

Частина II

МАТЕМАТИЧНА СТАТИСТИКА

Розділ 7

СТАТИСТИЧНІ РОЗПОДІЛИ ВИБІРКИ

Статистика (від італійського слова *stato* – держава) вивчає кількісний бік суспільних явищ і процесів у нерозривному зв'язку з їх якісним змістом.

Статистику поділяють на описову та пояснювальну. Прикладом описової статистики є книга рекордів Гіннеса. Пояснювальна статистика формулює певні висновки про ті чи інші явища, робить прогноз.

Основним змістом математичної статистики є систематизація, обробка і використання статистичної інформації для виявлення статистичних закономірностей ознаки або ознак певної сукупності елементів.

Математична статистика – розділ математики, в якому на основі дослідних даних вивчаються імовірнісні закономірності масових явищ.

Основними завданнями математичної статистики є статистична перевірка гіпотез, оцінка розподілу статистичних ймовірностей та його параметрів, вивчення статистичної залежності, визначення основних числових характеристик випадкових вибірок, якими є: вибіркове середнє, вибіркові дисперсії, стандартне відхилення.

Математична статистика виникла (XVII ст.) та почала розвиватись одночасно з теорією ймовірностей. Подальшим розвитком (кінець XIX – початок XX ст.) математичної статистики займалися П. Чебишов, А. Марков, О. Ляпунов, а також К. Гаусс, Ф. Гальтон, К. Пірсон та інші. У XX ст. найбільший внесок у математичну статистику зробили В. Романовський, Е. Слуцький, А. Колмогоров, Стюдент (псевдонім У. Госсета), Е. Пірсон, Ю. Нейман, А. Вальд, А. Скороход, В. Королюк та інші вчені.

Прикладом перевірки статистичних гіпотез є з'ясування питання про те, змінюється чи не змінюється виробничий процес з часом. Прикладом оцінки параметрів є оцінка середнього значення статистичної змінної за дослідними даними. Для вивчення статистичної залежності використовують методи теорії кореляції. Загальні методи математичної статистики є основою теорії похибок.

Предмет математичної статистики полягає в розробці методів збору та обробки статистичних даних для одержання наукових та практичних висновків.

7.1. Генеральна та вибіркова сукупність

Основні поняття математичної статистики:

- **Генеральна сукупність** – це вся сукупність об'єктів, які досліджуються.
- **Вибірка або вибіркова сукупність** – це об'єкти, довільно або випадково відібрані з генеральної сукупності для дослідження.
- **Обсяг (об'єм) сукупності** – це кількість об'єктів цієї сукупності.

Приклад 7.1. Плоди дерева (200 шт.) обстежують на наявність специфічного смаку цього сорту. Для цього відбирають 10 штук. Тут 200 – обсяг генеральної сукупності, 10 – обсяг вибірки.

Вибірка повинна бути репрезентативною (представницькою), тобто правильно відображати ті властивості генеральної сукупності, які вивчаються. Досліджувані явища мають бути масовими. Лише тоді статистичні дані будуть достовірними.

Вибірки бувають повторні та неповторні. *Повторною* називають вибірку, за якої відібраний об'єкт повертається до генеральної сукупності перед відбором іншого об'єкта. Вибірку називають *безповторною*, якщо взятий об'єкт до генеральної сукупності не повертається. Найчастіше використовують неповторні вибірки.

План проведення статистичних досліджень:

- 1) формулюють завдання дослідження та визначають обсяг вибірки (мета, об'єкти вивчення, їх кількість, певні ознаки, характеристики);
- 2) збирають потрібні дані, вибирають доцільну форму їх подання для подальшого дослідження (водночас застосовують такі методи, як спостереження, порівняння, усне та письмове анкетування);
- 3) проводять остаточну обробку статистичного матеріалу та його вивчення;
- 4) за результатами формулюють певні висновки.

Компактною та наочною формою подання даних, отриманих у результаті статистичного дослідження, є: таблиці, графіки, стовпчасті та кругові діаграми.

7.2. Способи відбору статистичного матеріалу

У практичній діяльності використовують різноманітні способи відбору об'єктів із генеральної сукупності.

1. Простий випадковий відбір.

Навмання вибирається певна кількість об'єктів з усієї генеральної сукупності. Наприклад, вибираємо m об'єктів із n об'єктів генеральної сукупності. Для цього нумеруємо картки від 1 до n на об'єктах і вибираємо по одній картці і т. д. Простий відбір може бути повторним або безповторним. Для великого обсягу вибірки використовують готові таблиці «випадкових чисел».

2. Типовий відбір.

Об'єкти вибирають не з усієї сукупності, а з певної її типової частини. Наприклад, соціологи вивчають різні групи людей. Типовими ознаками можуть бути: місце проживання, вік, професія...

3. Механічний відбір.

Генеральну сукупність поділяють на стільки груп, частин, скільки планується об'єктів у вибірці. З кожної частини вибирають один об'єкт. Наприклад, якщо обстежується 20% телевізорів з партії, то беруть кожен п'ятий, якщо 10% – кожен десятий. Щоб механічний відбір був репрезентативним, треба враховувати специфіку технологічного процесу.

4. Серійний відбір.

Об'єкти з генеральної сукупності вибирають не по одному, а серіями, партіями. Серійний відбір використовують тоді, коли ознака, яку досліджують, мало змінюється в різних серіях.

На практиці розглянуті способи відбору поєднують і комбінують довільним способом.

7.3. Статистичний розподіл вибірки

Об'єкти обстежують за їх певними характеристиками або ознаками (вік, професія, ...), які називають **варіантами**.

Сукупність значень ознаки (статистичної змінної), записаних у порядку їх зростання, називається **варіаційним рядом**.

Нехай X – статистична змінна, x_i ($i = 1, 2, \dots, n$) – її значення, то тоді x'_1, x'_2, \dots, x'_n – варіаційний ряд, якщо $x'_{i+1} \geq x'_i$.

Якщо кількість варіант велика, то сукупність їх значень для зручності розбивають на інтервали. Кожен інтервал характеризується одним числом – його серединою. Величину інтервалу ще називають інтервальною різницею. Частинні інтервали варіант, які розміщені у зростаючій послідовності, утворюють **інтервальний варіаційний ряд**.

Важлива характеристика – частота появи цієї варіанти.

Додатне число n_i , яке показує, скільки разів варіанта x_i ($i = 1, 2, \dots, m$) трапляється в таблиці даних, називається **частотою** варіанти x_i .

Ряд n_1, n_2, \dots, n_m називається **рядом частот**. Сума усіх частот повинна дорівнювати об'єму вибірки:

$$\sum_{i=1}^m n_i = n. \quad (7.1)$$

Статистичний розподіл вибірки встановлює зв'язок між рядом варіант, що зростає або спадає, і відповідними частотами. Він може бути зображений у вигляді таблиці

x_i	x_1	x_2	...	x_m
n_i	n_1	n_2	...	n_m

Відношення частоти n_i варіанти x_i до об'єму вибірки n називають **відносною частотою** і позначають

$$W_i = \frac{n_i}{n}. \quad (7.2)$$

Сума усіх відносних частот

$$\sum_{i=1}^m W_i = 1. \quad (7.3)$$

Залежність між упорядкованим рядом варіант і відповідними їм відносними частотами називається **статистичним розподілом відносних частот вибірки**:

x_i	x_1	x_2	...	x_m
W_i	$\frac{n_1}{n}$	$\frac{n_2}{n}$...	$\frac{n_m}{n}$

Приклад 7.2. Під час дослідження кількісної ознаки X із генеральної сукупності було отримано вибірку

4; 3; 6; 4; 7; 2; 5; 1; 2; 5; 4; 4; 3; 5; 6; 3; 4; 1; 3; 4.

Знайти обсяг вибірки, побудувати її варіаційний ряд, статистичні розподіли частот та відносних частот.

Розв'язання. Оскільки вибірка містить 20 значень, то об'єм вибірки $n = 20$.

Побудуємо варіаційний ряд вибірки, тобто запишемо всі її значення у порядку зростання:

1; 1; 2; 2; 3; 3; 3; 3; 4; 4; 4; 4; 4; 4; 5; 5; 5; 6; 6; 7.

У цій вибірці лише сім різних значень, тобто варіант:

1; 2; 3; 4; 5; 6; 7.

Знайдемо їх частоти:

$n_1 = 2$; $n_2 = 2$; $n_3 = 4$; $n_4 = 6$; $n_5 = 3$; $n_6 = 2$; $n_7 = 1$.

Запишемо шуканий статистичний розподіл частот вибірки:

x_i	1	2	3	4	5	6	7
n_i	2	2	4	6	3	2	1

Знайдемо відносні частоти варіант вибірки:

$W_1 = \frac{2}{20} = 0,1$; $W_2 = \frac{2}{20} = 0,1$; $W_3 = \frac{4}{20} = 0,2$; $W_4 = \frac{6}{20} = 0,3$;

$W_5 = \frac{3}{20} = 0,15$; $W_6 = \frac{2}{20} = 0,1$; $W_7 = \frac{1}{20} = 0,05$.

Отже, шуканий розподіл відносних частот має такий вигляд:

x_i	1	2	3	4	5	6	7
W_i	0,1	0,1	0,2	0,3	0,15	0,1	0,05

7.4. Емпірична функція розподілу та її властивості

За даними статистичного розподілу вибірки будується емпірична функція розподілу.

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

Математично це означення записується:

$$F^*(x) = \frac{n_x}{n}, \quad (7.4)$$

де n_x – кількість варіант, які менші від x , n – об'єм вибірки.

Емпірична функція розподілу $F^*(x)$ має такі властивості:

1. $0 \leq F^*(x) \leq 1$;
2. $F^*(x)$ – зростаюча функція;
3. $F^*(x) = \begin{cases} 0, & x \leq x_1; \\ 1, & x > x_m, \end{cases}$

де x_1 – найменше значення варіанти, x_m – найбільше значення варіанти.

Приклад 7.3. Нехай маємо таку частотну таблицю (розподіл частот):

x_i	3	5	7	10	15
n_i	2	4	7	4	3

Записати емпіричну функцію розподілу та побудувати її графік.

Розв'язання. Об'єм вибірки:

$$n = \sum_{i=1}^m n_i = 2 + 4 + 7 + 4 + 3 = 20.$$

Оскільки найменше значення варіанти $x_1 = 3$, то

$$F^*(x) = 0$$

для всіх $x \leq 3$.

Значення $x < 5$, а саме: $x_1 = 3$, спостерігається двічі, тому

$$F^*(x) = \frac{2}{20} = 0,1,$$

при $3 < x \leq 5$.

