

# ЛАБОРАТОРНА РОБОТА №1

## ПОБУДОВА СТАТИСТИЧНИХ РЯДІВ І ОБЧИСЛЕННЯ ЇХ СТАТИСТИЧНИХ ХАРАКТЕРИСТИК

### Статистичні ряди та їх основні характеристики

Вивчення розподілу випадкової величини переважно починають з отримання статистичного матеріалу та його впорядкування. Припустимо, що в результаті виконання  $n$  незалежних випробувань значення  $x_1$  випадкової величини  $X$  спостерігалось  $n_1$  разів,  $x_2$  —  $n_2$  разів, ...,  $x_k$  —  $n_k$  разів ( $x_1 < x_2 < \dots < x_k$ ),  $n = \sum_{i=1}^k n_i$ . Множину всіх можливих значень випадкової величини  $X$  називають генеральною сукупністю, а множину значень  $x_i$ ,  $i = 1, 2, \dots, k$ , отриманих у результаті випробувань, — вибіркою з генеральної сукупності,  $n$  — обсягом вибірки. Спостережувані значення  $x_i$ ,  $i = 1, 2, \dots, k$  називаються варіантами, послідовність варіант, записаних за зростанням — варіаційним рядом, числа  $n_i$  — частотами, відношення  $\omega_i = \frac{n_i}{n}$  — відносними частотами. Статистичним рядом (розподілом) вибірки називається відповідність між спостережуваними значеннями випадкової величини та їх частотами або відносними частотами.

Статистичний ряд можна подати у вигляді таких таблиць:

$x_i$	$x_1$	$x_2$	...	$x_k$
$n_i$	$n_1$	$n_2$	...	$n_k$

 — статистичний ряд частот, (1.1)

де  $n = \sum_{i=1}^k n_i$  — обсяг вибірки, або

$x_i$	$x_1$	$x_2$	...	$x_k$
$\omega_i$	$\omega_1$	$\omega_2$	...	$\omega_k$

 — статистичний ряд відносних частот, (1.2)

де  $\sum_{i=1}^k \omega_i = 1$ .

Поряд з розглянутим статистичним рядом, який називається дискретним, для опису вибірки використовують також і інтервальний статистичний ряд (інша назва — неперервний). Інтервальний статистичний ряд використовують, якщо обсяг вибірки великий, а також тоді, коли члени варіаційного ряду є значеннями неперервної випадкової величини.

Щоб побудувати інтервальний статистичний ряд, необхідно весь відрізок, на якому розміщені варіанти, поділити на скінченну кількість  $s$  про-

міжків без спільних точок:  $\Delta_1 = [z_0; z_1)$ ,  $\Delta_2 = [z_1; z_2)$ ,  $\dots$ ,  $\Delta_s = [z_{s-1}; z_s]$ ,  $x_1 = z_0 < z_1 < z_2 < \dots < z_s = x_k$  і підрахувати кількість  $n_j$  ( $j = 1, 2, \dots, s$ ), тих значень  $x_i$ , що в результаті спостережень потрапили на інтервал  $\Delta_j$ . Інтервальний статистичний ряд подають у вигляді таблиці

$\Delta_j$	$[z_0; z_1)$	$[z_1; z_2)$	$\dots$	$[z_{s-1}; z_s]$
$n_j$	$n_1$	$n_2$	$\dots$	$n_s$

(1.3)

Частинні проміжки  $\Delta_j$  можуть мати як різні, так і однакові довжини (останній випадок зручніший для подальших досліджень). Для визначення оптимальної кількості  $s$  проміжків, за якої інтервальний статистичний ряд не буде занадто громіздким, зберігаючи при цьому особливості генеральної сукупності, рекомендуємо за  $s$  вибрати ціле число, близьке до  $1 + 3, 2 \lg n$ :

$$s \approx 1 + 3, 2 \lg n,$$

де  $n$  – обсяг вибірки. Тоді у випадку проміжків однакової довжини довжиною проміжка буде  $h = \frac{x_k - x_1}{s}$ .

Замінюючи в (1.3)  $n_j$  на  $\omega_j = \frac{n_j}{n}$ , отримаємо інтервальний ряд відносних частот.

Інтервальний статистичний ряд (1.3) в разі необхідності можна замінити дискретним. Для цього за значення  $x_j$  відповідного дискретного статистичного ряду виду (1.1) досить вибрати середини  $x_j = \frac{z_{j-1} + z_j}{2}$  частинних проміжків  $\Delta_j = [z_{j-1}; z_j)$  ( $j = 1, 2, \dots, s$ ), а відповідні їм значення частот  $n_j$  залишити без змін.

Статистичний ряд можна задати графічно у вигляді полігона або гістограми.

Полігоном частот дискретного статистичного ряду (1.1) називають ламану, відрізки якої послідовно з'єднують точки  $(x_1; n_1)$ ,  $(x_2; n_2)$ ,  $\dots$ ,  $(x_k; n_k)$ .

Щоб побудувати полігон частот, на осі абсцис відкладають варіанти  $x_i$ , а на осі ординат — відповідні їм частоти  $n_i$ . Точки  $(x_i; n_i)$  з'єднують відрізками прямих і отримують полігон частот.

Полігоном відносних частот дискретного статистичного ряду (1.2) називають ламану, відрізки якої послідовно з'єднують точки  $(x_1; \omega_1)$ ,  $(x_2; \omega_2)$ ,  $\dots$ ,  $(x_k; \omega_k)$ .

Щоб побудувати полігон відносних частот, на осі абсцис відкладають варіанти  $x_i$ , а на осі ординат — відповідні їм відносні частоти  $\omega_i$ . Точки  $(x_i; \omega_i)$  з'єднують відрізками прямих і отримують полігон відносних частот.

Інтервальний статистичний ряд (1.3) можна графічно зображати у вигляді гістограми.

Гістограмою частот називається східчаста фігура, складена з прямокутників, основами яких є частинні інтервали  $[z_{i-1}; z_i)$  ( $i = 1, 2, \dots, s$ ), а площі яких дорівнюють  $n_i$ . Щоб побудувати гістограму частот, на осі абсцис відкладають частинні інтервали  $[z_{i-1}; z_i)$  і над кожним з них на відстані  $\frac{n_i}{z_i - z_{i-1}}$  проводять відрізок паралельно до осі абсцис. Отже, площа гістограми частот дорівнює обсягові вибірки  $n$ .

Якщо висоту кожного прямокутника, що входить до складу гістограми частот, зменшити в  $n$  разів, то дістанемо гістограму відносних частот. Отже, гістограму відносних частот складають прямокутники, основами яких є відрізки  $[z_{i-1}; z_i)$  ( $i = 1, 2, \dots, s$ ), а висоти прямокутників дорівнюють відповідно  $\frac{\omega_i}{z_i - z_{i-1}}$ . Тому площі цих прямокутників дорівнюють  $\omega_i$ , а площа гістограми відносних частот дорівнює 1.

Якщо дано статистичний розподіл (1.1), (1.2) або (1.3), то накопиченою частотою  $m_i$  (накопиченою відносною частотою  $\eta_i$ ) значення  $x_i$  або проміжку  $[z_{i-1}; z_i)$  називають суму значень частот (відповідно відносних частот) до цього значення чи проміжку, включаючи їх. Отже,

$$m_i = \sum_{j=1}^i n_j = n_1 + n_2 + \dots + n_i \quad \text{— накопичена частота,}$$

$$\eta_i = \sum_{j=1}^i \omega_j = \frac{n_1 + n_2 + \dots + n_i}{n} \quad \text{— накопичена відносна частота.}$$

Для кожного  $x \in \mathbb{R}$  позначимо  $m(x)$  суму частот тих варіант, що менші за  $x$ , тобто  $m(x) = \sum_{x_i < x} n_i$ .

Емпіричною функцією розподілу  $F^*(x)$  називають функцію, яка для кожного  $x \in \mathbb{R}$  дорівнює відношенню величини  $m(x)$  до обсягу вибірки:

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

Отже, емпірична функція розподілу встановлює для кожного значення  $x \in \mathbb{R}$  відносну частоту події  $\{X < x\}$ . Нагадаємо, що функція розподілу  $F(x) = P(X < x)$  (яку називають ще теоретичною функцією розподілу) визначає ймовірність тієї ж події  $\{X < x\}$ .

Для дискретного статистичного ряду (1.1) емпіричну функцію розподілу  $F^*(x)$  можна задати у вигляді таблиці:

$x$	$(-\infty; x_1]$	$(x_1; x_2]$	$(x_2; x_3]$	$\dots$	$(x_{k-1}; x_k]$	$(x_k; +\infty)$
$F^*(x)$	0	$\frac{n_1}{n}$	$\frac{n_1 + n_2}{n}$	$\dots$	$\frac{1}{n} \sum_{i=1}^{k-1} n_i$	1

На кожному з інтервалів  $(x_i; x_{i+1}]$ ,  $i = 1, 2, \dots, k-1$  та поза ними функція  $F^*(x)$  є сталою, її графік має вигляд, зображений на рис. 1.1.

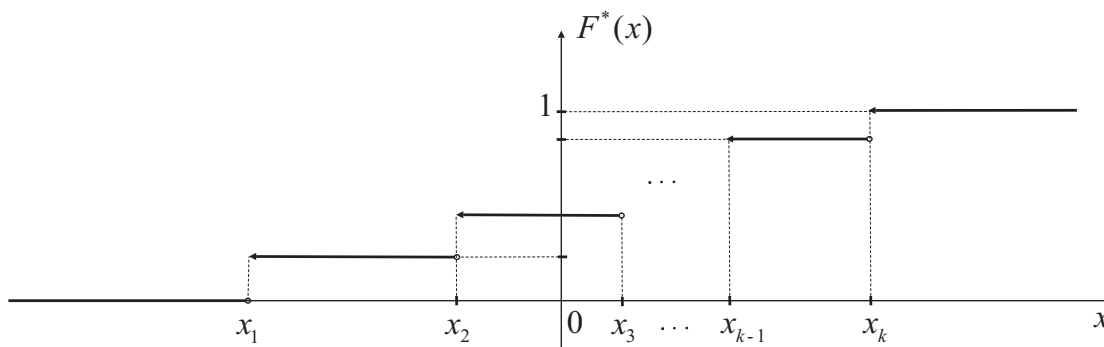


Рис.1.1. Емпірична функція розподілу для дискретного ряду

Для інтервального статистичного ряду (1.3) емпіричну функцію розподілу  $F^*(x)$  будують так: спочатку визначають функцію  $F^*(x)$  у точках, що відповідають кінцям проміжків  $\Delta_j$  ( $j = 1, 2, \dots, s$ ):

$$F(z_0) = 0, \quad F(z_1) = \frac{n_1}{n}, \quad F(z_2) = \frac{n_1 + n_2}{n}, \quad \dots, \quad F(z_s) = \frac{n_1 + n_2 + \dots + n_s}{n} = 1.$$

Далі покладають  $F^*(x) = 0$ , якщо  $x < z_0$ ,  $F^*(x) = 1$ , якщо  $x > z_s$ , а на кожен частковий проміжок  $\Delta_j$  функцію  $F^*(x)$  продовжують лінійно і неперервно. Так побудовану функцію називають емпіричною функцією розподілу  $F^*(x)$  для інтервального ряду (1.3); її графік показано на рис.1.2.

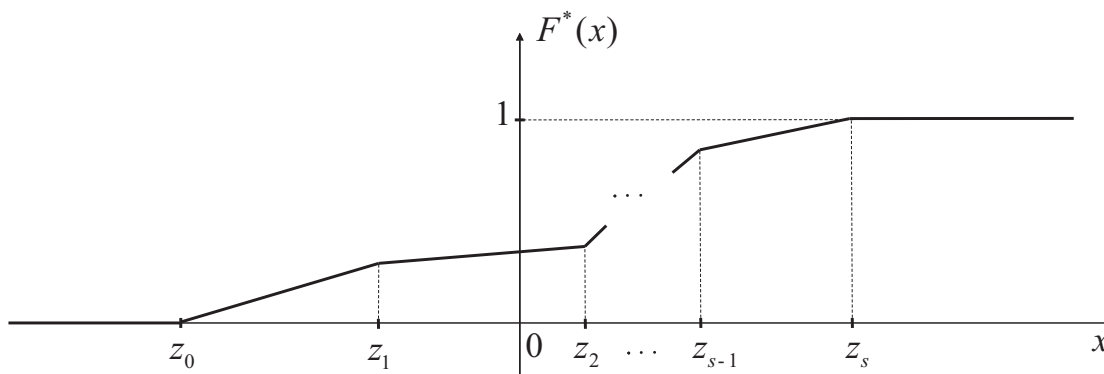


Рис.1.2. Емпірична функція розподілу для інтервального статистичного ряду

У практичних задачах часто замість повного вивчення даних вибірки достатньо буває обмежитися знаходженням деяких її числових характеристик. Назвемо основні з них.

Вибірковим середнім значенням  $\bar{x}_B$  дискретного статистичного ряду (1.1) називають середнє арифметичне варіант  $x_i$  із врахуванням їх частот:

$$\bar{x}_B = \frac{1}{n} \sum_{i=1}^k n_i x_i = \frac{1}{n} (n_1 x_1 + n_2 x_2 + \dots + n_k x_k). \quad (1.4)$$

Модою  $M_0(x)$  статистичного ряду (1.1) називають те значення  $x_i$ , якому відповідає найбільша частота  $n_i$ . У випадку інтервального ряду (1.3) визначають модальний інтервал, тобто інтервал, якому відповідає найбільша частота. В межах цього інтервалу можна встановити значення моди (див. [3], ст. 294).

Медіаною  $M_e(x)$  дискретного статистичного ряду (1.1) називають ту варіанту  $x_e$ , з обидвох боків від якої міститься не більш ніж половина членів статистичного ряду, враховуючи їх частоти.

Щоб знайти медіану інтервального статистичного ряду, необхідно спочатку виділити медіанний інтервал, тобто той частинний інтервал  $[z_{M-1}; z_M)$ , ліворуч і праворуч від якого знаходиться не більше ніж половина елементів вибірки, враховуючи їх частоти. За медіану можна наближено взяти середину медіанного інтервалу. Положення медіани можна уточнити, а саме: якщо  $[z_{M-1}; z_M)$  — медіанний інтервал,  $n_M$  — частота, що йому відповідає,  $m_{M-1}$  — накопичена частота попереднього інтервалу,  $n$  — обсяг вибірки, то медіану розраховують за формулою

$$M_e(x) = z_{M-1} + \frac{z_M - z_{M-1}}{n_M} \left( \frac{n}{2} - m_{M-1} \right).$$

Вибіркове середнє значення, мода і медіана — це характеристики положення випадкової величини, обчислені за даними вибірки.

Перейдемо до розгляду інших важливих характеристик випадкової величини — характеристик розсіяння.

Розмахом вибірки  $r$  називають різницю між найбільшим і найменшим значеннями її варіант:

$$r = x_k - x_1.$$

Вибірковою дисперсією  $D_B$  статистичного ряду називають середнє значення квадратів відхилень його членів від вибіркового середнього значення  $\bar{x}_B$  статистичного ряду:

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

Для обчислення вибіркової дисперсії зручно використовувати формулу

$$D_B = \overline{(x^2)} - (\bar{x}_B)^2 = \frac{1}{n} \sum_{i=1}^k x_i^2 n_i - (\bar{x}_B)^2. \quad (1.5)$$

Розмірність дисперсії  $D_B$  дорівнює квадрату розмірності заданої випадкової величини, що створює деяку незручність у дослідженнях. Цього недоліку позбавлене середнє квадратичне відхилення  $\sigma_B = \sqrt{D_B}$ , розмірність якого така ж, як і розмірність випадкової величини.

Характеристикою розсіяння даних вибірки відносно середнього значення є коефі-цієнт варіації статистичного ряду, що є відношенням вибіркового середнього квадратичного відхилення до вибіркового середнього:  $V = \frac{\sigma_B}{\bar{x}_B}$ .

У математичній статистиці використовуються ще й інші числові характеристики статистичних рядів. Найважливішими з них є центральні й початкові емпіричні моменти.

Центральним емпіричним моментом  $\mu_j$  порядку  $j$  називають середнє значення степенів порядку  $j$  відхилень членів статистичного ряду від його середнього значення  $\bar{x}_B$ :

$$\mu_j = \overline{(x - \bar{x}_B)^j} = \frac{1}{n} \sum_{i=1}^k (x_i - \bar{x}_B)^j n_i. \quad (1.6)$$

Зокрема,  $\mu_0 = 1$ ,  $\mu_1 = 0$ ,  $\mu_2 = D_B$ .

Числовими характеристиками, що використовуються для оцінки відхилення розподілу даного статистичного ряду від нормального розподілу, є асиметрія та ексцес.

Асиметрією статистичного розподілу вибірки називають відношення центрального емпіричного моменту третього порядку до куба середнього квадратичного відхилення:

$$A = \frac{\mu_3}{\sigma_B^3} = \frac{\frac{1}{n} \sum_{i=1}^k (x_i - \bar{x}_B)^3 n_i}{\sigma_B^3}. \quad (1.7)$$

Ексцесом статистичного розподілу вибірки називають число, що обчислюють за формулою

$$E = \frac{\mu_4}{\sigma_B^4} - 3 = \frac{\frac{1}{n} \sum_{i=1}^k (x_i - \bar{x}_B)^4 n_i}{\sigma_B^4} - 3. \quad (1.8)$$

Якщо випадкова величина  $X$  розподілена за нормальним законом, то її асиметрія та ексцес дорівнюють нулеві. Тому чим більше віддалені від нуля асиметрія та ексцес статистичного ряду, тим менше підстав сподіватися, що вибірка, з якої утворено статистичний ряд, зроблена з нормально розподіленої генеральної сукупності.

## ЗАВДАННЯ ДО ЛАБОРАТОРНОЇ РОБОТИ

За поданими нижче в табл. 1.1 даними дослідів необхідно:

- 1) побудувати дискретний статистичний ряд;
- 2) обчислити розмах вибірки; моду та медіану дискретного ряду;
- 3) побудувати полігон частот (або полігон відносних частот);
- 4) для даної вибірки утворити інтервальний статистичний ряд, знайти його моду та медіану;
- 5) побудувати гістограму частот (або гістограму відносних частот);
- 6) побудувати емпіричну функцію розподілу та її графік;
- 7) обчислити середні значення дискретного та інтервального статистичних рядів.

Завдання 8 – 11 виконати лише для інтервального статистичного ряду:

- 8) обчислити дисперсію і середньоквадратичне відхилення статистичного ряду;
- 9) обчислити коефіцієнт варіації;
- 10) обчислити центральні емпіричні моменти третього і четвертого порядків;
- 11) обчислити асиметрію та ексцес.

Вибірki для окремих варіантів формуються згідно з табл. 1.2.

Таблиця 1.1

Номери дослідів	Числові дані									
1-20	21	23	18	31	15	24	20	28	19	25
	32	16	25	13	30	26	22	23	17	21
21-40	24	10	29	37	23	11	34	15	27	13
	33	25	20	26	21	28	33	24	14	35
41-60	31	16	24	38	27	25	22	17	29	23
	23	30	13	34	26	18	31	25	19	36
61-80	26	15	27	35	24	12	23	18	28	32
	21	25	20	28	22	29	17	24	16	26
81-100	36	23	33	27	19	26	37	30	31	25
	25	17	30	24	32	38	27	18	23	22
101-120	21	27	10	28	33	29	36	15	26	24
	31	21	23	17	22	32	15	25	18	22
121-140	16	35	31	20	24	26	36	29	25	19
	27	24	30	22	27	21	28	24	12	23
141-160	20	19	15	23	25	38	26	31	27	37
	24	11	28	18	20	30	22	27	33	21
161-180	13	26	34	31	29	24	30	28	25	23
	32	21	10	22	16	27	25	24	23	24
181-200	35	29	20	17	30	23	32	18	26	31
	38	12	25	33	28	15	30	19	22	10

Таблиця 1.2

Номер варіанта	Номери дослідів	Номер варіанта	Номери дослідів	Номер варіанта	Номери дослідів
0	101-200				
1	91-190	11	6-105	21	3-102
2	81-180	12	16-115	22	13-112
3	71-170	13	26-125	23	23-122
4	61-160	14	36-135	24	33-132
5	51-150	15	46-145	25	43-142
6	41-140	16	56-155	26	53-152
7	31-130	17	66-165	27	63-162
8	21-120	18	76-175	28	73-172
9	11-110	19	86-185	29	83-182
10	1-100	20	96-195	30	93-192



## ЛАБОРАТОРНА РОБОТА № 2

### СТАТИСТИЧНЕ ОЦІНЮВАННЯ НЕВІДОМИХ ПАРАМЕТРІВ РОЗПОДІЛУ

#### Загальні відомості

Математична статистика займається описом та вивченням результатів дослідів та перевіркою відповідних їм математичних моделей, що містять поняття ймовірності. Відомо, що в математичній статистиці конструюються певні функції від випадкових величин  $x_1, x_2, \dots, x_n$ , поведінка яких із збільшенням  $n$  може бути досить точно передбачена. Такі функції  $Y = \varphi(x_1, x_2, \dots, x_n)$  називають статистиками. Наприклад, якщо  $x_1, x_2, \dots, x_n$  – послідовність випадкових величин, що мають однакові математичні сподівання  $m$  і однакові дисперсії  $D$ , то статистика

$$Y_n = \frac{x_1 + x_2 + \dots + x_n}{n}$$

із зростанням  $n$  для кожного  $\varepsilon > 0$  задовольняє нерівність

$$P(|Y_n - m| > \varepsilon) < \frac{D}{n\varepsilon^2}.$$

З цієї нерівності, яка складає суть однієї з теорем закону великих чисел (теорема Чебишова), видно, що із збільшенням  $n$  статистика  $Y_n$  поводить себе майже як не випадкова величина, причому її дисперсія  $D(Y_n) = \frac{D}{n} \rightarrow 0$  при  $n \rightarrow \infty$ , що ще яскравіше підкреслює сказане.

Зауважимо, що повною характеристикою випадкової величини є її закон розподілу. Для його дослідного встановлення потрібні значні витрати ресурсів і часу. Часто, однак, можна передбачити вигляд закону (пуассонівський, показниковий, нормальний і т.д.). Тоді постає питання про оцінку параметрів, що характеризують цей закон, на підставі значень  $x_1, x_2, \dots, x_n$ , отриманих в результаті спостережень. При цьому будується певна статистика (випадкова величина!), конкретна реалізація якої приймається за числове значення параметра, що оцінюється. Оцінки параметрів, отже, є певними статистиками, тобто випадковими величинами, які в результаті дослідів набувають конкретних числових значень (реалізацій).

Відомо, що, застосувавши вибіркового метод, обчислюють статистичні характеристики вибірки: середні значення (вибіркове середнє, середнє геометричне), міри розсіювання (вибіркиму дисперсію, середнє квадратичне відхилення,

розмах вибірки, коефіцієнт варіації) і т.д. За значенням цих характеристик робиться висновок про відповідні параметри генеральної сукупності (даються статистичні оцінки параметрів генеральної сукупності).

Нехай спостерігається випадкова величина  $X$ , теоретичний закон розподілу якої містить невідомий параметр  $\theta$ . Потрібно знайти відповідну оцінку для параметра  $\theta$  за результатами  $n$  незалежних випробувань, в кожному з яких величина  $X$  набуває певних значень  $x_i$  ( $i = 1, 2, \dots, n$ ). Позначимо через  $\theta^*$  оцінку для параметра  $\theta$ . Очевидно, що

$$\theta^* = \theta^*(x_1, x_2, \dots, x_n).$$

Розв'язання задачі про "найкращу" оцінку невідомого параметра і є предметом статистичного оцінювання. При цьому розрізняють два способи оцінки параметрів: точковий та інтервальний.

Точною називають оцінку, що визначається одним числом, навколо якого групуються значення досліджуваного параметра.

Інтервальною називають оцінку, що визначається двома числами – кінцями інтервалу, в якому з наперед заданою ймовірністю знаходиться невідоме значення параметра.

