежить теоретично-ймовірнісна схема:

$$x_{ij} = a + \alpha_i + \varepsilon_{ij}, \quad j = \overline{1, n_i}, \quad i = \overline{1, m}, \quad n = n_1 + n_2 + \dots + n_m,$$

де x_{ij} – спостереження величини, яку досліджують;

n_i – число спостережень при i -му значенні фактора;

a – загальне середнє;

α_i – ефект фактора;

ε_{ij} – похибка спостережень, незалежні випадкові величини, що мають нормальний розподіл $N(0, \sigma^2)$, тобто $M(\varepsilon_{ij}) = 0$, $D(\varepsilon_{ij}) = \sigma^2$.

Потрібно перевірити гіпотезу:

$H_0 : \alpha_i = 0, \quad i = \overline{1, m}$, тобто фактор не впливає на результат спостережень.

Нехай $\bar{x} = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij}$ – загальне середнє,

$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$ – групове середнє,

$$Q = \sum_{i=1}^m \sum_{j=1}^{n_j} (x_{ij} - \bar{x})^2 \quad \text{— сума квадратів відхилень спостережень від}$$

загального середнього,

$$Q = Q_1 + Q_2, \quad \text{де } Q_1 = \sum_{i=1}^m n_i (\bar{x}_i - \bar{x})^2 \quad \text{— сума квадратів відхилень між}$$

групами,

$$Q_2 = \sum_{i=1}^m \sum_{j=1}^{n_i} (\bar{x}_i - \bar{x})^2 \quad \text{— сума квадратів відхилень всередині групи.}$$

$$S_1^2 = \frac{Q_1}{m-1}, \quad S_2^2 = \frac{Q_2}{n-m}, \quad S^2 = \frac{Q}{n-1} \quad \text{— незміщена оцінка } \sigma^2.$$

$$\text{Статистика критерію: } F = \frac{S_1^2}{S_2^2}.$$

Для перевірки гіпотези H_0 маємо критерій: якщо $F \geq F_{(\alpha, k_1, k_2)}$, то гіпотеза H_0 відхиляється, а при $F < F_{(\alpha, k_1, k_2)}$ приймається. Величина $F_{(\alpha, k_1, k_2)}$ знаходиться за таблицею розподілу Фішера–Снедекора (табл. дод. 3).

$$k_1 = m - 1, k_2 = n - m.$$

Двофакторний дисперсійний аналіз

$$x_{ijs} = a + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijs}, \quad i = \overline{1, m}, \quad j = \overline{1, k}, \quad s = \overline{1, n_{ij}},$$

m — кількість рівнів фактора А;

k — кількість рівнів фактора В;

α_i — ефект фактора А;

β_j — ефект фактора В;

γ_{ij} — ефект взаємодії факторів.

Якщо всі $n_{ij} = 1$, тобто при кожному рівні фактора А і фактора В маємо лише одне спостереження, то:

$$x_{ij} = a + \alpha_i + \beta_j + \varepsilon_{ij}.$$

Розглянемо цю модель:

$$\bar{x}_{i*} = \frac{1}{k} \sum_{j=1}^k x_{ij}, \quad \bar{x}_{*j} = \frac{1}{m} \sum_{i=1}^m x_{ij}, \quad \bar{x} = \frac{1}{km} \sum_{i=1}^m \sum_{j=1}^k x_{ij}.$$

$$Q = \sum_{i=1}^m \sum_{j=1}^k (x_{ij} - \bar{x})^2 = Q_1 + Q_2 + Q_3,$$

де

$Q_1 = k \sum_{i=1}^m (\bar{x}_{i*} - \bar{x})^2$ – сума квадратів відхилень за фактором А,

$Q_2 = m \sum_{j=1}^k (\bar{x}_{*j} - \bar{x})^2$ – сума квадратів відхилень за фактором В,

$Q_3 = \sum_{i=1}^m \sum_{j=1}^k (x_{ij} - \bar{x}_{i*} - \bar{x}_{*j} + \bar{x})^2$ – залишкова сума квадратів.

$$H_0^A : \alpha_i = 0, i = \overline{1, m},$$

$$H_0^B : \beta_j = 0, j = \overline{1, k}.$$

$$S_1^2 = \frac{Q_1}{m-1}, S_2^2 = \frac{Q_2}{k-1}, S_3^2 = \frac{Q_3}{(k-1)(m-1)}, S^2 = \frac{Q}{km-1} \text{ – незміщена оцінка } \sigma^2.$$

Критерій перевірки гіпотези H_0^A :

якщо

$$F_A = \frac{S_1^2}{S_3^2} \geq F_{(\alpha, k_1, k_2)},$$

то гіпотеза H_0^A відхиляється, якщо $F_A < F_{(\alpha, k_1, k_2)}$ – приймається, за умови, що: $k_1 = m - 1, k_2 = (k - 1)(m - 1)$.

