МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ ПОЛІТЕХНІЧНИЙ УНІВЕРСИТЕТ
ІНСТИТУТ ЗАОЧНОЇ ОСВІТИ
КАФЕДРА «ПРИРОДНИЧО-НАУКОВОЇ ПІДГОТОВКИ»

ПОСІБНИК З МАТЕМАТИЧНОЇ СТАТИСТИКИ

Рекомендовано до друку вченою радою
Одеського національного політехнічного університету
Протокол № 2 від 13.01.2014 р.

Одеса 2015

Посібник з математичної статистики [Текст] / Одес. нац. політехн. ун-т; уклад.: Г. Ф. Сафонова, – Херсон: ХПТК ОНПУ, 2015. – 43 с.

Навчальне видання

ПОСІБНИК З МАТЕМАТИЧНОЇ СТАТИСТИКИ

Укладачі: **Сафонова** Ганна Феліксівна, доцент, канд. техн. наук

Рецензент: **Семакова** Тетяна Олексіївна, доцент, канд. пед. наук

За редакцією укладачів
Надруковано з оригінал-макета замовника

Підп. до друку 12.02.2014. Формат 600 х 840 М 1/16.
Папір офсетний. Ум. друк. арк. 2,62. Гарнітура Times.
Спосіб друку – ризографія. Тираж 4 прим. Зам. № 46.

Лабораторія організаційно-видавничої діяльності
ХПТК ОНПУ
73000, м. Херсон, вул. 40 років Жовтня, 23
тел. (0552) 22-55-38, тел./факс (0552) 22-27-43

ЗМІСТ

ПЕРЕДМОВА

Мета даного посібника – допомогти студентам самостійно з мінімальними витратами часу засвоїти основні розділи математичної статистики дисципліни «Теорія ймовірностей та математична статистика», які передбачені діючою програмою.

У навчальному виданні коротко викладено теоретичний матеріал з кожного розділу, надані приклади розв’язання типових задач.

Наведено список літератури, необхідний для більш детального вивчення розглянутих розділів.

В кінці методичних вказівок приведені таблиці значень функції:

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt\,,\ \ f\,(x,\lambda)=\frac{\lambda^x e^{-\lambda}}{x!}\,,$$

критичні точки *F*-розподілу Фішера-Снедекора при рівні значущості $\alpha=0,05$ та таблиця критичних точок розподілу χ^2.

1 ВИБІРКОВИЙ МЕТОД

1.1 Генеральна і вибіркова сукупності

Нехай потрібно вивчити сукупність однорідних об'єктів щодо деякої якісної або кількісної ознаки, яка характеризує ці об'єкти.

Іноді проводять суцільне обстеження, тобто обстежують кожний з об'єктів сукупності відносно ознаки, якою цікавляться. На практиці, проте суцільне обстеження застосовують порівняно рідко. Частіше випадково відбирають зі всієї сукупності обмежене число об'єктів і піддають їх вивченню.

Визначення. Вибірковою сукупністю або просто ***вибіркою*** називають сукупність випадково відібраних об'єктів.

Визначення. Генеральною сукупністю називають сукупність об'єктів, з яких проводиться вибірка.

Визначення. Об'ємом сукупності (вибірковою або генеральною) називають число об'єктів цієї сукупності.

Зауваження. Часто генеральна сукупність містить скінченне число об'єктів. Проте якщо це число достатньо велике, то іноді в цілях спрощення обчислень, або для полегшення теоретичних висновків, допускають, що генеральна сукупність складається з незліченної кількості об'єктів. Таке допущення виправдовується тим, що збільшення об'єму генеральної сукупності (достатньо великого об'єму) практично не позначається на результатах обробки даних вибірки.

1.2 Статистичний розподіл вибірки

Нехай з генеральної сукупності взята вибірка, причому x_1 спостерігалося n_1 разів, x_2 – n_2 разів, x_k – n_k разів, а $\sum n_i = n$ – об'єм вибірки. Спостережувані значення x, – називають варіантами, а послідовність варіант, записаних в зростаючому порядку, – варіаційним рядом. Числа спостережень називають частотами, а їх відношення до об'єму вибірки $\frac{n_i}{n} = w_i$ відносними частотами.

Визначення. Статистичним розподілом вибірки називають перелік варіант і відповідних їм частот або відносних частот. Статистичний розподіл можна задати також у вигляді послідовності інтервалів і відповідних їм частот (як частоту, яка відповідає інтервалу, приймають суму частот, що потрапили в цей інтервал).

Зауваження. В теорії ймовірності під ***розподілом*** розуміють співвідношення між можливими значеннями випадкової величини і їх ймовірностями, а в математичній статистиці співвідношення між спостережуваними варіантами та їх частотами, або відносними частотами.

1.3 Емпірична функція розподілу

Нехай відомий статистичний розподіл частот кількісної ознаки X. Введемо позначення: n_x – число спостережень, при яких спостерігалося значення ознаки, менше x; n – загальне число спостережень (об'єм вибірки). Відносна частота події $X < x$ дорівнює $\frac{n_x}{n}$. Якщо x змінюється, то змінюється і відносна частота, тобто відносна частота $\frac{n_x}{n}$ є функція від x. Оскільки ця функція знаходиться емпіричним (досвідним) шляхом, то її називають емпіричною.

Визначення. Емпіричною функцією розподілу (функцією розподіли вибірки) називають функцію $F^*(x)$, яка визначає для кожного значення x відносну частоту події $X < x$.

Отже, за визначенням

$$F^*(x) = \frac{n_x}{n}, \tag{1.1}$$

де n_x – число варіант, менших x; n – об'єм вибірки.

На відміну від емпіричної функції розподілу вибірки функцію розподілу $F\left(x\right)$ генеральної сукупності називають ***теоретичною функцією розподілу***. Відмінність між емпіричною і теоретичною функціями полягає в тому, що теоретична функція $F\left(x\right)$ визначає ймовірність $X < x$, а емпірична функція $F^*\left(x\right)$ визначає відносну частоту цієї ж події.

Емпірична функція має наступні ***властивості.***

1. Значення емпіричної функції належать відрізку [0; 1].

2. $F^*\left(x\right)$ – неспадна функція.

3. Якщо x_1 – найменша варіанта, а x_k – найбільша, то $F^*\left(x\right)=0$ при $x \le x_1$ і $F^*\left(x\right)=1$ при $x > x_k$.

1.4 Наочне зображення статистичних розподілів

Для наочного зображення статистичних розподілів використовують графіки та діаграми: ***полігон***, ***гістограму***, ***кумуляту***, ***огіву***.

Визначення. Полігон частот – многокутник (ламана), побудований в системі координат (x, n_i) або (x, w_i) (полігон частот або відносних частот). Для його побудови на осі абсцис відкладають варіанти x_i, а на осі ординат – відповідні їм n_i або w_i. Точки (x_i, n_i) або (x_i, w_i) з'єднують відрізками прямих і отримують полігон частот.

Визначення. Гістограма – діаграма в системі координат $(\Delta x, n_i)$, $\left(\Delta x, w_i\right)$. Її доцільно будувати у випадку неперервної ознаки, для цього інтервал, в якому містяться всі спостережувані значення ознаки розбивають на декілька часткових інтервалів довжиною $h = \Delta x = x_i - x_{i-1}$ і знаходять для кожного часткового інтервалу n_i – суму частот варіант, які потрапили в i-ий інтервал. Для її графіка будується ступінчата фігура, що складається з прямокутників, основами яких є частинні інтервали довжиною h, а висоти рівні відношенню $\frac{n_i}{h}$ або $\frac{w_i}{h}$. Отже, на осі абсцис відкладаються частинні інтервали, а над ними проводять відрізки, паралельні осі абсцис на висоті $\frac{n_i}{h}$, $\frac{w_i}{h}$. Тоді площа *i*-го частинного прямокутника дорівнює $\frac{n_i}{h}\cdot h = n_i$, $\frac{w_i}{h}\cdot h = w_i$ – сумі частот варіант (відносних частот) *i*-го інтервалу, а площа гістограми частот дорівнює об'єму вибірки $\sum_{i=1}^{k} n_i = n$ або $\sum_{i=1}^{k} w_i = 1$.

Визначення. Кумулята – ламана лінія в системі координат $(x, F^*(x))$ (для дискретного варіаційного ряду).

Визначення. Огіва – крива в системі координат $(x, F^*(x))$ (для інтервального ряду).

***Приклад* 1.1** Протягом дня магазином продане чоловіче взуття наступних розмірів: 39, 40, 41, 40, 43, 41, 44, 42, 40, 42, 41, 41, 43, 42, 39, 42, 43, 41, 42, 41, 38, 42, 42, 41, 40, 41, 43, 39, 40, 41;

- скласти таблицю статистичного розподілу розміру X проданого чоловічого взуття та побудувати полігон частот та;
- знайти емпіричну функцію розподілу $F^*(x)$ та побудувати кумуляту.

Розв'язання. Складемо таблицю статистичного розподілу розміру X проданого чоловічого взуття Підрахувавши кількість взуття кожного розміру в даній виборці отримаємо наступний розподіл частот вибірки:

x_i	38	39	40	41	42	43	44
n_i	1	3	5	9	7	4	1

Знайдемо об'єм вибірки, додавши всі варіанти:

$$n = 1+3+5+9+7+4+1 = 30.$$

Знайдемо відносні частоти, для чого розділимо частоти на об'єм вибірки:

$$w_i = \frac{n_i}{n}.$$

Таким чином таблиця розподілу дискретного ряду буде мати вигляд:

x_i	38	39	40	41	42	43	44
n_i	1	3	5	9	7	4	1
w_i	$\frac{1}{30}$	$\frac{1}{10}$	$\frac{1}{6}$	$\frac{3}{10}$	$\frac{7}{30}$	$\frac{2}{15}$	$\frac{1}{30}$

Контроль:

$$\frac{1}{30}+\frac{1}{10}+\frac{1}{6}+\frac{3}{10}+\frac{7}{30}+\frac{2}{15}+\frac{1}{30}=\frac{1+3+5+9+7+4+1}{30}=1$$

А полігон частот:

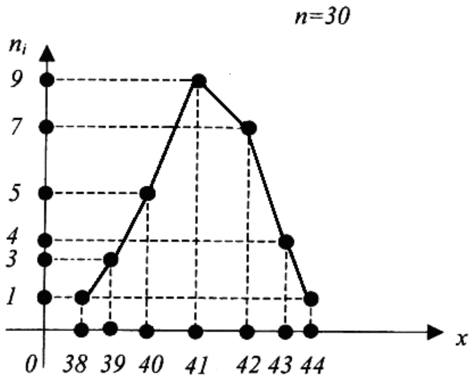


Рисунок 1.1

Знайдемо емпіричну функцію розподілу $F^*(x)$.