### Точкові оцінки

Часто на практиці трапляється так, що за точкову оцінку невідомого параметра можна взяти не одну, а декілька різних числових характеристик вибірки, тобто функцій від вибірки. Серед цих характеристик намагаються вибрати "найкращу" (що не завжди вдається), яка задовольняє умови незміщеності, ефективності та обґрунтованості.

Незміщеною називають статистичну оцінку  $\theta^*$ , математичне сподівання якої дорівнює оцінюваному параметру  $\theta$  незалежно від обсягу вибірки, тобто  $M(\theta^*) = \theta$ .

Ефективною називають статистичну оцінку  $\theta^*$ , яка за заданого обсягу вибірки має найменшу можливу дисперсію.

Обґрунтованою (а також умотивованою або конзистентною) називають статистичну оцінку, яка із необмеженим зростанням обсягу вибірки ( $n \rightarrow \infty$ ) прямує за ймовірністю до оцінюваного параметра:

$$\lim_{n \rightarrow \infty} P(|\theta^* - \theta| < \varepsilon) = 1 \quad (\varepsilon > 0).$$

Відомо, що найважливішими числовими характеристиками випадкової величини  $X$  є математичне сподівання  $M(X)$  (характеристика положення випад-

кової величини), дисперсія  $D(X)$  та середнє квадратичне відхилення  $\sigma(X)$  (міри розсіювання навколо середнього значення). Тому природно ставити задачу про знаходження найкращих статистичних оцінок для цих параметрів. Можна показати, що для вибірки, отриманої в результаті незалежних спостережень над випадковою величиною  $X$  і заданої дискретним статистичним рядом виду (1.1), "найкращими" статистичними оцінками для невідомого математичного сподівання  $a$  і середнього квадратичного відхилення  $\sigma$  генеральної сукупності є відповідно вибіркове середнє

$$a^* = \bar{x}_B = \frac{1}{n} \sum_{i=1}^k n_i x_i \quad (2.1)$$

і виправлене вибіркове середнє квадратичне відхилення  $\sigma^* = S = \sqrt{\frac{n}{n-1} D_B}$ , де  $D_B$  задається формулою (1.5) або

$$\sigma^* = S = \left( \frac{1}{n-1} \sum_{i=1}^k n_i (x_i - \bar{x}_B)^2 \right)^{\frac{1}{2}}. \quad (2.2)$$

### Інтервальні оцінки

Продовжуючи задачу оцінювання, припустимо, що знайдено точкову оцінку  $\theta^*$  невідомого параметра  $\theta$  деякого теоретичного розподілу. У деяких випадках доцільно передбачити ще й величину помилок, що можуть виникнути внаслідок заміни числа  $\theta$  його оцінкою  $\theta^*$ . Тоді використовують інтервальне оцінювання, яке будується на підставі поняття надійності й інтервалу надійності (інтервалу довіри).

Надійністю оцінки параметра  $\theta$  за точковою оцінкою  $\theta^*$  називають ймовірність  $\gamma$  того, що інтервал  $(\theta^* - \delta; \theta^* + \delta)$  накриє значення невідомого параметра  $\theta$ .

Випадковий інтервал  $(\theta^* - \delta; \theta^* + \delta)$ , який із заданою ймовірністю  $\gamma$  накриває значення параметра  $\theta$ , називається його інтервалом надійності або  $\gamma$ -надійним інтервалом.

Розглянемо побудову точкових оцінок та інтервалів надійності для невідомих параметрів неперервного (нормального) та дискретного (біномного) розподілів.

**Нормальний розподіл.** Нехай відомо, що випадкова величина  $X$  розподілена за нормальним законом зі щільністю

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right).$$

Подамо способи побудови точкових оцінок та інтервалів надійності для параметрів  $a$  (математичного сподівання) і  $\sigma$  (середньоквадратичного відхилення) генеральної сукупності, якими визначається нормальний закон розподілу.

1. Точковою оцінкою  $a^*$  для математичного сподівання генеральної сукупності  $a$  є вибіркове середнє  $\bar{x}_B$  (див. (2.1)).

Інтервал надійності для  $a$  за вибірковим середнім  $\bar{x}_B$ , якщо  $\sigma$  *відоме*, будується так:

$$\bar{x}_B - \frac{\sigma}{\sqrt{n}}t_\gamma < a < \bar{x}_B + \frac{\sigma}{\sqrt{n}}t_\gamma, \quad (2.3)$$

де  $\frac{\sigma}{\sqrt{n}}t_\gamma = \delta$  – точність оцінки,  $t_\gamma$  – таке значення аргумента  $t$  функції Лапласа  $\Phi(t)$ , для якого  $\Phi(t_\gamma) = \frac{\gamma}{2}$ . Значення  $t_\gamma$  знаходять з таблиць функції Лапласа (див. додаток, табл. 3).

Інтервал надійності для  $a$  за вибірковим середнім  $\bar{x}_B$ , якщо  $\sigma$  *невідоме*, знаходиться згідно з нерівністю:

$$\bar{x}_B - \frac{S}{\sqrt{n}}t_\gamma < a < \bar{x}_B + \frac{S}{\sqrt{n}}t_\gamma, \quad (2.4)$$

де  $S$  – виправлене вибіркове середньоквадратичне відхилення, що визначається (2.2),  $\frac{S}{\sqrt{n}}t_\gamma = \delta$  – точність оцінки,  $t_\gamma$  – таке значення аргумента  $t$  функції розподілу Стюдента  $S(t, n)$ , для якого  $S(t, n) = \frac{\gamma}{2}$ ,  $\gamma$  – задана ймовірність. Значення  $t_\gamma$  знаходять з відповідно складених таблиць для розподілу Стюдента (див. додаток, табл. 4).

2. Точковою оцінкою  $\sigma^*$  для середньоквадратичного відхилення  $\sigma$  генеральної сукупності є виправлене вибіркове середньоквадратичне відхилення  $S$  (див. (2.2)).

Інтервал надійності для  $\sigma$  за виправленим вибірковим середньоквадратичним відхиленням  $S$  будується так:

$$S(1 - q_\gamma) < \sigma < S(1 + q_\gamma), \quad \text{якщо } q_\gamma < 1, \quad (2.5)$$

або

$$0 < \sigma < S(1 + q_\gamma), \quad \text{якщо } q_\gamma > 1,$$

де параметр  $q_\gamma$  певним чином пов'язаний з розподілом  $\chi^2$  і може бути знайдений з відповідно складених таблиць (див. додаток, табл. 5).

**Біномний розподіл.** Нехай проводять незалежні випробування з невідомою ймовірністю  $p$  появи події  $A$  у кожному випробуванні. Випадкова величина  $X$  — кількість появ події  $A$  в  $n$  випробуваннях описується біномним законом розподілу

$$P(X = k) = C_n^k p^k q^{n-k} \quad (q = 1 - p, \quad k = \overline{0, n}).$$

Побудуємо точкову оцінку та інтервал надійності для параметра  $p$ , який визначає біномний розподіл.

Точною оцінкою параметра  $p$  є відносна частота появи події:

$$W_n = \frac{m}{n}, \quad (2.6)$$

де  $m$  — кількість появ події,  $n$  — загальна кількість випробувань.

Інтервал довіри для  $p$  за відносною частотою  $W_n$  з надійністю  $\gamma$  знаходиться згідно з нерівністю

$$\frac{n}{t_\gamma^2 + n} \left[ W_n + \frac{t_\gamma^2}{2n} - \Delta \right] < p < \frac{n}{t_\gamma^2 + n} \left[ W_n + \frac{t_\gamma^2}{2n} + \Delta \right], \quad (2.7)$$

де  $\Delta = t_\gamma \sqrt{\frac{W_n(1 - W_n)}{n} + \left(\frac{t_\gamma}{2n}\right)^2}$ ,  $t_\gamma$  — таке значення аргумента  $t$  функції Лапласа  $\Phi(t)$ , для якого  $\Phi(t_\gamma) = \frac{\gamma}{2}$  (див. табл. 3),  $\gamma$  — задана ймовірність.

## ЗАВДАННЯ ДО ЛАБОРАТОРНОЇ РОБОТИ

### Завдання А

Припускаючи, що у наведених нижче задачах вибірка отримана з нормально розподіленої генеральної сукупності, необхідно:

1) знайти точкові оцінки для математичного сподівання і середнього квадратичного відхилення спостережуваної випадкової величини;

2) знайти  $\gamma$ -надійні інтервали для математичного сподівання і середнього квадратичного відхилення спостережуваної випадкової величини.

Варіанти завдань формуються на підставі поданих нижче задач, для яких значення параметра  $k$  і надійності  $\gamma$  подані у таблиці:

варіант	0	1	2	3	4	5	6	7	8	9	10
$k$	80	40	60	70	30	50	45	1000	1500	2000	2500
$\gamma$	0.99	0.999	0.95	0.99	0.999	0.95	0.99	0.999	0.95	0.99	0.999

варіант	11	12	13	14	15	16	17	18	19	20
$k$	1800	3000	950	960	970	980	990	1000	2.5	2.0
$\gamma$	0.95	0.99	0.999	0.95	0.99	0.999	0.95	0.99	0.999	0.95

варіант	21	22	23	24	25	26	27	28	29	30
$k$	1.5	1.2	3.0	3.4	20	300	400	500	600	700
$\gamma$	0.99	0.999	0.95	0.99	0.999	0.95	0.99	0.999	0.95	0.99

Задача 1 (варіанти 0 – 6). Результати перевірки на міцність 300 зразків дроту одного діаметра, задаються у вигляді таблиці:

Розривне зусилля, кН	$[k; k+2)$	$[k+2; k+4)$	$[k+4; k+6)$	$[k+6; k+8)$
Кількість зразків	20	47	80	89

Розривне зусилля, кН	$[k+8; k+10)$	$[k+10; k+12)$	$[k+12; k+14]$
Кількість зразків	40	16	8

Задача 2 (варіанти 7 – 12). Час роботи 40 довільно вибраних електричних ламп одної потужності відображено в таблиці:

Час роботи, год	$[k; k+100)$	$[k+100; k+200)$	$[(k+200; k+300)$	$[(k+300; k+400)$
Кількість ламп	2	4	9	14

Час роботи, год	$[k+400; k+500)$	$[k+500; k+600)$	$[k+600; k+700]$
Кількість ламп	8	3	1

Задача 3 (варіанти 13 – 18). Результати зважування 300 довільним чином вибраних пакетів цукру, розфасованих на фасувальному пристрої певної моделі, для якого генеральне середнє квадратичне відхилення  $\sigma = 14$ , подано у

таблиці:

Маса пакету, г	$[k;k+10)$	$[k+10;k+20)$	$[k+20;k+30)$	$[k+30;k+40)$
Кількість пакетів	20	47	80	89

Маса пакету, г	$[k+40;k+50)$	$[k+50;k+60)$	$[k+60;k+70]$
Кількість пакетів	40	16	8

Задача 4 (варіанти 19 – 24). Результати вимірювання настригу вовни з 340 довільно вибраних овець фермерського господарства задано у таблиці:

Маса пакету, г	$[k;k+0,5)$	$[k+0,5;k+1,0)$	$[k+1,0;k+1,5)$	$[k+1,5;k+2,0)$
Кількість овець	15	20	30	80

Маса пакету, г	$[k+2,0;k+2,5)$	$[k+2,5;k+3,0)$	$[k+3,0;k+3,5]$
Кількість овець	120	50	25

Задача 5 (варіанти 25 – 30). Результати вимірювання висоти 100 довільно вибраних ялин одного віку, висаджених на ділянці (генеральне середнє квадратичне відхилення  $\sigma = 7$ ) подано у таблиці:

Висота ялини, см	$[k+5;k+10)$	$[k+10;k+15)$	$[k+15;k+20)$	$[k+20;k+25)$
Кількість дерев	6	8	15	40

Висота ялини, см	$[k+25;k+30)$	$[k+30;k+35)$	$[k+35;k+40]$
Кількість дерев	16	8	7

## Завдання Б

Припускаючи, що у наведених нижче задачах вибірка, отримана з генеральної сукупності, розподілена за біномним законом, розв'язати задачі для вказаних значень параметрів  $m, n, \gamma$ .

Задача 1 (варіанти 0 – 6). У деякій місцевості серед досліджених довільно вибраних  $n$  днів року виявилось  $m$  дощових. Знайти точкову оцінку ймовірності  $p$  того, що довільно вибраний день буде дощовий та інтервал надійності для оцінки  $p$  з надійністю  $\gamma$ , де

номер варіанта	k=0	k=1	k=2	k=3	k=4	k=5	k=6
$n$	15	20	25	30	20	25	30
$m$	6	5	7	12	6	4	14
$\gamma$	0.95	0.999	0.99	0.95	0.999	0.99	0.95

Задача 2 (варіанти 7 – 12). Під час епідемії грипу в деякому населеному пункті серед  $n$  досліджених людей зафіксовано  $m$  випадків захворювання грипом. Знайти точкову оцінку ймовірності  $p$  захворювання довільно обраного мешканця населеного пункту та інтервал надійності для оцінки  $p$  з надійністю  $\gamma$ , де

номер варіанта	k=7	k=8	k=9	k=10	k=11	k=12
$n$	40	50	30	50	40	30
$m$	4	6	6	4	6	3
$\gamma$	0.999	0.99	0.95	0.999	0.99	0.95

Задача 3 (варіанти 13 – 18). Після комп'ютерного перекладу тексту філологом виявлено  $m$  сторінок з помилками серед  $n$  довільно вибраних сторінок. Знайти точкову оцінку ймовірності  $p$  помилки на довільно взятій сторінці та інтервал надійності для оцінки  $p$  з надійністю  $\gamma$ , де

номер варіанта	k=13	k=14	k=15	k=16	k=17	k=18
$n$	25	50	75	100	50	40
$m$	3	5	10	15	8	4
$\gamma$	0.95	0.99	0.95	0.99	0.95	0.99

Задача 4 (варіанти 19 – 24). Серед  $n$  довільно вибраних книг бібліотеки виявлено  $m$  книг з пошкодженими сторінками. Знайти точкову оцінку ймовірності  $p$  того, що довільно вибрана книга виявиться з пошкодженнями та



інтервал надійності для оцінки  $p$  з надійністю  $\gamma$ , де

номер варіанта	k=19	k=20	k=21	k=22	k=23	k=24
$n$	40	50	60	75	80	100
$m$	20	25	40	50	50	80
$\gamma$	0.99	0.95	0.99	0.95	0.99	0.95

Задача 5 (варіанти 25 – 30). Під час тестування партії з  $n$  штук автомобілів певної марки успішно витримали тест  $m$  автомобілів. Знайти точкову оцінку ймовірності  $p$  того, що довільно вибраний автомобіль витримає тест та інтервал надійності для оцінки  $p$  з надійністю  $\gamma$ , де

номер варіанта	k=25	k=26	k=27	k=28	k=29	k=30
$n$	30	50	40	50	60	75
$m$	5	5	6	4	6	10
$\gamma$	0.95	0.95	0.99	0.95	0.99	0.95

# ЛАБОРАТОРНА РОБОТА № 3

## ПОБУДОВА ПРЯМОЇ ЛІНІЇ РЕГРЕСІЇ ЗА ДОСЛІДНИМИ ДАНИМИ

Нехай вивчається генеральна сукупність, що характеризується системою двох дискретних кількісних ознак  $(X; Y)$ . Для аналізу залежності між випадковими величинами  $X$  і  $Y$  зроблена вибірка, причому складова  $X$  набула значень  $x_1, x_2, \dots, x_k$ , складова  $Y$  –  $y_1, y_2, \dots, y_l$ , а подія  $\{X = x_i; Y = y_j\}$  мала частоту появи  $n_{ij}$  ( $i = \overline{1, k}; j = \overline{1, l}$ ). Результати цих спостережень зручно записувати у формі кореляційної таблиці.

*Таблиця 3.1.*

$Y \backslash X$	$x_1$	$x_2$	$\dots$	$x_k$	$m_j$
$y_1$	$n_{11}$	$n_{21}$	$\dots$	$n_{k1}$	$m_1$
$y_2$	$n_{12}$	$n_{22}$	$\dots$	$n_{k2}$	$m_2$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$y_l$	$n_{1l}$	$n_{2l}$	$\dots$	$n_{kl}$	$m_l$
$n_i$	$n_1$	$n_2$	$\dots$	$n_k$	$n$

де

$$n_i = n_{i1} + n_{i2} + \dots + n_{il}, \quad (i = 1, 2, \dots, k),$$

$$m_j = n_{1j} + n_{2j} + \dots + n_{kj}, \quad (j = 1, 2, \dots, l)$$

– частоти варіант  $x_i$  та  $y_j$  відповідно, а

$$n = \sum_{i=1}^k n_i = \sum_{j=1}^l m_j$$

– обсяг вибірки.

Регресійний і кореляційний аналізи системи випадкових величин  $(X; Y)$  передбачають розв'язання таких задач:

– встановлення наявності, форми і напрямку зв'язку між випадковими величинами  $X$  і  $Y$ ;

– виявлення аналітичної форми цього зв'язку за допомогою умовних математичних сподівань випадкових величин та функцій регресій (графіки останніх називаються лініями регресій);

– вивчення щільності зв'язку між випадковими величинами  $X$  і  $Y$  на підставі оцінки розсіювання значень випадкових величин відносно побудованих ліній регресій.

У лабораторній роботі вивчаються ці задачі за умови, що між випадковими величинами  $X$  і  $Y$  існує лінійний кореляційний зв'язок (про що свідчить кореляційне поле вибірки). Тоді функції регресії  $Y$  на  $X$  та  $X$  на  $Y$  є лінійними і мають вигляд:

– рівняння прямої лінії регресії  $Y$  на  $X$ :

$$\frac{y_p(x) - \bar{y}_B}{\sigma_B(Y)} = r \frac{x - \bar{x}_B}{\sigma_B(X)} \quad \text{або} \quad y_p(x) - \bar{y}_B = r \frac{\sigma_B(Y)}{\sigma_B(X)} (x - \bar{x}_B); \quad (3.1)$$

– рівняння прямої лінії регресії  $X$  на  $Y$ :

$$\frac{x_p(y) - \bar{x}_B}{\sigma_B(X)} = r \frac{y - \bar{y}_B}{\sigma_B(Y)} \quad \text{або} \quad x_p(y) - \bar{x}_B = r \frac{\sigma_B(X)}{\sigma_B(Y)} (y - \bar{y}_B). \quad (3.2)$$

Наведемо формули для визначення сталих  $\bar{x}_B$ ,  $\bar{y}_B$ ,  $\sigma_B(X)$ ,  $\sigma_B(Y)$ ,  $r$  в (3.1) і (3.2).

Коефіцієнт кореляції  $r$  обчислюють за формулою

$$r = \frac{\overline{(xy)} - \bar{x}_B \bar{y}_B}{\sigma_B(X) \sigma_B(Y)}, \quad (3.3)$$

де середні значення  $\bar{x}_B$ ,  $\bar{y}_B$  та  $\overline{(xy)}$  знаходять за означеннями:

$$\bar{x}_B = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^l x_i n_{ij} = \frac{1}{n} \sum_{i=1}^k x_i n_i = \frac{1}{n} (n_1 x_1 + n_2 x_2 + \dots + n_k x_k), \quad (3.4)$$

$$\bar{y}_B = \frac{1}{n} \sum_{j=1}^l \sum_{i=1}^k y_j n_{ij} = \frac{1}{n} \sum_{j=1}^l y_j m_j = \frac{1}{n} (m_1 y_1 + m_2 y_2 + \dots + m_l y_l). \quad (3.5)$$

$$\overline{(xy)} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^l x_i y_j n_{ij} = \frac{1}{n} (x_1 y_1 n_{11} + x_1 y_2 n_{12} + \dots + x_k y_l n_{kl}). \quad (3.6)$$

Середні квадратичні відхилення  $\sigma_B(X)$  і  $\sigma_B(Y)$  обчислюють із співвідношень

$$\sigma_B^2(X) = \overline{(x^2)} - (\bar{x}_B)^2, \quad \sigma_B^2(Y) = \overline{(y^2)} - (\bar{y}_B)^2, \quad (3.7)$$

$$\overline{(x^2)} = \frac{1}{n} \sum_{i=1}^k x_i^2 n_i = \frac{1}{n} (x_1^2 n_1 + x_2^2 n_2 + \dots + x_k^2 n_k), \quad (3.8)$$

$$\overline{(y^2)} = \frac{1}{n} \sum_{j=1}^l y_j^2 m_j = \frac{1}{n} (y_1^2 m_1 + y_2^2 m_2 + \dots + y_l^2 m_l). \quad (3.9)$$

Обчислення подвійної суми в (3.6)

$$S_{kl} = \sum_{i=1}^k \sum_{j=1}^l x_i y_j n_{ij}$$

для контролю виконують двома способами:

$$\begin{aligned} S_{kl} &= \sum_{i=1}^k x_i \sum_{j=1}^l y_j n_{ij} = \sum_{i=1}^k x_i a_i = \{ \text{тут } a_i = \sum_{j=1}^l y_j n_{ij} \} = x_1(y_1 n_{11} + \\ &+ y_2 n_{12} + \dots + y_l n_{1l}) + x_2(y_1 n_{21} + y_2 n_{22} + \dots + y_l n_{2l}) + \dots + \\ &+ x_k(y_1 n_{k1} + y_2 n_{k2} + \dots + y_l n_{kl}) = x_1 a_1 + x_2 a_2 + \dots + x_k a_k, \end{aligned} \quad (3.10)$$

та

$$\begin{aligned} S_{kl} &= \sum_{j=1}^l y_j \sum_{i=1}^k x_i n_{ij} = \sum_{j=1}^l y_j b_j = \{ \text{тут } b_j = \sum_{i=1}^k x_i n_{ij} \} = y_1(x_1 n_{11} + \\ &+ x_2 n_{21} + \dots + x_k n_{k1}) + y_2(x_1 n_{12} + x_2 n_{22} + \dots + x_k n_{k2}) + \dots + \\ &+ y_l(x_1 n_{1l} + x_2 n_{2l} + \dots + x_k n_{kl}) = y_1 b_1 + y_2 b_2 + \dots + y_l b_l. \end{aligned} \quad (3.11)$$

За результатами обчислень можна скласти таблицю (значення  $a_i$  та  $b_j$  ті ж самі, що в (3.10) і (3.11)):

$Y \backslash X$	$x_1$	$x_2$	$\dots$	$x_k$	$m_j$	$y_j m_j$	$y_j^2 m_j$	$b_j$
$y_1$	$n_{11}$	$n_{21}$	$\dots$	$n_{k1}$	$m_1$	$y_1 m_1$	$y_1^2 m_1$	$b_1$
$y_2$	$n_{12}$	$n_{22}$	$\dots$	$n_{k2}$	$m_2$	$y_2 m_2$	$y_2^2 m_2$	$b_2$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$y_l$	$n_{1l}$	$n_{2l}$	$\dots$	$n_{kl}$	$m_l$	$y_l m_l$	$y_l^2 m_l$	$b_l$
$n_i$	$n_1$	$n_2$	$\dots$	$n_k$	$n$	$\sum_{j=1}^l y_j m_j$	$\sum_{j=1}^l y_j^2 m_j$	$\sum_{j=1}^l y_j b_j$
$x_i n_i$	$x_1 n_1$	$x_2 n_2$	$\dots$	$x_k n_k$	$\sum_{i=1}^k x_i n_i$			
$x_i^2 n_i$	$x_1^2 n_1$	$x_2^2 n_2$	$\dots$	$x_k^2 n_k$	$\sum_{i=1}^k x_i^2 n_i$			
$a_i$	$a_1$	$a_2$	$\dots$	$a_k$	$\sum_{i=1}^k x_i a_i$			

Кутові коефіцієнти

$$\rho_{yx} = r \frac{\sigma_B(Y)}{\sigma_B(X)} \quad \text{і} \quad \rho_{xy} = r \frac{\sigma_B(X)}{\sigma_B(Y)} \quad (3.12)$$

прямих регресій (3.1) і (3.2) називають коефіцієнтами регресій відповідно  $\underline{Y}$  на  $\underline{X}$  і  $\underline{X}$  на  $\underline{Y}$ .