Значення $x < 7$, а саме: $x_1 = 3$, $x_2 = 5$, спостерігалися $2 + 4 = 6$ разів, тому

$$F^*(x) = \frac{6}{20} = 0,3,$$

при $5 < x \leq 7$.

Значення $x < 10$, а саме, $x_1 = 3$, $x_2 = 5$, $x_3 = 7$, спостерігалися $2 + 4 + 7 = 13$ разів, тому

$$F^*(x) = \frac{13}{20} = 0,65,$$

при $7 < x \leq 10$.

Значення $x < 15$, а саме, $x_1 = 3$, $x_2 = 5$, $x_3 = 7$, $x_4 = 10$, спостерігалися $2 + 4 + 7 + 4 = 17$ разів, тому

$$F^*(x) = \frac{17}{20} = 0,85,$$

при $10 < x \leq 15$.

Оскільки $x_5 = 15$ – найбільше значення варіанти, то

$$F^*(x) = 1$$

при $x > 15$.

Отже, запишемо шукану емпіричну функцію розподілу:

$$F^*(x) = \begin{cases} 0, & \text{при } x \leq 3; \\ 0,1, & \text{при } 3 < x \leq 5; \\ 0,3, & \text{при } 5 < x \leq 7; \\ 0,65, & \text{при } 7 < x \leq 10; \\ 0,85, & \text{при } 10 < x \leq 15; \\ 1, & \text{при } x > 15. \end{cases}$$

Графік цієї функції зображено на рисунку 7.1.

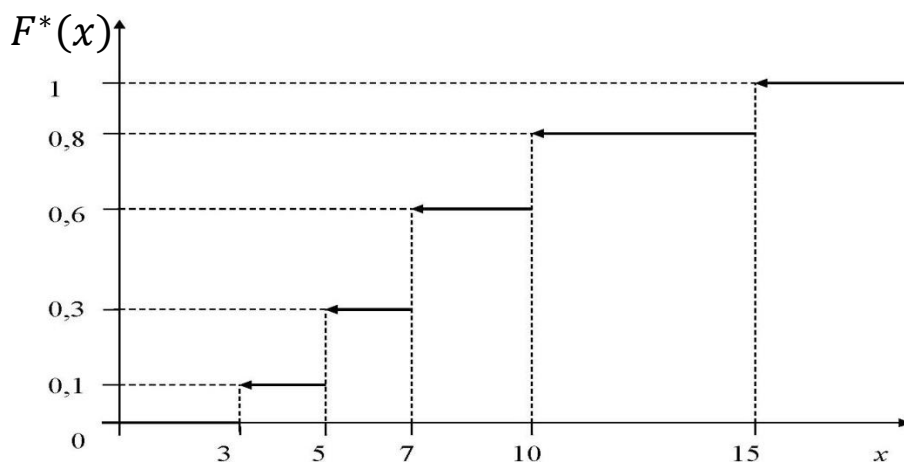


Рис. 7.1. Графік емпіричної функції розподілу

7.5. Згруповані розподіли вибірки

Часто для спрощення статистичного дослідження використовують згрупований розподіл вибірки.

Загальна схема побудови згрупованого розподілу частот

1. Визначити найбільше x_{max} та найменше x_{min} значення варіанти x_i і обчислити розмах варіант $R = x_{max} - x_{min}$.
2. Задати певне непарне число класів k . При загальному числі замірів $n \geq 100$ доцільно брати $9 \leq k \leq 15$, а при $n \leq 100$ можна вибрати $k = 7$.
3. Визначити ширину класу $h = \frac{R}{k}$.
4. Встановити границі класів і підрахувати кількість варіант у кожному класі. При підрахунку числа варіант значення x_i , що знаходиться на границі класів, слід відносити завжди до одного й того ж класу, а саме там, де це число трапилося вперше. Відтак воно стає нижньою границею класу.
5. Визначити частоту для кожного класу і записати ряд розподілу.

Згрупований розподіл накопиченої частоти та відносної частоти

Часто поряд з розподілом частот варіанти необхідно мати розподіл накопичених частот.

Розподіл накопиченої частоти одержують послідовним додаванням частот чергового інтервалу, починаючи з першого і закінчуючи останнім.

Розподіл накопиченої частоти (позначається F_i) дозволяє відповісти на питання: «Скільки існує варіант, які менші, ніж варіанта x_i ?».

Розподіл накопиченої відносної частоти одержують послідовним додаванням відносних частот чергового інтервалу, починаючи з першого і закінчуючи останнім.

Розподіл накопиченої відносної частоти (позначається $\frac{F_i}{n}$) дозволяє відповісти на питання: «Яка пропорція варіант, які менші, ніж x_i ?»

Згрупований розподіл щільності частот і щільності відносних частот

Якщо поділити всі частоти на ширину інтервалу h , то отримаємо **розподіл щільності частот вибірки** $\frac{n_i}{h}$.

Якщо поділити всі відносні частоти на ширину інтервалу h , то отримаємо **розподіл щільності відносних частот вибірки** $\frac{W_i}{h}$.

Приклад 7.4. У таблиці 1 наведена вибірка середньомісячної платні 100 співробітників фірми. Впорядкувати вибірку. Записати розподіл частот та відносних частот середньомісячної платні співробітників фірми; згрупований розподіл частот та відносних частот середньомісячної платні співробітників фірми; згрупований розподіл накопичених частот та накопичених відносних частот середньомісячної платні співробітників фірми; згрупований розподіл щільності частот та щільності відносних частот.

Таблиця 7.1

Вибірка середньомісячної платні 100 співробітників фірми									
338	348	304	314	326	314	324	304	342	308
336	304	302	338	314	304	320	321	322	321
312	323	336	324	312	312	364	356	362	302
322	310	334	292	362	381	304	366	298	304
381	368	304	298	368	290	340	328	316	322
302	314	292	342	321	322	290	332	298	296
296	298	324	338	352	326	318	304	332	322
360	312	331	331	304	316	332	282	342	338
342	322	324	325	302	328	354	330	316	324
334	350	334	324	332	340	324	314	326	323

Розв'язання. Запишемо розподіл частот та відносних частот цієї вибірки:

Таблиця 7.2

Розподіл частот та відносних частот середньомісячної платні співробітників фірми											
x_i	n_i	W_i	x_i	n_i	W_i	x_i	n_i	W_i	x_i	n_i	W_i
282	1	0,01	314	5	0,05	328	2	0,02	350	1	0,01
290	2	0,02	316	3	0,03	330	1	0,01	352	1	0,01
292	2	0,02	318	1	0,01	331	2	0,02	354	1	0,01
296	2	0,02	320	1	0,01	332	4	0,04	356	1	0,01
298	4	0,04	321	3	0,03	334	3	0,03	360	1	0,01
302	4	0,04	322	6	0,06	336	2	0,02	362	2	0,02
304	9	0,09	323	2	0,02	338	4	0,04	364	1	0,01
308	1	0,01	324	7	0,07	340	2	0,02	366	1	0,01
310	1	0,01	325	1	0,01	342	4	0,04	368	2	0,02
312	4	0,04	326	3	0,03	348	1	0,01	381	2	0,02

Наступним кроком в обробці даних, що веде до суттєвого спрощення досліджень, є групування.

Знаходимо максимальне та мінімальне значення варіанти вибірки:

$$x_{max} = 381, x_{min} = 282.$$

Розмах

$$R = x_{max} - x_{min} = 381 - 282 = 99.$$

Задамо число класів $k = 11$.

$$\text{Визначимо ширину класу } h = \frac{R}{k} = \frac{99}{11} = 9.$$

Просумуємо частоти для кожного класу інтервалів, значення x_i , що знаходиться на границі класів, заносимо до того класу, де це число трапилось вперше. Результат запишемо у вигляді таблиці.

Таблиця 7.3

Згрупований розподіл частот середньомісячної платні співробітників фірми		
Інтервали платні	Частоти n_i	
282 – 291	3	=(1+2)
291 – 300	8	=(2+2+4)
300 – 309	14	=(4+9+1)
309 – 318	14	=(1+4+5+3+1)
318 – 327	23	=(1+3+6+2+7+1+3)
327 – 336	14	=(2+1+2+4+3+2)
336 – 345	10	=(4+2+4)
345 – 354	4	=(1+1+1+1)
354 – 363	4	=(1+1+2)
363 – 372	4	=(1+1+2)
372 – 381	2	=(2)
Разом:	100	

Запишемо розподіл накопичених частот, який одержимо послідовним додаванням частот чергового інтервалу.

Таблиця 7.4

Згрупований розподіл накопичених частот середньомісячної платні співробітників фірми		
Платня	Накопичені частоти F_i	
менше ніж 291	3	(=3)
менше ніж 300	11	(=3+8)
менше ніж 309	25	(=11+14)
менше ніж 318	39	(=25+14)
менше ніж 327	62	(=39+23)
менше ніж 336	76	(=62+14)
менше ніж 345	86	(=76+10)
менше ніж 354	90	(=86+4)
менше ніж 363	94	(=90+4)
менше ніж 372	98	(=94+4)
менше ніж 381	100	(=98+2)

Якщо поділити частоти та накопичені частоти на об'єм вибірки $n = 100$, то отримаємо, відповідно, розподіл відносних частот W_i та накопичених відносних частот $\frac{F_i}{n}$.

Запишемо розподіл відносних частот W_i та накопичених відносних частот $\frac{F_i}{n}$ (таблиця 7.5).

Таблиця 7.5

Згрупований розподіл відносних та накопичених відносних частот середньомісячної платні співробітників фірми			
Інтервали платні	Відносні частоти W_i	Платня	Накопичені відносні частоти $\frac{F_i}{n}$
282 – 291	0,03	менше ніж 291	0,03
291 – 300	0,08	менше ніж 300	0,11
300 – 309	0,14	менше ніж 309	0,25
309 – 318	0,14	менше ніж 318	0,39
318 – 327	0,23	менше ніж 327	0,62
327 – 336	0,14	менше ніж 336	0,76
336 – 345	0,1	менше ніж 345	0,86
345 – 354	0,04	менше ніж 354	0,9
354 – 363	0,04	менше ніж 363	0,94
363 – 272	0,04	менше ніж 372	0,98
372 – 381	0,02	менше ніж 381	1

Запишемо згрупований розподіл щільності частот та щільності відносних частот.

Для цього поділимо частоти та відносні частоти на ширину інтервалу $h = 10$. Щоб підсумувати результати, які одержали в Прикладі 7.4, зведемо разом в одну таблицю розглянуті розподіли *Таблиця 7.6*.