Найменша варіанта $X = 38$, отже $F^*(x) = 0$ при $x \le 38$.

Значення $X < 39$ спостерігається 1 раз. Отже, $F^*(x) = \frac{1}{30}$ при $38 < x \le 39$.

$X < 40$ спостерігається $1+3=4$ рази. $F^*(x) = \frac{4}{30} = \frac{2}{15}$ при $39 < x \le 40$.

$X<41$ спостерігається $1+3+5=9$ разів. $F^*(x)=\dfrac{9}{30}=\dfrac{3}{10}$ при $40<x\le 41$.

$X<42$ спостерігається $1+3+5+9=18$ разів. $F^*(x)=\dfrac{18}{30}=\dfrac{3}{5}$ при $41<x\le 42$.

$X<43$ спостерігається $1+3+5+9+7=25$ разів. $F^*(x)=\dfrac{25}{30}=\dfrac{5}{6}$ при $42<x\le 43$.

$X<44$ спостерігається $1+3+5+9+7+4=29$ разів. $F^*(x)=\dfrac{29}{30}$ при $43<x\le 44$.

Оскільки $X=44$ найбільша варіанта, то $F^*(x)=1$ при $x>44$.

Таким чином емпірична функція має вигляд:

$$F^*(x)=\begin{cases} 0 & \textit{при} \quad x\le 38, \\ \dfrac{1}{30} & \textit{при} \quad 38<x\le 39, \\ \dfrac{2}{15} & \textit{при} \quad 39<x\le 40, \\ \dfrac{3}{10} & \textit{при} \quad 40<x\le 41, \\ \dfrac{3}{5} & \textit{при} \quad 41<x\le 42, \\ \dfrac{5}{6} & \textit{при} \quad 42<x\le 43, \\ \dfrac{29}{30} & \textit{при} \quad 43<x\le 44, \\ 1 & \textit{при} \quad x>44 \end{cases}$$

Графік цієї функції (кумулята) має вигляд (рис. 1.2):

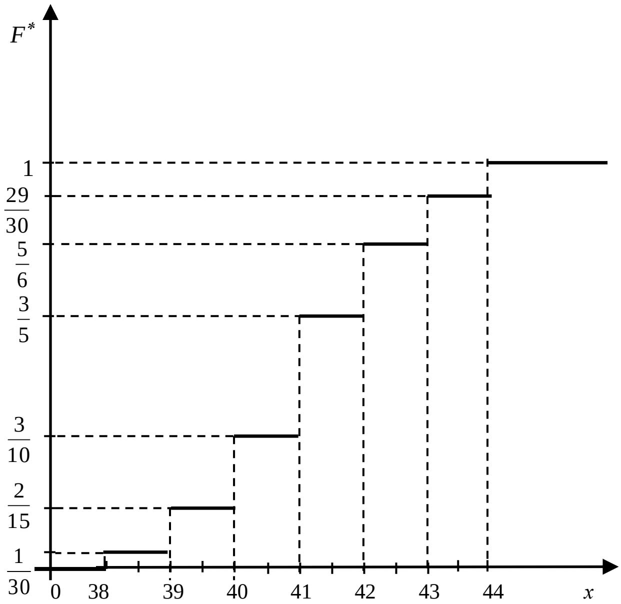


Рисунок 1.2

2 СТАТИСТИЧНІ ОЦІНКИ ПАРАМЕТРІВ РОЗПОДІЛУ

2.1. Точкові оцінки

Визначення. Статистичною оцінкою Q^* невідомого параметра Q теоретичного розподілу називають функцію $f\left(X_1,\ X_2...,\ X_n\right)$ від випадкових величин $X_1,\ X_2...,\ X_n$, які спостерігаються.

Визначення. Точковою називають статистичну оцінку, яка визначається одним числом $Q^* = f\left(X_1,\ X_2...,\ X_n\right)$, де $X_1,\ X_2...,\ X_n$ – результати *n* спостережень над кількісною ознакою *X* (вибірка).

Визначення. Незміщеною називають точкову оцінку, математичне сподівання якої дорівнює оцінюваному параметру при будь-якому об’ємі вибірки.

Визначення. ***Зміщеною*** називають точкову оцінку, математичне сподівання якої не дорівнює оцінюваному параметру.

Незміщеною оцінкою генеральної середньої (математичного сподівання) служить ***вибіркова середня*:**

$$\bar{x}_в = \frac{\left(\sum_{i=1}^{k} n_i x_i\right)}{n}, \qquad (2.1)$$

де x_i – варіанта вибірки, n_i – частота варіанти, $\sum n_i = n$ – об'єм вибірки.

Зміщеною оцінкою генеральної дисперсії служить ***вибіркова дисперсія***:

$$D_в = \frac{\sum_{i=1}^{k} n_i \left(x_i - \bar{x}_в\right)^2}{n} \qquad (2.2)$$

Зручніша формула

$$D_в = \bar{x}^2 - \left[\bar{x}\right]^2 = \frac{\sum n_i x_i^2}{n} - \left[\frac{\sum n_i x_i}{n}\right]^2. \qquad (2.3)$$

Зауваження. Якщо первинні варіанти є десятковими дробами з k десятковими знаками після коми, то щоб уникнути дій з дробами, множать первинні варіанти на постійне число $C = 10^к$, тобто переходять до умовних варіантів:

$$u_i = Cx_i \qquad (2.4)$$

При цьому дисперсія збільшиться в C^2 разів. Тому, знайшовши дисперсію умовних варіант, треба розділити її на C^2:

$$D_в(x) = \frac{D_в(u)}{C^2} \qquad (2.5)$$

Незміщеною оцінкою генеральної дисперсії служить ***виправлена вибіркова дисперсія***:

$$s^2 = \frac{n}{n-1} D_в = \frac{\sum n_i \left(x_i - x_в\right)^2}{n-1} \qquad (2.1)$$

Зручніша формула

$$s_x^2 = \frac{\sum n_i x_i^{\ 2} - \left[\sum n_i x_i\right]^2 / n}{n-1} \qquad (2.7)$$

***Приклад* 2.2** Для задачі з прикладу 2.1:

- обчислити вибіркову середню $\overline{x_в}$ та вибіркову дисперсію $D_в$ заданої вибірки;
- обчислити виправлене середньоквадратичне відхилення S;

Розв'язання. За формулою (2.1) $\bar{x}_в = \dfrac{\left(\sum\limits_{i=1}^{k} n_i x_i\right)}{n}$ маємо вибіркове середнє:

$$\bar{x}_в = \frac{1 \cdot 38 + 3 \cdot 39 + 5 \cdot 40 + 9 \cdot 41 + 7 \cdot 42 + 4 \cdot 43 + 1 \cdot 44}{30} \approx 41{,}1$$

За формулою (2.2) $D_в = \bar{x}^2 - [\bar{x}]^2 = \dfrac{\sum n_i x_i^{\ 2}}{n} - \left[\dfrac{\sum n_i x_i}{n}\right]^2$ знайдемо вибіркову дисперсію:

$$D_в = \frac{1 \cdot 38^2 + 3 \cdot 39^2 + 5 \cdot 40^2 + 9 \cdot 41^2 + 7 \cdot 42^2 + 4 \cdot 43^2 + 1 \cdot 44^2}{30} -$$

$$-(41{,}1)^2 \approx 1{,}92.$$

За формулою (2.7) $s_x^2 = \dfrac{\sum n_i x_i^{\;2} - \left[\sum n_i x_i\right]^2 / n}{n-1}$ знайдемо виправлену дисперсію:

$$s_x^{\;2} = \frac{1\cdot 38^2 + 3\cdot 39^2 + 5\cdot 40^2 + 9\cdot 41^2 + 7\cdot 42^2 + 4\cdot 43^2 + 1\cdot 44^2}{29} -$$

$$- \frac{[1\cdot 38 + 3\cdot 39 + 5\cdot 40 + 9\cdot 41 + 7\cdot 42 + 4\cdot 43 + 1\cdot 44]^2 / 30}{29} =$$

$$= \frac{50816 - 50758{,}5}{29} \approx 1{,}98$$

Таким чином середнє квадратичне відхилення (5.15):

$$S = \sqrt{S_x^2} \approx 1{,}41$$

***Приклад* 2.3** Знайти вибіркове середнє, вибіркову дисперсію, вибіркове та виправлене середнє квадратичне відхилення за даним розподілом вибірки об'єму $n = 20$*;*

x_i	0,1	0,5	0,7	0,9
n_i	6	12	1	1

***Розв'язання*:**

1. За формулою (2.1) та враховуючи (2.4) $u_i = Cx_i$ маємо вибіркове середнє:

$$\bar{x}_в = \frac{\left(\sum_{i=1}^{k} n_i x_i\right)}{10n} = \frac{1\cdot 6 + 5\cdot 12 + 7\cdot 1 + 9\cdot 1}{200} = 0{,}41.$$

2. За формулою (2.3) та враховуючи (2.5) знайдемо вибіркову дисперсію:

$$D_в(x) = \frac{\overline{u}^{\,2} - [\overline{u}]^2}{100} = \left(\frac{\sum n_i u_i^{\;2}}{n} - \left[\frac{\sum n_i u_i}{n} \right]^2 \right) / 100 =$$

$$= \left(\frac{6\cdot 1 + 12\cdot 25 + 1\cdot 49 + 1\cdot 81}{20} - (4.1)^2 \right) / 100 = 0.0499.$$

3. Враховуючи (2.4) вибіркове середнє квадратичне відхилення

$$\sigma_в(X) = \sqrt{D_в(X)} = \sqrt{0{,}0499} \approx 0{,}2234.$$

4. За формулою (2.7) враховуючи (2.5) знайдемо виправлену дисперсію:

$$s_x^{\ 2} = \frac{s_u^2}{100} = \frac{\sum n_i u_i^{\ 2} - \left[\sum n_i u_i\right]^2 / n}{100(n-1)} =$$

$$= \frac{(6 \cdot 1 + 12 \cdot 25 + 1 \cdot 49 + 1 \cdot 81) - (6 \cdot 1 + 12 \cdot 5 + 1 \cdot 7 + 1 \cdot 9)^2 / 20}{100 \cdot 19} =$$

$$= 0.05253.$$

Враховуючи (5.15) виправлене середнє квадратичне відхилення:

$$S = \sqrt{S_x^2} \approx 0.2292.$$

2.2 Інтервальні оцінки

Визначення. Інтервальною називають оцінку, яка визначається двома числами – кінцями інтервалу. Інтервальні оцінки дозволяють встановити точність і надійність оцінок.