Прямі регресій (3.1) і (3.2) проходять через точку  $C(\bar{x}_B, \bar{y}_B)$ .

Побудувавши рівняння прямих регресії, знаходять щільність розсіювання випадкових величин  $X$  і  $Y$  відносно цих прямих. Мірою розсіювання системи випадкових величин  $(X; Y)$  навколо прямої регресії можна вибирати різні характеристики. Зокрема, обчислюють дисперсію величини  $Y$  відносно лінії регресії  $y_p(x)$ :

$$\sigma^2(y, y_p(x)) = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^l (y_j - y_p(x_i))^2 n_{ij} = \overline{((y - y_p(x))^2)} \quad (3.13)$$

або відповідну суму квадратів

$$n\sigma^2(y, y_p(x)) = \sum_{i=1}^k \sum_{j=1}^l (y_j - y_p(x_i))^2 n_{ij} = n\overline{((y - y_p(x))^2)}. \quad (3.14)$$

Ці величини можна також обчислити за формулами

$$\sigma^2(y, y_p(x)) = \sigma_B^2(Y)(1 - r^2), \quad (3.15)$$

$$n\sigma^2(y, y_p(x)) = n\sigma_B^2(Y)(1 - r^2). \quad (3.16)$$

Дисперсію величини  $X$  відносно лінії регресії  $x_p(y)$  та відповідну суму квадратів відхилень можна знайти за формулами:

$$\sigma^2(x, x_p(y)) = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^l (x_i - x_p(y_j))^2 n_{ij} = \overline{((x - x_p(y))^2)}, \quad (3.17)$$

$$n\sigma^2(x, x_p(y)) = \sum_{i=1}^k \sum_{j=1}^l (x_i - x_p(y_j))^2 n_{ij} = n\overline{((x - x_p(y))^2)}, \quad (3.18)$$

або

$$\sigma^2(x, x_p(y)) = \sigma_B^2(X)(1 - r^2), \quad (3.19)$$

$$n\sigma^2(x, x_p(y)) = n\sigma_B^2(X)(1 - r^2). \quad (3.20)$$

Іншою можливою характеристикою розсіювання є сума квадратів відхилень  $\delta^2$  умовних середніх значень  $\bar{y}(x)$  величини  $Y$  від значень на лінії регресії  $y_p(x)$ :

$$\delta^2 = \sum_{i=1}^k \delta_i^2 n_i = \sum_{i=1}^k (\bar{y}(x_i) - y_p(x_i))^2 n_i. \quad (3.21)$$

Щоб обчислити  $\delta^2$ , спочатку знаходять середні значення  $\bar{y}(x_i)$  величини  $Y$  за умови, що  $X = x_i$  ( $i = 1, 2, \dots, k$ ):

$$\bar{y}(x_i) = \frac{1}{n_i} \sum_{j=1}^l n_{ij} y_j = \frac{1}{n_i} (y_1 n_{i1} + y_2 n_{i2} + \dots + y_l n_{il}) = \frac{a_i}{n_i}, \quad (3.22)$$

( $a_i$  такі ж, як у (3.10)), а далі

$$y_p(x_i) = \bar{y}_B + \rho_{yx}(x_i - \bar{x}_B), \quad (3.23)$$

$$\delta_i = \bar{y}(x_i) - y_p(x_i), \quad \delta^2 = \sum_{i=1}^k \delta_i^2 n_i. \quad (3.24)$$

Обчислення зручно супроводжувати таблицею:

$y_p(x_i)$	$y_p(x_1)$	$y_p(x_2)$	$\dots$	$y_p(x_k)$
$\bar{y}(x_i)$	$\bar{y}(x_1)$	$\bar{y}(x_2)$	$\dots$	$\bar{y}(x_k)$
$\delta_i$	$\delta_1$	$\delta_2$	$\dots$	$\delta_k$
$\delta_i^2 n_i$	$\delta_1^2 n_1$	$\delta_2^2 n_2$	$\dots$	$\delta_k^2 n_k$

За величиною  $\delta^2$  можна знайти також дисперсію

$$\sigma^2(\bar{y}(x), y_p(x)) = \overline{((\bar{y}(x) - y_p(x))^2)} = \frac{\delta^2}{n}$$

умовних середніх значень  $\bar{y}(x)$  величини  $Y$  від значень на лінії регресії  $y_p(x)$ .

Аналогічно до формул (3.21) – (3.24) можна записати формули для обчислення суми квадратів відхилень  $\varepsilon^2$  умовних середніх значень  $\bar{x}(y)$  величини  $X$  від значень на лінії регресії  $x_p(y)$ :

$$\varepsilon^2 = \sum_{j=1}^l \varepsilon_j^2 m_j = \sum_{j=1}^l (\bar{x}(y_j) - x_p(y_j))^2 m_j, \quad (3.25)$$

де

$$\bar{x}(y_j) = \frac{1}{m_j} \sum_{i=1}^k n_{ij} x_i = \frac{1}{m_j} (x_1 n_{1j} + x_2 n_{2j} + \dots + x_k n_{kj}) = \frac{b_j}{m_j}, \quad (3.26)$$

( $b_i$  такі ж, як у (3.11)),

$$x_p(y_j) = \bar{x}_B + \rho_{xy}(y_j - \bar{y}_B), \quad (3.27)$$

$$\varepsilon_j = \bar{x}(y_j) - x_p(y_j), \quad \varepsilon^2 = \sum_{j=1}^l \varepsilon_j^2 m_j. \quad (3.28)$$

Подібно складається таблиця умовних середніх значень  $\bar{x}(y)$  величини  $X$  від значень на лінії регресії  $x_p(y)$ :

$x_p(y_j)$	$x_p(y_1)$	$x_p(y_2)$	$\cdots$	$x_p(y_l)$
$\bar{x}(y_j)$	$\bar{x}(y_1)$	$\bar{x}(y_2)$	$\cdots$	$\bar{x}(y_l)$
$\varepsilon_j$	$\varepsilon_1$	$\varepsilon_2$	$\cdots$	$\varepsilon_l$
$\varepsilon_j^2 m_j$	$\varepsilon_1^2 m_1$	$\varepsilon_2^2 m_2$	$\cdots$	$\varepsilon_l^2 m_l$

За величиною  $\varepsilon^2$  знаходять також дисперсію

$$\sigma^2(\bar{x}(y), x_p(y)) = \overline{((\bar{x}(y) - x_p(y))^2)} = \frac{\varepsilon^2}{n}.$$

## ЗАВДАННЯ ДО ЛАБОРАТОРНОЇ РОБОТИ

1. Побудувати кореляційне поле (тобто зобразити варіанти  $(x_i; y_j)$  вибірки точками на площині  $Oxy$ ) і висунути припущення про вид зв'язку між ознаками  $X$  і  $Y$  (лінійний, квадратичний, гіперболічний і т. д.).

2. Припускаючи лінійну залежність між випадковими величинами  $X$  і  $Y$ , необхідно:

- обчислити коефіцієнт кореляції  $r$  (формули (3.3) – (3.11));
- скласти рівняння прямих регресії  $Y$  на  $X$  і  $X$  на  $Y$  (формули (3.1) і (3.2));
- знайти коефіцієнти регресій  $Y$  на  $X$  і  $X$  на  $Y$  (формули (3.12));
- обчислити суму квадратів відхилень випадкової величини  $Y$  від лінії регресії  $y_p(x)$  і суму квадратів відхилень випадкової величини  $X$  від лінії регресії  $x_p(y)$  (формули (3.16) та (3.21));
- обчислити суму квадратів відхилень умовних середніх значень  $\bar{y}(x)$  від лінії регресії  $y_p(x)$  і умовних середніх значень  $\bar{x}(y)$  від лінії регресії  $x_p(y)$  (формули (3.21) – (3.28)).

Задача 1 (для варіантів  $k = 0, 1, 2, \dots, 8$ ). Дано розподіл однотипних підприємств за обсягом продукції  $X$  та її собівартістю  $Y$  (в умовних одиницях):

$Y \backslash X$	2	3	4	5	6
100				$11-k$	$k+1$
200			$2k$	$k$	$10-k$
300			$10-k$	$k+1$	
400		$14-k$	$k$		
500	$12-k$	$k$			
600	$k+1$	$12-k$			

Задача 2 (для варіантів  $k = 9, 10, \dots, 15$ ). Дано розподіл підприємств за обсягом їх основних фондів  $X$  та готовою продукцією  $Y$  (в умовних одиницях):

$Y \backslash X$	20	30	40	50	60	70
15	$k+4$	$21-k$				
25	$18-k$	$2k$	$k+1$			
35		$k$	$22-k$	$17-k$		
45				$k+2$	$k$	$19-k$
55					$24-k$	$k+1$

Задача 3 (для варіантів  $k = 16, 17, \dots, 23$ ). Дано розподіл однотипних підприємств за обсягом їх основних фондів  $X$  і собівартістю продукції  $Y$  (в умовних одиницях):

$Y \backslash X$	1,5	2	2,5	3	3,5	4
8					$k-13$	$26-k$
13				$k-15$	$25-k$	$2k-29$
18			$25-k$	$k-13$	$29-k$	
23	$k-10$	$27-k$	$k-14$			
28	$26-k$	$k-7$				

Задача 4 (для варіантів  $k = 24, 25, \dots, 30$ ). Дано розподіл накладних витрат  $X$  і обсягу виконаних робіт  $Y$  (в умовних одиницях):

$Y \backslash X$	3	5	7	9	11
35	$k-19$	$35-k$			
45	$42-k$	$2k-48$	$k-22$		
55		$k-21$	$46-k$	$39-k$	
65			$k-17$	$k-20$	$33-k$
75				$36-k$	$k-23$
85				$k-18$	$40-k$



## ВКАЗІВКИ ДО ЛАБОРАТОРНИХ РОБІТ № 4, 5

### Про перевірку статистичних гіпотез

У багатьох випадках, досліджуючи вибірку з генеральної сукупності, вдається сформулювати деякі припущення щодо властивостей розподілу ознак всієї генеральної сукупності. Ці припущення мають здебільшого гіпотетичний характер і називаються статистичними гіпотезами. У математичній статистиці розглядаються різні види статистичних гіпотез, наприклад, гіпотези про вигляд розподілу, гіпотези однорідності, незалежності, випадковості, гіпотези про значення невідомого параметра розподілу за умови, що вид закону розподілу відомий (останні гіпотези називають параметричними).

Результатом перевірки статистичної гіпотези є висновок про її істинність або хибність, причому вказується також ймовірність того, що рішення є правильним або помилковим. Проблеми зменшення ймовірності того, що прийняте рішення є помилковим, також розглядають в теорії перевірки статистичних гіпотез.

Для виконання лабораторних робіт необхідно ознайомитися з теоретичним матеріалом про перевірку статистичних гіпотез, наведеним [1, 2, 3].

Перевірку статистичної гіпотези виконують за такою схемою.

1. Формулюють нульову гіпотезу  $H_0$ , альтернативну їй  $H_1$  та задають рівень значущості  $\alpha$ .
2. Вибирають критерій  $K$  для перевірки гіпотези.
3. Відповідно до альтернативної гіпотези встановлюють вид критичної області. Для побудови цієї області за рівнем значущості  $\alpha$  знаходять критичні точки  $K_{кр}$ .
4. За вибіркою обчислюють емпіричне (спостережуване) значення критерію  $K_{емп}$ .
5. Приймають рішення: якщо  $K_{емп}$  потрапляє в критичну область, то  $H_0$  відхиляють; якщо  $K_{емп}$  потрапляє в область прийняття рішення, то  $H_0$  приймають.

### ЛАБОРАТОРНА РОБОТА № 4

#### ПЕРЕВІРКА ГІПОТЕЗ ПРО ЗАКОН

#### РОЗПОДІЛУ ВИПАДКОВОЇ ВЕЛИЧИНИ

У даній лабораторній роботі розв'язується задача про встановлення можливого закону розподілу випадкової величини шляхом перевірки статисти-

чних гіпотез. Припустимо, що для вивчення закону розподілу генеральної сукупності отримана вибірка. Припустимо також, що, враховуючи попередні відомості про генеральну сукупність і результати обробки даних вибірки, можна висловити припущення про характер розподілу генеральної сукупності. Отже, нульова гіпотеза  $H_0$  полягає в тому, що функцією розподілу генеральної сукупності є функція  $F(x)$ . Критерії, за допомогою яких перевіряють гіпотези про вигляд функції розподілу, називають критеріями згоди. Розглянемо застосування критерію згоди  $\chi^2$  (хі-квадрат), що був запропонований К.Пірсоном.

Згідно з цим критерієм для перевірки гіпотези  $H_0$  вводиться випадкова величина (характеристика)

$$K = \chi^2 = \sum_{i=1}^m \frac{(n_i - n'_i)^2}{n'_i} = n \sum_{i=1}^m \frac{(\omega_i - p_i)^2}{p_i}, \quad (4.1)$$

де  $m$  – кількість груп у статистичному розподілі вибірки (під групами розуміють інтервали в інтервальному статистичному ряді чи варіанти у дискретному статистичному ряді);  $n_i$  (чи  $\omega_i$ ) – частота (чи відносна частота), що відповідає  $i$ -й групі;  $p_i$  (чи  $n'_i = np_i$ ) – теоретична ймовірність (чи частота) того, що значення випадкової величини  $X$  належить  $i$ -й групі. Ця ймовірність розрахована на підставі гіпотетичних функції розподілу  $F(x)$  чи щільності  $f(x)$ .

### **Перевірка гіпотези про закон розподілу ймовірності неперервної випадкової величини**

1. Статистичні дані (результати вибірки) записуємо як інтервальний статистичний ряд (перші два рядки таблиці 4.1).
2. Будуємо гістограму частот, на підставі якої висуваємо гіпотезу про закон розподілу.
3. Якщо серед параметрів розподілу є невідомі параметри, то шукаємо їх точкові оцінки.
4. Для кожного інтервалу  $[z_{i-1}; z_i)$  знаходимо теоретичні ймовірності  $p_i$  і теоретичні частоти  $n'_i = np_i$  того, що випадкова величина  $X$  потрапить у цей інтервал. Ці величини шукаємо за гіпотетичною функцією розподілу ймовірностей  $F(x)$  чи щільністю  $f(x)$ , у яких невідомі параметри замінюємо їх точковими оцінками:

$$p_i = P(z_{i-1} \leq X < z_i) = F(z_i) - F(z_{i-1}) \quad (4.2)$$

або

$$p_i = \int_{z_{i-1}}^{z_i} f(x) dx, \quad (i = 1, 2, \dots, m), \quad (4.3)$$

причому  $z_0 = -\infty, z_m = +\infty$  і  $\sum_{i=1}^m p_i = 1$ . Результати обчислень записуємо у третій та четвертий рядки таблиці 4.1.

5. На підставі знайдених  $n'_i$  визначаємо  $\frac{(n_i - n'_i)^2}{n'_i}$  (заповнюємо останній рядок таблиці 4.1) та емпіричне значення критерію

$$K_{emn} = \chi_{emn}^2 = \sum_{i=1}^m \frac{(n_i - n'_i)^2}{n'_i}. \quad (4.4)$$

Таблиця 4.1

$[z_{i-1}; z_i)$	$(z_0; z_1)$	$[z_1; z_2)$	$\dots$	$[z_{m-1}; z_m)$
$n_i$	$n_1$	$n_2$	$\dots$	$n_m$
$p_i$	$p_1$	$p_2$	$\dots$	$p_m$
$n'_i$	$n'_1$	$n'_2$	$\dots$	$n'_m$
$\frac{(n_i - n'_i)^2}{n'_i}$	$\frac{(n_1 - n'_1)^2}{n'_1}$	$\frac{(n_2 - n'_2)^2}{n'_2}$	$\dots$	$\frac{(n_m - n'_m)^2}{n'_m}$

6. Для даного рівня значущості  $\alpha$  і числа ступенів вільності  $k = m - s - 1$  (де  $s$  – кількість оцінених в п. 3 невідомих параметрів гіпотетичного розподілу) з таблиць критичних значень розподілу  $\chi^2$  (додаток, табл. 6) знаходимо критичне значення критерію  $K_{kp} = \chi_{kp}^2(\alpha; k)$ .

7. Зіставляємо значення  $\chi_{emn}^2$  і  $\chi_{kp}^2$ :

- якщо  $\chi_{emn}^2 \geq \chi_{kp}^2$ , то гіпотезу  $H_0$  відхиляють;
- якщо  $\chi_{emn}^2 < \chi_{kp}^2$ , то гіпотезу  $H_0$  приймають.

**Зауваження.** Критерій  $\chi^2$  використовують для досить великих обсягів вибірок  $n \geq 50$  та емпіричних частот  $n_i \geq 5$  і теоретичних частот  $n'_i \geq 5$  в окремих групах. Якщо кількість спостережень в групах мала, то сусідні групи об'єднують.

### Перевірка гіпотези про закон розподілу ймовірності дискретної випадкової величини

Ця перевірка здійснюється за подібним до наведеного вище алгоритмом, але з такими особливостями:

- в п. 1 статистичні дані записуємо як дискретний статистичний ряд (два перші рядки таблиці 4.2);

– в п. 2 будуємо полігон частот, на підставі якого висуваємо гіпотезу про закон розподілу випадкової величини  $X$ ;

– в п. 4 на підставі гіпотетичного закону розподілу шукаємо теоретичні ймовірності  $p_i$  і теоретичні частоти  $n'_i$  для окремих значень  $x_i$  випадкової величини  $X$ . Всі дані записуємо у таблицю 4.2.

Таблиця 4.2

$x_i$	$x_1$	$x_2$	$\dots$	$x_m$
$n_i$	$n_1$	$n_2$	$\dots$	$n_m$
$p_i$	$p_1$	$p_2$	$\dots$	$p_m$
$n'_i$	$n'_1$	$n'_2$	$\dots$	$n'_m$
$\frac{(n_i - n'_i)^2}{n'_i}$	$\frac{(n_1 - n'_1)^2}{n'_1}$	$\frac{(n_2 - n'_2)^2}{n'_2}$	$\dots$	$\frac{(n_m - n'_m)^2}{n'_m}$

Виділимо окремі випадки застосування критерію  $\chi^2$ .

1. Якщо перевіряється гіпотеза про нормальний закон розподілу, функція розподілу ймовірностей якого

$$F(x, a, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-a)^2}{2\sigma^2}} dt,$$

де  $a, \sigma$  – невідомі параметри (причому  $a = M(X)$ ,  $\sigma = \sigma(X)$ ), то вони оцінюються відповідно вибірковою середнім

$$a^* = \bar{x}_B = \frac{1}{n} \sum_{i=1}^m x_i n_i, \quad (4.5)$$

де  $x_i = \frac{z_{i-1} + z_i}{2}$ , та виправленим вибірковою середнім квадратичним відхиленням

$$\sigma^* = S = \left( \frac{1}{n-1} \sum_{i=1}^m n_i (x_i - \bar{x}_B)^2 \right)^{1/2}. \quad (4.6)$$

Тоді теоретичні ймовірності  $p_i$  потрапляння нормально розподіленої випадкової величини  $X$  в проміжок  $[z_{i-1}; z_i)$  обчислюється за формулою

$$p_i = P(z_{i-1} \leq X < z_i) = \Phi\left(\frac{z_i - \bar{x}_B}{S}\right) - \Phi\left(\frac{z_{i-1} - \bar{x}_B}{S}\right) \quad (4.7)$$

за допомогою таблиці значень функції Лапласа  $\Phi(x)$  (додаток, табл. 3).

При обчисленні критичного значення критерію  $\chi^2_{kr}$  кількість ступенів вільності  $k = m - s - 1 = m - 3$  (де  $m$  – число груп статистичного ряду).

2. Якщо перевіряється гіпотеза про показниковий закон розподілу, функція розподілу ймовірностей якого

$$F(x, \lambda) = \begin{cases} 0, & x < 0, \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases}$$

і  $\lambda > 0$  – невідомий параметр (причому  $M(X) = \frac{1}{\lambda}$ ), то його точковою оцінкою  $\lambda^*$  є

$$\lambda^* = \frac{1}{\bar{x}_B}, \quad (4.8)$$

де  $\bar{x}_B$  визначається (4.5). Тоді теоретична ймовірність  $p_i$  потрапляння випадкової величини  $X$  на проміжок  $[z_{i-1}; z_i)$  обчислюється згідно з формулою

$$p_i = P(z_{i-1} \leq X < z_i) = F(z_i) - F(z_{i-1}) = e^{-\frac{z_{i-1}}{\bar{x}_B}} - e^{-\frac{z_i}{\bar{x}_B}} \quad (4.9)$$

(за умови, що  $z_{i-1}, z_i$  – невід’ємні). Якщо  $z_{i-1} \leq 0$  чи  $z_i \leq 0$ , то відповідне значення  $F(z_{i-1}) = 0$  чи  $F(z_i) = 0$ .

При обчисленні критичного значення критерію  $\chi^2_{kr}$  кількість ступенів вільності  $k = m - s - 1 = m - 2$ .

3. Якщо перевіряється гіпотеза про рівномірний закон розподілу, функція розподілу ймовірностей якого

$$F(x, a, b) = \begin{cases} 0, & x < a, \\ \frac{x - a}{b - a}, & a \leq x < b, \\ 1, & x \geq b, \end{cases} \quad (4.10)$$

де  $a, b$  – невідомі параметри, то зважаючи на те, що  $M(X) = \frac{a + b}{2}$ ,  $\sigma(X) = \frac{b - a}{2\sqrt{3}}$ , матимемо систему для знаходження точкових оцінок  $a^*$  і  $b^*$ :

$$\begin{cases} \frac{a^* + b^*}{2} = \bar{x}_B, \\ \frac{b^* - a^*}{2\sqrt{3}} = \sigma_B(X), \end{cases}$$

розв’язками якої є

$$\begin{aligned} a^* &= \bar{x}_B - \sqrt{3}\sigma_B(X), \\ b^* &= \bar{x}_B + \sqrt{3}\sigma_B(X), \end{aligned} \quad (4.11)$$

Тоді теоретичні ймовірності  $p_i$  потрапляння випадкової величини  $X$  на проміжок  $[z_{i-1}; z_i)$  обчислюється

$$p_i = P(z_{i-1} \leq X < z_i) = F(z_i) - F(z_{i-1}),$$

де  $F(z)$  задається виразом (4.10), у якому значення параметрів  $a, b$  замінені знайденими згідно (4.11) їх точковими оцінками  $a^*$  і  $b^*$  відповідно.

Кількість ступенів вільності при цьому  $k = m - 3$ . Якщо за умовою задачі значення параметрів  $a$  та  $b$  можна вважати відомими (див. наприклад задачу 9 завдань до лабораторної роботи), то формули (4.11) для оцінки параметрів  $a$  та  $b$  не застосовують, а кількість ступенів вільності  $k = m - 1$ .

Нехай статистичні дані подано у вигляді дискретного статистичного ряду

Таблиця 4.3

$x_i$	$x_1$	$x_2$	$\cdots$	$x_i$	$\cdots$	$x_m$
$n_i$	$n_1$	$n_2$	$\cdots$	$n_i$	$\cdots$	$n_m$

де  $n = \sum_{i=1}^m n_i$  – обсяг вибірки.

4. Якщо перевіряють гіпотезу про біномний закон розподілу, то випадкова величина  $X$  може набувати цілих значень  $0, 1, 2, \dots, m$  з ймовірностями

$$p_i = P(X = i) = P_m(i) = C_m^i p^i q^{m-i}, \quad (4.12)$$

де  $p$  – параметр розподілу ( $0 < p < 1$ ), який, якщо його значення невідоме, необхідно попередньо оцінити;  $C_m^i = \frac{m!}{i!(m-i)!}$ .