Ця таблиця містить усі важливі статистичні розподіли вибірки.

Таблиця 7.6

Статистичні розподіли вибірки							
x_i	n_i	W_i	$\frac{n_i}{h}$	$\frac{W_i}{h}$	x_i	F_i	$\frac{F_i}{n}$
282 – 291	3	0,03	$\frac{1}{3}$	$\frac{1}{300}$	менше, ніж 291	3	0,03
291 – 300	8	0,08	$\frac{8}{9}$	$\frac{2}{225}$	менше, ніж 300	11	0,11
300 – 309	14	0,14	$\frac{14}{9}$	$\frac{7}{450}$	менше, ніж 309	25	0,25
309 – 318	14	0,14	$\frac{14}{9}$	$\frac{7}{450}$	менше, ніж 318	39	0,39
318 – 327	23	0,23	$\frac{23}{9}$	$\frac{23}{900}$	менше, ніж 327	62	0,62
327 – 336	14	0,14	$\frac{14}{9}$	$\frac{7}{450}$	менше, ніж 336	76	0,76
336 – 345	10	0,1	$\frac{10}{9}$	$\frac{1}{90}$	менше, ніж 345	86	0,86
345 – 354	4	0,04	$\frac{4}{9}$	$\frac{1}{225}$	менше, ніж 354	90	0,9
354 – 363	4	0,04	$\frac{4}{9}$	$\frac{1}{225}$	менше, ніж 363	94	0,94
363 – 372	4	0,04	$\frac{4}{9}$	$\frac{1}{225}$	менше, ніж 372	98	0,98
372 – 381	2	0,02	$\frac{2}{9}$	$\frac{1}{450}$	менше, ніж 381	100	1

7.6. Полігон частот та відносних частот

Статистичні розподіли вибірки можуть бути представлені також графічно. Завдяки цьому ми можемо побачити характерні зміни ряду розподілу, не користуючись аналізом цифрових даних.

Якщо в результаті вибірки ми одержали статистичний розподіл ознаки X , яку треба дослідити, то будемо мати перелік варіант ознаки

$$x_1, x_2, x_3, \dots, x_m$$

та відповідних їм частот

$$n_1, n_2, n_3, \dots, n_m$$

або відносних частот

$$W_1, W_2, W_3, \dots, W_m.$$

Значення варіант та частот або відносних частот можна розглядати як координати точок

$$K_1(x_1; n_1), K_2(x_2; n_2), K_3(x_3; n_3), \dots, K_m(x_m; n_m)$$

або

$$L_1(x_1; W_1), L(x_2; W_2), L_3(x_3; W_3), \dots, L_m(x_m; W_m).$$

Полігоном частот називають ламану, відрізки якої з'єднують точки з координатами

$$K_1(x_1; n_1), K_2(x_2; n_2), K_3(x_3; n_3), \dots, K_m(x_m; n_m).$$

Полігоном відносних частот називають ламану, відрізки якої з'єднують точки з координатами

$$L_1(x_1; W_1), L(x_2; W_2), L_3(x_3; W_3), \dots, L_m(x_m; W_m).$$

Для побудови полігону частот на осі абсцис відкладають варіанти x_i ознаки X , а на осі ординат – відповідні їм частоти n_i . Точки $K_i(x_i; n_i)$ ($i = 1, 2, 3, \dots, m$) з'єднують відрізками прямих і одержують полігон частот.

Для побудови полігону відносних частот на осі абсцис відкладають варіанти x_i ознаки X , а на осі ординат – відповідні їм відносні частоти W_i ($i = 1, 2, 3, \dots, m$), а потім точки $L_i(x_i; W_i)$ з'єднують відрізками прямих.

При неперервному розподілі ознаки X у разі великої кількості спостережень увесь інтервал, в якому розміщені спостережні значення ознаки, як правило, розбивають на кілька частинних інтервалів однакової довжини h і знаходять

n_i – суму частот варіант, що потрапили в i -й інтервал. Для вибору оптимальної величини інтервалу рекомендовано використовувати формулу:

$$h \approx \frac{x_{max} - x_{min}}{1 + 3,2 \cdot \lg n},$$

де x_{max} , x_{min} – відповідно найбільше і найменше значення вибірки, n – об'єм вибірки.

Якщо задано інтервальний статистичний розподіл вибірки, то для побудови полігону частот за даними вибірки з'єднують точки, абсциси яких є значення середин частинних інтервалів, а ординатами – відповідні їм значення частот.

Приклад 7.5. Вибірку задано інтервальним розподілом частот:

$(x_i ; x_{i+1})$	(1; 3)	(3; 5)	(5; 7)	(7; 9)	(9; 11)
n_i	13	9	5	16	7

Побудувати полігон відносних частот.

Розв'язання. Знайдемо об'єм вибірки:

$$n = 13 + 9 + 5 + 16 + 7 + 7 = 50.$$

Для побудови інтервального статистичного розподілу відносних частот визначимо відносні частоти:

$$W_1 = \frac{n_1}{n} = \frac{13}{50} = 0,26; \quad W_2 = \frac{n_2}{n} = \frac{9}{50} = 0,18;$$

$$W_3 = \frac{n_3}{n} = \frac{5}{50} = 0,1; \quad W_4 = \frac{n_4}{n} = \frac{16}{50} = 0,32;$$

$$W_5 = \frac{n_5}{n} = \frac{7}{50} = 0,14.$$

Побудуємо інтервальний статистичний розподіл відносних частот:

$(x_i ; x_{i+1})$	(1; 3)	(3; 5)	(5; 7)	(7; 9)	(9; 11)
W_i	0,26	0,18	0,1	0,32	0,14

Знайдемо середини частинних інтервалів:

$$x_1^{\text{сер}} = \frac{x_1+x_2}{2} = \frac{1+3}{2} = 2;$$

$$x_2^{\text{сер}} = \frac{x_2+x_3}{2} = \frac{3+5}{2} = 4;$$

$$x_3^{\text{сер}} = \frac{x_3+x_4}{2} = \frac{5+7}{2} = 6;$$

$$x_4^{\text{сер}} = \frac{x_4+x_5}{2} = \frac{7+9}{2} = 8;$$

$$x_5^{\text{сер}} = \frac{x_5+x_6}{2} = \frac{9+11}{2} = 10.$$

Відкладемо на осі абсцис значення середин частинних інтервалів $x_i^{\text{сер}}$, а на осі ординат – значення відповідних їм відносних частот W_i . Послідовно з'єднуючи між собою точки $(x_i^{\text{сер}}; W_i)$ відрізками, отримаємо полігон відносних частот (рис. 7.2).

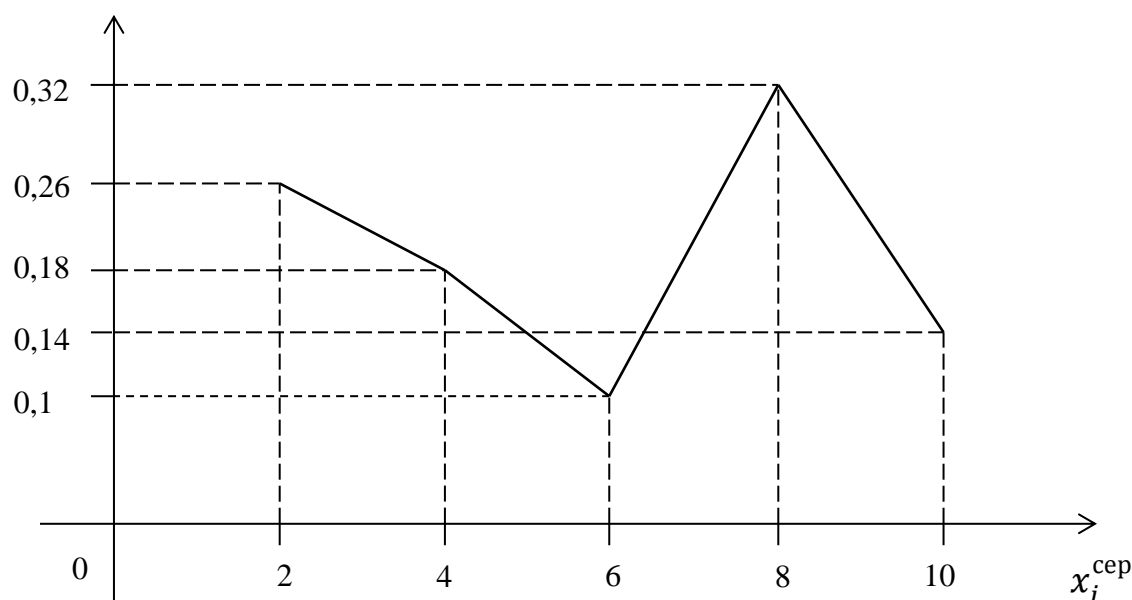


Рис. 7.2. Графік побудови полігону відносних частот

7.7. Гістограма частот та відносних частот

Гістограмою частот називають ступінчасту фігуру, яка складається з прямокутників, основами яких є частинні інтервали варіант довжиною $h = x_i - x_{i-1}$, а висоти дорівнюють $\frac{n_i}{h}$ (щільність частоти).

Гістограмою відносних частот називають ступінчасту фігуру, яка складається з прямокутників, основами яких є частинні інтервали варіант довжиною $h = x_i - x_{i-1}$, а висоти дорівнюють $\frac{W_i}{h}$ (щільність відносної частоти).

Площа гістограми частот дорівнює об'єму вибірки, а площа гістограми відносних частот – одиниці.

Для побудови гістограми частот (відносних частот) проміжок $[x_{min}; x_{max}]$, тобто від найменшого значення x_{min} , що спостерігали, до найбільшого значення x_{max} , розбивають на декілька відрізків рівної довжини h . Потім підраховують суму частот (відносних частот) тих значень варіант, які належать кожному із одержаних відрізків.

Якщо в i -му відрізку ($i = 1, 2, 3, \dots, m$) кількість варіант, що спостерігали, з врахуванням їх частот дорівнює n_i , то будують прямокутник Π_i , основою якого буде i -ий відрізок довжиною h , а висотою буде відрізок довжиною $\frac{n_i}{h}$ (для відносних частот висота – $\frac{W_i}{h}$). Площа такого прямокутника дорівнює $h \cdot \frac{n_i}{h} = n_i$ (у випадку відносних частот площа дорівнює $h \cdot \frac{W_i}{h} = W_i$). Тому площа усіх прямокутників буде дорівнювати сумі n_i (W_i), тобто $S = \sum_{i=1}^m n_i = n$ об'єму вибірки.