Нехай знайдена за даними вибірки статистична характеристика $\theta*$ служить оцінкою невідомого параметра θ. Вважатимемо θ постійним числом (θ може бути і випадковою величиною). Зрозуміло, що $\theta*$ тим точніше визначає параметр θ, чим менше абсолютна величина різниці $|\theta - \theta*|$. Іншими словами, якщо $\delta > 0$ і $|\theta - \theta*| < \delta$, то чим менше δ, тим оцінка точніша. Таким чином, додатне число δ характеризує точність оцінки.

Проте статистичні методи не дозволяють категорично стверджувати, що оцінка $\theta*$ задовольняє нерівність $|\theta - \theta*| < \delta$; можна лише говорити про ймовірність γ, з якою ця нерівність виконується.

Визначення. Надійністю (надійною ймовірністю) оцінки γ за $\theta*$ називають ймовірність γ, з якою виконується нерівність $|\theta - \theta*| < \delta$. Зазвичай надійність оцінки задається наперед, причому в якості γ беруть число, близьке до одиниці. Найчастіше задають надійність, рівну 0,95; 0,99 і 0,999.

Нехай ймовірність того, що $|\theta - \theta*| < \delta$, дорівнює γ, тоді:

$$P\left[\left|\theta-\theta^*\right|<\delta\right]=\gamma. \qquad (2.8)$$

Замінивши нерівність $|\theta-\theta^*|<\delta$ на рівносильну їй подвійну нерівність $-\delta<\theta-\theta^*<\delta$, або $\theta^*-\delta<\theta<\theta^*+\delta$, маємо:

$$P\left[\theta^*-\delta<\theta<\theta^*+\delta\right]=\gamma. \qquad (2.9)$$

Це співвідношення слід розуміти так:

Ймовірність того, що інтервал $\left(\theta^*-\delta;\ \theta^*+\delta\right)$ містить в собі (покриває) невідомий параметр θ, дорівнює γ.

Визначення. Надійним називають інтервал $\left(\theta^*-\delta;\ \theta^*+\delta\right)$, який покриває невідомий параметр із заданою надійністю γ.

2.3 Надійний інтервал для оцінки математичного сподівання нормального розподілу

Нехай кількісна ознака X генеральної сукупності розподілена нормально, причому середнє квадратичне відхилення σ цього розподілу відомо.

Потрібно оцінити відоме математичне сподівання a за вибірковою середньою $\bar{x}_в$. Поставимо своїм завданням знайти надійні інтервали, які покривають параметр a з надійністю γ.

Розглядатимемо вибіркову середню $\bar{x}_в$ як випадкову величину $\overline{X}_в$ ($\bar{x}_в$ змінюється від вибірки до вибірки) і вибіркові значення ознаки $x_1, x_2..., x_n$ – як однаково розподілені незалежні випадкові величини $X_1, X_2..., X_n$ (ці числа також змінюються від вибірки до вибірки). Іншими словами, математичне сподівання кожної з цих величин дорівнює a і середнє квадратичне відхилення – σ.

Приймемо без доведення, що якщо випадкова величина X розподілена нормально, то вибіркова середня $\overline{X}_в$, знайдена в незалежних спостереженнях, також розподілена нормально. Параметри розподілу $\overline{X}_в$ такі

$$M(\overline{X_в}) = a\,,\ \sigma\left(\overline{X_в}\right) = \frac{\sigma}{\sqrt{n}}\,.$$

Необхідно, щоб виконувалося співвідношення

$$P(|\overline{X_в} - a| < \delta) = \gamma\,.$$

Користуючись формулою (5.51, Ч.2)

$$P(|X - m| < \Delta) = 2\Phi\left[\frac{\Delta}{\sigma}\right]$$

замінивши *X* на $\overline{X}_в$, *m* на *a* і *σ* на $\sigma\left(\overline{X_в}\right) = \frac{\sigma}{\sqrt{n}}$, отримаємо

$$P(|\overline{X_в} - a| < \delta) = 2\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right) = 2\Phi(t)\,, \qquad (2.10)$$

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

Знайшовши з останньої рівності $\delta = \frac{t\sigma}{\sqrt{n}}$, можемо записати

$$P(|\overline{X_в} - a| < \frac{t\sigma}{\sqrt{n}}) = 2\Phi\left(t\right). \qquad (2.11)$$

Прийнявши до уваги, що ймовірність *P* задана і дорівнює *γ*, остаточно маємо (щоб отримати робочу формулу, вибіркову середню знов позначимо через $\bar{x}_в$).

$$P(\overline{x_в} - \frac{t\sigma}{\sqrt{n}} < a < \overline{x_в} + \frac{t\sigma}{\sqrt{n}}) = 2\Phi\left(t\right) = \gamma \qquad (2.12)$$

Зміст отриманого співвідношення такий: з надійністю *γ* можна стверджувати, що надійний інтервал $(\bar{x}_в - \frac{t\sigma}{\sqrt{n}}\,;\bar{x}_в + \frac{t\sigma}{\sqrt{n}})$ покриває невідомий параметр *a*; точність оцінки $\delta = \frac{t\sigma}{\sqrt{n}}$

Число t визначається з рівності $2\Phi(t)=\gamma$ або $\Phi(t)=\frac{\gamma}{2}$; за таблицею функції Лапласа (див. додаток) знаходиться аргумент t, якому відповідає значення функції Лапласа, рівне $\frac{\gamma}{2}$.

Зауваження. Надійну ймовірність не слід пов’язувати з оцінюваним параметром; вона зв’язана лише з границями надійного інтервалу, які, як вже було вказано, змінюються від вибірки до вибірки.

***Приклад* 2.4** Знайти надійний інтервал для математичного сподівання a нормального розподілу з надійністю 0,95, знаючи вибіркову середню $\bar{x}_в$, об’єм вибірки n і середнє квадратичне відхилення σ:

$$\bar{x}_в=75{,}18;\ n=256,\ \sigma=16.$$

Розв’язання. Надійний інтервал для математичного сподівання $(\bar{x}_в-\delta\,;\bar{x}_в+\delta)$, де $\delta=\frac{t\sigma}{\sqrt{n}}$. Параметр t визначаємо з умови

$$\Phi(t)=\frac{\gamma}{2}=0{,}95/2=0{,}475.$$

З додатку визначаємо, що $\Phi(t)=0{,}475$ при $t=1{,}96$.

Тоді $\delta=\frac{16\cdot 1{,}96}{\sqrt{256}}=1{,}96$.

$\bar{x}_в-\delta=75{,}18-1{,}96=73{,}22$,

$\bar{x}_в+\delta=75{,}18+1{,}96=77{,}14$.

(73,22; 77,14) – шуканий надійний інтервал.

***Приклад* 2.5** Дослідження часу безвідмовної роботи 50 лазерних принтерів ($n=50$). З апріорних досліджень відомо, що середнє квадратичне відхилення часу безвідмовної роботи $\sigma=16$ годин. За результатами досліджень отримано середній час безвідмовної роботи $\overline{X}=1000$ годин. Побудуйте 95 % довірчий інтервал (γ = 0,95) для середнього часу безвідмовної роботи.

Розв'язання. $2\Phi(u)=0{,}95$. Звідки $\Phi(u)=\frac{0{,}95}{2}=0{,}475$. За таблицею функції Лапласа знаходимо значення $u=1{,}95$. Далі знаходимо точність оцінки: $\varepsilon=\frac{u\cdot\sigma}{\sqrt{n}}=\frac{1{,}96\cdot 16}{\sqrt{50}}=4{,}43$. Знаходимо границі інтервалу:

$\overline{X}-\varepsilon=1000-4{,}43=995{,}57;\quad \overline{X}+\varepsilon=1000+4{,}43=1004{,}43$.

Одержимо надійний інтервал: $(995{,}57;1004{,}43)$.

3 МЕТОДИ СТАТИСТИЧНИХ ПЕРЕВІРОК ГІПОТЕЗ

3.1 Статистичні гіпотези. Статистичний критерій перевірки нульової гіпотези

На практиці часто потрібно на основі результатів випробувань (вибірки) знайти закон розподілу генеральної сукупності. Якщо закон розподілу невідомий, але є підстава вважати, що він має певний вигляд (наприклад назвемо його R), то висувають гіпотезу: генеральна сукупність розподілена за законом R, тобто в цій гіпотезі мова йтиме про вигляд передбаченого розподілу.

Можливі випадки, коли закон розподілу відомий, але його параметри невідомі. Якщо є підстава припустити, що його невідомий параметр a рівний певному значенню a_0, то висувають гіпотезу: $a = a_0$; в цьому випадку гіпотеза припускає оцінку параметру конкретного розподілу.

Можливі й інші гіпотези: про рівність параметрів двох або декількох розподілів, про незалежність вибірок, про значимість вибіркового коефіцієнта кореляції тощо.

Визначення. Статистичною називають гіпотезу про вигляд невідомого розподілу або про параметри невідомих розподілів. Наприклад, статистичними є гіпотези:

1) генеральна сукупність розподілена за нормальним законом;

2) коефіцієнт кореляції генеральної сукупності системи (x, y), розподіленої нормально, відмінний від нуля.

Визначення. Перевірку гіпотез на основі вибіркових статистичних даних називають ***статистичною перевіркою гіпотез***.

Одну з висунутих гіпотез виділяють в ролі основної і позначають, як правило H_0 (нульова), поряд з нею висувають альтернативну (конкуруючу) гіпотезу, яка суперечить основній і позначають H_1.