Зважаючи на те, що для біномного закону розподілу  $M(X) = mp$ , де  $m$  – кількість незалежних випробувань, а точковою оцінкою математичного сподівання  $M(X)$  є вибіркове середнє  $\bar{x}_B = \frac{1}{n} \sum_{i=1}^m x_i n_i$ , обчислене згідно з таблицею 4.3, матимемо  $M(X) = \bar{x}_B$ , звідки для оцінки невідомого параметра  $p$  отримуємо формулу

$$p = \frac{\bar{x}_B}{m} = \frac{1}{mn} \sum_{i=1}^m x_i n_i. \quad (4.13)$$

Таким чином, теоретичні ймовірності  $p_i$  обчислюються згідно з (4.12), для якої значення  $p$  знайдені відповідно до (4.13), а теоретичні частоти  $n'_i = np_i$ .

При обчисленні критичного значення критерію  $\chi^2_{kr}$  кількість ступенів вільності така:  $k = m$ , якщо  $p$  вважається відомим;  $k = m - 1$ , якщо  $p$  оцінюється відповідно до (4.13).

5. Якщо висувають гіпотезу про розподіл Пуассона, то випадкова величина  $X$  може набувати значень  $0, 1, 2, 3, \dots, m, \dots$  з ймовірностями

$$p_i = P(X = i) = \frac{\lambda^i e^{-\lambda}}{i!}, \quad (4.14)$$

де  $\lambda > 0$  – параметр розподілу, який, якщо його значення невідоме, необхідно попередньо оцінити.

Враховуючи, що для розподілу Пуассона математичне сподівання  $M(X) = \lambda$  і  $M(X) = \bar{x}_B$  (див. п. 4), матимемо оцінку невідомого параметра  $\lambda$ , отриману на підставі дискретного статистичного ряду (табл. 4.3)

$$\lambda = \frac{1}{n} \sum_{i=1}^m x_i n_i \quad (4.15)$$

Отже, теоретичні ймовірності  $p_i$  знаходяться згідно з (4.14), для якої значення невідомого параметра  $\lambda$  обчислюють відповідно до (4.15), а теоретичні частоти  $n'_i = np_i$ .

При обчисленні критичного значення критерію  $\chi^2_{kr}$  кількість ступенів вільності така:  $k = m - 1$ , якщо параметр  $\lambda$  відомий;  $k = m - 2$ , якщо параметр  $\lambda$  вважається невідомим.

## ЗАВДАННЯ ДО ЛАБОРАТОРНОЇ РОБОТИ

У поданих нижче задачах наведені результати досліджень вибірок з деяких генеральних сукупностей. Необхідно:

- 1) побудувати полігон чи гістограму частот;
- 2) сформулювати гіпотезу про закон розподілу досліджуваної ознаки генеральної сукупності (у задачах 1 – 6 рекомендуємо перевіряти вибірки на нормальний закон, а в задачах 7 – 12 – на інші розподіли, наприклад, рівномірний, показниковий, біномний, закон розподілу Пуассона);
- 3) перевірити сформульовану гіпотезу за критерієм  $\chi^2$ , вибравши рівень значущості  $\alpha = 0,05$ .

Задача 1 (варіанти 0 – 5). Для вивчення технічних властивостей нової марки бетону досліджувалися окремі його зразки. Розподіл кількості  $n_i$  зразків бетону і відповідного їм стискування  $X$  (тобто такого, що веде до руйнування зразка) наведено в таблиці

$X$ , кГ/см <sup>2</sup>	170– –180	180– –190	190– –200	200– –210	210– –220	220– –230	230– –240	240– –250	250– –260	260– –270
$n_i$ (варіант 0)	4	9	32	54	72	65	50	25	12	7
$n_i$ (варіант 1)	4	8	28	52	70	66	52	24	10	6
$n_i$ (варіант 2)	4	8	26	52	76	66	48	24	11	6
$n_i$ (варіант 3)	4	7	28	56	70	60	52	26	10	6
$n_i$ (варіант 4)	4	9	28	48	70	72	52	22	10	6
$n_i$ (варіант 5)	4	8	30	52	64	66	56	24	9	6

Задача 2 (варіанти 6 – 10). Для контролю за готовою продукцією відібрано деталі, що виготовляються на однотипних верстатах-автоматах. Розподіл кількості  $n_i$  відібраних деталей залежно від їх контрольованого розміру  $X$  наведено в таблиці

$X$ , мм	23,2– –23,4	23,4– –23,6	23,6– –23,8	23,8– –24,0	24,0– –24,2	24,2– –24,4	24,4– –24,6	24,6– –24,8	24,8– –25,0	25,0– –25,2
$n_i$ (варіант 6)	1	3	23	79	141	146	75	25	4	1
$n_i$ (варіант 7)	1	3	23	75	141	154	75	23	4	1
$n_i$ (варіант 8)	1	3	21	79	149	146	71	25	4	1
$n_i$ (варіант 9)	1	3	25	79	133	146	79	25	4	1
$n_i$ (варіант 10)	1	3	23	83	141	138	75	27	4	1

Задача 3 (варіанти 11 – 15). Для вдосконалення організації праці на підприємствах торгівлі були зібрані дані про реалізацію за місяць товарів у магазинах міста. Розподіл кількості  $n_i$  магазинів залежно від обсягу реалізації  $X$  товарів наведено в таблиці

$X$ , ум. од.	28– –30	30– –32	32– –34	34– –36	36– –38	38– –40	40– –42	42– –44	44– –46	46– –48
$n_i$ (варіант 11)	1	2	10	51	88	85	45	15	3	1
$n_i$ (варіант 12)	1	2	10	48	88	91	45	13	3	1
$n_i$ (варіант 13)	1	2	8	51	94	85	42	15	3	1
$n_i$ (варіант 14)	1	2	12	51	82	85	48	15	3	1
$n_i$ (варіант 15)	1	2	10	54	88	79	45	17	3	1



Задача 4 (варіанти 16 – 20). З метою пошуку шляхів підвищення продуктивності праці на підприємстві реєструвався час, витрачений робітниками на виготовлення однотипних деталей. Розподіл кількості робітників  $n_i$  залежно від часу  $T$ , витраченого ними на виконання трудових операцій, наведено в таблиці

$T$ , хв.	4,0– –4,5	4,5– –5,0	5,0– –5,5	5,5– –6,0	6,0– –6,5	6,5– –7,0	7,0– –7,5	7,5– –8,0	8,0– –8,5	8,5– –9,0
$n_i$ (варіант 16)	1	3	16	66	112	116	60	23	2	1
$n_i$ (варіант 17)	1	3	16	62	112	122	60	21	2	1
$n_i$ (варіант 18)	1	3	14	66	118	116	56	23	2	1
$n_i$ (варіант 19)	1	3	18	66	106	116	64	23	2	1
$n_i$ (варіант 20)	1	3	16	70	112	110	60	25	2	1

Задача 5 (варіанти 21 – 25). Для вироблення рекомендацій щодо економного витрачання енергоресурсів вивчалися добові витрати електроенергії приватними споживачами в одному з районів міста. Розподіл кількості споживачів  $n_i$  залежно від кількості спожитої електроенергії  $X$  наведено в таблиці

$X$ , квт.-год.	1,0– –1,5	1,5– –2,0	2,0– –2,5	2,5– –3,0	3,0– –3,5	3,5– –4,0	4,0– –4,5	4,5– –5,0	5,0– –5,5	5,5– –6,0
$n_i$ (варіант 21)	2	24	136	470	868	864	469	143	23	1
$n_i$ (варіант 22)	2	24	134	470	876	864	465	143	24	1
$n_i$ (варіант 23)	2	23	136	474	868	856	469	145	23	1
$n_i$ (варіант 24)	2	25	136	466	868	872	469	142	23	1
$n_i$ (варіант 25)	2	24	138	470	860	864	473	143	22	1

Задача 6 (варіанти 26 – 30). У відділі технічного контролю підприємства перевірено партію виготовлених електроламп. Розподіл кількості  $n_i$  перевірених електроламп залежно від часу  $T$  їх справно функціонування наведено в таблиці

$T$ , год.	1000– –1100	1100– –1200	1200– –1300	1300– –1400	1400– –1500	1500– –1600	1600– –1700	1700– –1800	1800– –1900	1900– –2000
$n_i$ (варіант 26)	2	7	50	153	286	291	159	42	9	1
$n_i$ (варіант 27)	2	7	48	153	292	291	155	42	10	1
$n_i$ (варіант 28)	2	6	50	157	286	285	159	44	9	1
$n_i$ (варіант 29)	2	8	50	149	286	297	159	40	9	1
$n_i$ (варіант 30)	2	7	52	153	280	291	163	42	8	1

Задача 7 (варіанти 1 – 5). Для покращання обслуговування сільськогосподарських машин були зібрані дані про вихід з ладу техніки у господарствах району за період весняно-польових робіт. Розподіл кількості  $n_i$  перевірених одиниць техніки залежно від кількості  $X$  поломок наведено у таблиці

$X$ , к-ть поломок	0	1	2	3	4	5
$n_i$ (варіант 0)	2462	1767	605	141	20	5
$n_i$ (варіант 1)	2483	1783	608	142	25	4
$n_i$ (варіант 2)	2483	1767	608	146	25	3
$n_i$ (варіант 3)	2451	1783	616	142	23	4
$n_i$ (варіант 4)	2515	1783	600	142	27	4
$n_i$ (варіант 5)	2483	1799	608	138	25	5

Задача 8 (варіанти 6 – 10). Щоб вивчити проростання насіння кукурудзи, піддослідну ділянку поділено на квадрати і на кожен квадрат висіяно по 8 насінин. У поданій таблиці наведено кількості  $n_i$  квадратів ділянки, на яких зійшло по однаковій кількості  $X$  насінин.

$X$ , к-ть насінин	0	1	2	3	4	5	6	7	8
$n_i$ (варіант 6)	0	3	10	47	136	263	295	198	51
$n_i$ (варіант 7)	0	3	10	45	136	267	295	195	51
$n_i$ (варіант 8)	0	3	9	47	140	263	289	198	52
$n_i$ (варіант 9)	0	3	11	47	132	263	301	198	50
$n_i$ (варіант 10)	0	3	10	49	136	259	295	201	51

Задача 9 (варіанти 11 – 15). У деякій місцевості було зібрано дані про народжуваність дітей впродовж року. Розподіл кількості новонароджених  $n_i$  залежно від місяця року відображено в таблиці

місяці року	1	2	3	4	5	6	7	8	9	10	11	12
$n_i$ (варіант 11)	1438	1380	1393	1421	1430	1389	1390	1433	1444	1407	1395	1376
$n_i$ (варіант 12)	1395	1400	1411	1380	1425	1438	1381	1378	1409	1442	1407	1385
$n_i$ (варіант 13)	1398	1405	1420	1379	1385	1413	1437	1440	1388	1409	1417	1443
$n_i$ (варіант 14)	1423	1410	1380	1396	1377	1421	1430	1441	1379	1442	1393	1409
$n_i$ (варіант 15)	1381	1397	1417	1423	1438	1375	1390	1407	1424	1399	1440	1432

Задача 10 (варіанти 16 – 20). Щоб встановити, яка частка виготовлених приладів потребує додаткового регулювання, утворено вибірки по 7 зразків продукції в кожній вибірці. Розподіл кількості вибірок  $n_i$  за кількістю  $X$  приладів, які необхідно регулювати, подано в таблиці

$X$ приладів	0	1	2	3	4	5	6	7
$n_i$ (варіант 16)	25	74	95	68	29	8	1	0
$n_i$ (варіант 17)	25	71	95	71	29	7	1	0
$n_i$ (варіант 18)	24	74	100	68	27	8	1	0
$n_i$ (варіант 19)	26	74	90	68	30	8	1	0
$n_i$ (варіант 20)	25	77	95	65	29	9	1	0

Задача 11 (варіанти 21 – 25). Для вироблення рекомендацій щодо покращення обслуговування абонентів на декількох АТС міста були зібрані відомості про випадки несправної роботи апаратури автоматичного телефонного зв'язку. Розподіл одиниць апаратури  $n_i$ , в яких було виявлено по  $X$  несправностей, наведено в таблиці

$X$	0	1	2	3	4	5	6	7 і більше
$n_i$ (варіант 21)	25	75	112	112	84	50	25	5
$n_i$ (варіант 22)	25	71	112	114	84	48	25	6
$n_i$ (варіант 23)	24	75	114	112	86	50	27	5
$n_i$ (варіант 24)	26	75	110	112	88	50	23	5
$n_i$ (варіант 25)	25	79	112	110	84	52	25	4

Задача 12 (варіанти 26 – 30). Для вироблення рекомендацій щодо покращення роботи портів реєструвався час очікування кораблів на розвантаження в декількох портах країни. Розподіл за один рік кількості кораблів  $n_i$  залежно від часу  $T$ , який вони очікували на розвантаження, відображено в таблиці

$T$ , год.	0–6	6–12	12–18	18–24	24–30	30–36	36–42	42–48	48–54	$T \geq 54$
$n_i$ (варіант 26)	518	384	284	211	156	116	86	63	44	34
$n_i$ (варіант 27)	518	376	284	217	156	112	86	65	44	33
$n_i$ (варіант 28)	502	384	292	211	150	116	90	63	42	35
$n_i$ (варіант 29)	534	384	276	211	162	116	82	63	46	34
$n_i$ (варіант 30)	518	392	284	205	156	120	86	61	44	34

## ЛАБОРАТОРНА РОБОТА № 5

### ПЕРЕВІРКА ГІПОТЕЗ ПРО ПАРАМЕТРИ НОРМАЛЬНОГО РОЗПОДІЛУ

У даній роботі перевіряються гіпотези відносно параметрів функції розподілу відомого вигляду. Гіпотези такого типу можуть виникати, коли після застосування критеріїв згоди (див. лабораторну роботу 4) встановлено вигляд функції розподілу  $F(x, \theta_1, \theta_2, \dots, \theta_k)$  випадкової величини, але самі параметри  $\theta_1, \theta_2, \dots, \theta_k$ , від яких залежить функція розподілу, невідомі й потребують уточнення. Статистична гіпотеза, в якій сформульовані припущення щодо значень параметрів функції розподілу, називається параметричною гіпотезою. Перевіряють параметричні гіпотези за загальним правилом перевірки гіпотез, а статистики, за допомогою яких вони перевіряються, називають критеріями значущості.

У лабораторній роботі розглядаються критерії значущості для перевірки гіпотез у припущенні, що випадкова величина розподілена за нормальним законом розподілу. Оскільки нормальний закон розподілу визначається двома параметрами:  $\mu$  (математичне сподівання) і  $\sigma$  (середньоквадратичне відхилення), то гіпотези формулюються відносно цих параметрів.

Нульові параметричні гіпотези зазвичай записують у вигляді рівностей, наприклад

$$H_0 : \mu = \mu_0, \quad \text{або} \quad H_0 : \sigma = \sigma_0, \quad \text{або} \quad H_0 : \mu = \mu_0, \quad \sigma = \sigma_0. \quad (5.1)$$

Кожна нульова гіпотеза розглядається тільки разом з альтернативною. До однієї нульової гіпотези залежно від умов задачі треба підібрати лише одну альтернативну (хоча одній нульовій гіпотезі можуть логічно протиставлятися декілька різних альтернативних гіпотез). Наприклад, можливими альтернативними до першої з гіпотез (5.1) можуть бути такі гіпотези  $H_1$ :

$$H_1 : \mu \neq \mu_0, \quad \text{або} \quad H_1 : \mu < \mu_0, \quad \text{або} \quad H_1 : \mu > \mu_0, \quad \text{або} \quad H_1 : \mu = \mu_1 \neq \mu_0.$$

Якщо за вибірками оцінюються дві генеральні сукупності  $X$  та  $Y$ , кожна з яких має нормальний розподіл з невідомими параметрами відповідно  $\mu_x, \sigma_x$  і  $\mu_y, \sigma_y$ , то, порівнюючи ці сукупності, можна сформулювати такі параметричні гіпотези:

$$H_0 : \mu_x = \mu_y, \quad \text{або} \quad H_0 : \sigma_x = \sigma_y, \quad \text{або} \quad H_0 : \mu_x = \mu_y, \quad \sigma_x = \sigma_y. \quad (5.2)$$

Кожну з цих нульових гіпотез треба розглянути у парі з однією альтернативною. До першої з гіпотез у (5.2) альтернативними можуть бути гіпотези

$$H_1 : a_x \neq a_y, \text{ або } H_1 : a_x < a_y, \text{ або } H_1 : a_x > a_y.$$

Для критеріїв значущості при перевірці параметричних гіпотез про нормальний розподіл використовують випадкові величини, що мають нормальний розподіл, а також розподіли  $\chi^2$ , Стюдента і Фішера. Функції розподілів цих випадкових величин табульовані (див. додаток, табл. 3,6,7,8).

### Гіпотези про математичні сподівання

Задача А. Перевірка гіпотези про значення генеральної середньої нормально розподіленої сукупності.

Розглянемо вибірку зі значеннями  $x_1, x_2, \dots, x_n$  (серед яких не обов'язково всі різні) з генеральної сукупності, що має нормальний розподіл з параметрами  $a, \sigma$ . Перевіряючи нульову гіпотезу  $H_0 : a = a_0$ , виділяють два випадки: 1) середнє квадратичне відхилення  $\sigma$  випадкової величини  $X$  відоме; 2) середнє квадратичне відхилення  $\sigma$  невідоме.

У випадку 1 використовують статистику

$$Z = \frac{\bar{x}_B - a_0}{\sigma} \sqrt{n},$$

яка має нормований нормальний розподіл, де  $\bar{x}_B$  – вибіркове середнє,  $n$  – обсяг вибірки.

У випадку 2 використовують статистику

$$T = \frac{\bar{x}_B - a_0}{S} \sqrt{n},$$

де  $S$  – виправлене вибіркове середнє квадратичне відхилення. Випадкова величина  $T$  має розподіл Стюдента з  $k = n - 1$  ступенями вільності.

На рис. 5.1 подано графіки щільності нормованого нормального розподілу та розподілу Стюдента. Залежно від виду конкуруючої гіпотези будують критичну область.

У випадку 1:

1) якщо конкуруюча гіпотеза  $H_1 : a \neq a_0$ , то критична область двостороння, а критичне значення  $Z_{kp}$  шукають за таблицею значень функції Лапласа  $\Phi(x)$  (додаток, табл. 3) з умови

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

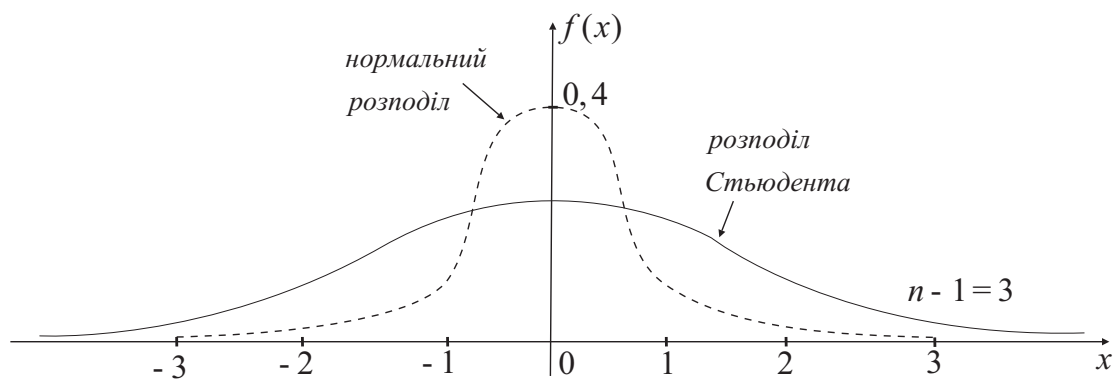


Рис.5.1. Розподіли нормальний і Стьюдента

тоді:

- якщо  $|Z_{емп}| < Z_{кр}$ , то  $H_0$  приймають,
- якщо  $|Z_{емп}| \geq Z_{кр}$ , то  $H_0$  відхиляють;

2) якщо конкуруюча гіпотеза  $H_1 : a > a_0$ , то критична область – правостороння, а критичне значення  $Z_{кр}$  визначають з рівняння

$$\Phi(Z_{кр}) = \frac{1 - 2\alpha}{2}, \quad (5.3)$$

тоді:

- якщо  $Z_{емп} < Z_{кр}$ , то  $H_0$  приймають,
- якщо  $Z_{емп} \geq Z_{кр}$ , то  $H_0$  відхиляють;

3) якщо конкуруюча гіпотеза  $H_1 : a < a_0$ , то критична область – лівостороння, а критичне значення  $Z_{кр}$  шукають з (5.3), причому:

- якщо  $Z_{емп} > -Z_{кр}$ , то  $H_0$  приймають,
- якщо  $Z_{емп} \leq -Z_{кр}$ , то  $H_0$  відхиляють.

У випадку 2 аналогічно будуються критичні області, а їх критичні значення  $t_{кр}$  визначають за таблицею розподілу Стьюдента (додаток, табл. 7).

Задача Б. Перевірка гіпотези про значення дисперсії нормально розподіленої сукупності.

Перевіряючи нульову гіпотезу  $H_0 : \sigma = \sigma_0$ , розрізняють два випадки:

- 1) математичне сподівання  $a$  випадкової величини  $X$  відоме;
- 2) математичне сподівання  $a$  невідоме.

У випадку 1 використовують статистику

$$\chi^2(n) = \frac{nD_B}{\sigma_0^2},$$

де  $D_B$  – вибіркова дисперсія,  $n$  – обсяг вибірки.

У випадку 2 використовують статистику

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

де  $S^2$  – виправлена вибіркова дисперсія, що визначається (2.2) та (1.5).

Обидві статистики мають розподіл  $\chi^2$ , але  $\chi^2(n)$  має  $n$  ступенів вільності, а  $\chi^2(n-1)$  –  $(n-1)$  ступенів вільності. На рис. 5.2 подано графік щільності розподілу  $\chi^2$ . Критичні області будуються подібно до того, як було описано вище, а критичні значення знаходять з таблиці критичних точок розподілу  $\chi^2$  (додаток, табл. 6) згідно з числом ступенів вільності  $k$  та рівнем значущості  $\alpha$ .

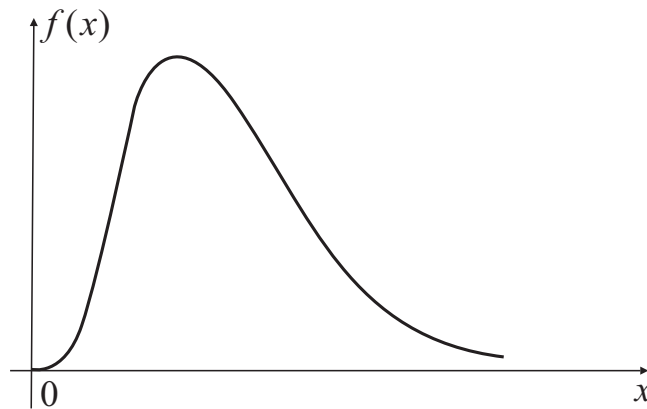


Рис.5.2.  $\chi^2$ -розподіл

Задача В. Перевірка гіпотези про рівність середніх квадратичних відхилень двох нормально розподілених сукупностей.

Розглянемо дві вибірки  $x_1, x_2, \dots, x_n$  та  $y_1, y_2, \dots, y_m$  з генеральних сукупностей  $X$  і  $Y$ , які розподілені за нормальним законом з параметрами  $a_x, \sigma_x$  та  $a_y, \sigma_y$ , відповідно.

Перевіряючи нульову гіпотезу  $H_0 : \sigma_x = \sigma_y$ , виділяють два випадки: 1) математичні сподівання  $a_x$  і  $a_y$  відомі; 2) математичні сподівання  $a_x$  і  $a_y$  невідомі.