У випадку гістограми відносних частот площа прямокутників буде дорівнювати сумі усіх відносних частот

$$S = \sum_{i=1}^m W_i = 1.$$

Приклад 7.6. Вибірку задано розподілом частот:

x_i	2	4	6	8	10	12	14
n_i	2	3	5	1	4	2	3

Побудувати гістограму відносних частот.

Розв'язання. Щоб побудувати гістограму відносних частот на основі цього статистичного розподілу частот, складемо інтервальний статистичний розподіл відносних частот і знайдемо їх щільність.

Обсяг вибірки:

$$n = 2 + 3 + 5 + 1 + 4 + 2 + 3 = 20.$$

Відносні частоти:

$$W_1 = \frac{n_1}{n} = \frac{2}{20} = 0,1;$$

$$W_2 = \frac{n_2}{n} = \frac{3}{20} = 0,15;$$

$$W_3 = \frac{n_3}{n} = \frac{5}{20} = 0,25;$$

$$W_4 = \frac{n_4}{n} = \frac{1}{20} = 0,05;$$

$$W_5 = \frac{n_5}{n} = \frac{34}{20} = 0,2;$$

$$W_6 = \frac{n_6}{n} = \frac{2}{20} = 0,1;$$

$$W_7 = \frac{n_7}{n} = \frac{1}{20} = 0,05.$$

Знайдемо частинні інтервали, їх довжину і щільність відносних частот. Частинні інтервали визначимо з умови: задані в дискретному статистичному розподілі варіанти мають бути серединами частинних інтервалів.

Отже, довжина частинних інтервалів

$$h = x_i - x_{i-1} = 2.$$

Шуканий інтервальний статистичний розподіл відносних частот має вигляд:

$(x_i; x_{i+1})$	(1; 3)	(3; 5)	(5; 7)	(7; 9)	(9; 11)	(11; 13)	(13; 15)
W_i	0,1	0,15	0,25	0,05	0,2	0,1	0,15

Щільності відносних частот:

$$\frac{W_1}{h} = \frac{0,1}{2} = 0,05;$$

$$\frac{W_2}{h} = \frac{0,15}{2} = 0,075;$$

$$\frac{W_3}{h} = \frac{0,25}{2} = 0,125;$$

$$\frac{W_4}{h} = \frac{0,05}{2} = 0,025;$$

$$\frac{W_5}{h} = \frac{0,2}{2} = 0,1;$$

$$\frac{W_6}{h} = \frac{0,1}{2} = 0,05;$$

$$\frac{W_7}{h} = \frac{0,15}{2} = 0,075.$$

Побудуємо гістограму відносних частот (рис. 7.3).

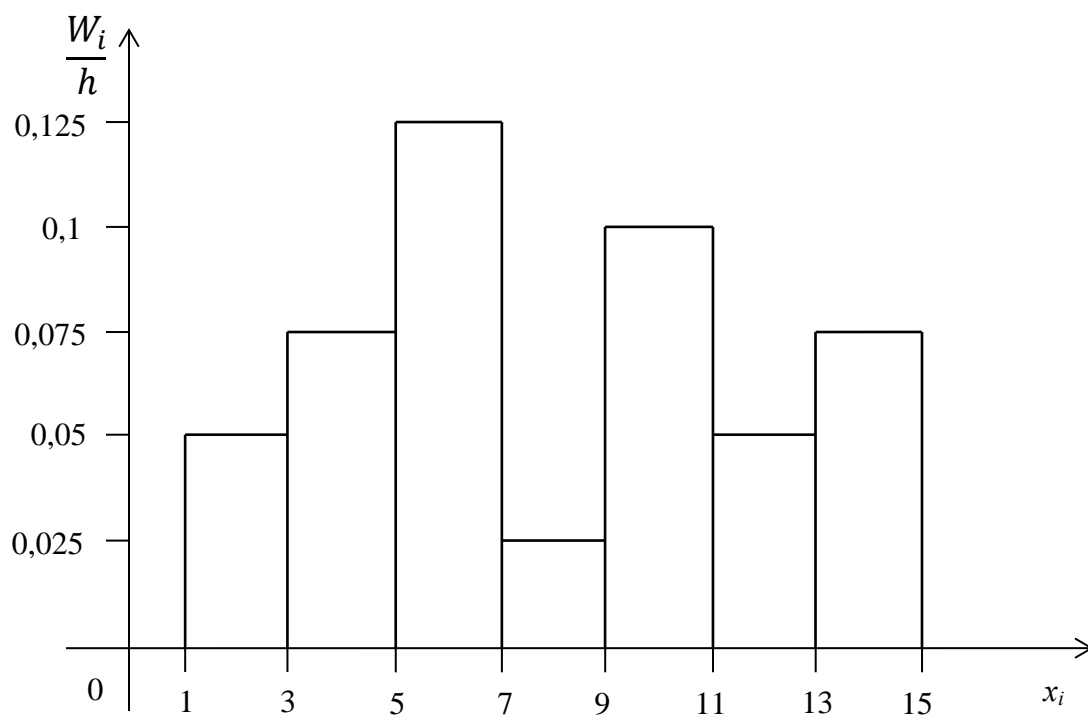


Рис. 7.3. Гістограма відносних частот

Запитання для самоконтролю

1. Що є предметом математичної статистики?
2. Вказати основні завдання математичної статистики.
3. Що називають генеральною та вибірковою сукупністю, об'ємом цих сукупностей?
4. Навести приклади генеральної сукупності та вибірки.
5. Способи відбору статистичного матеріалу (проілюструвати на прикладах).
6. Що називається варіантою, варіаційним рядом?
7. Що таке частота, відносна частота варіант?
8. Що називається емпіричною функцією розподілу?
9. Властивості емпіричної функція розподілу та її графік.
10. Статистичний розподіл частот вибірки.
11. Статистичний розподіл відносних частот вибірки.
12. Згрупований розподіл вибірки, загальна схема його побудови.
13. Що називають згрупованим розподілом накопичених частот вибірки; накопичених відносних частот вибірки?
14. Що називають згрупованим розподілом щільності частот; щільності відносних частот?

15. Що називається інтервальним статистичним розподілом вибірки? Які рекомендації до вибору числа класів інтервалів?
16. Що називається полігоном частот і відносних частот?
17. Що називається гістограмою частот і відносних частот?

Задачі для самостійного розв'язування

1. Для заданої вибірки із генеральної сукупності побудувати варіаційний ряд, скласти розподіли частот та відносних частот.

а) 7, 4, 4, 8, 12, 12, 12, 7, 8, 12, 8, 12, 4, 12, 12, 4, 12, 4, 12, 12, 6, 9, 5, 3, 6, 6, 9, 3, 5, 6, 9, 5, 6, 6, 9, 6, 9, 6, 6, 6.

б) 0,45; 0,21; 0,14; 0,15; 1,52; 0,1; 0,52; 1,59; 3,38; 2,25; 0,8; 1,26; 2,31; 0,84; 3,72; 2,11; 1,02; 4,2; 2,53; 0,78; 2,92; 0,71; 4,7; 3,02; 1,58; 4,12; 2,59; 0,88; 0,96; 1,76; 1,93; 4,9; 2,82; 1,14; 5,7; 1,21; 1,47; 3,52; 0,36; 0,64.

в) 4; 4,3; 5,68; 6,2; 5,64; 5,8; 4,25; 5,4; 5,3; 5,2; 4,55; 5,32; 6; 6,15; 4,56; 6,64; 6,5; 4,7; 6,8; 6,15; 5,6; 5,1; 4,2; 4,8; 6,9; 7; 4,9; 5; 5,25; 6,2.

г) 40,25; 40,29; 40,46; 40,33; 40,37; 40,27; 40,39; 40,34; 40,33; 40,35; 40,38; 40,32; 40,28; 40,41; 40,45; 40,39; 40,29; 40,3; 40,44; 40,37; 40,41; 40,33; 40,35; 40,35; 40,35; 40, 40; 40, 40; 40,3; 40,28; 40,34; 40,45; 40,44.

2. Знайти емпіричну функцію заданого розподілу вибірки та побудувати її графік.

а)

x_i	2	5	7	8
n_i	1	3	2	4

б)

x_i	4	8	12	16
n_i	15	10	25	30

в)

x_i	2	3	5	6	6
n_i	6	4	5	5	5

г)

x_i	-2	-1	0	1	2
n_i	7	13	14	15	11

3. П'ятдесят абітурієнтів на вступних іспитах з інформатики отримали таку кількість балів:

12, 14, 19, 15, 14, 18, 13, 16, 17, 12, 20, 17, 15, 13, 17, 16, 20, 14, 14, 13, 17, 16, 15, 19, 16, 15, 18, 17, 15, 14, 16, 15, 15, 18, 15, 15, 19, 14, 16, 18, 18, 15, 15, 17, 15, 16, 16, 14, 14, 17.

Побудувати дискретний статистичний розподіл частот та відносних частот. Знайти емпіричну функцію розподілу та побудувати її графік.

4. 200 однотипних деталей x_i після шліфування були піддані контрольним вимірюванням, результати яких наведено в таблиці:

x_i , мм	3,7	3,8	3,9	4	4,1	4,2	4,3	4,4
n_i	1	22	40	79	27	26	4	1

Побудувати полігон відносних частот, емпіричну функцію розподілу.

5. Побудувати полігон частот та відносних частот заданого розподілу вибірки.

а)

x_i	-2	-1	0	1
n_i	1	3	2	4

б)

x_i	2	3	5	6
n_i	15	10	25	30

в)

x_i	2	5	7	8	10
n_i	6	4	5	5	5

г)

x_i	-2	0	3	5	7
n_i	7	13	14	15	11

Побудувати гістограму частот та відносних частот.

6. У таблиці наведена вибірка оцінок за екзамен з дисципліни «Теорія ймовірностей та математична статистика» 100 студентів економічного факультету. Впорядкувати вибірку. Записати розподіл частот та відносних частот оцінок за екзамен; згрупований розподіл частот та відносних частот оцінок за екзамен; згрупований розподіл накопичених частот та накопичених відносних частот оцінок за екзамен; згрупований розподіл щільності частот та щільності відносних частот.