Наприклад, якщо нульова гіпотеза полягає в припущенні, що математичне сподівання певного розподілу m_x дорівнює 5, то альтернативна гіпотеза, зокрема, може полягати в тому, що $m_x \neq 5$. Коротко це записують так:

$$H_0:\ m_x = 5;\ H_1 : m_x \neq 5\,.$$

Розрізняють також гіпотези за кількістю припущень.

Визначення. Простою називається гіпотеза, що має лише одне припущення, інакше гіпотеза є ***складною***, тобто складається зі скінченного або нескінченного числа простих гіпотез.

Наприклад. Якщо λ – параметр показникового розподілу, то гіпотеза $H_0 : \lambda = 2$ – проста, гіпотеза $H_0 : \lambda > 5$ – складна, бо складається з нескінченної множини простих гіпотез: $H_1 : \lambda = a_i$, де a_i – довільне число, більше λ.

Очевидно, що на основі статистичних даних дуже важко, іноді і неможливо, робити безпомилкові висновки щодо гіпотез. В підсумку може бути зроблений неправильний висновок, тобто можуть бути допущені помилки двох родів.

Визначення. Помилка першого роду полягає в тому, що буде відхилена правильна гіпотеза.

Визначення. Помилка другого роду полягає в тому, що буде прийнята неправильна гіпотеза.

Правильний висновок може бути зроблений також у двох випадках:

а) гіпотеза приймається, причому і в дійсності вона правильна;

б) гіпотеза відхиляється, причому і в дійсності вона неправильна.

Визначення. Ймовірність зробити помилку першого роду позначають через α і називають її ***рівнем значущості***. Число α задають малим і найчастіше використовують значення α, які дорівнюють 0,05; 0,001 і т.д. Якщо, наприклад, $a = 0{,}01$, то це означає, що в одному випадку зі 100 є ризик допустити помилку першого роду (відхилити гіпотезу H_0).

Для перевірки нульової гіпотези використовують спеціально підібрану випадкову величину, точний або наближений розподіл якої відомо. Цю величину позначають через Φ, якщо вона розподілена нормально, F – за законом Фішера-Снедекора, T – за законом Ст'юдента, χ^2 – за законом "хі квадрат" і т.д. Оскільки зараз конкретний вигляд розподілу до уваги не береться, то позначають цю величину взагалі через K.

Визначення. Статистичним критерієм (критерієм) називають випадкову величину K, що використовується для перевірки нульової гіпотези. Для різних гіпотез ці критерії є різними і поділяються на параметричні й непараметричні.

Визначення. Параметричні критерії використовуються в завданнях перевірки параметричних гіпотез і включають у свій розрахунок показники розподілу, наприклад, середні, дисперсії тощо. Параметричні критерії дозволяють прямо оцінити рівень основних параметрів генеральних сукупностей, різниці середніх і відмінності в дисперсіях. Критерії спроможні виявити тенденції зміни ознаки при переході від умови до умови, оцінити взаємодію двох і більш факторів у впливі на зміни ознаки. Параметричні критерії вважаються дещо більш потужними, ніж непараметричні, за умов, якщо ознака виміряна за інтервальною шкалою і нормально розподілена. Проте з інтервальною шкалою можуть виникнути певні проблеми, якщо дані, представлено не в стандартизованих оцінках. До того ж перевірка розподілу "на нормальність" вимагає досить складних розрахунків, результат яких заздалегідь невідомий. Найчастіше розподіли ознак відрізняються від нормального, тоді доводиться звертатися до непараметричних критеріїв.

Визначення. Непараметричні критерії позбавлені перерахованих вище обмежень. Проте вони не дозволяють здійснити пряму оцінку рівня таких важливих параметрів, як середнє або дисперсія, з їхньою допомогою неможливо оцінити взаємодію двох і більше умов або факторів, що впливають на зміну ознаки. Непараметричні критерії дозволяють вирішити деякі важливі завдання, які супроводжують дослідження в психології і педагогіці: виявлення відмінностей у рівні досліджуваної ознаки, оцінка зсуву значень досліджуваної ознаки, виявлення відмінностей у розподілах ознак.

Наприклад,

а) коли перевіряють гіпотезу про рівність дисперсій двох нормальних генеральних сукупностей, то в ролі параметричного критерію K беруть відношення виправлених вибіркових дисперсій:

$$K = F = \frac{S_1^2}{S_2^2}\,. \tag{3.1}$$

Ця величина випадкова, тому в різних випробуваннях дисперсії приймають різні, наперед невідомі значення і розподілені за законом Фішера-Снедекора.

б) найбільш розповсюдженим критерієм перевірки гіпотези H_0 про закон розподілу ознаки генеральної сукупності є непараметричний критерій узгодженості:

$$\chi^2 = \sum_{i=1}^{m} \frac{\left(n_i - np_i\right)^2}{np_i}, \tag{3.2}$$

де m – число інтервалів, на які розбита вибірка, n – об'єм вибірки, n_i – частота i-го інтервалу, ρ_i – ймовірність попадання значень ознаки в i-ий інтервал, яка обчислюється для теоретичного закону розподілу.

Визначення. Спостережуваним значенням K_{cn} називається значення критерію, обчислене за результатами вибірки.

3.2 Критична область. Загальна методика побудови критичних областей

Всю множину значень статистичного критерію K можна розбити на дві підмножини, які не перетинаються A і $\bar{A}$.

Визначення. Значення статистичного критерію підмножини $A \in \Omega$, при яких нульова гіпотеза приймається, називаються ***областю прийняття гіпотези***, а підмножина значень $\bar{A}$, при яких гіпотеза H_0 відхиляється – ***критичною областю***.

Основний принцип перевірки статистичних гіпотез формулюється так: якщо спостережуване значення критерію K_{cn} належить області прийняття гіпотези A – гіпотезу приймають, якщо K_{cn} належить критичній області $\bar{A}$ гіпотезу відхиляють.

Оскільки критерій K – одномірна випадкова величина, то всі її можливі значення належать деякому інтервалу. Тому область прийняття гіпотези A і критична область $\bar{A}$ також є інтервальними, а, значить, існують точки, які їх розділяють. Ці точки називаються критичними і позначаються $k_{кр}$.

Розрізняють односторонню (правосторонню або лівосторонню) і двосторонню критичні області (див. рис. 3.1).

Визначення. Правосторонньою називають критичну область, яка визначається нерівністю $K > k_{кр}$, де $k_{кр}$ – додатне число (рис. 1, а).

Визначення. Лівосторонньою називають критичну область, яка визначається нерівністю $K < k_{кр}$, де $k_{кр} < 0$ (рис. 3.1, б).

Визначення. Двосторонньою називають критичну область, яка визначається нерівністю $K < k_{1кр}$, $K > k_{2кр}$, де $k_{2кр} > k_{1кр}$ (рис. 3.1, в). Зокрема, якщо критичні точки симетричні відносно нуля, двостороння критична область визначається нерівностями: $K < -k_{кр}$, $K > k_{кр}$, або $|K| > \kappa_{кр}$ ($k_{кр} > 0$).

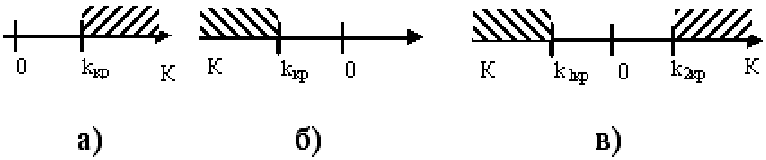


Рисунок 3.1 – Критичні області

***Перевірка статистичних гіпотез будь-якої природи здійснюється за такою схемою*:**

1. Формулюється статистична гіпотеза H_0.
2. Обирається статистичний критерій відповідно до сформульованої нульової гіпотези H_0.
3. Залежно від гіпотези H_0 і альтернативної H_1 вибирається одностороння або двостороння критична область.

Щоб побудувати критичні області, необхідно знайти значення критичних точок.

В основі побудови критичної області лежить принцип практичної неможливості здійснитися малоймовірній випадковій події при одній спробі. Тому задається мала величина ймовірності α ($a = 0,01$; $a = 0,05$) (рівень значущості) критерію перевірки правильності гіпотези H_0: на основі відомого розподілу ймовірності критерію K визначається за допомогою спеціальних таблиць (див. додаток) критична точка $k_{кр}$. За знайденою $k_{кр}$ відповідно відбудеться лівостороння, правостороння або двостороння критична область.

4. За результатами вибірки обчислюється спостережене значення критерію $K_{сп}$.

5. Виходячи з вимоги, що при правильності гіпотези H_0 ймовірність того, що $K_{сп}$ потрапить у критичну область, має дорівнювати прийнятому рівню значущості α, перевіряється статистична гіпотеза.

Це твердження подають для лівосторонньої критичної області так: $P(K < \kappa_{кр}) = \alpha$, для правосторонньої: $P(K > \kappa_{кр}) = \alpha$ для двосторонньої критичної області: $P(K < \kappa_{1кр}) + P(K > \kappa_{2кр}) = \alpha$.

На практиці двосторонню критичну область будують симетрично розміщену відносно нуля, розділяючи при цьому α порівну між кінцями критичних областей, тобто $P(K < \kappa_{1кр}) = P(K > \kappa_{2кр}) = \dfrac{\alpha}{2}$.

Якщо K потрапляє у критичну область, а ця подія малоймовірна і вона все-таки здійснилася, то нульова гіпотеза H_0 відхиляється. У протилежному випадку – приймається.

3.3 Перевірка гіпотези про нормальний розподіл генеральній сукупності за критерієм Пірсона

Нехай емпіричний розподіл заданий у вигляді послідовності рівновіддалених варіант і відповідних ним частот.

x_i	x_1	x_2	…	x_N
n_i	n_1	n_2	….	n_N

Необхідно, використовуючи критерій Пірсона, перевірити гіпотезу про те, що генеральна сукупність X розподілена нормально.

Правило. Для того, щоб при заданому рівні значимості α перевірити гіпотезу про нормальний розподіл генеральної сукупності, треба:

1. Обчислити вибіркову середню $\bar{x}_в$ і вибіркове середнє квадратичне відхилення $\sigma_в$.

2. Обчислити теоретичні частоти

$$n_i' = \frac{nh}{\sigma_в} \cdot \varphi(u_i), \qquad (3.3)$$

де n – об'єм вибірки (сума всіх частот), h – крок (різниця між двома сусідніми варіантами),

$$u_i = \frac{x_i - \overline{x_в}}{\sigma_в}, \quad \varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \qquad (3.4)$$

Значення функції $\varphi(u_i)$ знаходяться в додатку.