У випадку 1 використовують статистику

$$F(n, m) = \frac{D_x}{D_y},$$

причому приймають такі позначення, що  $D_x$  – більша дисперсія,  $D_y$  – менша дисперсія:

$$D_x = \frac{1}{n} \sum_{i=1}^n (x_i - a_x)^2, \quad D_y = \frac{1}{m} \sum_{j=1}^m (y_j - a_y)^2, \quad D_x > D_y.$$

У випадку 2 застосовують статистику

$$F(n-1, m-1) = \frac{S_x^2}{S_y^2},$$

де  $S_x^2$  – більша виправлена дисперсія,  $S_y^2$  – менша виправлена дисперсія:

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - a_x)^2, \quad S_y^2 = \frac{1}{m-1} \sum_{j=1}^m (y_j - a_y)^2, \quad S_x^2 > S_y^2. \quad (5.4)$$

Обидві статистики розподілені за законом Фішера, але  $F(n, m)$  має  $n$  і  $m$  ступенів вільності, а  $F(n-1, m-1)$  – відповідно  $n-1$  і  $m-1$  ступенів вільності. На рис. 5.3 зображено графік щільності розподілу Фішера.

Якщо конкуруюча гіпотеза  $H_1 : \sigma_x > \sigma_y$ , то критична область правостороння, а критичну точку  $F_{кр}(\alpha, k_1, k_2)$  знаходять за таблицею критичних точок розподілу Фішера-Снедекора (табл. 8 додатку).

Якщо конкуруюча гіпотеза  $H_1 : \sigma_x \neq \sigma_y$ , то будують двосторонню критичну область, критичні точки якої  $F_{кр}(\frac{\alpha}{2}, k_1, k_2)$ .

Задача Г. Перевірка гіпотези про рівність математичних сподівань двох нормально розподілених сукупностей.

При перевірці нульової гіпотези  $H_0 : a_x = a_y$ , розрізняють два випадки: 1) середні квадратичні відхилення  $\sigma_x$  і  $\sigma_y$  відомі; 2) середні квадратичні відхилення  $\sigma_x$  і  $\sigma_y$  невідомі.

У випадку 1 за критерій перевірки гіпотези вибирають статистику

$$Z = \frac{\bar{x}_B - \bar{y}_B}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}},$$

де  $\bar{x}_B$  – вибіркве середнє першої вибірки,  $n$  – її обсяг,  $\bar{y}_B$  – вибіркве середнє другої вибірки,  $m$  – її обсяг.

У випадку 2 застосовують статистику

$$T = \frac{\bar{x}_B - \bar{y}_B}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}},$$

де  $S_x^2$ ,  $S_y^2$  – виправлені вибіркві дисперсії відповідно першої та другої вибірок, що визначаються (5.4).

У випадку 1:



1) якщо конкуруюча гіпотеза  $H_1 : a_x \neq a_y$ , то критична область двостороння і критичну точку шукають з рівності

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

причому:

- якщо  $|Z_{емп}| < Z_{кр}$ , то  $H_0$  приймають,
- якщо  $|Z_{емп}| \geq Z_{кр}$ , то  $H_0$  відхиляють.

2) якщо конкуруюча гіпотеза  $H_1 : a_x > a_y$ , то критична область правостороння і

$$\Phi(Z_{кр}) = \frac{1 - 2\alpha}{2}, \quad (5.5)$$

причому:

- якщо  $Z_{емп} < Z_{кр}$ , то  $H_0$  приймають,
- якщо  $Z_{емп} \geq Z_{кр}$ , то  $H_0$  відхиляють.

3) якщо конкуруюча гіпотеза  $H_1 : a_x < a_y$ , то критична область лівостороння, а критична точка знаходиться з умови (5.5), причому:

- якщо  $Z_{емп} > -Z_{кр}$ , то  $H_0$  приймають,
- якщо  $Z_{емп} \leq -Z_{кр}$ , то  $H_0$  відхиляють.

У випадку 2 побудова критичної області здійснюється аналогічно, але критичні значення  $t_{кр}$  визначаються за таблицею критичних точок розподілу Стюдента (додаток, табл. 7).

## ЗАВДАННЯ ДО ЛАБОРАТОРНОЇ РОБОТИ

**Задача А.** Покращення якості готової продукції дає підстави передбачити зменшення витрат підприємства на ремонтне обслуговування приладів за час їх гарантійного строку служби. Дані про щомісячні витрати на гарантійне обслуговування приладів за останні 12 місяців наведено у вибірках  $x_i$ ,  $i = 1, 2, \dots, 12$  (див. нижче табл. 5.1). Вважається, що всі вибірки зроблено з нормально розподілених генеральних сукупностей.

### Завдання до задачі А

1. Для варіантів 1 – 15. Перевірити гіпотезу про те, що зміни у процесі виробництва не призвели до зменшення затрат  $m$  на гарантійний ремонт приладів за умови, що планові щомісячні відрахування підприємства на ремонт становлять  $a$  одиниць, а середньоквадратичне відхилення цих затрат відоме і дорівнює  $\sigma_0$  одиниць. Отже,

- нульова гіпотеза  $H_0 : m = a$ ,
- альтернативна гіпотеза  $H_1 : m < a$ ,  $\sigma_0$  – відоме,
- рівень значущості  $\alpha = 0,05$ .

Числові дані взяти з таблиці 5.1.

Для варіантів 16 – 30. Перевірити ті самі нульову  $H_0$  і альтернативну  $H_1$  гіпотези, що і у варіантах 1 – 15, але за умови, що середнє квадратичне відхилення затрат  $\sigma_0$  вважається невідомим:

- нульова гіпотеза  $H_0 : m = a$ ,
- альтернативна гіпотеза  $H_1 : m < a$ ,  $\sigma_0$  – невідоме,
- рівень значущості  $\alpha = 0,05$ .

Числові дані взяти з таблиці 5.1.

2. Для варіантів 16 – 30. Перевірити гіпотезу про те, що середньоквадратичні відхилення  $\sigma$  затрат на ремонт мають заданий рівень  $\sigma_0$  за умови, що відомі середні щомісячні затрати  $a$  на ремонтне обслуговування. Отже,

- нульова гіпотеза  $H_0 : \sigma = \sigma_0$ ,
- альтернативна гіпотеза  $H_1 : \sigma \neq \sigma_0$ ,  $a$  – відоме,
- рівень значущості  $\alpha = 0,05$ .

Числові дані взяти з таблиці 5.1.

Для варіантів 1 – 15. Перевірити ті самі нульову  $H_0$  і альтернативну  $H_1$  гіпотези, що і для варіантів 16 – 30, але за умови, що середньомісячні затрати  $a$  на ремонтне обслуговування вважаються невідомими:

- нульова гіпотеза  $H_0 : \sigma = \sigma_0$ ,

- альтернативна гіпотеза  $H_1 : \sigma \neq \sigma_0$ ,  $a$  – невідоме,
- рівень значущості  $\alpha = 0,05$ .

Числові дані взяти з таблиці 5.1.

**Задача Б.** На підприємстві впроваджено нову технологію з метою економії використовуваних у виробничому процесі ресурсів. Дані про вартість сировини (в умовних одиницях), затрачені на виготовлення одиниці продукції старим і новим способами наведено відповідно у двох вибірках  $x_i, i=1, 2, \dots, 9$  та  $y_j, j=1, 2, \dots, 12$  (див. нижче табл. 5.2). Вважається, що вибірки зроблено з нормально розподілених генеральних сукупностей.

Завдання до задачі Б

1. Для варіантів 1 – 15. Перевірити гіпотезу про те, що середньоквадратичні відхилення  $\sigma_x$  і  $\sigma_y$  величини затрат сировини на виготовлення одиниці продукції є однаковими для обидвох генеральних сукупностей за умови, що середні затрати сировини у цих сукупностях відомі і становлять  $a_x$  і  $a_y$ . Отже,

- нульова гіпотеза  $H_0 : \sigma_x = \sigma_y$ ,
- альтернативна  $H_1 : \sigma_x \neq \sigma_y$ ,  $a_x$  і  $a_y$  – відомі,
- рівень значущості  $\alpha = 0,05$ .

Числові дані взяти з таблиці 5.2.

Для варіантів 16 – 30. Перевірити ті самі нульову  $H_0$  і альтернативну  $H_1$  гіпотези, що і для варіантів 1 – 15, але за умови, що середні затрати  $a_x$  і  $a_y$  сировини у вибірках є невідомими:

- нульова гіпотеза  $H_0 : \sigma_x = \sigma_y$ ,
- альтернативна  $H_1 : \sigma_x \neq \sigma_y$ ,  $a_x$  і  $a_y$  – невідомі,
- рівень значущості  $\alpha = 0,05$ .

Числові дані взяти з таблиці 5.2.

2. Для варіантів 1 – 30. Залежно від результатів виконання завдань 1, 2 встановити, чи можна вважати, що середні затрати сировини  $m_x$  і  $m_y$  у цих генеральних сукупностях однакові. Отже, виконати перевірку:

- нульова гіпотеза  $H_0 : m_x = m_y$ ,
- альтернативна гіпотеза  $H_1 : m_x \neq m_y$ .

Причому, якщо за результатами виконання завдань 1, 2 виявилося, що  $\sigma_x = \sigma_y$ , то  $\sigma_x$  і  $\sigma_y$  слід вважати невідомими, а якщо  $\sigma_x \neq \sigma_y$ , то їх потрібно вважати відомими. Гіпотези перевіряти на рівні значущості  $\alpha = 0,05$ , а числові дані взяти з таблиці 5.2.

Числові дані до задачі 1

Таблиця 5.1

Варі- ант	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$	$x_{11}$	$x_{12}$	$a$	$\sigma_0$
0	56,4	52,43	49,3	46,7	50,6	53,1	46,9	47,3	54,0	49,3	48,0	47,6	50,5	3,2
1	51,5	63,3	57,1	51,9	59,7	64,7	52,3	52,9	67,3	57,1	54,5	53,7	59,4	4,5
2	26,5	32,4	29,3	26,7	30,6	33,1	26,9	27,3	34,4	29,3	28,0	27,6	30,5	3,2
3	14,3	17,2	15,7	13,9	16,3	17,6	14,5	14,6	18,2	15,7	15,0	14,8	16,0	2
4	27,5	33,4	30,3	27,7	31,6	34,1	27,9	28,3	35,4	30,3	29,0	28,6	31,5	3,2
5	53,5	65,3	59,1	53,9	61,7	66,7	54,3	54,9	69,3	59,1	56,5	55,7	61,4	4,5
6	28,5	34,4	31,3	28,7	32,6	35,1	28,9	29,3	36,4	31,3	30,0	29,6	32,5	3,2
7	16,3	19,2	17,7	15,9	18,3	19,6	16,5	16,6	20,2	17,7	17,0	16,8	18,2	2
8	29,5	35,4	32,3	29,7	33,6	36,1	29,9	30,3	37,4	32,3	31,0	30,6	33,5	3,2
9	55,5	67,3	61,1	55,9	63,7	68,7	56,3	56,9	71,3	61,1	58,5	57,7	63,4	4,5
10	30,5	36,4	33,3	30,7	34,6	37,1	30,9	31,3	38,4	33,3	32,0	31,6	34,5	3,2
11	18,3	21,2	19,7	17,9	20,3	21,6	18,5	18,6	22,2	19,7	10,0	18,8	20,0	2
12	31,5	37,4	34,3	31,7	35,6	38,1	31,9	32,3	39,4	34,3	33,0	32,6	35,5	3,2
13	57,5	69,3	63,1	57,9	65,7	70,7	58,3	58,9	73,3	63,1	60,5	59,7	65,4	4,5
14	32,5	38,4	35,3	32,7	36,6	39,1	32,9	33,3	40,4	35,3	34,0	33,6	36,5	3,2
15	20,3	23,2	21,7	19,9	22,3	23,6	20,5	20,6	24,2	21,7	21,0	20,8	22,0	2
16	33,5	39,4	36,3	33,7	37,6	40,1	33,9	34,3	41,4	36,3	35,0	34,6	37,5	3,2
17	59,5	71,3	65,1	59,9	67,7	72,7	60,3	60,9	75,3	65,1	62,5	61,7	67,4	4,5
18	34,5	40,4	37,3	4,7	38,6	41,1	34,9	35,3	42,4	37,3	36,0	35,6	38,5	3,2
19	22,3	25,2	23,7	21,9	24,3	25,6	22,5	22,6	26,2	23,7	23,0	22,8	24,0	2
20	35,4	41,4	38,3	35,7	39,6	42,1	35,9	36,3	43,4	38,3	37,0	36,6	39,5	3,2
21	61,5	73,3	67,1	61,9	69,7	74,7	62,3	62,9	77,3	67,1	64,5	63,7	69,4	4,5
22	36,5	42,4	39,3	36,7	40,6	43,1	36,9	37,3	44,4	39,3	38,0	37,6	40,5	3,2
23	21,3	27,2	25,7	23,9	26,3	27,6	24,5	24,6	28,2	25,7	25,0	24,8	26,0	2
24	37,5	43,4	40,3	37,7	41,6	44,1	37,0	38,3	45,4	40,3	39,0	38,6	41,5	3,2
25	63,5	75,3	69,1	63,9	71,7	76,7	64,3	64,9	79,3	69,1	66,5	65,7	71,4	4,5
26	38,5	44,4	41,3	38,7	42,6	45,1	38,9	39,3	46,4	41,3	40,0	39,6	42,5	3,2
27	26,3	29,2	27,7	25,9	28,3	29,6	26,5	26,6	30,2	27,7	27,0	26,8	28,0	2
28	39,5	45,4	42,3	39,7	43,6	46,1	39,9	40,3	47,4	42,3	41,0	40,6	43,5	3,2
29	65,5	77,3	71,1	65,9	73,7	78,7	66,3	66,9	81,3	71,1	68,5	67,7	73,4	4,5
30	40,5	46,4	43,3	40,7	44,6	47,1	40,9	41,3	48,4	43,3	42,0	41,6	44,5	3,2

Числові дані до задачі 2

Таблиця 5.2

Варі- -ант	$x_1$ $y_1$	$x_2$ $y_2$	$x_3$ $y_3$	$x_4$ $y_4$	$x_5$ $y_5$	$x_6$ $y_6$	$x_7$ $y_7$	$x_8$ $y_8$	$x_9$ $y_9$	$x_{10}$ $y_{10}$	$x_{11}$ $y_{11}$	$x_{12}$ $y_{12}$	$a_x$ $a_y$	$\sigma_x$ $\sigma_y$
0	10,9	10,8	10,6	10,0	10,6	10,2	10,9	10,7	9,7	—	—	—	—	—
	11,2	11,6	10,2	10,9	11,7	10,4	11,3	9,9	10,1	10,4	11,0	11,2	—	—
1	8,7	8,6	8,4	7,8	8,4	8,0	8,7	8,5	7,5	—	—	—	8,1	0,4
	22,4	23,4	19,9	21,6	23,6	20,4	22,6	19,1	19,6	20,4	21,9	22,4	21,4	1,5
2	8,8	8,7	8,5	7,9	8,5	8,1	8,8	8,6	7,6	—	—	—	8,2	0,4
	13,6	14,2	12,1	13,1	14,3	12,4	13,7	11,6	11,9	12,4	13,3	13,6	13,0	0,9
3	8,9	8,8	8,6	8,0	8,6	8,2	8,9	8,7	7,7	—	—	—	8,3	0,4
	9,2	9,6	8,2	8,9	9,7	8,4	9,3	7,9	8,1	8,4	9,0	9,2	8,8	0,6
4	9,0	8,9	8,7	8,1	8,7	8,3	9,0	8,8	7,8	—	—	—	8,4	0,4
	13,8	14,4	12,3	13,3	14,5	12,6	13,9	11,8	12,1	12,6	13,5	13,8	13,2	0,9
5	9,1	9,0	8,8	8,2	8,8	8,4	9,1	8,9	7,9	—	—	—	8,5	0,4
	22,8	23,8	20,3	22,0	24,0	20,8	23,0	19,5	20,0	20,8	22,3	22,8	21,8	1,5
6	9,2	9,1	8,9	8,3	8,9	8,5	9,2	9,0	8,0	—	—	—	8,6	0,4
	14,0	14,6	12,5	13,5	14,7	12,8	14,1	12,0	12,3	12,8	13,7	14,0	13,4	0,9
7	9,3	9,2	9,0	8,4	9,0	8,6	9,3	9,1	8,1	—	—	—	8,7	0,4
	9,6	10,0	8,6	9,3	10,1	8,8	9,7	8,3	8,5	8,8	9,4	9,6	9,2	0,6
8	9,4	9,3	9,1	8,5	9,1	8,7	9,4	9,2	8,2	—	—	—	8,8	0,4
	14,2	14,8	12,7	13,7	14,9	13,0	14,3	12,2	12,5	13,0	13,9	14,2	13,6	0,9
9	9,5	9,4	9,2	8,6	9,2	8,9	9,5	9,3	8,3	—	—	—	8,9	0,4
	23,2	24,2	20,7	22,4	24,4	21,2	23,4	19,9	20,4	21,2	22,7	23,2	22,2	1,5
10	9,6	9,5	9,3	8,7	9,3	8,9	9,6	9,4	8,4	—	—	—	9,0	0,4
	14,4	15,0	12,9	13,9	15,1	13,2	14,5	12,4	12,7	13,2	14,1	14,4	13,8	0,9
11	9,7	9,6	9,4	8,8	9,4	9,0	9,7	9,5	8,5	—	—	—	9,1	0,4
	10,0	10,4	9,0	9,7	10,5	9,2	10,1	8,7	8,9	9,2	9,8	10,0	9,6	0,6
12	9,8	9,7	9,5	8,9	9,5	9,1	9,8	9,6	8,6	—	—	—	9,2	0,4
	14,6	15,2	13,1	14,1	15,3	13,4	14,7	12,6	12,9	13,4	14,3	14,6	14,0	0,9
13	9,9	9,8	9,6	9,0	9,6	9,2	9,9	9,7	8,7	—	—	—	9,3	0,4
	23,6	24,6	21,1	22,8	24,8	21,6	23,8	20,3	20,8	21,6	23,1	23,6	22,6	1,5
14	10,0	9,9	9,7	9,1	9,7	9,3	10,0	9,8	8,8	—	—	—	9,4	0,4
	14,8	15,4	13,3	14,3	15,5	13,6	14,9	12,8	132,1	13,6	14,5	14,8	14,2	0,9
15	10,1	10,0	9,8	9,2	9,8	9,4	10,1	9,9	8,9	—	—	—	9,5	0,4
	10,4	10,8	9,4	10,1	10,9	9,6	10,5	9,1	9,3	9,6	10,2	10,4	10,0	0,6
16	11,3	11,5	11,6	10,7	11,3	10,9	11,6	11,4	10,4	—	—	—	11,0	0,4
	14,9	17,0	16,4	15,9	17,1	15,2	16,5	14,4	14,7	16,1	16,4	15,2	15,8	0,9
17	10,3	10,2	10,0	9,4	10,0	9,6	10,3	10,1	9,1	—	—	—	9,7	0,4
	24,0	25,0	21,5	23,2	25,2	22,0	24,2	20,7	21,0	22,0	23,5	24,0	23,0	1,5
18	10,4	10,3	10,1	9,5	10,1	9,7	10,4	10,2	9,2	—	—	—	9,8	0,4
	15,2	15,8	13,7	14,7	15,9	14,0	15,3	13,2	13,5	14,0	14,9	15,2	14,6	0,9
19	10,5	10,4	10,2	9,6	10,2	9,8	10,5	10,3	9,3	—	—	—	9,9	0,4
	10,8	11,2	9,8	10,5	11,3	10,0	10,9	9,5	9,7	10,0	10,6	10,8	10,4	0,6

варі- ант	$x_1$ $y_1$	$x_2$ $y_2$	$x_3$ $y_3$	$x_4$ $y_4$	$x_5$ $y_5$	$x_6$ $y_6$	$x_7$ $y_7$	$x_8$ $y_8$	$x_9$ $y_9$	$x_{10}$ $y_{10}$	$x_{11}$ $y_{11}$	$x_{12}$ $y_{12}$	$a_x$ $a_y$	$\sigma_x$ $\sigma_y$
20	10,6	10,5	10,3	9,7	10,3	9,9	10,6	10,4	9,4	—	—	—	10,0	0,4
	15,4	16,0	13,9	14,9	16,1	14,2	15,5	13,4	13,7	14,2	15,1	15,4	14,8	0,9
21	10,7	10,6	10,4	9,8	10,4	10,0	10,7	10,5	9,5	—	—	—	10,1	0,4
	24,4	25,4	21,9	23,6	25,6	22,4	24,6	21,1	21,6	22,4	23,9	24,4	23,4	1,5
22	10,8	10,7	10,5	9,9	10,5	10,1	10,8	10,6	9,6	—	—	—	10,2	0,4
	15,6	16,2	14,1	15,1	16,3	14,4	15,7	13,6	13,9	14,4	15,3	15,6	15,0	0,9
23	10,9	10,8	10,6	10,0	10,6	10,2	10,9	10,7	9,7	—	—	—	10,3	0,4
	11,2	11,6	10,2	10,9	11,7	10,4	11,3	9,9	10,1	10,4	11,0	11,2	10,8	0,6
24	11,0	10,9	10,7	10,1	10,7	10,3	11,0	10,8	9,8	—	—	—	10,4	0,4
	15,8	16,4	14,3	15,3	16,5	14,6	15,9	13,8	14,1	14,6	15,5	15,8	15,2	0,9
25	11,1	11,0	10,8	10,2	10,8	10,4	11,1	10,9	9,9	—	—	—	10,5	0,4
	24,8	25,8	22,3	24,0	26,0	22,8	25,0	21,5	22,0	22,8	24,3	24,8	23,8	1,5
26	11,2	11,1	10,9	10,3	10,9	10,5	11,2	14,0	10,0	—	—	—	10,6	0,4
	16,0	16,6	14,5	15,5	16,7	14,8	16,1	14,0	14,3	14,8	15,7	16,0	15,4	0,9
27	11,3	11,2	11,0	10,4	11,0	10,6	11,3	11,1	10,1	—	—	—	10,7	0,4
	11,6	12,0	10,6	11,3	12,1	10,8	11,7	10,3	10,5	10,8	11,4	11,6	11,2	0,6
28	11,4	11,3	11,1	10,5	11,1	10,7	11,4	11,2	10,2	—	—	—	10,8	0,4
	16,2	16,8	14,7	15,7	16,9	15,0	16,3	14,2	14,5	15,0	15,9	16,2	15,6	0,9
29	11,5	11,4	11,2	10,6	11,2	10,8	11,5	11,3	10,3	—	—	—	10,9	0,4
	25,2	26,2	22,7	24,4	26,4	23,2	25,4	21,9	22,4	23,2	24,7	25,2	24,2	1,5
30	11,6	11,5	11,3	10,7	11,3	10,9	11,6	11,4	10,4	—	—	—	11,0	0,4
	16,4	17,0	14,9	15,9	17,1	15,2	16,5	14,4	14,7	15,2	16,1	16,4	15,8	0,9

# ПРИКЛАД РОЗВ'ЯЗУВАННЯ ТИПОВОГО ВАРІАНТУ

## Лабораторна робота № 1

Дано вибірку (вона отримана згідно з таблицею 1.1 для варіанту  $k = 0$ ):  
 21, 27, 10, 28, 33, 29, 36, 15, 26, 24, 31, 21, 23, 17, 22, 32, 15, 25, 18, 22, 16, 35,  
 31, 20, 24, 26, 36, 29, 25, 19, 27, 24, 30, 22, 27, 21, 28, 24, 12, 23, 20, 19, 15, 23,  
 25, 38, 26, 31, 27, 37, 24, 11, 28, 18, 20, 30, 22, 17, 33, 21, 13, 26, 34, 31, 29, 24,  
 30, 28, 25, 23, 32, 21, 10, 22, 16, 27, 25, 24, 23, 24, 35, 29, 20, 17, 30, 23, 32, 18, 26,  
 31, 38, 12, 25, 33, 28, 15, 30, 19, 22, 10.