<i>Вибірка оцінок за екзамен 100 студентів</i>									
52	54	70	73	75	52	82	93	95	82
60	62	50	79	92	91	81	53	80	70
52	82	93	95	82	55	60	62	50	90
52	54	70	73	75	52	82	93	95	82
82	82	93	55	60	62	50	79	28	100
70	73	75	52	58	73	75	52	82	84
52	54	70	73	75	52	82	93	95	82
100	66	84	93	95	82	82	93	95	82
84	60	62	50	79	52	54	70	73	73
73	75	52	82	66	70	73	75	52	54

7. Із генеральної сукупності зроблено вибірку обсягом $n = 32$. Здобуто такі реалізації випадкової величини: 2,2; 7,1; 6,3; 3,9; 5,9; 5,6; 5,6; 4,7; 7,9; 3,2; 6,1; 5,5; 6,4; 6,0; 6,9; 4,7; 6,4; 6,9; 6,7; 7,9; 4,2; 6,7; 6,0; 9,2; 5,5; 6,5; 3,5; 4,9; 7,2; 4,9; 8,9; 5,7. Скласти інтервальний ряд і побудувати гістограму.

8. Було виміряно глибину виробу. Результати вимірювання, мм: 2,41; 2,62; 2,73; 2,52; 2,54; 2,41; 2,81; 2,53; 2,64; 2,61; 2,72; 2,45; 2,52; 2,1; 2,64; 2,52; 2,5; 2,33; 2,24; 2,4; 2,72; 2,51; 2,4; 2,61; 2,42; 2,43; 2,65; 2,54; 2,35; 2,54; 2,62; 2,9; 2,75; 2,24; 2,65; 2,45; 2,53; 2,32; 2,24; 2,55. Скласти інтервальний ряд і побудувати гістограму.

9. Для дослідження розподілу маси новонароджених x_i була зібрана інформація про 100 дітей. Ця інформація подана як інтервальний статистичний розподіл, що має такий вигляд у табличній формі:

$(x_i; x_{i+1})$	(1; 1,5)	(1,5; 2)	(2; 2,5)	(2,5; 3)	(3; 3,5)	(3,5; 4)	(4; 4,5)	(4,5; 5)
n_i	2	8	10	30	40	6	3	1

10. Побудувати емпіричну функцію розподілу, гістограму та полігон частот вибірки, поданої у вигляді таблиці частот:

а)	$[x_i; x_{i+1})$	[-3;-2)	[-2; -1)	[-1; 0)	[0;1)	[1;2)	[2;3)	[3;4)	[4;5)
	n_i	3	10	15	24	25	13	7	3
б)	$[x_i; x_{i+1})$	[0,2; 2,2)	[2,2; 4,2)	[4,2; 6,2)	[6,2; 8,2)	[8,2; 10,2)			
	n_i	70	20	4	3	3			

11. Із партії однотипних сталевих болтів, виготовлених заводом, була здійснена вибірка обсягом 200 шт. і результати вимірювання їх діаметрів x_i наведено у вигляді інтервального статистичного розподілу:

x_i , мм $h = 2$ мм	14,40–14,42	14,42–14,44	14,44–14,46	14,46–14,48	14,48–14,50	14,50–14,52	14,52–14,54	14,54–14,56	14,56–14,58	14,58–14,60	14,60–14,62	14,62–14,64
n_i	2	2	8	9	9	14	41	76	21	11	4	3

Побудувати гістограму частот і емпіричну функцію розподілу.

12. Відсоток виконання плану підприємства за рік та кількість підприємств, що виконують цей план, наведено у вигляді інтервального статистичного розподілу:

$x_i, \%$ $h = 10\%$	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100	100-110	110-120
n_i	2	6	13	16	25	12	10	8	5	3	1

Побудувати гістограму частот і емпіричну функцію розподілу.

Розділ 8

ЧИСЛОВІ ХАРАКТЕРИСТИКИ СТАТИСТИЧНОГО МАТЕРІАЛУ

8.1. Числові характеристики вибірки

Як доповнення до табличних та графічних методів представлення даних, ще одним найважливішим засобом обробки даних є обчислення їх числових характеристик. Найважливіші з них: *вибіркове середнє, дисперсія, середньоквадратичне відхилення*.

Ці характеристики можуть бути обчислені за даними, що знаходяться у вибірці, або за даними, що входять у кінцеву генеральну сукупність.

Числові характеристики, обчислені за вибіркою або ті, що використовуються для опису даних вибірки, називають **статистиками**.

Вибірковим середнім статистичного матеріалу називають суму всіх значень вибірки, поділену на обсяг вибірки n :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (8.1)$$

Якщо маємо частотну таблицю, то вибіркове середнє вираховуємо за формулою:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^m n_i x_i, \quad (8.2)$$

де n – об'єм вибірки, m – число різних варіант, n_1, n_2, \dots, n_m – частоти варіант ($n = n_1 + n_2 + \dots + n_m$), x_i – значення i -ої варіанти.

Вибіркове середнє для інтервального статистичного ряду знаходиться за формулою

$$\bar{x} = \frac{1}{n} \sum_{i=1}^s n_i z_i, \quad (8.3)$$

де n – об'єм вибірки, s – кількість інтервалів, z_i – середина i -ого інтервалу, n_1, n_2, \dots, n_s – число елементів вибірки, які потрапили в i -ий інтервал.

Вибіркове середнє є аналогом математичного сподівання і використовується дуже часто. Воно може набувати різних числових значень при різних вибірках однакового об'єму.

Тому можна розглядати розподіли (теоретичний та емпіричний) вибіркового середнього та числові характеристики цього розподілу (цей розподіл називають *вибірковим*).

Основні властивості вибіркового середнього

1. При множенні усіх варіант вибірки на однаковий множник вибіркоче середнє також множиться на цей множник

$$\frac{1}{n} \sum_{i=1}^m n_i (cx_i) = \frac{c}{n} \sum_{i=1}^m n_i x_i = c\bar{x}.$$

2. Якщо додати (відняти) до всіх варіант вибірки однакове число, то вибіркоче середнє зростає (зменшується) на це число

$$\frac{1}{n} \sum_{i=1}^m n_i (x_i \pm c) = \frac{1}{n} \sum_{i=1}^m n_i x_i \pm \frac{1}{n} c \sum_{i=1}^m n_i = \bar{x} \pm c.$$

Степеневим середнім вибірки називають середнє значення, яке знаходять за формулою

$$\bar{x}_\alpha = \left(\sum_{i=1}^m \frac{n_i}{n} x_i^\alpha \right)^{\frac{1}{\alpha}}. \quad (8.4)$$

При $\alpha = 1$ одержимо вибіркоче середнє.

При $\alpha = 2$ одержимо середньоквадратичне значення вибірки

$$\bar{x}_2 = \left(\sum_{i=1}^m \frac{n_i}{n} x_i^2 \right)^{\frac{1}{2}}.$$

При $\alpha = -1$ одержимо середнє гармонічне

$$\bar{x}_{-1} = \bar{x}_{\text{гарм}} = \frac{n}{\sum_{i=1}^m \frac{n_i}{x_i}}.$$

Середнім геометричним вибірки називають середнє значення, яке обчислюється за формулою

$$\bar{x}_\Gamma = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \dots \cdot x_m^{n_m}}. \quad (8.5)$$

Це середнє обчислюється лише за умови, що всі значення варіанти є додатні, тобто $x_i > 0, i = 1, 2, \dots, m$.

Середнє геометричне застосовується у статистиці для визначення темпу зростання під час дослідження змін ознаки з часом.

Зауваження 1. Обрання того чи іншого середнього для характеристики розподілу пов'язане з якісним аналізом цього розподілу.

Зауваження 2. Крім вказаних степеневих середніх, у статистиці застосовують ще структурні середні, які не залежать від значень варіант, які розташовані на краях розподілу, а пов'язані з рядом частот.

До структурних середніх відносять моду та медіану.

Мода M_o – це елемент, який найчастіше трапляється у вибірці.

Для дискретних статистичних рядів

$$M_o = x_i, \text{ якщо } n_i = \max_i n_i$$

Для інтервальних статистичних рядів

$$M_o = x_i + h \cdot \frac{n_i - n_{i-1}}{2n_i - n_{i-1} - n_{i+1}}, \quad (8.6)$$

де x_i – початок інтервалу з найбільшою частотою, n_i – частота i -го інтервалу.

Медіаною M_e називають варіанту, яка ділить варіаційний ряд на дві частини, які рівні за числом варіант.

Для дискретних статистичних рядів

$$M_e = \begin{cases} x_m, & \text{при } n = 2m - 1, \\ \frac{x_m + x_{m+1}}{2}, & \text{при } n = 2m, \end{cases} \quad (8.7)$$

де x_m – середина варіаційного ряду.

Для інтервальних статистичних рядів

$$M_e = x_i + h \cdot \frac{\frac{n}{2} - \sum_{j=1}^{i-1} n_j}{n_i}, \quad (8.8)$$

де x_i – початок медіанного інтервалу, тобто такого, якому відповідає перша з нагромаджених частот, що перевищує половину всіх спостережень, h – довжина i -ого інтервалу, n_i – частота медіанного інтервалу.

Вибірковою дисперсією D_B називають середнє значення квадратів відхилення варіант від вибіркового середнього з урахуванням відповідних частотей.

$$D_B = \frac{1}{n} \sum_{i=1}^m n_i (x_i - \bar{x})^2. \quad (8.9)$$

Зауваження. Обчислення вибіркової дисперсії спрощується, якщо її знаходити за формулою:

$$D_B = \frac{1}{n} \sum_{i=1}^m n_i x_i^2 - \left(\frac{1}{n} \sum_{i=1}^m n_i x_i \right)^2 = \frac{1}{n} \sum_{i=1}^m n_i x_i^2 - (\bar{x})^2. \quad (8.10)$$

Виправленою вибірковою дисперсією (варіансою) називають суму квадратів відхилень елементів від вибіркового середнього, поділену на $(n-1)$:

$$S^2 = \frac{n}{n-1} D_B = \frac{1}{n-1} \sum_{i=1}^m n_i (x_i - \bar{x})^2. \quad (8.11)$$

Вибіркову дисперсію та виправлену вибірку дисперсію для інтервального розподілу знаходять, відповідно, за формулами:

$$D_B = \frac{1}{n} \sum_{i=1}^s n_i (z_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^s n_i z_i^2 - (\bar{x})^2, \quad (8.12)$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^s n_i (z_i - \bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^s n_i z_i^2 - (\bar{x})^2, \quad (8.13)$$

$$\text{або } S^2 = \frac{n}{n-1} D_B. \quad (8.14)$$

Вибірковим середньоквадратичним відхиленням називають квадратний корінь із вибіркової дисперсії:

$$\sigma_B = \sqrt{D_B}. \quad (8.15)$$

Стандарт – це арифметичний квадратний корінь із варіанси:

$$S = \sqrt{S^2}. \quad (8.16)$$

Розмах – це різниця між найбільшим і найменшим значенням варіаційного ряду (між крайніми елементами) і позначається

$$R = x_{max} - x_{min}. \quad (8.17)$$

Приклад 8.1. Обчислити вибіркоче середнє, вибіркочу дисперсію, вибіркоче середньоквадратичне відхилення, варіанту, стандарт, моду, медіану та розмах вибірки, заданої статистичним розподілом частот:

x_i	-1	2	5	8	10
n_i	5	10	20	5	10

Розв'язання. Об'єм вибірки $n = 5 + 10 + 20 + 5 + 10 = 50$.