Аналогічно H_0^B відхиляється, якщо $F_B = \frac{S_2^2}{S_3^2} \geq F_{(\alpha, k_1, k_2)}$, приймається

якщо $F_B < F_{(\alpha, k_1, k_2)}$, за умови, що:

$$k_1 = k - 1, k_2 = (k - 1)(m - 1).$$

Приклад розв'язання задачі

Задача. Перевірити гіпотези значущості факторів для двофакторного комплексу з одним спостереженням, результати якого подано в таблиці. Рівень значущості $\alpha = 0,95$.

	B ₁	B ₂	B ₃	\bar{x}_{i*}
A ₁	1	2	3	2
A ₂	5	6	10	7
\bar{x}_{*j}	3	4	6,5	4,5

Розв'язання. $\bar{x}_{1*} = \frac{1+2+3}{3} = 2$, загальне середнє $\bar{x} = 4,5$.

Тоді

$$Q_1 = 3((2 - 4,5)^2 + (7 - 4,5)^2) = 3((-2,5)^2 + 2,5^2) = 37,5,$$

$$Q_2 = 2((3 - 4,5)^2 + (4 - 4,5)^2 + (6,5 - 4,5)^2) = 13,$$

$$Q_3 = (1 - 2 - 3 + 4,5)^2 + (2 - 2 - 4 + 4,5)^2 + (3 - 2 - 6,5 + 4,5)^2 + \\ + (3 - 2 - 6,5 + 4,5)^2 + (5 - 7 - 3 + 4,5)^2 + \\ + (6 - 7 - 4 + 4,5)^2 + (10 - 7 - 6,5 + 4,5)^2 = 3, \\ Q = 37,5 + 13,0 + 3,0 = 53,5$$

$$S_1^2 = \frac{37,5}{1} = 37,5, S_2^2 = \frac{13}{2} = 6,5, S_3^2 = \frac{3}{1 \cdot 2} = 1,5.$$

Отримані дані можна записати у вигляді таблиці:

Компонента дисперсій	Сума квадратів	Число степенів свободи	Оцінка дисперсії
Між середніми в рядках (фактор А)	37,5	1	37,5
Між середніми в стовпчиках (фактор В)	13,0	2	6,5
Залишкова	3,0	2	1,5
Повна	53,5	5	10,7

Обчислимо F_A та F_B :

$$F_A = \frac{S_1^2}{S_3^2} = \frac{37,5}{1,5} = 25,$$

$$F_B = \frac{S_2^2}{S_3^2} = \frac{6,5}{1,5} = 4,3.$$

Для рівня значущості $\alpha = 0,05$ та $k_2 = 2$, $k_1 = 1$ степенів свободи за таблицею Фішера–Снедекора (табл. дод. 3) знаходимо:

$$F_{\alpha(A)} = 18,51, F_A = 25,$$

$$F_{\alpha(B)} = 19,0, F_B = 4,3.$$

Порівнюючи табличні значення з обчисленими, маємо:

$$F_A > F_{\alpha(A)}, F_B < F_{\alpha(B)}.$$

Отримані результати дають змогу дійти висновку: нульова гіпотеза про рівність середніх у рядках не підтверджується, тобто вплив фактора А на досліджувану ознаку значний; нульова гіпотеза про рівність середніх у стовпчиках не заперечується, тобто вплив фактора В на досліджувану ознаку незначний.