3. Порівняти емпіричні і теоретичні частоти за допомогою критерію Пірсона. Для цього:

а) складають розрахункову таблицю, по якій знаходять спостережуване значення критерію

$$\chi^2_{сп} = \sum \frac{(n_i - n_i')^2}{n_i'} \qquad (3.5)$$

б) згідно таблиці критичних точок розподілу χ^2 (додаток), за заданим рівнем значущості α і числу степенів свободи $k = s - 3$ (s – число груп вибірки) знаходять критичну точку $\chi^2_{кр}(\alpha; k)$ правосторонньої критичної області.

Якщо $\chi^2_{сп} < \chi^2_{кр}$ – немає підстав відкидати гіпотезу про нормальний розподіл генеральної сукупності. Іншими словами емпіричні і теоретичні частоти розрізняються не значущо (випадково).

Якщо $\chi^2_{сп} > \chi^2_{кр}$ – гіпотезу відкидають. Іншими словами, емпіричні і теоретичні частоти розрізняються значущо.

Зауваження. Малочислені частоти ($n_i < 5$) потрібно об’єднати; в цьому випадку і відповідні їм теоретичні частоти також треба додати. Якщо проводилося об’єднання частот, то при визначенні числа степенів свободи по формулі $k = s - 3$ потрібно в якості s прийняти число груп вибірки, що залишилися після об’єднання частот.

***Приклад* 3.1** За допомогою критерію Пірсона з рівнем значущості $\alpha = 0,01$ перевірити гіпотезу, що сукупність Y має нормальний розподіл з параметрами рівними їх оцінкам $a = \overline{Y}$ та $\sigma^2 = D$:

Yi	0	1	2	3	4	5	6
n_i	4	22	30	22	17	4	1

Розв’язання.

1. За формулою $\bar{y}_в = \dfrac{\left(\sum\limits_{i=1}^{k} n_i y_i\right)}{n}$ маємо вибіркове середнє:

$$\bar{y}_в = \frac{0\cdot 4 + 1\cdot 22 + 2\cdot 30 + 3\cdot 22 + 4\cdot 17 + 5\cdot 4 + 6\cdot 1}{100} \approx 2,42 .$$

За формулою $D_в = \bar{y}^2 - \left[\bar{y}\right]^2 = \dfrac{\sum n_i y_i^{\,2}}{n} - \left[\dfrac{\sum n_i y_i}{n}\right]^2$ знайдемо вибіркову дисперсію:

$$D_в = \frac{0^2\cdot 4 + 1^2\cdot 22 + 2^2\cdot 30 + 3^2\cdot 22 + 4^2\cdot 17 + 5^2\cdot 4 + 6^2\cdot 1}{100} - (2,42)^2 \approx 1,62.$$

$$\sigma_в = \sqrt{D_в} = 1,27$$

2. Обчислюємо теоретичні частоти, враховуючи, що $n = 100$, $h = 1$, $\sigma_в = 1,27$.

$$n_i = \frac{nh}{\sigma_в}\cdot\varphi(u_i) = \frac{100}{1,27}\cdot\varphi(u_i) = 78,7\cdot\varphi(u_i)$$

i	y_i	$u_i = \dfrac{y_i - \overline{y_в}}{\sigma_в}$	$\phi(u_i)$	$n_i' = 78,7\cdot\varphi(u_i)$
1	0	-1,91	0,0644	5,1

2	1	-1,12	0,2131	16,8
3	2	-0,33	0.3778	29,7
4	3	0,46	0,3589	28,2
5	4	1,24	0,1849	14,6
6	5	2,03	0,0508	4,0
7	6	2,82	0.0075	0,6

3. Порівняємо емпіричні та теоретичні частоти.

а) Складемо розрахункову таблицю, згідно якої знайдемо значення критерію, що спостерігається (3.5).

$$\chi^2_{сп} = \sum \frac{(n_i - n_i')^2}{n_i'}$$

i	n_i	n_i'	$(n_i - n_i')$	$(n_i - n_i')^2$	$\frac{(n_i - n_i')^2}{n_i'}$
1	26	21,9	4,1	16,81	0,768
2	30	29,7	0,3	0,09	0,003
3	22	28,2	-6,2	38,44	1,363
4	17	14,6	2,4	5,76	0,395
5	5	4,6	0,4	0,16	0,035
					$\chi^2_{набл} = 2,6$

Згідно зауваження останні дві частоти було об'єднано в одну.

б) За таблиці критичних точок розподілу χ^2 (додаток), за рівнем значущості $\alpha = 0,01$ і числом степенів свободи

$k = s - 3 = 5 - 3 = 2$ знаходимо критичну точку правосторонньої критичній області

$\chi^2_{крит}\left(0,01;2\right) = 9,2$ Так як $\chi^2_{сп} < \chi^2_{крит}$, то нема підстав відкидати гіпотезу про нормальний розподіл генеральної сукупності.

3.4 Перевірка гіпотези про рівність дисперсій двох нормальних сукупностей за критерієм Фішера

Нехай знайдені виправлені вибіркові дисперсії S_1^2 та S_2^2 двох незалежних вибірок об'ємів n_1 та n_2 із нормальних сукупностей X та Y.

Необхідно, використовуючи критерій Фішера, перевірити гіпотезу про рівність дисперсій генеральних сукупностей, тобто $D(X) = D(Y)$.

Правило. Для того, щоб при заданому рівні значимості α перевірити гіпотезу про рівність дисперсій генеральних сукупностей, треба:

1. Обчислити відношення більшої виправленої дисперсії до меншої за формулою (3.1):

$$F = \frac{S_1^2}{S_2^2} ,$$

де S_1^2 – більша виправлена дисперсія, S_2^2 – менша виправлена дисперсія.

2. По таблиці розподілу Фішера (додаток), за рівнем значущості α і числам ступенів свободи $k_1 = n_1 - 1$ і $k_2 = n_2 - 1$ знайти критичну точку $F_{крит}$.

3. Порівняти емпіричні і теоретичні значення критерію.

Якщо $F < F_{кр}$ – немає підстав відкинути нульову гіпотезу про рівність генеральних дисперсій.

Якщо $F > F_{кр}$ – гіпотезу відкидають.

***Приклад* 3.2** За даними двох незалежних вибірок об'єму $n_1 = 10$ та $n_2 = 15$ із нормальних сукупностей X та Y знайдені виправлені вибіркові дисперсії $S_1^2 = 15,42$ та $S_2^2 = 11,36$. При рівні значущості $\alpha = 0,05$ перевірити гіпотезу $H_0:\ D(X) = D(Y)$ при альтернативній $H_1:\ D(X) > D(Y)$.

Розв'язання. Знайдемо відношення більшої виправленої дисперсії до меншої (3.1):

$$F = \frac{S_1^2}{S_2^2} = \frac{15,42}{11,36} = 1,36 .$$

За таблицею розподілу Фішера (додаток), за рівнем значущості $\alpha = 0,05$ і числам ступенів свободи $k_1 = n_1 - 1 = 10 - 1 = 9$ і $k_2 = n_2 - 1 = 15 - 1 = 14$ знаходимо критичну точку

$F_{крит} = F_{крит}(0,05; 9; 14) = 2,65$. Враховуючи, що $F < F_{крит}$, робимо висновок, що немає підстав відкинути нульову гіпотезу про рівність генеральних дисперсій.

***Приклад* 3.3** Дисперсія такого показника, як стресостійкість для вчителів складає 6,17 ($n_1 = 32$), а для менеджерів 4,41 ($n_2 = 33$). Визначити, чи можна вважати рівень дисперсій приблизно однаковим для даних вибірок при рівні значущості $\alpha = 0,05$.

Розв'язання. Знайдемо емпіричне значення критерію (3.1):

$$F = \frac{6,17}{4,41} \approx 1,4 \ .$$

За таблицею розподілу Фішера (додаток), критичне значення критерію – $F_{крит} = F_{крит}(0,05; 31; 32) = 2$

Враховуючи, що $F < F_{крит}$, робимо висновок, що немає підстав відкинути нульову гіпотезу про рівність генеральних дисперсій при рівні значущості 0,05.

4 ЕЛЕМЕНТИ КОРЕЛЯЦІЙНО-РЕГРЕСІЙНОГО АНАЛІЗУ

В багатьох задачах потрібно встановити залежність випадкової величини Y від однієї або декількох інших величин. Залежності між величинами можна поділити на функціональні і статистичні. В природничих, технічних науках здебільшого зустрічаються функціональні залежності, при яких кожному значенню аргументу x за певним законом відповідає одне значення функції y.

Строга функціональна залежність здійснюється рідко, так як обидві величини x та y, або одна з них підпадає під дію випадкових впливів (факторів), причому деякі з них можуть бути спільними для обох величин x та y.

Між змінними, що характеризують економічні величини, здебільшого існують залежності, які проявляються в тому, що одна з них реагує на зміну іншої зміною свого закону розподілу. Наприклад, урожайність сільськогосподарських культур залежить від кількості внесеного добрива, але ця залежність не буде функціональна, оскільки на врожайність, крім того, впливатимуть кліматичні умови, технологія землі та посіву тощо.

Визначення. Статистичною називають залежність, при якій зміна однієї з величин веде до зміни розподілу іншої, зокрема ***кореляційним*** називається зв'язок між статистичними змінними X і Y, за якими при зміні ознаки X змінюється середнє значення ознаки Y. Причому при кореляційній залежності одному значенню незалежної змінної X відповідає не одне, а декілька значень залежної змінної Y. Наведений приклад показує, що середня врожайність є функцією від кількості внесеного добрива, тобто Y зв'язаний з X кореляційною залежністю.

4.1 Рівняння парної регресії. Лінійна кореляція

В ролі оцінки умовних математичних сподівань беруть умовні середні, які знаходять за даними вибірки.

Визначення. Умовною середньою $\overline{y_x}$ називають середнє арифметичне із значень *Y*, що відповідають одному і тому ж значенню $X = x$.