1. Побудуємо дискретний статистичний ряд частот та відносних частот

Таблиця 1

$x_i$	10	11	12	13	15	16	17	18	19	20	21	22	23	24
$n_i$	3	1	2	1	4	2	3	3	3	4	5	6	6	8
$w_i$	0,03	0,01	0,02	0,01	0,04	0,02	0,03	0,03	0,03	0,04	0,05	0,06	0,06	0,08

$x_i$	25	26	27	28	29	30	31	32	33	34	35	36	37	38
$n_i$	6	5	5	5	4	5	5	3	3	1	2	2	1	2
$w_i$	0,06	0,05	0,05	0,05	0,04	0,05	0,05	0,03	0,03	0,01	0,02	0,02	0,01	0,02

Обсяг вибірки  $n = 100$ .

2. Розмах вибірки  $r = 38 - 10 = 28$ .

Мода цього ряду  $M_0(x) = 24$ , оскільки цьому значенню відповідає найбільша частота 8. Такий статистичний ряд є одномодальний.

Знайдемо медіану  $M_e(x)$  дискретного статистичного ряду (табл. 1). Обсяг вибірки є парним числом, тому медіану визначають як середнє арифметичне варіант з номерами  $\frac{n}{2} = 50$  та  $\frac{n+2}{2} = 51$ . Тоді

$$M_e(x) = \frac{x_{50} + x_{51}}{2} = \frac{24 + 24}{2} = 24.$$

3. Полігон частот зображений на рис. 1.

4. Для побудови інтервального статистичного ряду визначимо оптимальну кількість проміжків  $s = 1 + 3, 2 \cdot \lg 100 = 7, 4$ . Оскільки розмах вибірки  $r = 28$ , то доцільно весь відрізок, на якому задана вибірка, розбити на 7 однакових проміжків, довжина кожного з яких  $h = \frac{28}{7} = 4$ .

Інтервальний статистичний ряд має вигляд

Таблиця 2

$\Delta_j$	[10; 14)	[14; 18)	[18; 22)	[22; 26)	[26; 30)	[30; 34)	[34; 38)
$n_j$	7	9	15	26	19	16	8

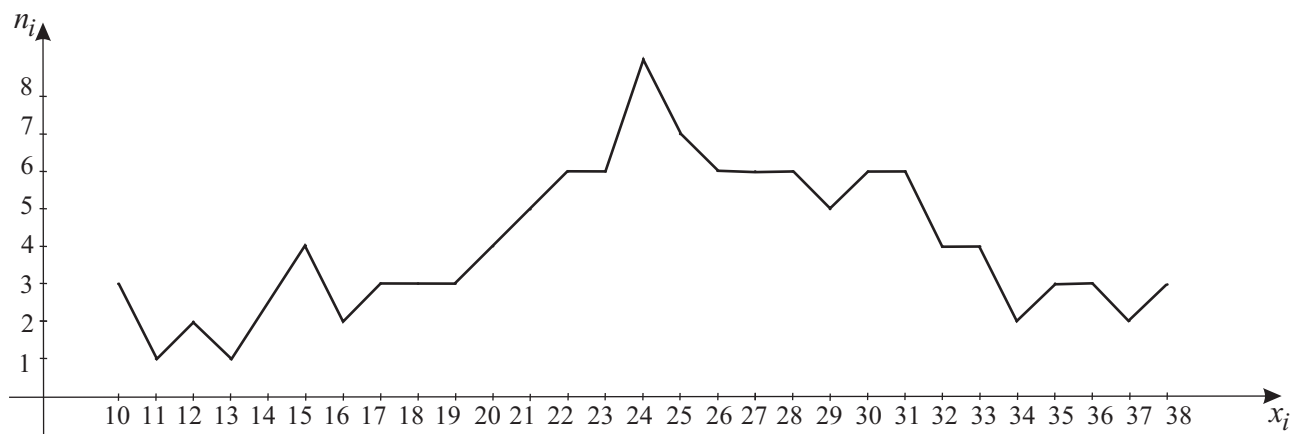


Рис. 1. Полігон частот

5. Побудуємо гістограму частот. Обчислимо висоту кожного з прямокутників:

$$h_1 = \frac{7}{4} = 1,75; \quad h_2 = \frac{9}{4} = 2,25; \quad h_3 = \frac{15}{4} = 3,75; \quad h_4 = \frac{26}{4} = 6,5;$$

$$h_5 = \frac{19}{4} = 4,75; \quad h_6 = \frac{16}{4} = 4; \quad h_7 = \frac{8}{4} = 2.$$

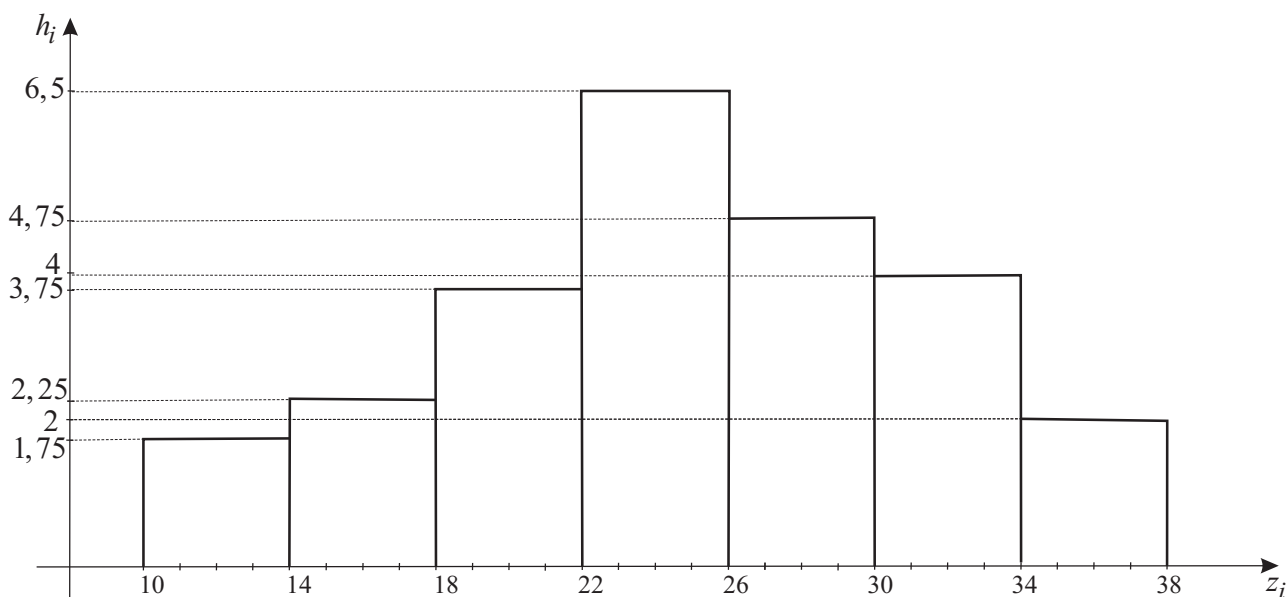


Рис. 2. Гістограма частот

Модальним інтервалом для інтервального статистичного ряду є  $[22; 26)$ , оскільки йому відповідає найбільша щільність частоти  $h_4 = 6,5$ .

Для визначення медіани для інтервального статистичного ряду (табл. 2) встановимо медіанний інтервал:  $[z_{M-1}; z_M) = [22; 26)$ , для якого  $n_M = 26$ ,  $m_{M-1} = 7 + 9 + 15 = 31$ , звідки  $M_e = 22 + \frac{26 - 22}{26} \left( \frac{100}{2} - 31 \right) \approx 24,92$ .

6. Побудуємо графік емпіричної функції розподілу  $F^*(x)$  для інтервального статистичного ряду. Визначимо функцію  $F^*(x)$  у точках, що відповідають



кінцям проміжків  $\Delta_j$ ,  $j = 1, 2, \dots, 7$ :

$$F^*(10) = 0; \quad F^*(14) = \frac{7}{100} = 0,07; \quad F^*(18) = \frac{7+9}{100} = 0,16;$$

$$F^*(22) = \frac{7+9+15}{100} = 0,31; \quad F^*(26) = \frac{7+9+15+26}{100} = 0,57;$$

$$F^*(30) = \frac{7+9+15+26+19}{100} = 0,76;$$

$$F^*(34) = \frac{7+9+15+26+19+16}{100} = 0,92;$$

$$F^*(38) = \frac{7+9+15+26+19+16+8}{100} = 1,$$

а на самих проміжках функцію  $F^*(x)$  продовжимо лінійно та неперервно.

Якщо  $x \leq 10$ , то  $F^*(x) = 0$ , а для  $x \geq 38$  матимемо  $F^*(x) = 1$ .

Графік функції  $F^*(x)$  зображений на рис. 3.

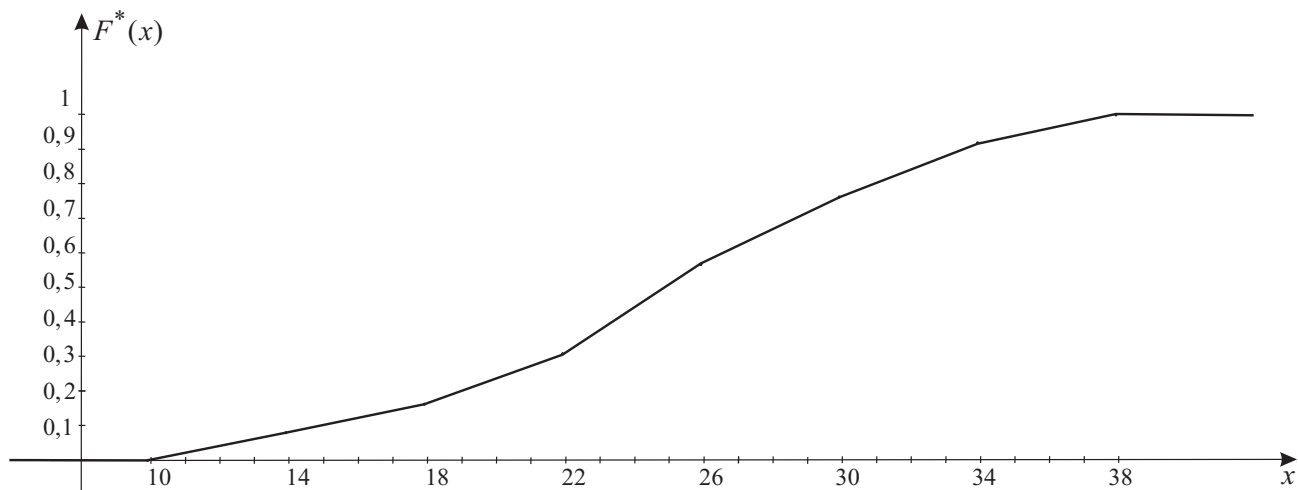


Рис. 3. Графік емпіричної функції розподілу

7. Вибіркове середнє значення  $\bar{x}_B$  дискретного статистичного ряду (табл. 1) обчислюється:

$$\bar{x}_B = \frac{1}{100}(30+11+24+13+60+32+51+54+57+80+105+132+138+192+ \\ +150+130+135+140+116+150+155+96+99+34+70+72+37+76) = 24,39.$$

Для знаходження вибіркового середнього значення  $\bar{x}_B$  інтервального статистичного ряду побудуємо спочатку відповідний дискретний статистичний ряд

Таблиця 3

$x_i$	12	16	20	24	28	32	36
$n_i$	7	9	15	26	19	16	8

Тоді

$$\bar{x}_B = \frac{1}{100}(84 + 144 + 300 + 624 + 532 + 512 + 288) = 24,84.$$

8. Дисперсію  $D_B$  інтервального статистичного ряду обчислюємо за даними відповідного дискретного ряду (табл. 3.)

$$\begin{aligned} D_B = \frac{1}{100} & ((12 - 24,84)^2 \cdot 7 + (16 - 24,84)^2 \cdot 9 + (20 - 24,84)^2 \cdot 15 + \\ & + (24 - 24,84)^2 \cdot 26 + (28 - 24,84)^2 \cdot 19 + (32 - 24,84)^2 \cdot 16 + \\ & + (36 - 24,84)^2 \cdot 8) \approx 42,33. \end{aligned}$$

**Зауваження.** Для обчислення дисперсії можна використовувати і формулу (1.5), згідно якої

$$\begin{aligned} D_B = \frac{1}{100} & (12^2 \cdot 7 + 16^2 \cdot 9 + 20^2 \cdot 15 + 24^2 \cdot 26 + 28^2 \cdot 19 + 32^2 \cdot 16 + \\ & + 36^2 \cdot 8) - 24,84^2 = 42,32. \end{aligned}$$

Середнє квадратичне відхилення  $\sigma_B = \sqrt{42,33} \approx 6,51$ .

9. Коефіцієнтом варіації  $V$  для інтервального ряду є:

$$V = \frac{6,51}{24,84} \approx 0,26.$$

10. Центральні емпіричні моменти третього та четвертого порядків обчислюються згідно з (1.6):

$$\begin{aligned} \mu_3 = \frac{1}{100} & ((12 - 24,84)^3 \cdot 7 + (16 - 24,84)^3 \cdot 9 + (20 - 24,84)^3 \cdot 15 + \\ & + (24 - 24,84)^3 \cdot 26 + (28 - 24,84)^3 \cdot 19 + (32 - 24,84)^3 \cdot 16 + \\ & + (36 - 24,84)^3 \cdot 8) \approx -57,70; \\ \mu_4 = \frac{1}{100} & ((12 - 24,84)^4 \cdot 7 + (16 - 24,84)^4 \cdot 9 + (20 - 24,84)^4 \cdot 15 + \\ & + (24 - 24,84)^4 \cdot 26 + (28 - 24,84)^4 \cdot 19 + (32 - 24,84)^4 \cdot 16 + \\ & + (36 - 24,84)^4 \cdot 8) \approx 4215,08. \end{aligned}$$

11. Враховуючи знайдені вище центральні емпіричні моменти третього та четвертого порядків обчислимо асиметрію та ексцес:

$$A = \frac{-57,70}{(6,51)^3} \approx -0,21, \quad E = \frac{4215,08}{(6,51)^4} - 3 = -0,65.$$

## Лабораторна робота № 2

### Завдання А

Результати перевірки на міцність 300 зразків дроту одного діаметра, задано у вигляді таблиці (її одержали згідно з таблицею задачі 1 при  $k = 0$ ):

Розривне зусилля,кН	[80, 82)	[82, 84)	[84, 86)	[86, 88)	[88, 90)	[90, 92)	[92, 94]
К-сть зразків	20	47	80	89	40	16	8

Припускаючи, що вибірка отримана з нормально розподіленої генеральної сукупності, знайдемо точкові оцінки для математичного сподівання  $a$  та середнього квадратичного відхилення  $\sigma$  спостережуваної випадкової величини.

Замінімо інтервальний статистичний ряд відповідним дискретним рядом

Розривне зусилля,кН	81	83	85	87	89	91	93
К-сть зразків	20	47	80	89	40	16	8

Точковою оцінкою для математичного сподівання  $a$  є вибіркове середнє, яке обчислюється за формулою (2.1), тому

$$\begin{aligned}\bar{x}_B &= \frac{1}{300} (81 \cdot 20 + 83 \cdot 47 + 85 \cdot 80 + 87 \cdot 89 + 89 \cdot 40 + 91 \cdot 16 + 93 \cdot 8) = \\ &= \frac{1}{300} (1620 + 3901 + 6800 + 7743 + 3560 + 1456 + 744) = \frac{1}{300} \cdot 25824 \approx 86,08.\end{aligned}$$

Точковою оцінкою для дисперсії  $\sigma^2$  є виправлена вибіркова дисперсія  $S^2$ , яка знаходиться відповідно до формули (2.2):

$$\begin{aligned}S^2 &= \frac{1}{299} (81^2 \cdot 20 + 83^2 \cdot 47 + 85^2 \cdot 80 + 87^2 \cdot 89 + 89^2 \cdot 40 + 91^2 \cdot 16 + 93^2 \cdot 8) - \\ &- \frac{300}{299} \cdot (86,08)^2 = \frac{1}{299} (131220 + 323783 + 578000 + 673641 + 316840 + \\ &+ 132496 + 69192) - \frac{300}{299} \cdot 7409,77 = \frac{1}{299} \cdot 2225172 - \frac{300}{299} \cdot 7409,77 \approx \\ &\approx 7442,05 - 7434,55 \approx 7,50.\end{aligned}$$

Звідси виправлене вибіркове середнє квадратичне відхилення

$$S = \sqrt{7,50} \approx 2,74.$$

Знайдемо  $\gamma$  - надійний інтервал для математичного сподівання  $a$  і середнього квадратичного відхилення  $\sigma$  спостережуваної випадкової величини.

Інтервал надійності для математичного сподівання  $a$  за вибірковим середнім  $\bar{x}_B$ , якщо  $\sigma$  невідоме, обчислюється за формулою (2.4). Із таблиці 4

додатку за відомим  $\gamma = 0,99$  та  $n = 300$  знаходимо  $t_\gamma = 2,576$ . Тоді точність оцінки

$$\delta = \frac{S}{\sqrt{n}} t_\gamma = \frac{2,74}{\sqrt{300}} \cdot 2,576 \approx \frac{2,74}{17,32} \cdot 2,576 \approx 0,41.$$

Отже,

$$86,08 - 0,41 < a < 86,08 + 0,41,$$

звідки  $\gamma$  - надійним інтервалом ( $\gamma = 0,99$ ) для математичного сподівання  $a \in (85,67; 86,49)$ .

Інтервал надійності для  $\sigma$  за виправленим вибірковою середнім квадратичним відхиленням  $S$ , обчислюється за формулою (2.5).

Значення  $q_\gamma$  за відомим  $\gamma = 0,99$ ,  $n = 30$  знаходимо із таблиці 5 додатку: маємо  $q_\gamma = 0,12$ . Оскільки  $q_\gamma < 1$ , то

$$2,74(1 - 0,12) < \sigma < 2,74(1 + 0,12),$$

звідки  $2,41 < \sigma < 3,07$ . Отже,  $\gamma$  - надійним інтервалом ( $\gamma = 0,99$ ) для середнього квадратичного відхилення  $\sigma \in (2,41; 3,07)$ .

## Завдання Б

У деякій місцевості серед досліджених довільно вибраних 15 днів року виявилось 6 дощових. Знайти точкову оцінку ймовірності  $p$  того, що вибраний день буде дощовий та інтервал надійності для оцінки  $p$  з надійністю  $\gamma = 0,95$ .

Розв'яжемо задачу, припускаючи, що вибірка отримана з генеральної сукупності, розподіленої за біномним законом.

Точковою оцінкою параметра  $p$  є відносна частота появи події  $\omega = \frac{6}{15} = 0,4$ .

Інтервал надійності для  $p$  за відносною частотою  $\omega$  з надійністю  $\gamma$  визначимо згідно формулою (2.6).

Значення  $t_\gamma$ , для якого  $\Phi(t_\gamma) = \frac{\gamma}{2} = \frac{0,95}{2} = 0,475$  знаходимо за таблицею 3 додатку і маємо  $t_\gamma = 1,96$ .

Обчислимо

$$\begin{aligned} \Delta &= t_\gamma \sqrt{\frac{\omega(1-\omega)}{n} + \left(\frac{t_\gamma}{2n}\right)^2} = 1,96 \sqrt{\frac{0,4(1-0,4)}{15} + \left(\frac{1,96}{2 \cdot 15}\right)^2} \approx \\ &\approx 1,96 \sqrt{0,016 + 0,004} \approx 1,96 \sqrt{0,020} \approx 1,96 \cdot 0,14 \approx 0,28. \end{aligned}$$

Отже, інтервалом надійності для ймовірності  $p \in$

$$\frac{15}{(1,96)^2 + 15} \left[ 0,4 + \frac{(1,96)^2}{2 \cdot 15} - 0,28 \right] < p < \frac{15}{(1,96)^2 + 15} \left[ 0,4 + \frac{(1,96)^2}{2 \cdot 15} + 0,28 \right],$$

$$0,80 \cdot 0,25 < p < 0,80 \cdot 0,81,$$

$$0,20 < p < 0,65.$$

Таким чином,  $\gamma$  - надійним інтервалом ( $\gamma = 0,95$ ) для ймовірності  $p \in (0,20; 0,65)$ .

### Лабораторна робота № 3

Дано розподіл однотипних підприємств за обсягом продукції  $X$  та її собівартістю  $Y$  (в умовних одиницях) (його одержали згідно з таблицею задачі 1 при  $k = 0$ ).

$Y \backslash X$	2	3	4	5	6
100				11	1
200					10
300			10	1	
400		14			
500	12				
600	1	12			

Виходячи з цього розподілу, будуюмо кореляційне поле (пропонуємо вибрати масштаб по осі  $OX$  1 : 1, а по осі  $OY$  100 : 1).

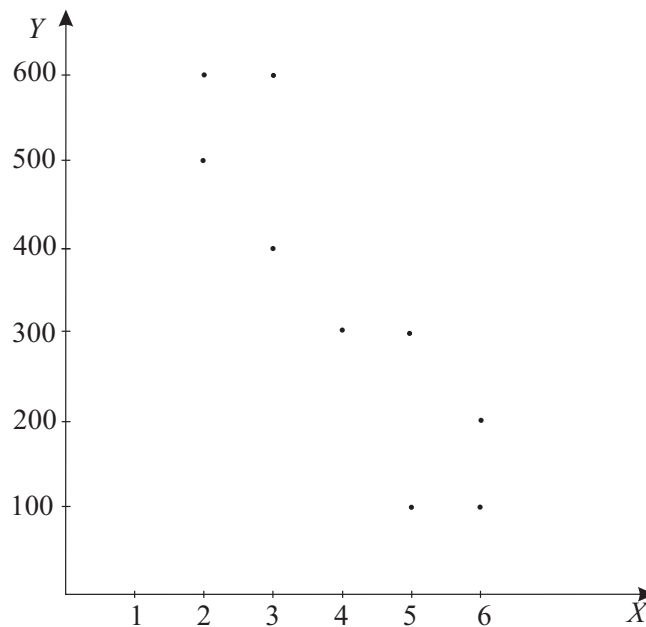


Рис. 4. Кореляційне поле

Результати проміжних обчислень, необхідних для знаходження коефіцієнта кореляції  $r$ , запишемо у вигляді таблиці, де  $n = \sum_{i=1}^5 n_i = 13 + 26 + 10 + 12 +$

$$+11 = 72 \text{ або } n = \sum_{j=1}^6 m_j = 12 + 10 + 11 + 14 + 12 + 13 = 72.$$

$Y \backslash X$	2	3	4	5	6	$m_j$	$y_j m_j$	$y_j^2 m_j$	$b_j$
100				11	1	12	1200	120000	61
200					10	10	2000	400000	60
300			10	1		11	3300	990000	45
400		14				14	5600	2240000	42
500	12					12	6000	3000000	24
600	1	12				13	7800	4680000	38
$n_i$	13	26	10	12	11	72	25900	11430000	83200
$x_i n_i$	26	78	40	60	66	270			
$x_i^2 n_i$	52	234	160	300	396	1142			
$a_i$	6600	12800	3000	1400	2100	83200			

Елементи таблиці  $a_i$  і  $b_j$  було обчислено згідно з формулами (3.10) – (3.11):

$$a_i = \sum_{j=1}^6 y_j n_{ij}, \quad b_j = \sum_{i=1}^5 x_i n_{ij}.$$

Згідно зі зробленими у таблиці обрахунками та формулами (3.4)–(3.11) матимемо

$$\bar{x}_B = \frac{1}{n} \sum_{i=1}^5 x_i n_i = \frac{270}{72} = 3,75;$$

$$\bar{y}_B = \frac{1}{n} \sum_{j=1}^6 y_j m_j = \frac{25900}{72} \approx 359,72;$$

$$\overline{x^2} = \frac{1}{n} \sum_{i=1}^5 x_i^2 n_i = \frac{1142}{72} \approx 15,86;$$

$$\overline{y^2} = \frac{1}{n} \sum_{j=1}^6 y_j^2 m_j = \frac{11430000}{72} = 158750;$$

$$\overline{xy} = \frac{S_{kl}}{n} = \frac{83200}{72} \approx 1155,56;$$

$$\sigma_B^2(X) = \overline{x^2} - (\bar{x}_B)^2 = 15,86 - (3,75)^2 \approx 1.80;$$

$$\sigma_B^2(Y) = \overline{y^2} - (\bar{y}_B)^2 = 158750 - (359,72)^2 \approx 29351.52.$$

Тоді коефіцієнт кореляції згідно з (3.3) дорівнює

$$r = \frac{\overline{(xy)} - \bar{x}_B \bar{y}_B}{\sigma_B(X) \sigma_B(Y)} = \frac{1155,56 - 3,75 \cdot 359,72}{1,34 \cdot 171,32} \approx \frac{-193,39}{228,54} \approx -0,85,$$

а кутові коефіцієнти прямих ліній регресії згідно з (3.12) набувають значень

$$\mu_{yx} = r \frac{\sigma_B(Y)}{\sigma_B(X)} \approx -0,85 \cdot \frac{171,32}{1,34} \approx -108,67;$$

$$\mu_{xy} = r \frac{\sigma_B(X)}{\sigma_B(Y)} \approx -0,85 \cdot \frac{1,34}{171,32} \approx -0,0066.$$

Рівняння прямої лінії регресії  $Y$  на  $X$  згідно з (3.1) матиме вигляд

$$\frac{y_p(x) - 359,72}{171,32} = -0,85 \frac{x - 3,75}{1,34},$$

або після спрощень

$$y_p(x) = -108,67x + 767,23.$$

Аналогічно згідно з (3.2) одержують рівняння прямої лінії регресії  $X$  на  $Y$

$$\frac{x_p(y) - 3,75}{1,34} = -0,85 \frac{y - 359,72}{171,32},$$

яке після спрощень набуває остаточного вигляду

$$x_p(y) = -0,0066y + 6,12.$$

Точкою перетину знайдених прямих як відомо є точка, абсцисою і ординатою якої є вибіркові середні відповідно випадкових величин  $X$  та  $Y$ , тобто точка  $C(3,75; 359,72)$ .