Список літератури

1. *Міхайленко В.М.* Теорія ймовірностей, ймовірнісні процеси та математична статистика: навч. посіб. / В.М. Міхайленко, С.А. Теренчук, О.О. Кубайчук. – К.: Вид-во Європейського університету, 2007. – 163 с.
2. *Очинников П.Ф.* Высшая математика: учебн. пособие/ П.Ф. Очинников, В.М. Михайленко, Б.М. Лисицын. – К.: Вища шк., 1989. – 679 с.
3. *Федоренко Н.Д.* Теорія ймовірностей та математична статистика: навч. посіб. / Н.Д. Федоренко, О.І. Баліна, І.С. Безклубенко та ін. – К.: КНУБА, 2007. – 104 с.
4. *Барковський В.В.* Теорія ймовірностей та математична статистика: навч. посіб. / В.В. Барковський, Н.В. Барковська, О.К. Лопатін. – К.:ЦУЛ, 2002. – 448 с.
5. *Міхайленко В.М.* Теорія ймовірностей, ймовірнісні процеси та математична статистика. Збірник задач: навч. посібник / В.М. Міхайленко, С.А. Теренчук, О.О. Кубайчук. – К.: Вид-во Європейського університету, 2007. – 116 с.
6. *Федоренко Н.Д.* Теорія ймовірностей і математична статистика. Модуль І. Теорія ймовірностей: методичні вказівки та завдання для самостійної роботи / Н.Д. Федоренко, С.А. Теренчук, О.В. Доля. – К.: КНУБА, 2008. - 48 с.
7. *Кулинич Г.Л.* Вища математика: підручник: У 2 кн. – 2-ге вид., перер. і доп. / Г.Л. Кулинич, Є.Ю. Таран, В.М. Бурим та ін. – К.: Либідь, 2003. – 368 с.
8. *Андрухаев Ч.М.* Сборник задач по теории вероятностей. – М.: Просвещение, 1985.-160 с.
9. *Солодовников А.С.* Теория вероятностей: учебник / А.С. Солодовников. – М.: Просвещение, 1983.-207с.

ДОДАТКИ

Додаток 1

Значення $\chi^2_{k,\alpha}$ залежно від імовірності $P\{\chi^2(k) \geq \chi^2_{k,\alpha}\} = \alpha$ і числа степенів свободи k . Щільність розподілу $\chi^2(k)$ дорівнює:

$$P_{\chi^2(k)}(x) = \frac{x^{\frac{k}{2}-1} e^{-\frac{x}{2}}}{\Gamma\left(\frac{k}{2}\right) 2^{\frac{k}{2}}}, x > 0$$

$k \backslash 1 - \alpha$	0,01	0,05	0,10	0,90	0,95	0,99
1	0,00016	0,0039	0,016	2,7	3,8	6,6
2	0,020	0,103	0,211	4,6	6,0	9,2
3	1,115	0,352	0,584	6,3	7,8	11,3
4	0,30	0,71	1,06	7,8	9,5	13,3
5	0,55	1,14	1,61	9,2	11,1	15,1
6	1,187	1,63	2,20	10,6	12,6	16,8
7	1,24	2,17	2,83	12,0	14,1	18,5
8	1,65	2,73	3,49	13,4	15,5	20,1
9	2,09	3,32	4,17	14,7	16,9	21,7
10	2,56	3,94	4,86	16,0	18,3	23,2
11	3,1	4,6	5,6	17,3	19,7	24,7
12	3,6	5,2	6,3	18,5	21,0	26,2
13	4,1	5,9	7,0	19,8	22,4	27,7
14	4,7	6,6	7,8	21,1	23,7	29,1
15	5,2	7,3	8,5	22,3	25,0	30,6
16	5,8	8,0	9,3	23,5	26,3	32,0
17	6,4	8,7	10,1	24,8	27,6	33,4
18	7,0	9,4	10,9	26,0	28,9	34,8
19	7,6	10,1	11,7	27,2	30,1	36,2
20	8,3	10,9	12,4	28,4	31,4	37,6
21	8,9	11,6	13,2	29,6	32,7	38,9
22	9,5	12,3	14,0	30,8	33,9	40,3
23	10,2	13,1	14,8	32,0	35,2	41,6
24	10,9	13,8	15,7	33,2	36,4	43,0
25	11,5	14,6	16,5	34,4	37,7	44,3
26	12,2	15,4	17,3	35,6	38,8	45,6
27	12,9	16,2	18,1	36,7	40,1	47,0
28	13,6	16,9	18,9	37,9	41,3	48,3
29	14,3	17,7	19,8	39,1	42,6	49,6
30	15,0	18,5	20,6	40,3	43,8	50,9

Значення t_α для розподілу Стюдента залежно від імовірності $P\{|t_k| < t_\alpha\} = 1 - \alpha$ і числа степенів свободи k . Щільність розподілу

дорівнює:
$$P_{t_k}(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

$k \backslash 1 - \alpha$	0,90	0,95	0,99
1	6,31	12,71	63,7
2	2,92	4,30	9,92
3	2,35	3,18	5,84
4	2,13	2,77	4,60
5	2,02	2,57	4,03
6	1,943	2,45	3,71
7	1,895	2,36	3,50
8	1,860	2,31	3,36
9	1,833	2,26	3,25
10	1,812	2,23	3,17
11	1,796	2,20	3,11
12	1,782	2,18	3,06
13	1,771	2,16	3,01
14	1,761	2,14	2,98
15	1,753	2,13	2,95
16	1,746	2,12	2,92
17	1,740	2,11	2,90
18	1,734	2,10	2,88
19	1,729	2,09	2,86
20	1,725	2,09	2,84
25	1,708	2,06	2,79
30	1,697	2,04	2,46
80	1,659	1,991	2,640
100	1,651	1,984	2,627
∞	1,645	1,960	2,576