Знайдемо, відповідно, вибіркоче середнє, вибіркочу дисперсію, вибіркоче середньоквадратичне відхилення, варіансу, стандарт, моду та медіану вибірки:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^m n_i x_i = \frac{1}{50} (-1 \cdot 5 + 2 \cdot 10 + 5 \cdot 20 + 8 \cdot 5 + 10 \cdot 10) = \frac{255}{50} = 5,1;$$

$$D_B = \frac{1}{n} \sum_{i=1}^m n_i x_i^2 - (\bar{x})^2 = \frac{1}{50} (1 \cdot 5 + 4 \cdot 10 + 25 \cdot 20 + 64 \cdot 5 + 100 \cdot 10) - 5,1^2 = \frac{1}{50} \cdot 1865 - 26,01 = 37,3 - 26,01 = 11,29;$$

$$\sigma_B = \sqrt{D_B} = \sqrt{11,29} \approx 3,36;$$

$$S^2 = \frac{n}{n-1} D_B = \frac{50}{49} \cdot 11,29 = 11,52;$$

$$S = \sqrt{S^2} = \sqrt{11,52} \approx 3,39;$$

$$M_o = 5;$$

$$M_e = 5;$$

$$R = x_{max} - x_{min} = 10 - (-1) = 11.$$

Приклад 8.2. Обчислити вибіркоче середнє, вибіркочу дисперсію, вибіркоче середньоквадратичне відхилення, варіанту, стандарт, моду, медіану та розмах вибірки, заданої інтервальним розподілом:

інтервал	[2; 4)	[4; 6)	[6; 8)	[8; 10)	[10; 12)
n_i	2	8	35	40	15

Розв'язання. Об'єм вибірки $n = 2 + 8 + 35 + 40 + 15 = 100$.

Перетворимо цей інтервальний статистичний розподіл вибірки на точковий, знайшовши середини частинних інтервалів:

$$z_1 = \frac{x_1+x_2}{2} = \frac{2+4}{2} = 3; \quad z_2 = \frac{x_2+x_3}{2} = \frac{4+6}{2} = 5;$$

$$z_3 = \frac{x_3+x_4}{2} = \frac{6+8}{2} = 7; \quad z_4 = \frac{x_4+x_5}{2} = \frac{8+10}{2} = 9;$$

$$z_5 = \frac{x_5+x_6}{2} = \frac{10+12}{2} = 11.$$

Отже, ми отримали такий статистичний розподіл вибірки:

z_i	3	5	7	9	11
n_i	2	8	35	40	15

Знайдемо, відповідно, вибіркове середнє, вибіркочу дисперсію, вибіркочу середньоквадратичне відхилення, варіансу, стандарт, моду та медіану вибірки:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^s n_i z_i = \frac{1}{100} (3 \cdot 2 + 5 \cdot 8 + 7 \cdot 35 + 9 \cdot 40 + 11 \cdot 15) =$$

$$= \frac{1}{100} \cdot 816 = 8,16;$$

$$D_B = \frac{1}{n} \sum_{i=1}^s n_i z_i^2 - (\bar{x})^2 = \frac{1}{100} (9 \cdot 2 + 25 \cdot 8 + 49 \cdot 35 + 81 \cdot 40 + 121 \cdot 15) - 8,16^2 =$$

$$\frac{1}{100} \cdot 6988 - 66,5856 = 69,88 - 66,59 = 3,29;$$

$$\sigma_B = \sqrt{D_B} = \sqrt{3,29} \approx 1,81;$$

$$S^2 = \frac{n}{n-1} D_B = \frac{100}{99} \cdot 3,29 = 3,32;$$

$$S = \sqrt{S^2} = \sqrt{3,32} \approx 1,82;$$

$$M_o = 8 + 2 \cdot \frac{40-35}{2 \cdot 40 - 35 - 15} = 8 + 2 \cdot \frac{5}{30} = 8,33;$$

$$M_e = 8 + 2 \cdot \frac{\frac{100}{2} - (8+2+35)}{40} = 8 + 2 \cdot \frac{50-45}{40} = 8,25;$$

$$R = x_{max} - x_{min} = 12 - 2 = 10.$$

8.2. Метод добутків обчислення вибіркового середнього та вибіркової дисперсії

Нехай вибірку задано у вигляді розподілу рівновіддалених варіант і відповідних їм частот. У цьому разі вибіркоче середнє та вибіркочу дисперсію зручно знаходити методом добутків.

Алгоритм методу добутків

1. У перший стовпчик таблиці записують рівновіддалені варіанти x_i вибірки, розміщуючи їх у порядку зростання.

2. У другий стовпчик таблиці записують відповідні частоти n_i відповідних варіант. Суму елементів цього стовпчика (об'єм вибірки n) записують в останню його клітинку.

3. Третій стовпчик містить умовні варіанти u_i вибірки. Для знаходження умовних варіант вибірки треба:

а) значення варіанти вибірки, яка знаходиться приблизно посередині варіаційного ряду C обрати за умовний нуль;

б) знайти різницю h між будь-якими двома сусідніми варіантами;

в) обчислити умовні варіанти $u_i = \frac{x_i - C}{h}$. Зауважимо, що умовні варіанти завжди будуть цілими числами.

4. У четвертий стовпчик записують добутки частот та відповідних їм умовних варіант $n_i \cdot u_i$. Суму елементів стовпчика $\sum_{i=1}^m n_i \cdot u_i$ записують в його останню клітинку.

5. Знаходять добутки частот та квадратів умовних варіант $n_i \cdot u_i^2$ і записують їх у п'ятий стовпчик. Суму елементів стовпчика $\sum_{i=1}^m n_i \cdot u_i^2$ записують в його останню клітинку.

6. Знаходять добутки частот та квадратів умовних варіант, збільшених на одиницю, $n_i \cdot (u_i + 1)^2$ і записують їх у шостий стовпчик. Суму елементів стовпчика $\sum_{i=1}^m n_i \cdot (u_i + 1)^2$ записують в його останню клітинку.

7. Перевіряють обчислення так: сума елементів шостого стовпчика повинна задовольняти тотожність:

$$\sum_{i=1}^m n_i \cdot (u_i + 1)^2 = \sum_{i=1}^m n_i \cdot u_i^2 + 2 \cdot \sum_{i=1}^m n_i \cdot u_i + n.$$

8. Обчислюють умовні моменти за формулами:

$$M_1^* = \frac{1}{n} \sum_{i=1}^m n_i \cdot u_i; \quad M_2^* = \frac{1}{n} \sum_{i=1}^m n_i \cdot u_i^2.$$

9. Обчислюють вибіркове середнє та вибіркoву дисперсiю за формулами:

$$\bar{x} = M_1^* \cdot h + C, \quad D_B = (M_2^* - (M_1^*)^2) \cdot h^2.$$

Зауваження. Якщо варіанти вибірки не є рівновіддаленими, то інтервал, у якому розміщені всі варіанти вибірки, ділять на кілька однакових за довжиною частинних інтервалів, кожен з яких має містити не менше 8–10 варіант. Потім знаходять середини частинних інтервалів, які й утворюють послідовність рівновіддалених варіант. За частоту кожної середини частинного інтервалу беруть суму частот варіант, які потрапили у відповідний частинний інтервал.

Під час обчислення вибіркової дисперсії з метою зменшення похибки, зумовленої групуванням (особливо при малій кількості інтервалів), роблять поправку Шеппарда, а саме: обчислюють дисперсію за формулою:

$$D'_B = D_B - \frac{1}{12} \cdot h^2.$$

Приклад 8.3. Знайти методом добутків вибіркoве середнє та вибіркoву дисперсію для статистичного розподілу вибірки об'єму $n = 100$.

x_i	2	8	9	13	15	18	20	21	24	27
n_i	8	1	15	13	9	3	4	10	22	15

Розв'язання. Розіб'ємо інтервал $[2; 27]$ на п'ять частинних інтервалів довжиною $h = 5$:

$$[2; 7], \quad (7; 12], \quad (12; 17], \quad (17; 22], \quad (22; 27].$$

Новими варіантами будуть середини цих частинних інтервалів:

$$y_1 = \frac{2+7}{2} = 4,5; \quad y_2 = \frac{7+12}{2} = 9,5; \quad y_3 = \frac{12+17}{2} = 14,5;$$

$$y_4 = \frac{17+22}{2} = 19,5; \quad y_5 = \frac{22+27}{2} = 24,5.$$

За частоти m_i варіант y_i візьмемо суму частот варіант, які потрапили у відповідний i -ий інтервал:

$$m_1 = 8; \quad m_2 = 1 + 15 = 16; \quad m_3 = 13 + 9 = 22;$$

$$m_4 = 3 + 4 + 10 = 17; \quad m_5 = 22 + 15 = 37.$$

Отже, ми отримали такий статистичний розподіл рівновіддалених варіант:

y_i	4,5	9,5	14,5	19,5	24,5
m_i	8	16	22	17	37

Складемо розрахункову таблицю:

y_i	m_i	u_i	$m_i u_i$	$m_i u_i^2$	$m_i (u_i + 1)^2$
4,5	8	-2	-16	32	8
9,5	16	-1	-16	16	0
14,5	22	0	0	0	22
19,5	17	1	17	17	68
24,5	37	2	74	148	333
	$n = 100$		$\sum_{i=1}^5 m_i u_i = 59$	$\sum_{i=1}^5 m_i u_i^2 = 213$	$\sum_{i=1}^5 m_i (u_i + 1)^2 = 431$

Перевірка:

$$\sum_{i=1}^5 m_i (u_i + 1)^2 = 431;$$

$$\sum_{i=1}^5 m_i u_i^2 + 2 \cdot \sum_{i=1}^5 m_i u_i + n = 213 + 2 \cdot 59 + 100 = 431.$$

Обчислимо умовні моменти першого і другого порядків:

$$M_1^* = \frac{1}{n} \sum_{i=1}^m n_i \cdot u_i = \frac{59}{100} = 0,59; \quad M_2^* = \frac{1}{n} \sum_{i=1}^m n_i \cdot u_i^2 = \frac{213}{100} = 2,13.$$

Методом добутків знайдемо вибіркове середнє й дисперсію, враховуючи, що хибний нуль $C = 14,5$:

$$\bar{x} = M_1^* \cdot h + C = 0,59 \cdot 5 + 14,5 = 17,45,$$

$$D_B = (M_2^* - (M_1^*)^2) \cdot h^2 = (2,13 - 0,59^2) \cdot 5^2 = 44,5475.$$

Оскільки кількість частинних інтервалів мала, то скористаємося поправкою Шеппарда:

$$D'_B = D_B - \frac{1}{12} \cdot h^2 = 44,5475 - \frac{1}{12} \cdot 5^2 \approx 42,46.$$

8.3. Метод сум обчислення вибіркового середнього та вибіркової дисперсії

Нехай вибірку задано у вигляді розподілу рівновіддалених варіант і відповідних їм частот. Тоді вибіркоче середнє та вибіркочу дисперсію зручно знаходити методом сум.