Перейдемо до обчислення числових характеристик відхилення значень випадкових величин від точок ліній регресії.

Сума квадратів відхилень випадкової величини  $Y$  від лінії регресії  $y_p(x)$ , обчислена згідно з формулою (3.16), дорівнює

$$n\sigma^2(y, y_p(x)) = 72 \cdot 29350,54 \cdot (1 - (-0,85)^2) \approx 3640054,27,$$

а сума квадратів відхилень випадкової величини  $X$  від лінії регресії  $x_p(y)$ , обчислена згідно з формулою (3.20), дорівнює

$$n\sigma^2(x, x_p(y)) = 72 \cdot 1,79 \cdot (1 - (-0,85)^2) \approx 221,99.$$

Обчислимо дві інші характеристики розсіювання значень випадкових величин  $X$  та  $Y$  від прямої регресії, а саме суми квадратів відхилень  $\sigma^2$  умовних

середніх значень  $\bar{y}(x)$  величини  $Y$  від значень на лінії регресії  $y_p(x)$  та умовних середніх значень  $\bar{x}(y)$  величини  $X$  від значень на лінії регресії  $x_p(y)$ . Результати проміжних обчислень зручно впорядковувати за допомогою таблиці

$y_p(x_i)$	548,68	440,7	332,72	224,74	116,76
$\bar{y}(x_i)$	507,69	492,31	300	116,67	190,91
$\delta_i$	-40,99	51,61	-32,72	-108,07	74,15
$\delta_i^2 n_i$	21842,34	69253,39	10705,98	140149,48	60480,45

Елементи першого рядка таблиці одержано підстановкою у рівняння прямої лінії регресії  $Y$  на  $X$  значень  $x_i$ ,  $i = \overline{1, 5}$ :

$$y_p(x_1) = y_p(2) = -107,98 \cdot 2 + 764,64 = 548,68;$$

$$y_p(x_2) = y_p(3) = -107,98 \cdot 3 + 764,64 = 440,70;$$

$$y_p(x_3) = y_p(4) = -107,98 \cdot 4 + 764,64 = 332,72;$$

$$y_p(x_4) = y_p(5) = -107,98 \cdot 5 + 764,64 = 224,74;$$

$$y_p(x_5) = y_p(6) = -107,98 \cdot 6 + 764,64 = 116,76.$$

Другий рядок таблиці заповнюють згідно з формулою (3.22), де значення  $a_i$  та  $n_i$  знаходять з другої таблиці

$$\bar{y}(x_1) = \bar{y}(2) = \frac{a_1}{n_1} = \frac{6600}{13} \approx 507,69;$$

$$\bar{y}(x_2) = \bar{y}(3) = \frac{a_2}{n_2} = \frac{12800}{26} \approx 492,31;$$

$$\bar{y}(x_3) = \bar{y}(4) = \frac{a_3}{n_3} = \frac{3000}{10} = 300,00;$$

$$\bar{y}(x_4) = \bar{y}(5) = \frac{a_4}{n_4} = \frac{1400}{12} \approx 116,67;$$

$$\bar{y}(x_5) = \bar{y}(6) = \frac{a_5}{n_5} = \frac{2100}{11} \approx 190,91.$$

Рядок значень  $\delta_i$  заповнюють згідно з першою з формул (3.24), тобто його елементами є відповідні різниці елементів другого та першого рядків. Останній четвертий рядок складається із добутків квадратів величин  $\delta_i$  на відповідні частоти  $n_i$ .

Остаточно сума  $\delta^2$  квадратів відхилень умовних середніх значень  $\bar{y}(x)$  випадкової величини  $Y$  від ординат відповідних точок на лінії регресії  $y_p(x)$



згідно з другою з формул (3.24) дорівнює

$$\delta^2 = \sum_{i=1}^5 \delta_i^2 n_i = 21842,34 + 69253,39 + 10705,98 + \\ + 140149,48 + 60480,45 = 302431,64.$$

Аналогічно обчислюють суму  $\varepsilon^2$  квадратів відхилень умовних середніх значень  $\bar{x}(y)$  випадкової величини  $X$  від ординат відповідних точок на лінії регресії  $x_p(y)$ . Відповідна таблиця проміжних обчислень буде такою

$x_p(y_j)$	5,46	4,8	4,14	3,48	2,82	2,16
$\bar{x}(y_j)$	5,08	6	4,09	3	2	2,92
$\varepsilon_j$	-0,38	1,2	-0,05	-0,48	-0,82	0,76
$\varepsilon_j^2 m_j$	1,73	14,4	0,03	3,23	8,07	7,51

Елементи першого рядка було обчислено згідно рівняння прямої лінії регресії  $X$  на  $Y$ , другий – за формулою (3.26), а третій – (3.28).

Остаточно

$$\varepsilon^2 = 1,73 + 14,4 + 0,03 + 3,23 + 8,07 + 7,51 = 34,97.$$

Суттєва відмінність від нуля обчислених відхилень свідчать про те, що характер зв'язку між величинами  $X$  та  $Y$  не є лінійним.

## Лабораторна робота № 4

**Задача А.** Для вивчення технічних властивостей нової марки бетону досліджувались окремі його зразки. Розподіл кількості  $n_i$  зразків бетону і відповідного їм стискування  $X$  (тобто такого, що веде до руйнування зразка) наведено в таблиці (її дістали згідно з умовою задачі 1 для варіанту 0):

$X, \text{кг/см}^2$	170-180	180-190	190-200	200-210	210-220
$n_i$	4	9	32	54	72

$X, \text{кг/см}^2$	220-230	230-240	240-250	250-260	260-270
$n_i$	72	65	50	25	12

1. Оскільки досліджувана ознака має неперервний характер, то для даного статистичного розподілу побудуємо гістограму частот (рис. 5).

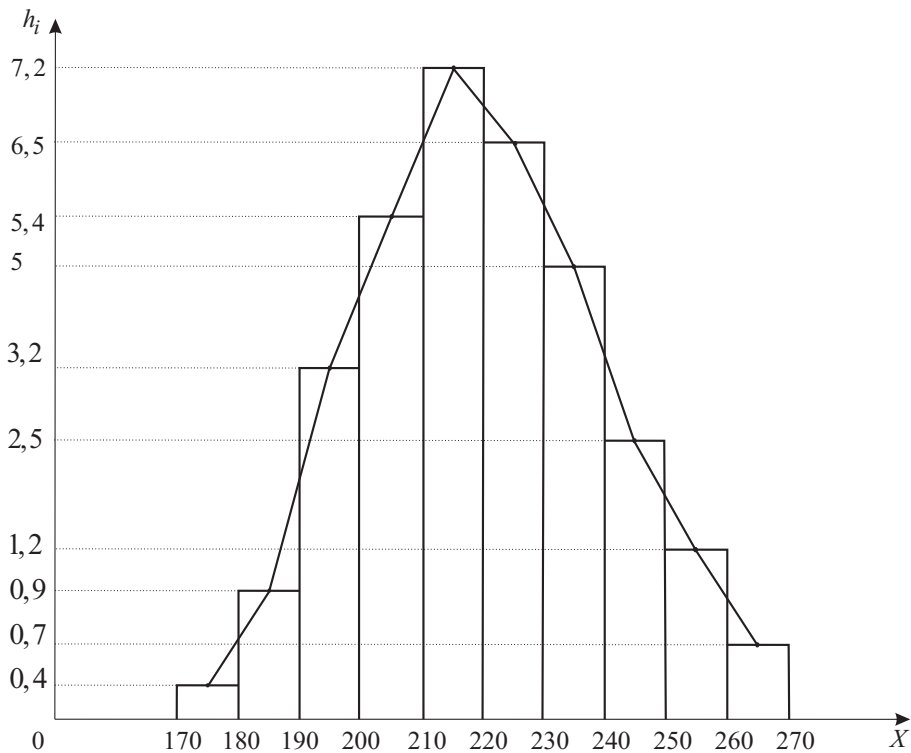


Рис. 5. Гістограма статистичного розподілу частот

2. Сформулюємо гіпотезу про закон розподілу досліджуваної ознаки.

За формою гістограми частот можемо припустити, що ознака  $X$  має нормальний закон розподілу. Отже, висуваємо нульову гіпотезу  $H_0$  : ознака  $X$  має нормальний закон розподілу ймовірностей.

3. Для перевірки правильності гіпотези  $H_0$  використаємо критерій Пірсона (критерій  $\chi^2$ ), який передбачає порівняння  $\chi_{емп}^2$  та  $\chi_{кр}^2$ .

Оскільки щільність нормального розподілу

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad -\infty < x < \infty,$$

визначається двома параметрами (математичним сподіванням  $a$  та середнім квадратичним відхиленням  $\sigma$ ), які є невідомими, то знайдемо їх точкові оцінки  $a^*$  та  $\sigma^*$ , які відповідно дорівнюють вибірковому середньому  $\bar{x}_B$  та виправленому вибірковому середньому квадратичному відхиленню  $S$  (див. л.р. № 2). Побудувавши дискретний розподіл за заданим інтервальним рядом, а саме

$X, \text{кг/см}^2$	175	185	195	205	215	225	235	245	255	265
$n_i$	4	9	32	54	72	65	50	25	12	7

обчислимо вибіркове середнє  $\bar{x}_B \approx 220 \text{ кг/см}^2$  та виправлене середнє квадратичне відхилення  $S \approx 18,32 \text{ кг/см}^2$ .

Оскільки випадкову величину  $X$  вважаємо нормально розподіленою, то теоретичні ймовірності  $p_i$  обчислюємо за формулою

$$p_i = P(z_{i-1} \leq X < z_i) = \Phi\left(\frac{z_i - \bar{x}_B}{S}\right) - \Phi\left(\frac{z_{i-1} - \bar{x}_B}{S}\right).$$

Введемо позначення  $\beta_{i-1} = \frac{z_{i-1} - \bar{x}_B}{S}$ ,  $\beta_i = \frac{z_i - \bar{x}_B}{S}$ .

Тоді, попередньо об'єднавши перший та другий інтервали (оскільки частота першого інтервалу менша за 5), та взявши лівий кінець першого інтервалу за  $-\infty$ , а правий кінець останнього інтервалу за  $+\infty$  обчислення теоретичних частот  $n'_i$

$$n'_i = n \cdot (\Phi(\beta_i) - \Phi(\beta_{i-1}))$$

можна подати у таблиці:

$z_{i-1}$	$z_i$	$n_i$	$\beta_{i-1}$	$\beta_i$	$\Phi(\beta_{i-1})$	$\Phi(\beta_i)$	$n'_i$
$-\infty$	190	13	$-\infty$	-1,64	-0,5	-0,4495	17
190	200	32	-1,64	-1,09	-0,4495	-0,3621	29
200	210	54	-1,09	-0,55	-0,3621	-0,2088	51
210	220	72	-0,55	0,00	-0,2088	0,00	69
220	230	65	0,00	0,55	0,00	0,2088	69
230	240	50	0,55	1,09	0,2088	0,3621	51
240	250	25	1,09	1,64	0,3621	0,4495	29
250	260	12	1,64	2,18	0,4495	0,4854	12
260	$+\infty$	7	2,18	2,73	0,4854	0,5	5

Для знаходження емпіричного значення  $\chi^2_{\text{емп}} = \sum_{i=0}^9 \frac{(n_i - n'_i)^2}{n'_i}$  скористаємось таблицею:

$n_i$	$n'_i$	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$
13	17	-4	16	0,9412
32	29	3	9	0,3103
54	51	3	9	0,1765
72	69	3	9	0,1304
65	69	-4	16	0,2319
50	51	-1	1	0,0196
25	29	-4	16	0,5517
12	12	0	0	0
7	5	2	4	0,8
330	99	-	-	$\chi^2_{\text{емп}} = 3,1616$

За таблицею додатка 6 знаходимо критичне значення критерію  $\chi^2$

$$\chi^2_{kp}(\alpha = 0,05; k = 9 - 2 - 1) = \chi^2_{kp}(\alpha = 0,05; k = 6) = 12,6.$$

Оскільки,  $\chi^2_{\text{емп}} = 3,1616 < \chi^2_{kp} = 12,6$ , то нульову гіпотезу приймаємо, тобто стискування  $X$  зразків бетону розподілене за нормальним законом.

**Задача Б.** Для покращення обслуговування сільськогосподарських машин були зібрані дані про вихід з ладу техніки у господарствах району за період весняно-польових робіт. Розподіл кількості  $X$  поломок у перевірених  $n_i$  одиницях техніки відображено у таблиці (її отримали згідно з умовою задачі 7 для варіанту 0):

$X$ , к-сть поломок	0	1	2	3	4	5
$n_i$	2462	1767	605	141	20	5

1. Нехай випадкова величина  $X$  – кількість поломок у перевірених одиницях  $n_i$  техніки. Оскільки, досліджувана ознака  $X$  є дискретною, то побудуємо полігон частот (рис. 6).

2. З вигляду полігону частот та змісту випадкової величини  $X$  робимо припущення, що  $X$  розподілена за законом Пуассона.

Отже, висуваємо нульову гіпотезу:

$$H_0 : P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots,$$

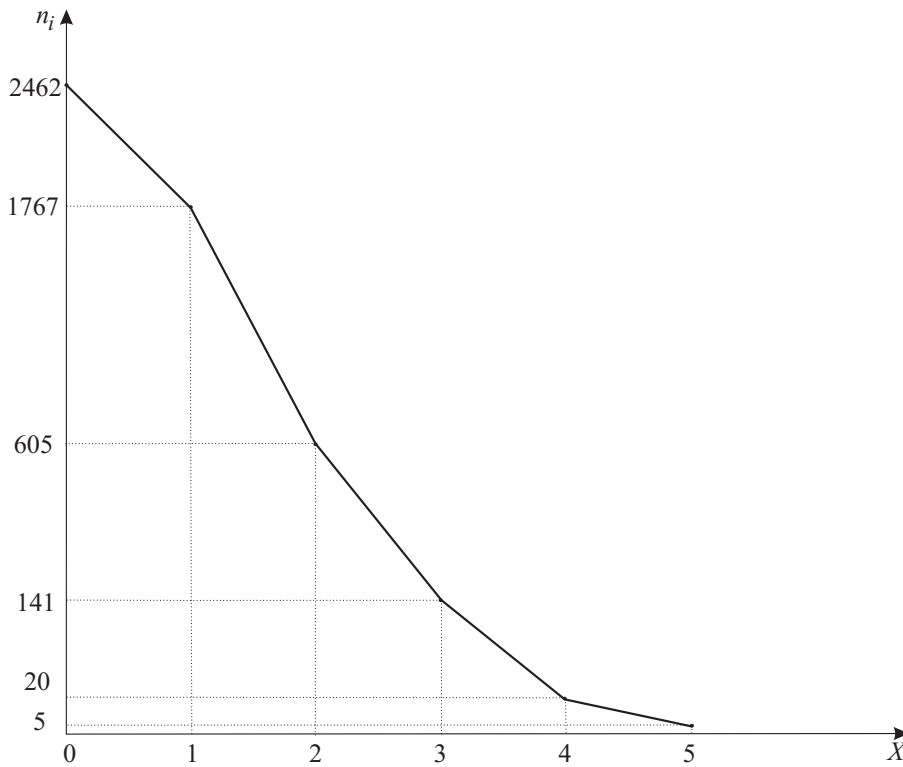


Рис. 6. Полігон частот

де  $\lambda = M(X)$  – невідомий параметр, що визначає цей розподіл.

Точковою оцінкою параметра  $\lambda$  розподілу Пуассона є вибіркове середнє  $\bar{x}_B$ . У даному випадку

$$\begin{aligned}\lambda^* = \bar{x}_B &= \frac{1}{5000} (2462 \cdot 0 + 1767 \cdot 1 + 605 \cdot 2 + 141 \cdot 3 + 20 \cdot 4 + 5 \cdot 5) = \\ &= \frac{1}{5000} \cdot 3505 = 0,701 \approx 0,7.\end{aligned}$$

Обчислимо теоретичні ймовірності  $p_k$  за формулою Пуассона, покладаючи в ній  $\lambda = 0,7$ :

$$p_k = \frac{0,7^k}{k!} e^{-0,7}, \quad k = 0, 1, 2, 3, 4, 5.$$

Слід зауважити також, що остання ймовірність  $p_5$  буде визначатися як доповнення суми попередніх ймовірностей до одиниці:

$$p_0 = P(X = 0) = \frac{0,7^0}{0!} e^{-0,7} = \frac{1}{e^{0,7}} \approx 0,4966;$$

$$p_1 = P(X = 1) = \frac{0,7^1}{1!} e^{-0,7} = \frac{0,7}{e^{0,7}} \approx 0,3476;$$

$$p_2 = P(X = 2) = \frac{0,7^2}{2!} e^{-0,7} = \frac{0,49}{2e^{0,7}} \approx 0,1217;$$

$$p_3 = P(X = 3) = \frac{0,7^3}{3!} e^{-0,7} = \frac{0,343}{6e^{0,7}} \approx 0,0284;$$

$$p_4 = P(X = 4) = \frac{0,7^4}{4!} e^{-0,7} = \frac{0,2401}{24e^{0,7}} \approx 0,0050;$$

$$p_5 = P(X = 5) \approx 1 - (0,4966 + 0,3476 + 0,1217 + 0,0284 + 0,0050) = 0,0007.$$

Для обчислення  $\chi_{\epsilon mn}^2$  використаємо формулу (4.4):

$$\chi_{\epsilon mn}^2 = \sum_{i=0}^5 \frac{(n_i - n'_i)^2}{n'_i},$$

де теоретичні частоти  $n'_i$  визначаються з рівності  $n'_i = np_i$ :

$$n'_0 = 5000 \cdot 0,4966 = 2483; \quad n'_1 = 5000 \cdot 0,3476 = 1738;$$

$$n'_2 = 5000 \cdot 0,1217 \approx 609; \quad n'_3 = 5000 \cdot 0,0284 = 142;$$

$$n'_4 = 5000 \cdot 0,005 = 25; \quad n'_5 = 5000 \cdot 0,0007 \approx 4.$$

Подальші обчислення запишемо у формі таблиці:

$n_i$	$n'_i$	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$
2462	2483	-21	441	0,1776
1767	1738	29	841	0,4839
605	609	-4	16	0,0262
141	142	-1	1	0,0070
20	25	-5	25	1
5	4	1	1	0,25
$\sum_{i=1}^6 n_i = 5000$	$\sum_{i=1}^6 n'_i = 5001$	-	-	$\sum_{i=1}^6 \frac{(n_i - n'_i)^2}{n'_i} = 1,9447$

За таблицею додатка 6 знаходимо критичне значення критерію  $\chi^2$ :

$$\chi_{kp}^2(\alpha = 0,05; k = 6 - 1 - 1) = \chi_{kp}^2(\alpha = 0,05; k = 4) = 9,5.$$

Оскільки  $\chi_{\epsilon mn}^2 = 1,9447 < \chi_{kp}^2 = 9,5$ , то сформульована гіпотеза  $H_0$  про те, що кількість полумок у перевірених одиницях сільськогосподарської техніки має розподіл Пуассона за рівня значущості  $\alpha = 0,05$ , приймається, бо вона не суперечить статистичним даним.

## Лабораторна робота № 5

**Задача А.** Покращення якості готової продукції дає підстави передбачити зменшення витрат підприємства на ремонтне обслуговування приладів за час їх гарантійного строку служби. Дані про місячні витрати на гарантійне обслуговування приладів за останні 12 місяців наведено у таблиці (її отримали з таблиці 5.2 для варіанту 0):

В-т	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$	$x_{11}$	$x_{12}$	$a$	$\sigma_0$
0	56.4	52.4	49.3	46.7	50.6	53.1	46.9	47.3	54	49.3	48	47.6	50.5	3.2

Вважається, що вибірка зроблена з нормально розподіленої сукупності.

1. Перевіримо гіпотезу про те, що зміни у процесі виробництва не призвели до зменшення затрат  $m$  на гарантійний ремонт приладів за умови, що планові щомісячні відрахування підприємства на ремонт становлять  $a = 50,5$  одиниць, а середньоквадратичне відхилення цих затрат відоме і дорівнює  $\sigma_0 = 3,2$  одиниць. Отже, потрібно перевірити нульову гіпотезу  $H_0 : m = 50,5$  за альтернативної гіпотези  $H_1 : m < 50,5$  при рівні значущості  $\alpha = 0,05$ .

За даною вибіркою обчислимо вибіркове середнє

$$\bar{x}_B = \frac{1}{12}(56,5 + 52,4 + 49,3 + 46,7 + 50,6 + 53,1 + 46,9 + 47,3 + 54,4 + 49,3 + 48 + 47,6) = 50,175.$$

Оскільки середнє квадратичне відхилення  $\sigma$  затрат на ремонт відоме, то за критерій вибираємо статистику задачі А (випадок 1), тобто

$$Z = \frac{\bar{x}_B - a_0}{\sigma} \sqrt{n}.$$

Тоді

$$Z_{emn} = \frac{50,175 - 50,5}{3,2} \sqrt{12} \approx -0,352$$

Оскільки альтернативна гіпотеза має вигляд  $H_1 : m < 50,5$ , то критична область буде односторонньою і її визначатиме інтервал  $(-\infty; -z_{\kappa p})$ . За формулою (5.3) обчислюємо

$$\Phi(z_{\kappa p}) = (1 - 2\alpha)/2 = (1 - 2 \cdot 0,05)/2 = 0,45.$$

З таблиць значень функції Лапласа (додаток 3) знаходимо  $z_{\kappa p} = 1,65$ . Звідси випливає, що критична область складається з інтервалу  $(-\infty; -1,65)$ . Значення  $Z_{emn} = -0,352$  не потрапляє в критичну область, тому гіпотеза

$H_0 : m = 50,5$  приймається. Отже, зміни у процесі виробництва не призвели до зменшення затрат  $m$  на гарантійний ремонт приладів.

2. Перевіримо гіпотезу про те, що середньоквадратичне відхилення  $\sigma$  затрат на ремонт мають заданий рівень  $\sigma_0$  за умови, що середньомісячні затрати  $a$  на ремонтне обслуговування вважаються невідомими. Отже, потрібно перевірити нульову гіпотезу  $H_0 : \sigma = 3,2$  за альтернативної гіпотези  $H_1 : \sigma \neq 3,2$  при рівні значущості  $\alpha = 0,05$ .