***Приклад* 4.1** Нехай *X* – статистична величина, що характеризує вагу людини в *кг*, а *Y* – відповідно зріст в *см*, і двовимірний статистичний розподіл задається таблицею:

Y	*X*			
	70	75	80	n_y
170	15	10	5	30
175	–	10	–	10
180	–	5	5	10
n_x	15	25	10	$n = 50$

Наприклад, вазі 75 *кг* відповідає середній зріст:

$$\overline{Y_{75}} = \frac{170 \cdot 10 + 175 \cdot 10 + 180 \cdot 5}{25} = 174\ см\,.$$

Аналогічно вводиться умовна середня .

Використовуючи поняття умовної середньої, введемо означення кореляційної залежності.

Визначення. Кореляційною називається залежність умовної середньої від аргументів і записується в такому вигляді: ,

якщо *n* змінних: .

Дані рівняння називають вибірковими рівняннями регресії *Y* на *X;* функцію – вибірковою регресією *Y* на *X,* а її графік – вибірковою лінією регресії *Y* на *X.*

Рівняння регресії найчастіше використовують як різновид статистичних моделей, що застосовують, наприклад, в економічному аналізі, де за допомогою рівнянь регресії є можливість виміряти вплив окремих факторів-аргументів на залежну змінну. Цим самим аналіз стає конкретним і цінність його суттєво збільшується. Крім регресивного аналізу, рівняння регресії використовують у прогнозних дослідженнях. В економічних дослідженнях кореляційні дослідження ввійшли під поняттям виробничі функції.

Визначення. Найпростішою буде кореляційна залежність, коли є один аргумент і вона називається ***парною***. Якщо ж аргументів більше, ніж один, то залежність називається ***множинною***.

Вигляд рівняння визначає тип кореляційної залежності. Найбільш поширеним і простим є рівняння лінійної регресії, коли всі параметри входять в першому степені:

$$. \tag{4.1}$$

Прикладами можуть бути: залежність між витратами на рекламу та обсягом реалізованої продукції, витратами на споживання та валовим національним продуктом (ВНП), зміною ВНП в залежності від часу і т.д.

Вибіркове рівняння прямої лінії регресії *Y* на *X* має вигляд:

$$\overline{y}_x - \overline{y} = r_B \frac{\sigma_y}{\sigma_x}(x - \overline{x}), \tag{4.2}$$

де $\overline{y}_x$ – умовна середня; $\overline{x}$ і $\overline{y}$ – вибіркові середні ознак *X* і *Y*; σ_x і σ_y – вибіркові середні квадратичні відхилення; r_B – вибірковий коефіцієнт кореляції.

Вибіркове рівняння прямої лінії регресії *X* на *Y* має вигляд

$$\overline{x}_y - \overline{x} = r_B \frac{\sigma_x}{\sigma_y}(y - \overline{y}). \tag{4.3}$$

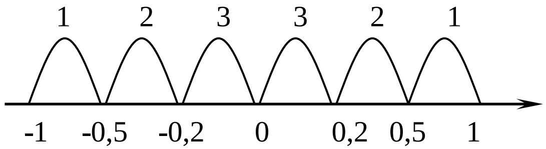
4.2 Вибірковий коефіцієнт кореляції та його властивості, методика знаходження

Число r_B є вибірковим коефіцієнтом кореляції, тобто оцінкою коефіцієнта кореляції

$$r = \frac{\mu_{xy}}{\sigma_x \sigma_y} = \frac{\mathrm{M}(X,Y) - \mathrm{M}(X)\cdot \mathrm{M}(Y)}{\sigma_x \sigma_y}. \tag{4.3}$$

Сила кореляційної залежності у випадку прямої регресії оцінюється коефіцієнтом кореляції *r*. Так як $|r| \le 1$, то чим *r* ближче до ± 1, тим щільніший зв’язок *Y* та *X*, який переходить у функціональну (лінійну) залежність при $r = \pm 1$. Якщо $r < 0$, то зв’язок між величинами обернений, якщо $r > 0$, то прямий, якщо $r = 0$, то зв’язок відсутній.

Залежність щільності зв’язку між явищами від величини коефіцієнта кореляції *r* можна зобразити графічно.



1 – зв’язок тісний,
2 –зв’язок середній,
3 – зв’язок слабкий.

Рисунок 4.1 – Величини коефіцієнта кореляції *r*

Вибірковий коефіцієнт r_B є оцінкою коефіцієнта кореляції *r* генеральної сукупності і тому також служить для вимірювання лінійного зв’язку між величинами *X* та *Y*. Якщо вибірка має досить великий об’єм і добре представляє генеральну сукупність, то висновок про щільність лінійної залежності між ознаками, отриманий по даним вибірки, в певній мірі може бути поширений і на генеральну сукупність. Наприклад, для оцінки коефіцієнта кореляції *r* нормально розподіленої сукупності (при $n \ge 50$), можна користуватись формулою

$$r_B - 3\frac{1 - r_B^2}{\sqrt{n}} \le r \le r_B + 3\frac{1 + r_B^2}{\sqrt{n}}. \tag{4.4}$$

Вибірковий коефіцієнт кореляції будемо знаходити за формулою:

$$r_B = \frac{\sum n_{xy} xy - n\overline{x}\overline{y}}{n\sigma_x \sigma_y}, \tag{4.5}$$

або можна використати формулу:

$$r_B = \frac{K(X,Y)}{\sigma_x \cdot \sigma_y} = \frac{\overline{X \cdot Y} - \overline{X} \cdot \overline{Y}}{\sigma_x \cdot \sigma_y}, \text{ де} \tag{4.6}$$

середні значення величин:

$$\overline{X} = \frac{\sum x \cdot n_x}{n}; \tag{4.7}$$

$$\overline{Y} = \frac{\sum y \cdot n_y}{n}; \tag{4.8}$$

кореляцію між величинами X і Y:

$$\overline{X \cdot Y} = \frac{\sum x \cdot y \cdot n_{xy}}{n}. \tag{4.9}$$

Приклад **4.1** Знайти вибіркове рівняння прямої лінії регресії Y на X за даними кореляційної таблиці; перевірити значущість параметрів і тісноту кореляційного зв'язку.

Y	X					
	-1,1	-1,0	-0,9	-0,8	-0,7	n_y
2,1	–	–	2	21	1	24
2,2	2	4	12	14	–	32
2,3	–	2	3	–	–	5
2,4	10	9	–	–	–	19
n_x	12	15	17	35	1	$n=80$

Розв'язання. Знайдемо середні значення величин X і Y (4.7, 4.8):

$$\overline{X}=\frac{\sum x\cdot n_x}{n}=\frac{-1{,}1\cdot 12-1{,}0\cdot 15-0{,}9\cdot 17-0{,}8\cdot 35-0{,}7\cdot 1}{80}=\frac{-72{,}2}{80}=$$
$$=-0{,}9025;$$
$$\overline{Y}=\frac{\sum y\cdot n_y}{n}=\frac{2{,}1\cdot 24+2{,}2\cdot 32+2{,}3\cdot 5+2{,}4\cdot 19}{80}=\frac{177{,}9}{80}=2{,}22375.$$

Знайдемо середні квадратичні відхилення величин X і Y:

$$\overline{X^2}=\frac{\sum x^2\cdot n_x}{n}=\frac{(-1{,}1)^2\cdot 12+(-1{,}0)^2\cdot 15+(-0{,}9)^2\cdot 17+(-0{,}8)^2\cdot 35+(-0{,}7)^2\cdot 1}{80}=$$
$$=\frac{66{,}18}{80}=0{,}82725.$$
$$D_x=\overline{X^2}-\overline{X}^2=0{,}82725-(-0{,}9025)^2=0{,}012744.$$
$$\sigma_x=\sqrt{D_x}=\sqrt{0{,}01274}=0{,}112888.$$
$$\overline{Y^2}=\frac{\sum y^2\cdot n_y}{n}=\frac{2{,}1^2\cdot 24+2{,}2^2\cdot 32+2{,}3^2\cdot 5+2{,}4^2\cdot 19}{80}=\frac{396{,}61}{80}=4{,}957625.$$
$$D_y=\overline{Y^2}-\overline{Y}^2=4{,}957625-2{,}22375^2=0{,}012561.$$
$$\sigma_y=\sqrt{D_y}=\sqrt{0{,}012561}=0{,}112076.$$

Знайдемо кореляцію між величинами X і Y (4.9):

$$\overline{X\cdot Y}=\frac{\sum x\cdot y\cdot n_{xy}}{n}=\frac{1}{80}\cdot(-1{,}1\cdot(2{,}2\cdot 2+2{,}4\cdot 10)-$$
$$-1{,}0\cdot(2{,}2\cdot 4+2{,}3\cdot 2+2{,}4\cdot 9)-0{,}9\cdot(2{,}1\cdot 2+2{,}2\cdot 12+2{,}3\cdot 3)-$$
$$-0{,}8\cdot(2{,}1\cdot 21+2{,}2\cdot 14)-0{,}7\cdot 2{,}1\cdot 1)=\frac{-161{,}38}{80}=-2{,}01725.$$
$$K(X,Y)=\overline{X\cdot Y}-\overline{X}\cdot\overline{Y}=-2{,}01725-(-0{,}9025)\cdot 2{,}22375=-0{,}0103156.$$

Знайдемо коефіцієнт кореляції між величинами X і Y (4.6):

$$r=\frac{K(X,Y)}{\sigma_x\cdot\sigma_y}=\frac{-0{,}0103156}{0{,}112888\cdot 0{,}112076}=-0{,}815335.$$

Знайдемо параметри лінійного рівняння (4.1) $y=k\cdot x+b$ регресії Y на X.

Коефіцієнт регресії (4.2): $k = r \cdot \dfrac{\sigma_y}{\sigma_x} = -0,815335 \cdot \dfrac{0,112076}{0,112888} = -0,8095$.

Вільний член регресії:

$b = \overline{Y} - k \cdot \overline{X} = 2,22375 - (-0,8095) \cdot (-0,9025) = 1,4932$.