Розподіл Фішера–Снедекора. Значення $F_{(\alpha, k_1, k_2)}$ залежно від імовірності $P\{F_{k_1, k_2} \geq F_{(\alpha, k_1, k_2)}\} = \alpha$ і числа степенів свободи (k_1, k_2) і щільність розподілу F_{k_1, k_2} дорівнює:

$$p(x) = \frac{\Gamma\left(\frac{k_1 + k_2}{2}\right)}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)} \left(\frac{k_1}{k_2}\right)^{\frac{k_1}{2}} \frac{x^{\frac{k_1}{2}-1}}{\left(\frac{k_1 x}{k_2} + 1\right)^{\frac{k_1 + k_2}{2}}}, x > 0, \alpha = 0.05$$

$k_1 \backslash k_2$	1	2	3	4	5	9	10	20	50	100	∞
1	161	200	216	225	230	241	242	248	252	253	254
2	18,51	19,00	19,16	19,25	19,30	19,38	19,39	19,44	19,47	19,49	19,50
3	10,13	9,55	9,28	9,12	9,01	8,81	8,78	8,66	8,58	8,56	8,53
4	7,71	6,94	6,59	6,39	6,26	6,00	5,96	5,80	5,70	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,78	4,74	4,56	4,44	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,10	4,06	3,87	3,75	3,71	3,67
7	5,59	4,74	4,35	4,12	3,97	3,68	3,63	3,44	3,32	3,28	3,23
8	5,32	4,46	4,07	3,84	3,69	3,39	3,34	3,15	3,03	2,98	2,93
9	5,12	4,26	3,86	3,63	3,48	3,18	3,13	2,93	2,80	2,76	2,71
10	4,96	4,10	3,71	3,48	3,33	3,02	2,97	2,77	2,64	2,59	2,54
11	4,84	3,98	3,56	3,36	3,20	2,90	2,86	2,65	2,50	2,45	2,40
12	4,75	3,88	3,49	3,26	3,11	2,80	2,76	2,54	2,40	2,25	2,30
14	4,60	3,74	3,34	3,11	2,96	2,65	2,60	2,39	2,24	2,19	2,13
16	4,49	3,63	3,24	3,01	2,85	2,54	2,49	2,28	2,13	2,07	2,01
20	4,35	3,49	3,10	2,87	2,71	2,40	2,35	2,12	1,96	1,90	1,84
30	4,17	3,32	2,92	2,69	2,53	2,21	2,16	1,93	1,76	1,69	1,62
50	4,03	3,18	2,79	2,56	2,40	2,07	2,02	1,78	1,60	1,52	1,44
100	3,94	3,09	2,70	2,46	2,30	1,97	1,92	1,68	1,48	1,39	1,28
∞	3,84	2,99	2,60	2,37	2,21	1,88	1,83	1,57	1,35	1,24	1,00

Критичні значення λ_α для розподілу Колмогорова

α	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
----------	------	------	------	------	------	------	------	------	-------

λ_α	0,828	0,895	0,974	1,073	1,224	1,358	1,520	1,627	1,950
------------------	-------	-------	-------	-------	-------	-------	-------	-------	-------

Навчальне видання

ДОЛЯ Олена Вікторівна
ТЕРЕНЧУК Світлана Анатоліївна

ТЕОРІЯ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНА СТАТИСТИКА

Модуль II. Математична статистика

Конспект лекцій
 для студентів, які навчаються за напрямом підготовки
 6.030510 «Товарознавство і торговельне підприємництво»
 (усі форми навчання)

Редагування та коректура *А.О. Бакієвої*
 Комп'ютерне верстання *А.П. Морозюк*

Підписано до друку 2012. Формат 60х84
 Ум. друк. арк. 2,56. Обл.-вид. арк. 2,75,
 Тираж 50 прим. Вид. № 11/I-12 Зам. №

КНУБА, Повітрофлотський проспект, 31, Київ, Україна, 03680

E-mai: red_isdat@ua.fm

Видруковано в редакційно-видавничому відділі
 Київського національного університету будівництва і архітектури

Свідоцтво про внесення до Державного реєстру суб'єктів видавничої справи
ДК № 808 від 13.02.2002р.