Алгоритм методу сум

1. У першій стовпчик таблиці записують рівновіддалені варіанти x_i вибірки, розміщуючи їх у порядку зростання.

2. У другий стовпчик таблиці записують відповідні частоти n_i варіант. Суму елементів цього стовпчика (об'єм вибірки n) записують в його останню клітинку.

3. Третій стовпчик містить умовні варіанти u_i вибірки. Для знаходження умовних варіант вибірки треба:

а) значення варіанти вибірки з найбільшою частотою C обрати за умовний нуль, це значення називають модою;

б) у порожніх клітинках третього стовпця, які розміщені над нулем (крім верхньої), зверху вниз записують послідовно нагромаджені частоти. Додавши всі нагромаджені частоти зверху до нуля, отримують число b_1 , яке записують у верхній клітинці;

в) у порожніх клітинках третього стовпця, які розміщені під нулем (крім нижньої), знизу вверх записують послідовно нагромаджені частоти. Додавши всі нагромаджені частоти знизу до нуля, отримують число a_1 , яке записують у його нижню клітинку.

4. У четвертому стовпці знову записують хибний нуль, над і під ним записують ще по одному нулю. Аналогічно до третього, заповнюють четвертий стовбець, причому нагромаджуються частоти третього стовпця. Суму нагромаджених частот, розміщених над нулем, позначають b_2 і записують у верхній клітинці четвертого стовпця. Суму нагромаджених частот, розміщених під нулем, позначають a_2 і записують у нижню клітинку четвертого стовпця.

5. Обчислюють d_1, s_1, s_2 :

$$d_1 = a_1 - b_1; \quad s_1 = a_1 + b_1; \quad s_2 = a_2 + b_2.$$

6. Обчислюють умовні моменти за формулами:

$$M_1^* = \frac{d_1}{n}; \quad M_2^* = \frac{s_1 + 2s_2}{n}.$$

7. Обчислюють вибіркоче середнє та вибіркочу дисперсїю за формулами:

$$\bar{x} = M_1^* \cdot h + C,$$

$$D_B = (M_2^* - (M_1^*)^2) \cdot h^2.$$

Приклад 8.4. Знайти методом сум вибіркоче середнє та вибіркочу дисперсїю для статистичного розподїлу вибірки об'єму $n = 100$.

x_i	52	58	64	70	76	82	88	94	100	106
n_i	4	8	10	13	18	21	14	9	2	1

Розв'язання. Складемо розрахункочу таблицю

1	2	3	4
x_i	n_i	$b_1 = 126$	$b_2 = 131$
52	4	4	4
58	8	12	16
64	10	22	38
70	13	35	73
76	18	53	0
82	21	0	0
88	14	26	0
94	9	12	16
100	2	3	4
106	1	1	1
		$a_1 = 42$	$a_2 = 21$

Знайдемо d_1, s_1, s_2 :

$$d_1 = a_1 - b_1 = 42 - 126 = -84;$$

$$s_1 = a_1 + b_1 = 42 + 126 = 168;$$

$$s_2 = a_2 + b_2 = 21 + 131 = 152.$$

Обчислимо умовні моменти першого і другого порядку:

$$M_1^* = \frac{d_1}{n} = \frac{-84}{100} = -0,84;$$

$$M_2^* = \frac{s_1 + 2s_2}{n} = \frac{168 + 2 \cdot 152}{100} = 4,72.$$

Знайдемо вибіркове середнє та дисперсію, врахувавши, що хибний нуль $C = 82$, а крок (відстань між двома сусідніми варіантами) $h = 6$, отримаємо:

$$\bar{x} = M_1^* \cdot h + C = -0,84 \cdot 6 + 82 = 76,96;$$

$$D_B = (M_2^* - (M_1^*)^2) \cdot h^2 = (4,72 - (-0,84)^2) \cdot 6^2 = 144,5184.$$

Запитання для самоконтролю

1. Що відноситься до числових характеристики вибірки?
2. Які числові характеристики називаються статистичними?
3. Як визначається вибіркова середня для статистичного розподілу вибірки?
4. Як визначається вибіркова середня для інтервального розподілу вибірки?
5. Які властивості має вибіркова середня?
6. Що таке степеневе середнє; середнє геометричне?
7. Як визначається дисперсія для статистичного розподілу вибірки?
8. Як визначається дисперсія для інтервального розподілу вибірки?
9. Як визначається середньоквадратичне відхилення?
10. Що таке мода і як її знайти?
11. Що таке медіана і як її знайти?
12. Що таке варіанса, стандарт, розмах і як їх обчислити?
13. Метод добутків для знаходження вибіркового середнього та вибіркової дисперсії, його алгоритм.
14. Метод сум для знаходження вибіркового середнього та вибіркової дисперсії, його алгоритм.
15. У яких випадках використовують виправлену вибірккову дисперсію і як вона пов'язана із вибірковою дисперсією?
16. В яких випадках обчислюють характеристики вибірки методом добутків; методом сум?

Задачі для самостійного розв'язування

1. Знайти вибіркове середнє, дисперсію, моду і медіану для вибірки. Побудувати емпіричну функцію розподілу та полігон частот.

а)

x_i	-2	5	7	8
n_i	12	12	16	20

б)

x_i	4	8	12	16
n_i	5	10	5	20

в)

x_i	1	2	3	4	5
n_i	5	3	2	8	2

г)

x_i	-2	-1	0	1	2
n_i	7	15	23	31	39

г)

x_i	0	1	2	3	4	5	6	7	8	9	10	11
n_i	45	41	34	54	39	43	41	33	37	41	47	39

2. Через кожен годину вимірювалась напруга в електромережі. Результати вимірювання напруги у вольтах наведені у вигляді статистичного ряду:

222, 219, 224, 220, 218, 217, 221, 220, 215, 218, 223, 225, 220, 226, 221, 216, 211, 219, 220, 221, 222, 218, 221, 219.

Побудувати статистичний розподіл, полігон частот. Обчислити числові характеристики вибірки.

3. Знайти вибіркове середнє, дисперсію, середньоквадратичне відхилення, варіансу, стандарт, моду, медіану і розмах для вибірки, заданої інтервальним розподілом. Побудувати гістограму та полігон частот.

а)

$[x_i; x_{i+1})$	[4; 6)	[6; 8)	[8; 10)	[10; 12)	[12; 14)
n_i	5	3	2	8	2

б)

$[x_i; x_{i+1})$	[3; 5)	[5; 7)	[7; 11)	[11; 13)	[13; 15)
n_i	4	3	6	3	4

в)

$[x_i; x_{i+1})$	[-3; -1)	[-1; 0)	[0; 1)	[1; 2)	[2; 3)	[3; 5)
n_i	13	15	24	25	13	10

г)

$[x_i; x_{i+1})$	[0; 1)	[1; 2)	[2; 3)	[3; 4)	[4; 5)	[5; 6)	[6; 7)	[7; 8)	[8; 9)	[9; 10)
n_i	35	16	15	17	17	19	11	16	30	24

4. Рівень води x_i в річці щодо номіналу вимірювався протягом 50-ти років навесні. Результати вимірювання наведено у формі інтервального статистичного розподілу.

x_i , см $h = 24$	0-24	24-48	48-72	72-96	96-120	120-144	144-168	168-192	192-216
n_i	0	2	4	6	12	16	6	3	1

Побудувати гістограму частот. Обчислити числові характеристики вибірки.

5. Методом добутоків знайти вибірконе середнє та вибірконе дисперсію для статистичного розподілу вибірки.

a)

x_i	12	14	16	18	20	22
n_i	5	15	50	16	10	4

б)

x_i	65	70	75	80	85
n_i	2	5	25	15	3

6. Методом сум знайти вибірконе середнє та вибірконе дисперсію для статистичного розподілу вибірки.

a)

x_i	40	46	52	68	74	80
n_i	10	14	12	5	11	8

б)

x_i	20	40	60	80	100
n_i	5	10	15	20	10

7. Середню температуру повітря у червні в Києві та Львові вимірювали протягом 40 років. Дані вимірювання наведено в таблиці (X – середня температура повітря у червні в Києві; Y – середня температура повітря у червні у Львові):

X	Y	X	Y	X	Y	X	Y
12	10,8	14	10	16	13	18	14
12	11,3	14	12	15,9	15	18	14
12	12	13,9	12,4	16	16	18,1	14,8
12,8	11	15	11	16,9	12,9	18,4	17,8
12	12	14,9	13	17,2	13,9	19,2	15
13,8	9,8	14,9	14,2	16,9	14,8	19,3	16
13,1	11,5	14,2	13,8	16,9	15	20	17
13	13	15	16	17	16	20	17,7
13,9	10,1	15,5	13,9	16,8	17	14	14,8
14,2	10	15,9	14,7	17,5	16	14	15,2

Знайти числові характеристики вибірки X та Y . Зробити висновки про те, як змінювалася середня температура повітря у червні в Києві та Львові.

8. Записати цю вибірку у вигляді варіаційного ряду; скласти статистичний ряд частот; скласти статистичний ряд відносних частот; згрупований розподіл частот та відносних частот; згрупований розподіл накопичених частот та накопичених відносних частот. Знайти емпіричну функцію розподілу та побудувати її графік. Побудувати гістограму та полігон частот та відносних частот. Обчислити вибіркові числові характеристики (вибіркове середнє, варіансу, стандарт, моду, медіану, розмах).