Оскільки середньомісячні затрати  $a$  на ремонтне обслуговування вважаються невідомими, то для перевірки гіпотези про те, що середньоквадратичне відхилення  $\sigma$  затрат на ремонт мають заданий рівень  $\sigma_0$ , використаємо статистику

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

Спочатку знайдемо виправлену дисперсію  $S^2$ . За заданою в умові вибіркою маємо

$$S^2 = \frac{12}{11} \left( \frac{1}{12} (56,5^2 + 52,4^2 + 49,3^2 + 46,7^2 + 50,6^2 + 53,1^2 + 46,9^2 + 47,3^2 + 54,4^2 + 49,3^2 + 48^3 + 47,6^2) - 50,175^2 \right) \approx 10,053.$$

Тоді

$$\chi_{emn}^2(n-1) = \frac{(12-1) \cdot 10,053}{3,2^2} \approx 10,799.$$

Критична область буде симетричною двосторонньою, її утворюють інтервали  $(-\infty; t_{kp}^{(1)}) \cup (t_{kp}^{(2)}; +\infty)$ . З таблиць критичних точок розподілу  $\chi^2$  (додаток 6) знаходимо  $t_{kp}^{(2)} = t_{kp}(\alpha = 0,005; n-1 = 11) = 19,7$ . Точка  $t_{kp}^{(1)}$  розташована симетрично і дорівнює  $-19,7$ . Звідси випливає, що критична область складається з інтервалів  $(-\infty; -19,7) \cup (19,7; +\infty)$ . Значення  $\chi_{emn}^2(n-1) = 10,799$  не потрапляє в критичну область, тому гіпотеза  $H_0 : \sigma = 3,2$  приймається. Отже, середнє квадратичне відхилення  $\sigma$  затрат на ремонт дорівнює  $3,2$  (ум.од).

**Задача Б.** На підприємстві впроваджено нову технологію з метою економії використаних у виробничому процесі ресурсів. Дані про вартість сировини (в умовних одиницях), затрачені на виготовлення одиниці продукції старим і новим способами наведено у двох вибірках  $x_i, i = 1, 2, \dots, 9$  та  $y_j, j = 1, 2, \dots, 12$ , (таблиця 5.2, варіант 0).



$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$	$x_{11}$	$x_{12}$
$y_1$	$y_2$	$y_3$	$y_4$	$y_5$	$y_6$	$y_7$	$y_8$	$y_9$	$y_{10}$	$y_{11}$	$y_{12}$
10,9	10,8	10,6	10,0	10,6	10,2	10,9	10,7	9,7	—	—	—
11,2	11,6	10,2	10,9	11,7	10,4	11,3	9,9	10,1	10,4	11,0	11,2

1. Перевіримо гіпотезу про те, що середньоквадратичні відхилення  $\sigma_x$  та  $\sigma_y$  величини затрат сировини на виготовлення одиниці продукції є однаковими для обидвох генеральних сукупностей за умови, що середні затрати  $a_x$  та  $a_y$  сировини у цих сукупностях невідомі. Тобто, перевіримо нульову гіпотезу  $H_0 : \sigma_x = \sigma_y$  за альтернативної гіпотези  $H_1 : \sigma_x \neq \sigma_y$  при рівні значущості  $\alpha = 0,05$ .

Обчислимо виправлені вибіркові дисперсії  $S_x^2$ ,  $S_y^2$ . Для цього спочатку знайдемо вибіркові середні  $\bar{x}_B$  та  $\bar{y}_B$ :

$$\bar{x}_B = \frac{1}{9} (10,9 + 10,8 + 10,6 + 10,0 + 10,6 + 10,2 + 10,9 + 10,7 + 9,7) \approx 10,489,$$

$$\begin{aligned} \bar{y}_B = \frac{1}{12} (11,2 + 11,6 + 10,2 + 10,9 + 11,7 + 10,4 + 11,3 + 9,9 + \\ + 10,1 + 10,4 + 11,0 + 11,2) = 10,825. \end{aligned}$$

Тоді

$$\begin{aligned} S_x^2 = \frac{9}{8} \left( \frac{1}{9} (10,9^2 + 10,8^2 + 10,6^2 + 10,0^2 + 10,6^2 + 10,2^2 + 10,9^2 + \right. \\ \left. + 10,7^2 + 9,7^2) - 10,489^2 \right) = 0,179; \end{aligned}$$

$$\begin{aligned} S_y^2 = \frac{12}{11} \left( \frac{1}{12} (11,2^2 + 11,6^2 + 10,2^2 + 10,9^2 + 11,7^2 + 10,4^2 + 11,3^2 + \right. \\ \left. + 9,9^2 + 10,1^2 + 10,4^2 + 11,0^2 + 11,2^2) - 10,825^2 \right) \approx 0,367. \end{aligned}$$

Враховуючи, що  $S_x^2 < S_y^2$ , маємо

$$F_{emn} = \frac{0,367}{0,179} \approx 2,05.$$

Оскільки  $k_1 = 12 - 1 = 11$ ,  $k_2 = 9 - 1 = 8$ , то критичне значення  $F_{kp}$  знаходимо з умови

$$P\{F(k_1 = 11, k_2 = 8) > F_{kp}\} = \frac{\alpha}{2} = 0,025.$$

За таблицею розподілу Фішера-Снедекора (додаток 8) визначаємо  $F_{kp} = 3,31$ . Оскільки число  $F_{emn} = 2,05$  не потрапляє в критичну область  $(3,31; +\infty)$ , то гіпотезу про рівність середньоквадратичних відхилень  $\sigma_x$  та

$\sigma_y$  величини затрат сировини на виготовлення одиниці продукції в обидвох генеральних сукупностей приймаємо.

2. Встановимо, чи можна вважати, що середні затрати сировини  $m_x$  та  $m_y$  у цих генеральних сукупностях однакові. Тобто, виконаємо перевірку нульової гіпотези  $H_0 : m_x = m_y$  за альтернативної гіпотези  $H_1 : m_x \neq m_y$  при рівні значущості  $\alpha = 0,05$ .

Для перевірки гіпотези про рівність середніх затрат сировини у двох заданих генеральних сукупностях вважатимемо, що математичне сподівання і середнє квадратичне відхилення законів розподілу невідомі. Тоді скористаємось статистикою

$$T = \frac{\bar{x}_B - \bar{y}_B}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}}.$$

Випадкова величина  $T$  підпорядкована розподілу Стюдента з  $k = m + n - 2$  ступенями вільності.

Обчислюємо емпіричне значення критерію

$$T_{emn} = \frac{10,489 - 10,825}{\sqrt{\frac{0,179}{9} + \frac{0,367}{12}}} \approx 1,493.$$

За таблицею розподілу (додаток 7) для  $k = 9 + 12 - 2 = 19$  і  $\alpha = 0,05$  рівня значущості (для двосторонньої критичної області) знаходимо  $t_{kp} = 2,09$ . Це означає, що критична область є об'єднанням інтервалів  $(-\infty; -2,09) \cup (2,09; +\infty)$ . Отримане значення критерію  $T_{emn}$  не належить критичній області. Звідси випливає, що середні затрати сировини  $m_x$  та  $m_y$  у цих генеральних сукупностях практично однакові.

# Д О Д А Т О К

Таблиця 1

Значення функції  $\frac{\lambda^k}{k!} e^{-\lambda}$

$k$	$\lambda=0,1$	$\lambda=0,2$	$\lambda=0,3$	$\lambda=0,4$	$\lambda=0,5$	$\lambda=0,6$	$\lambda=0,7$	$\lambda=0,8$	$\lambda=0,9$	
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066	
1	0,0905	0,1637	0,2222	0,2681	0,3033	0,3293	0,3476	0,3595	0,3659	
2	0,0045	0,0164	0,0333	0,0536	0,0758	0,0988	0,1217	0,1438	0,1647	
3	0,0002	0,0011	0,0033	0,0072	0,0126	0,0198	0,0284	0,0383	0,0494	
4	0,0000	0,0001	0,0003	0,0007	0,0016	0,0030	0,0050	0,0077	0,0111	
5		0,0000	0,0000	0,0001	0,0002	0,0004	0,0007	0,0012	0,0020	
6				0,0000	0,0000	0,0000	0,0001	0,0002	0,0003	
$k$	$\lambda=1$	$\lambda=2$	$\lambda=3$	$\lambda=4$	$\lambda=5$	$\lambda=6$	$\lambda=7$	$\lambda=8$	$\lambda=9$	$\lambda=10$
0	0,3679	0,1353	0,0498	0,0183	0,0067	0,0025	0,0009	0,0003	0,0001	0,0000
1	0,3679	0,2707	0,1494	0,0733	0,0337	0,0149	0,0064	0,0027	0,0011	0,0005
2	0,1839	0,2707	0,2240	0,1465	0,0842	0,0446	0,0223	0,0107	0,0050	0,0023
3	0,0613	0,1804	0,2240	0,1954	0,1404	0,0892	0,0521	0,0286	0,0150	0,0076
4	0,0153	0,0902	0,1680	0,1954	0,1755	0,1339	0,0912	0,0572	0,0337	0,0189
5	0,0031	0,0361	0,1008	0,1563	0,1755	0,1606	0,1277	0,0916	0,0607	0,0378
6	0,0005	0,0120	0,0504	0,1042	0,1462	0,1606	0,1490	0,1221	0,0911	0,0631
7	0,0001	0,0034	0,0216	0,0595	0,1044	0,1377	0,1490	0,1396	0,1171	0,0901
8	0,0000	0,0009	0,0081	0,0298	0,0653	0,1033	0,1304	0,1396	0,1318	0,1126
9		0,0002	0,0027	0,0132	0,0363	0,0688	0,1014	0,1241	0,1318	0,1251
10		0,0000	0,0008	0,0053	0,0181	0,0413	0,0710	0,0993	0,1186	0,1251
11			0,0002	0,0019	0,0082	0,0213	0,0452	0,0722	0,0970	0,1137
12			0,0001	0,0006	0,0034	0,0126	0,0263	0,0481	0,0728	0,0948
13			0,0000	0,0002	0,0013	0,0052	0,0142	0,0296	0,0504	0,0729
14				0,0001	0,0005	0,0022	0,0071	0,0169	0,0324	0,0521
15				0,0000	0,0002	0,0009	0,0033	0,0090	0,0194	0,0347
16					0,0000	0,0003	0,0014	0,0045	0,0109	0,0217
17						0,0001	0,0006	0,0021	0,0058	0,0128
18						0,0000	0,0002	0,0009	0,0029	0,0071
19							0,0001	0,0004	0,0014	0,0037
20							0,0000	0,0002	0,0006	0,0019
21								0,0001	0,0003	0,0009
22								0,0000	0,0001	0,0004
23									0,0000	0,0002
24										0,0001

Таблиця 2

Значення функції  $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3652	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025

Продовж. табл. 2

	0	1	2	3	4	5	6	7	8	9
3, 2	0,0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3, 3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3, 4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3, 5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3, 6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3, 7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3, 8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3, 9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Таблиця 3

Значення функції Лапласа  $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-z^2/2} dz$

$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$
0,00	0,0000	0,18	0,0714	0,36	0,1406	0,54	0,2054	0,72	0,2642
0,01	0,0040	0,19	0,0753	0,37	0,1443	0,55	0,2088	0,73	0,2673
0,02	0,0080	0,20	0,0793	0,38	0,1480	0,56	0,2123	0,74	0,2703
0,03	0,0120	0,21	0,0832	0,39	0,1517	0,57	0,2157	0,75	0,2734
0,04	0,0160	0,22	0,0871	0,40	0,1554	0,58	0,2190	0,76	0,2764
0,05	0,0199	0,23	0,0910	0,41	0,1591	0,59	0,2224	0,77	0,2794
0,06	0,0239	0,24	0,0948	0,42	0,1628	0,60	0,2257	0,78	0,2823
0,07	0,0279	0,25	0,0987	0,43	0,1664	0,61	0,2291	0,79	0,2852
0,08	0,0319	0,26	0,1026	0,44	0,1700	0,62	0,2324	0,80	0,2881
0,09	0,0359	0,27	0,1064	0,45	0,1736	0,63	0,2357	0,81	0,2910
0,10	0,0398	0,28	0,1103	0,46	0,1772	0,64	0,2389	0,82	0,2939
0,11	0,0438	0,29	0,1141	0,47	0,1808	0,65	0,2422	0,83	0,2967
0,12	0,0478	0,30	0,1179	0,48	0,1844	0,66	0,2454	0,84	0,2995
0,13	0,0517	0,31	0,1217	0,49	0,1879	0,67	0,2486	0,85	0,3023
0,14	0,0557	0,32	0,1255	0,50	0,1915	0,68	0,2517	0,86	0,3051
0,15	0,0596	0,33	0,1293	0,51	0,1950	0,69	0,2549	0,87	0,3078
0,16	0,0636	0,34	0,1331	0,52	0,1985	0,70	0,2580	0,88	0,3106
0,17	0,0675	0,35	0,1368	0,53	0,2019	0,71	0,2611	0,89	0,3133

$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$
0,90	0,3159	1,24	0,3925	1,58	0,4429	1,92	0,4726	2,50	0,4938
0,91	0,3186	1,25	0,3944	1,59	0,4441	1,93	0,4732	2,52	0,4941
0,92	0,3212	1,26	0,3962	1,60	0,4452	1,94	0,4738	2,54	0,4945
0,93	0,3238	1,27	0,3980	1,61	0,4463	1,95	0,4744	2,56	0,4948
0,94	0,3264	1,28	0,3997	1,62	0,4474	1,96	0,4750	2,58	0,4951
0,95	0,3289	1,29	0,4015	1,63	0,4484	1,97	0,4756	2,60	0,4953
0,96	0,3315	1,30	0,4032	1,64	0,4495	1,98	0,4761	2,62	0,4956
0,97	0,3340	1,31	0,4049	1,65	0,4505	1,99	0,4767	2,64	0,4959
0,98	0,3365	1,32	0,4066	1,66	0,4515	2,00	0,4772	2,66	0,4961
0,99	0,3389	1,33	0,4082	1,67	0,4525	2,02	0,4783	2,68	0,4963
1,00	0,3413	1,34	0,4099	1,68	0,4535	2,04	0,4793	2,70	0,4965
1,01	0,3438	1,35	0,4115	1,69	0,4545	2,06	0,4803	2,72	0,4967
1,02	0,3461	1,36	0,4131	1,70	0,4554	2,08	0,4812	2,74	0,4969
1,03	0,3485	1,37	0,4147	1,71	0,4564	2,10	0,4821	2,76	0,4971
1,04	0,3508	1,38	0,4162	1,72	0,4573	2,12	0,4830	2,78	0,4973
1,05	0,3531	1,39	0,4177	1,73	0,4582	2,14	0,4838	2,80	0,4974
1,06	0,3554	1,40	0,4192	1,74	0,4591	2,16	0,4846	2,82	0,4976
1,07	0,3577	1,41	0,4207	1,75	0,4599	2,18	0,4854	2,84	0,4977
1,08	0,3599	1,42	0,4222	1,76	0,4608	2,20	0,4861	2,86	0,4979
1,09	0,3621	1,43	0,4236	1,77	0,4616	2,22	0,4868	2,88	0,4980
1,10	0,3643	1,44	0,4251	1,78	0,4625	2,24	0,4875	2,90	0,4981
1,11	0,3665	1,45	0,4265	1,79	0,4633	2,26	0,4881	2,92	0,4982
1,12	0,3686	1,46	0,4279	1,80	0,4641	2,28	0,4887	2,94	0,4984
1,13	0,3708	1,47	0,4292	1,81	0,4649	2,30	0,4893	2,96	0,4985
1,14	0,3729	1,48	0,4306	1,82	0,4656	2,32	0,4898	2,98	0,4986
1,15	0,3749	1,49	0,4319	1,83	0,4664	2,34	0,4904	3,00	0,49865
1,16	0,3770	1,50	0,4332	1,84	0,4671	2,36	0,4909	3,20	0,49931
1,17	0,3790	1,51	0,4345	1,85	0,4678	2,38	0,4913	3,40	0,49966
1,18	0,3810	1,52	0,4357	1,86	0,4686	2,40	0,4918	3,60	0,499841
1,19	0,3830	1,53	0,4370	1,87	0,4693	2,42	0,4922	3,80	0,499928
1,20	0,3849	1,54	0,4382	1,88	0,4699	2,44	0,4927	4,00	0,499968
1,21	0,3869	1,55	0,4394	1,89	0,4706	2,46	0,4931	4,50	0,499997
1,22	0,3883	1,56	0,4406	1,90	0,4713	2,48	0,4934	5,00	0,499997
1,23	0,3907	1,57	0,4418	1,91	0,4719				

Таблиця 4

Значення  $t_\gamma = t(\gamma, n)$ 

$n$	$\gamma$		
	0,95	0,99	0,999
5	2,78	4,60	8,61
6	2,57	4,03	6,86
7	2,45	3,71	5,96
8	2,37	3,50	5,41
9	2,31	3,36	5,04
10	2,26	3,25	4,78
11	2,23	3,17	4,59
12	2,20	3,11	4,44
13	2,18	3,06	4,32
14	2,16	3,01	4,22
15	2,15	2,98	4,14
16	2,13	2,95	4,07
17	2,12	2,92	4,02
18	2,11	2,90	3,97
19	2,10	2,88	3,92
20	2,093	2,861	3,883
25	2,064	2,797	3,745
30	2,045	2,756	3,659
35	2,032	2,720	3,600
40	2,023	2,708	3,558
45	2,016	2,692	3,527
50	2,009	2,679	3,502
60	2,001	2,662	3,464
70	1,996	2,649	3,439
80	1,991	2,640	3,418
90	1,987	2,633	3,403
100	1,984	2,627	3,392
120	1,980	2,617	3,374
$\infty$	1,960	2,576	3,291

Таблиця 5

Значення  $q_\gamma = q(\gamma, n)$ 

$n$	$\gamma$		
	0,95	0,99	0,999
5	1,37	2,67	5,64
6	1,09	2,01	3,88
7	0,92	1,62	2,98
8	0,80	1,38	2,42
9	0,71	1,20	2,06
10	0,65	1,08	1,80
11	0,59	0,98	1,60
12	0,55	0,90	1,45
13	0,52	0,83	1,33
14	0,48	0,78	1,23
15	0,46	0,73	1,15
16	0,44	0,70	1,07
17	0,42	0,66	1,01
18	0,40	0,63	0,96
19	0,39	0,60	0,92
20	0,37	0,58	0,88
25	0,32	0,49	0,73
30	0,28	0,43	0,63
35	0,26	0,38	0,56
40	0,24	0,35	0,50
45	0,22	0,32	0,46
50	0,21	0,30	0,43
60	0,188	0,269	0,38
70	0,174	0,245	0,34
80	0,161	0,226	0,31
90	0,151	0,211	0,29
100	0,143	0,198	0,27
150	0,115	0,160	0,211
200	0,099	0,136	0,185
250	0,089	0,120	0,162

Таблиця 6

Критичні точки розподілу  $\chi^2$ 

Число ступенів вільності $k$	Рівень значущості $\alpha$					
	0, 01	0, 025	0, 05	0, 95	0, 975	0, 99
1	6, 6	5, 0	3, 8	0, 0039	0, 00098	0, 00016
2	9, 2	7, 4	6, 0	0, 103	0, 051	0, 020
3	11, 3	9, 4	7, 8	0, 352	0, 216	0, 115
4	13, 3	11, 1	9, 5	0, 711	0, 484	0, 297
5	15, 1	12, 8	11, 1	1, 15	0, 831	0, 554
6	16, 8	14, 4	12, 6	1, 64	1, 24	0, 872
7	18, 5	16, 0	14, 1	2, 17	1, 69	1, 24
8	20, 1	17, 5	15, 5	2, 73	2, 18	1, 65
9	21, 7	19, 0	16, 9	3, 33	2, 70	2, 09
10	23, 2	20, 5	18, 3	3, 94	3, 25	2, 56
11	24, 7	21, 9	19, 7	4, 57	3, 82	3, 05
12	26, 2	23, 3	21, 0	5, 23	4, 40	3, 57
13	27, 7	24, 7	22, 4	5, 89	5, 01	4, 11
14	29, 1	26, 1	23, 7	6, 57	5, 63	4, 66
15	30, 6	27, 5	25, 0	7, 26	6, 26	5, 23
16	32, 0	28, 8	26, 3	7, 96	6, 91	5, 81
17	33, 4	30, 2	27, 6	8, 67	7, 56	6, 41
18	34, 8	31, 5	28, 9	9, 39	8, 23	7, 01
19	36, 2	32, 9	30, 1	10, 1	8, 91	7, 63
20	37, 6	34, 2	31, 4	10, 9	9, 59	8, 26
21	38, 9	35, 5	32, 7	11, 6	10, 3	8, 90
22	40, 3	36, 8	33, 9	12, 3	11, 0	9, 54
23	41, 6	38, 1	35, 2	13, 1	11, 7	10, 2
24	43, 0	39, 4	36, 4	13, 8	12, 4	10, 9
25	44, 3	40, 6	37, 7	14, 6	13, 1	11, 5
26	45, 6	41, 9	38, 9	15, 4	13, 8	12, 2
27	47, 0	43, 2	40, 1	16, 2	14, 6	12, 9
28	48, 3	44, 5	41, 3	16, 9	15, 3	13, 6
29	49, 6	45, 7	42, 6	17, 7	16, 0	14, 3
30	50, 9	47, 0	43, 8	18, 5	16, 8	15, 0



Таблиця 7

## Критичні точки розподілу Стюдента

Число ступенів вільності $k$	Рівень значущості $\alpha$ (двостороння критична область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,31	12,70	31,82	63,70	318,30	637,00
2	2,92	4,30	6,97	9,92	22,33	31,60
3	2,35	3,18	4,54	5,84	10,22	12,90
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,40
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,96
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,51	3,79
23	1,71	2,07	2,50	2,81	3,49	3,77
24	1,71	2,06	2,49	2,80	3,47	3,74
25	1,71	2,06	2,49	2,79	3,45	3,72
26	1,71	2,06	2,48	2,78	3,44	3,71
27	1,71	2,05	2,47	2,77	3,42	3,69
28	1,70	2,05	2,46	2,76	3,40	3,66
29	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
$\infty$	1,64	1,96	2,33	2,58	3,09	3,29
	0,05	0,025	0,01	0,005	0,001	0,0005
	Рівень значущості $\alpha$ (одностороння критична область)					

Таблиця 8

**Критичні точки розподілу  $F$  Фішера-Снедекора** $(k_1$  — число ступенів вільності більшої дисперсії, $k_2$  — число ступенів вільності меншої дисперсії)

Рівень значущості $\alpha = 0,01$												
$k_2$	$k_1$											
	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6082	6106
2	98,49	99,01	99,17	99,25	99,30	99,33	99,34	99,36	99,38	99,40	99,41	99,42
3	34,12	30,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	9,86	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,45

Рівень значущості $\alpha = 0,05$												
$k_2$	$k_1$											
	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38

## СПИСОК ЛІТЕРАТУРИ

1. Гмурман В.Е. Теория вероятностей и математическая статистика. Изд. 5-е, перераб. и доп. – М.: Высшая школа, – 1977. – 479 с.
2. Булдык Г.М. Теория вероятностей и математическая статистика. – Минск: Высшая школа, 1989. – 285 с.
3. Дрогомирецька Х.Т. Теория ймовірностей та математична статистика: навч. посібник / Х.Т.Дрогомирецька, О.М.Рибицька, О.З.Слюсарчук, Н.В.Пабірівська, Л.В.Гошко, О.В.Веселовська, Д.В.Білонога.– Львів: Видавництво Львівської політехніки, 2012.– 396 с.
4. Карасев А.И. Теория вероятностей и математическая статистика. – М.: Статистика, 1979. – 279 с.
5. Сборник задач по математике для втузов. Ч.3. Теория вероятностей и математическая статистика. / Под ред. А.В.Ефимова. – М.: Наука, 1990. – 428 с.