Одержимо рівняння регресії у вигляді: $y = -0,8095 \cdot x + 1,4932$

Побудуємо графік лінії регресії та експериментальних точок вибірки:

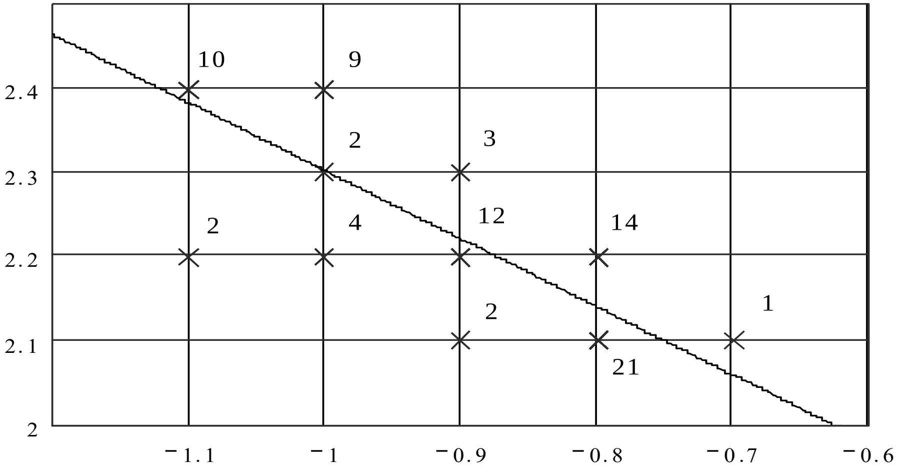


Рисунок 4.2 – Графік лінії регресії

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Агапон Г.И. Сборник задач по теории вероятностей [Текст]. – М.: Высшая школа, – 1986. – 55 с.

2. Гмурман В.Е. Теория вероятностей и математическая статистика [Текст]. – М.: Высшая школа, 1977. – 120 с.

3. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике [Текст]. – М.: Высшая школа, 1975. – 147 с.

4. Карасёв А.И. Курс высшей математики для экономических вузов [Текст]. – М.: Высшая школа, 1982. – 358 с.

5. Кудрявцев В.А., Демидович Б.П. Краткий курс высшей математики [Текст]. – М.: Наука, 1986. – 462 с.

6. Чистяков Б.Н. Курс теории вероятностей [Текст]. – М.: Наука, 1987. – 212 с.

7. Методические указания к выполнению типового расчета по высшей математики «Теория вероятностей» для студентов всех спеціальностей [Текст]: уч. пособие / Л.И. Малыгина, Н.В. Крапива, Д.В. Буряк, И.Б. Подсевалова. – Одесса : ОНПУ, 1995 – 57 с.

ДОДАТКИ

Значення функції Лапласа: $\Phi = \dfrac{1}{\sqrt{2\pi}} \cdot \int_0^x e^{-\frac{t^2}{2}}\,\mathrm{d}t$

x	*Φ*(*x*)	***x***	*Φ*(*x*)	***x***	*Φ*(*x*)	***x***	*Φ*(*x*)	***x***	*Φ*(*x*)
0,00	0,0000	**0,31**	0,1217	**0,62**	0,2324	**0,93**	0,3238	**1,24**	0,3925
0,01	0,0040	**0,32**	0,1255	**0,63**	0,2357	**0,94**	0,3264	**1,25**	0,3944
0,02	0,0080	**0,33**	0,1293	**0,64**	0,2389	**0,95**	0,3289	**1,26**	0,3962
0,03	0,0120	**0,34**	0,1331	**0,65**	0,2422	**0,96**	0,3315	**1,27**	0,3980
0,04	0,0160	**0,35**	0,1368	**0,66**	0,2454	**0,97**	0,3340	**1,28**	0,3997
0,05	0,0199	**0,36**	0,1406	**0,67**	0,2486	**0,98**	0,3365	**1,29**	0,4015
0,06	0,0239	**0,37**	0,1443	**0,68**	0,2517	**0,99**	0,3389	**1,30**	0,4032
0,07	0,0279	**0,38**	0,1480	**0,69**	0,2549	**1,00**	0,3413	**1,31**	0,4049
0,08	0,0319	**0,39**	0,1517	**0,70**	0,2580	**1,01**	0,3438	**1,32**	0,4066
0,09	0,0359	**0,40**	0,1554	**0,71**	0,2611	**1,02**	0,3461	**1,33**	0,4082
0,10	0,0398	**0,41**	0,1591	**0,72**	0,2642	**1,03**	0,3485	**1,34**	0,4099
0,11	0,0438	**0,42**	0,1628	**0,73**	0,2673	**1,04**	0,3508	**1,35**	0,4115
0,12	0,0478	**0,43**	0,1664	**0,74**	0,2704	**1,05**	0,3531	**1,36**	0,4131
0,13	0,0517	**0,44**	0,1700	**0,75**	0,2734	**1,06**	0,3554	**1,37**	0,4147
0,14	0,0557	**0,45**	0,1736	**0,76**	0,2764	**1,07**	0,3577	**1,38**	0,4162

0,15	0,0596	**0,46**	0,1772	**0,77**	0,2794	**1,08**	0,3599	**1,39**	0,4177
0,16	0,0636	**0,47**	0,1808	**0,78**	0,2823	**1,09**	0,3621	**1,40**	0,4192
0,17	0,0675	**0,48**	0,1844	**0,79**	0,2852	**1,10**	0,3643	**1,41**	0,4207
0,18	0,0714	**0,49**	0,1879	**0,80**	0,2881	**1,11**	0,3665	**1,42**	0,4222
0,19	0,0753	**0,50**	0,1915	**0,81**	0,2910	**1,12**	0,3686	**1,43**	0,4236
0,20	0,0793	**0,51**	0,1950	**0,82**	0,2939	**1,13**	0,3708	**1,44**	0,4251
0,21	0,0832	**0,52**	0,1985	**0,83**	0,2967	**1,14**	0,3729	**1,45**	0,4265
0,22	0,0871	**0,53**	0,2019	**0,84**	0,2995	**1,15**	0,3749	**1,46**	0,4279
0,23	0,0910	**0,54**	0,2054	**0,85**	0,3023	**1,16**	0,3770	**1,47**	0,4292
0,24	0,0948	**0,55**	0,2088	**0,86**	0,3051	**1,17**	0,3790	**1,48**	0,4306
0,25	0,0987	**0,56**	0,2123	**0,87**	0,3078	**1,18**	0,3810	**1,49**	0,4319
0,26	0,1026	**0,57**	0,2157	**0,88**	0,3106	**1,19**	0,3830	**1,50**	0,4332
0,27	0,1064	**0,58**	0,2190	**0,89**	0,3133	**1,20**	0,3849	**1,51**	0,4345
0,28	0,1103	**0,59**	0,2224	**0,90**	0,3159	**1,21**	0,3869	**1,52**	0,4357
0,29	0,1141	**0,60**	0,2257	**0,91**	0,3186	**1,22**	0,3888	**1,53**	0,4370
0,30	0,1179	**0,61**	0,2291	**0,92**	0,3212	**1,23**	0,3907	**1,54**	0,4382

Значення функції Лапласа: $\Phi = \frac{1}{\sqrt{2\pi}} \cdot \int_0^x e^{-\frac{t^2}{2}}\,\mathrm{d}t$

x	$\Phi(x)$	**x**	$\Phi(x)$	**x**	$\Phi(x)$	**x**	$\Phi(x)$	**x**	$\Phi(x)$
1,55	0,4394	**1,86**	0,4686	**2,17**	0,4850	**2,48**	0,4934	**2,79**	0,4974
1,56	0,4406	**1,87**	0,4693	**2,18**	0,4854	**2,49**	0,4936	**2,80**	0,4974
1,57	0,4418	**1,88**	0,4699	**2,19**	0,4857	**2,50**	0,4938	**2,82**	0,4976
1,58	0,4429	**1,89**	0,4706	**2,20**	0,4861	**2,51**	0,4940	**2,84**	0,4977
1,59	0,4441	**1,90**	0,4713	**2,21**	0,4864	**2,52**	0,4941	**2,86**	0,4979
1,60	0,4452	**1,91**	0,4719	**2,22**	0,4868	**2,53**	0,4943	**2,88**	0,4980
1,61	0,4463	**1,92**	0,4726	**2,23**	0,4871	**2,54**	0,4945	**2,90**	0,4981
1,62	0,4474	**1,93**	0,4732	**2,24**	0,4875	**2,55**	0,4946	**2,92**	0,4982
1,63	0,4484	**1,94**	0,4738	**2,25**	0,4878	**2,56**	0,4948	**2,94**	0,4984
1,64	0,4495	**1,95**	0,4744	**2,26**	0,4881	**2,57**	0,4949	**2,96**	0,4985
1,65	0,4505	**1,96**	0,4750	**2,27**	0,4884	**2,58**	0,4951	**2,98**	0,4986
1,66	0,4515	**1,97**	0,4756	**2,28**	0,4887	**2,59**	0,4952	**3,00**	0,4987
1,67	0,4525	**1,98**	0,4761	**2,29**	0,4890	**2,60**	0,4953	**3,05**	0,4989
1,68	0,4535	**1,99**	0,4767	**2,30**	0,4893	**2,61**	0,4955	**3,10**	0,4990
1,69	0,4545	**2,00**	0,4772	**2,31**	0,4896	**2,62**	0,4956	**3,15**	0,4992
1,70	0,4554	**2,01**	0,4778	**2,32**	0,4898	**2,63**	0,4957	**3,20**	0,4993
1,71	0,4564	**2,02**	0,4783	**2,33**	0,4901	**2,64**	0,4959	**3,25**	0,4994

1,72	0,4573	**2,03**	0,4788	**2,34**	0,4904	**2,65**	0,4960	**3,30**	0,4995
1,73	0,4582	**2,04**	0,4793	**2,35**	0,4906	**2,66**	0,4961	**3,35**	0,4996
1,74	0,4591	**2,05**	0,4798	**2,36**	0,4909	**2,67**	0,4962	**3,40**	0,4997
1,75	0,4599	**2,06**	0,4803	**2,37**	0,4911	**2,68**	0,4963	**3,45**	0,4997
1,76	0,4608	**2,07**	0,4808	**2,38**	0,4913	**2,69**	0,4964	**3,50**	0,4998
1,77	0,4616	**2,08**	0,4812	**2,39**	0,4916	**2,70**	0,4965	**3,55**	0,4998
1,78	0,4625	**2,09**	0,4817	**2,40**	0,4918	**2,71**	0,4966	**3,60**	0,4998
1,79	0,4633	**2,10**	0,4821	**2,41**	0,4920	**2,72**	0,4967	**3,65**	0,4999
1,80	0,4641	**2,11**	0,4826	**2,42**	0,4922	**2,73**	0,4968	**3,70**	0,4999
1,81	0,4649	**2,12**	0,4830	**2,43**	0,4925	**2,74**	0,4969	**3,75**	0,4999
1,82	0,4656	**2,13**	0,4834	**2,44**	0,4927	**2,75**	0,4970	**3,80**	0,4999
1,83	0,4664	**2,14**	0,4838	**2,45**	0,4929	**2,76**	0,4971	**3,85**	0,4999
1,84	0,4671	**2,15**	0,4842	**2,46**	0,4931	**2,77**	0,4972	**3,90**	0,5000
1,85	0,4678	**2,16**	0,4846	**2,47**	0,4932	**2,78**	0,4973	**3,95**	0,5000

Критичні точки *F*-розподілу Фішера-Снедекора при рівні значущості $\alpha = 0,05$

менш дисп.	Число ступенів вільності більшої дисперсії								
	1	2	3	4	5	6	7	8	9
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38

3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93
400	3,86	3,02	2,63	2,39	2,24	2,12	2,03	1,96	1,90
600	3,86	3,01	2,62	2,39	2,23	2,11	2,02	1,95	1,90
1000	3,85	3,00	2,61	2,38	2,22	2,11	2,02	1,95	1,89

Критичні точки *F*-розподілу Фішера-Снедекора при рівні значущості $\alpha = 0,05$

менш	Число ступенів вільності більшої дисперсії												
дисп.	10	11	12	13	14	16	20	24	30	40	50	100	1000
1	241,9	243,0	243,9	244,7	245,4	246,5	248,0	249,1	250,1	251,1	251,8	253,0	254,2
2	19,40	19,40	19,41	19,42	19,42	19,43	19,45	19,45	19,46	19,47	19,48	19,49	19,49
3	8,79	8,76	8,74	8,73	8,71	8,69	8,66	8,64	8,62	8,59	8,58	8,55	8,53
4	5,96	5,94	5,91	5,89	5,87	5,84	5,80	5,77	5,75	5,72	5,70	5,66	5,63
5	4,74	4,70	4,68	4,66	4,64	4,60	4,56	4,53	4,50	4,46	4,44	4,41	4,37
6	4,06	4,03	4,00	3,98	3,96	3,92	3,87	3,84	3,81	3,77	3,75	3,71	3,67
7	3,64	3,60	3,57	3,55	3,53	3,49	3,44	3,41	3,38	3,34	3,32	3,27	3,23
8	3,35	3,31	3,28	3,26	3,24	3,20	3,15	3,12	3,08	3,04	3,02	2,97	2,93
9	3,14	3,10	3,07	3,05	3,03	2,99	2,94	2,90	2,86	2,83	2,80	2,76	2,71
10	2,98	2,94	2,91	2,89	2,86	2,83	2,77	2,74	2,70	2,66	2,64	2,59	2,54
11	2,85	2,82	2,79	2,76	2,74	2,70	2,65	2,61	2,57	2,53	2,51	2,46	2,41
12	2,75	2,72	2,69	2,66	2,64	2,60	2,54	2,51	2,47	2,43	2,40	2,35	2,30
13	2,67	2,63	2,60	2,58	2,55	2,51	2,46	2,42	2,38	2,34	2,31	2,26	2,21
14	2,60	2,57	2,53	2,51	2,48	2,44	2,39	2,35	2,31	2,27	2,24	2,19	2,14
15	2,54	2,51	2,48	2,45	2,42	2,38	2,33	2,29	2,25	2,20	2,18	2,12	2,07
16	2,49	2,46	2,42	2,40	2,37	2,33	2,28	2,24	2,19	2,15	2,12	2,07	2,02
17	2,45	2,41	2,38	2,35	2,33	2,29	2,23	2,19	2,15	2,10	2,08	2,02	1,97
18	2,41	2,37	2,34	2,31	2,29	2,25	2,19	2,15	2,11	2,06	2,04	1,98	1,92
19	2,38	2,34	2,31	2,28	2,26	2,21	2,16	2,11	2,07	2,03	2,00	1,94	1,88
20	2,35	2,31	2,28	2,25	2,22	2,18	2,12	2,08	2,04	1,99	1,97	1,91	1,85
21	2,32	2,28	2,25	2,22	2,20	2,16	2,10	2,05	2,01	1,96	1,94	1,88	1,82
22	2,30	2,26	2,23	2,20	2,17	2,13	2,07	2,03	1,98	1,94	1,91	1,85	1,79
23	2,27	2,24	2,20	2,18	2,15	2,11	2,05	2,01	1,96	1,91	1,88	1,82	1,76
24	2,25	2,22	2,18	2,15	2,13	2,09	2,03	1,98	1,94	1,89	1,86	1,80	1,74
26	2,22	2,18	2,15	2,12	2,09	2,05	1,99	1,95	1,90	1,85	1,82	1,76	1,70
28	2,19	2,15	2,12	2,09	2,06	2,02	1,96	1,91	1,87	1,82	1,79	1,73	1,66
30	2,16	2,13	2,09	2,06	2,04	1,99	1,93	1,89	1,84	1,79	1,76	1,70	1,63
40	2,08	2,04	2,00	1,97	1,95	1,90	1,84	1,79	1,74	1,69	1,66	1,59	1,52
50	2,03	1,99	1,95	1,92	1,89	1,85	1,78	1,74	1,69	1,63	1,60	1,52	1,45
60	1,99	1,95	1,92	1,89	1,86	1,82	1,75	1,70	1,65	1,59	1,56	1,48	1,40
80	1,95	1,91	1,88	1,84	1,82	1,77	1,70	1,65	1,60	1,54	1,51	1,43	1,34
100	1,93	1,89	1,85	1,82	1,79	1,75	1,68	1,63	1,57	1,52	1,48	1,39	1,30
200	1,88	1,84	1,80	1,77	1,74	1,69	1,62	1,57	1,52	1,46	1,41	1,32	1,21
400	1,85	1,81	1,78	1,74	1,72	1,67	1,60	1,54	1,49	1,42	1,38	1,28	1,15
600	1,85	1,80	1,77	1,74	1,71	1,66	1,59	1,54	1,48	1,41	1,37	1,27	1,13
1000	1,84	1,80	1,76	1,73	1,70	1,65	1,58	1,53	1,47	1,41	1,36	1,26	1,11

Критичні точки розподілу χ^2

k/α	0,01	0,025	0,05	0,95	0,975	0,99
1	6,63490	5,02389	3,84146	0,00393	0,00098	0,00016
2	9,21034	7,37776	5,99146	0,10259	0,05064	0,02010
3	11,34487	9,34840	7,81473	0,35185	0,21580	0,11483
4	13,2767	11,14329	9,48773	0,71072	0,48442	0,29711
5	15,08627	12,8325	11,0705	1,14548	0,83121	0,55430
6	16,81189	14,44938	12,59159	1,63538	1,23734	0,87209
7	18,47531	16,01276	14,06714	2,16735	1,68987	1,23904
8	20,09024	17,53455	15,50731	2,73264	2,17973	1,64650
9	21,66599	19,02277	16,91898	3,32511	2,70039	2,08790
10	23,20925	20,48318	18,30704	3,94030	3,24697	2,55821
11	24,72497	21,92005	19,67514	4,57481	3,81575	3,05348
12	26,21697	23,33666	21,02607	5,22603	4,40379	3,57057
13	27,68825	24,7356	22,36203	5,89186	5,00875	4,10692
14	29,14124	26,11895	23,68479	6,57063	5,62873	4,66043
15	30,57791	27,48839	24,99579	7,26094	6,26214	5,22935
16	31,99993	28,84535	26,29623	7,96165	6,90766	5,81221
17	33,40866	30,19101	27,58711	8,67176	7,56419	6,40776
18	34,80531	31,52638	28,86930	9,39046	8,23075	7,01491
19	36,19087	32,85233	30,14353	10,11701	8,90652	7,63273
20	37,56623	34,16961	31,41043	10,85081	9,59078	8,26040
21	38,93217	35,47888	32,67057	11,59131	10,2829	8,89720
22	40,28936	36,78071	33,92444	12,33801	10,98232	9,54249
23	41,63840	38,07563	35,17246	13,09051	11,68855	10,19572
24	42,97982	39,36408	36,41503	13,84843	12,40115	10,85636
25	44,31410	40,64647	37,65248	14,61141	13,11972	11,52398
26	45,64168	41,92317	38,88514	15,37916	13,84391	12,19815
27	46,96294	43,19451	40,11327	16,15140	14,57338	12,87850
28	48,27824	44,46079	41,33714	16,92788	15,30786	13,56471
29	49,58788	45,72229	42,55697	17,70837	16,04707	14,25645
30	50,89218	46,97924	43,77297	18,49266	16,79077	14,95346
31	52,19139	48,23189	44,98534	19,28057	17,53874	15,65546
32	53,48577	49,48044	46,19426	20,07191	18,29076	16,36222
33	54,77554	50,72508	47,39988	20,86653	19,04666	17,07351
34	56,06091	51,96600	48,60237	21,66428	19,80625	17,78915
35	57,34207	53,20335	49,80185	22,46502	20,56938	18,50893
36	58,61921	54,43729	50,99846	23,26861	21,33588	19,23268
37	59,89250	55,66797	52,19232	24,07494	22,10563	19,96023
38	61,16209	56,89552	53,38354	24,8839	22,87848	20,69144
39	62,42812	58,12006	54,57223	25,69539	23,65432	21,42616

k/α	0,01	0,025	0,05	0,95	0,975	0,99
40	63,69074	59,34171	55,75848	26,5093	24,43304	22,16426
41	64,95007	60,56057	56,94239	27,32555	25,21452	22,90561
42	66,20624	61,77676	58,12404	28,14405	25,99866	23,65009
43	67,45935	62,99036	59,30351	28,96472	26,78537	24,39760
44	68,70951	64,20146	60,48089	29,78748	27,57457	25,14803
45	69,95683	65,41016	61,65623	30,61226	28,36615	25,90127
46	71,20140	66,61653	62,82962	31,43900	29,16005	26,65724
47	72,44331	67,82065	64,00111	32,26762	29,95620	27,41585
48	73,68264	69,02259	65,17077	33,09808	30,75451	28,17701
49	74,91947	70,22241	66,33865	33,93031	31,55492	28,94065
50	76,15389	71,42020	67,50481	34,76425	32,35736	29,70668

ДЛЯ НОТАТОК