2, 5, 8, 6, 4, 5, 8, 6, 8, 4, 8, 5, 8, 10, 2, 5, 8, 6, 8, 6, 8, 4, 8, 5, 8, 6, 8, 4, 8, 5, 8, 10, 5, 8, 6, 8, 6, 5, 8, 6, 8, 4, 8, 5, 8, 10, 2, 5, 8, 10, 4, 8, 5, 8, 10, 5, 8, 6, 8, 6, 5, 8, 6, 8, 4, 8, 5, 8, 10, 2, 5, 8, 10, 2, 5, 8, 6, 4, 5, 8, 6, 8, 4, 8, 5, 8, 10, 2, 5, 8, 6, 8, 6, 8, 4, 8, 5, 8, 6, 8.

Розділ 9

СТАТИСТИЧНІ ОЦІНКИ ПАРАМЕТРІВ РОЗПОДІЛУ

9.1. Точкові оцінки

Статистичною оцінкою θ^* невідомого параметра θ теоретичного розподілу називають функцію $f(X_1, X_2, \dots, X_n)$, де X_1, X_2, \dots, X_n – спостережувальні випадкові величини.

Точковою оцінкою називають статистичну оцінку, яка визначається одним числом $\theta^* = f(x_1, x_2, \dots, x_n)$, де x_1, x_2, \dots, x_n – результати n спостережень над кількісною ознакою випадкової величини X .

Наприклад, вибіркова середня \bar{x} , вибіркова дисперсія D_B та вибіркове середньоквадратичне σ_B – точкові оцінки відповідних числових характеристик генеральної сукупності.

Точкові оцінки параметрів розподілу є випадковими величинами, їх можна вважати первинними результатами обробки вибірки тому, що невідомо, з якою точністю кожна з них оцінює відповідну числову характеристику генеральної сукупності.

Якщо об'єм вибірки доволі великий, то точкові оцінки задовольняють практичні потреби точності.

Щоб статистичні оцінки давали найкращі наближення, вони повинні задовольняти певним вимогам. Розглянемо ці вимоги.

Нехай θ^* – статистична оцінка невідомого параметра θ теоретичного розподілу. Припустимо, що за вибіркою об'єму n знайдена оцінка θ_1^* . Для інших вибірок того ж об'єму одержано оцінки $\theta_2^*, \theta_3^*, \dots, \theta_m^*$. Оцінку θ^* можна розглядати як випадкову величину, а числа $\theta_1^*, \theta_2^*, \dots, \theta_m^*$ – як її можливі значення.

Якщо θ_k^* ($k = 1, 2, \dots, m$) будуть більші від значення θ , то оцінка θ^* дає наближене значення θ з надлишком. У цьому випадку математичне сподівання випадкової величини θ^* є більшим від θ ($M(\theta^*) > \theta$).

Якщо θ^* дає оцінку з недостачею, математичне сподівання випадкової величини θ^* є меншим від θ ($M(\theta^*) < \theta$).

Відтак використання статистичної оцінки, математичне сподівання якої не дорівнює параметру θ , спричиняє систематичні похибки. Вимога $M(\theta^*) = \theta$ застерігає від таких похибок.

Незміщеною називають точкову оцінку θ^* , математичне сподівання якої дорівнює оцінюваному параметру θ для будь-якого обсягу вибірки:

$$M(\theta^*) = \theta.$$

Зміщеною називають точкову оцінку θ^* , математичне сподівання якої відмінне від оцінюваного параметра θ .

Якщо $M(\theta^*) = \theta + b(\theta)$, то $b(\theta)$ називають **зсувом оцінки θ^*** .

Вимога про незміщеність оцінки θ^* є недостатньою, оскільки можливі значення θ^* можуть бути дуже розсіяні від середнього значення, дисперсія $D(\theta^*)$ може бути великою. Тоді, знайдена за даними однієї вибірки оцінка θ_k^* може значно відрізнятись від середнього значення θ^* , а отже, і від параметра θ .

Якщо дисперсія $D(\theta^*)$ буде незначною, то можливість допустити помилку виключена. Тому для статистичної оцінки виникає вимога про її ефективність.

Ефективною називають таку оцінку θ^* , яка при заданому об'ємі n має найменшу можливу дисперсію.

Обґрунтованою називають статистичну оцінку, яка при $n \rightarrow \infty$ прямує за ймовірністю до оцінюваного параметра.

Послідовність оцінок θ_k^* ($k = 1, 2, \dots, m$) параметра θ називається **спроможною**, якщо $\theta_k^* \xrightarrow{P} \theta, n \rightarrow \infty$.

Послідовність оцінок θ_k^* ($k = 1, 2, \dots, m$) параметра θ називається **сильно спроможною**, якщо $\theta_k^* \rightarrow \theta$ з ймовірністю одиниця при $n \rightarrow \infty$.

Незміщеною оцінкою генерального середнього (математичного сподівання) є вибіркове середнє

$$\bar{x} = \frac{1}{n} \sum_{i=1}^m n_i x_i,$$

де n_i – частота варіанти x_i ; x_i – варіанта вибірки; $n = n_1 + n_2 + \dots + n_m$ – об'єм вибірки.

Зауваження 1. Якщо варіанти x_i вибірки є дуже великими або дуже малими (близькими до нуля) числами, то для спрощення розрахунків доцільно відняти (у разі великих від'ємних чисел – додати) від кожної варіанти одне й те саме число C (як C можна вибрати будь – яке число, розміщене приблизно посередині варіаційного ряду), потім поділити (у разі близьких до нуля чисел – помножити) на одне й те саме число b (як b можна вибрати найбільше спільне кратне), тобто перейти до умовних варіант

$$u_i = \frac{x_i \pm C}{b} \quad (u_i = x_i b). \quad (9.1)$$

Тоді

$$\bar{x} = \frac{b}{n} \sum_{i=1}^m n_i u_i \mp C = b \bar{u}_b \mp C \quad \left(\bar{x} = \frac{\sum_{i=1}^m n_i u_i}{bn} = \frac{\bar{u}_b}{b} \right). \quad (9.2)$$

Зміщеною оцінкою генеральної дисперсії є вибіркова дисперсія

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

Ця оцінка зміщена, оскільки

$$M(D_B) = \frac{n-1}{n} D_\Gamma.$$

Для обчислення вибіркової дисперсії можна скористатися зручнішою формулою

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

Зауваження 2. Якщо варіанти x_i вибірки є дуже великими або дуже малими (близькими до нуля) числами, то для спрощення розрахунків доцільно відняти (у разі великих від'ємних чисел – додати) від кожної варіанти одне й те саме число C (як C можна вибрати будь-яке число, розміщене приблизно посередині варіаційного ряду), потім поділити (у разі близьких до нуля чисел – помножити) на одне й те саме число b (як b можна вибрати найбільше спільне кратне), тобто перейти до умовних варіант

$$u_i = \frac{x_i \pm C}{b} \quad (u_i = x_i b).$$

Тоді

$$D_B = b^2 \left(\frac{1}{n} \sum_{i=1}^m n_i u_i^2 - \left(\frac{1}{n} \sum_{i=1}^m n_i u_i \right)^2 \right) = b^2 (\overline{u^2} - (\bar{u})^2). \quad (9.3)$$

$$\left(D_B = \left(\frac{1}{b^2 n} \sum_{i=1}^m n_i u_i^2 - \left(\frac{1}{bn} \sum_{i=1}^m n_i u_i \right)^2 \right) = \frac{(\overline{u^2} - (\bar{u})^2)}{b^2} \right). \quad (9.4)$$

Незміщеною оцінкою генеральної дисперсії є виправлена вибіркова дисперсія

$$S^2 = \frac{n}{n-1} D_B.$$

Поправку $\frac{n}{n-1}$ називають *поправкою Бесселя*.

Приклад 9.1. З генеральної сукупності отримано деяку вибірку об'єму $n = 100$:

x_i	0,002	0,004	0,006	0,008	0,01	0,012	0,014
n_i	7	29	35	12	9	5	3

Знайти зміщену оцінку генеральної дисперсії.

Розв'язання. Зміщеною оцінкою генеральної дисперсії D_T є вибіркова дисперсія D_B . Оскільки варіанти вибірки є малими, близькими до нуля числами, перейдемо до умовних варіант

$$u_i = x_i \cdot 1000$$

(як b вибрано число 1000, оскільки в такому разі ми отримаємо цілі числа):

$$u_1 = x_1 \cdot 1000 = 0,002 \cdot 1000 = 2;$$

$$u_2 = x_2 \cdot 1000 = 0,004 \cdot 1000 = 4;$$

$$u_3 = x_3 \cdot 1000 = 0,006 \cdot 1000 = 6;$$

$$u_4 = x_4 \cdot 1000 = 0,008 \cdot 1000 = 8;$$

$$u_5 = x_5 \cdot 1000 = 0,01 \cdot 1000 = 10;$$

$$u_6 = x_6 \cdot 1000 = 0,012 \cdot 1000 = 12;$$

$$u_7 = x_7 \cdot 1000 = 0,014 \cdot 1000 = 14.$$

Тепер можна знайти вибіркову дисперсію:

$$\begin{aligned} D_B &= \left(\frac{1}{b^2 n} \sum_{i=1}^7 n_i u_i^2 - \left(\frac{1}{bn} \sum_{i=1}^7 n_i u_i \right)^2 \right) = \\ &= \frac{7 \cdot 2^2 + 29 \cdot 4^2 + 35 \cdot 6^2 + 12 \cdot 8^2 + 9 \cdot 10^2 + 5 \cdot 12^2 + 3 \cdot 14^2}{1000^2 \cdot 100} - \end{aligned}$$

$$\begin{aligned}
& - \left[\frac{7 \cdot 2 + 29 \cdot 4 + 35 \cdot 6 + 12 \cdot 8 + 9 \cdot 10 + 5 \cdot 12 + 3 \cdot 14}{1000 \cdot 100} \right]^2 = \\
& = \frac{28 + 464 + 1260 + 768 + 900 + 720 + 588}{100000000} - \\
& - \left[\frac{14 + 116 + 210 + 96 + 90 + 60 + 42}{100000} \right]^2 = \\
& = 4728 \cdot 10^{-8} - 394384 \cdot 10^{-10} = 78416 \cdot 10^{-10} = 0,0000078416.
\end{aligned}$$

Приклад 9.2. У результаті статистичних досліджень випадкової величини X отримано таку вибірку: 47, 45, 46, 46, 46, 45, 47, 44, 46, 45, 45, 46, 46, 44, 46, 48, 46, 46, 45, 46, 44, 46, 45, 47, 46, 46, 47, 46, 46, 48, 44, 46, 45, 46, 45, 46, 44, 47, 46, 46, 45, 47, 48, 44, 46, 46, 45, 46, 47, 45.

Знайти незміщені оцінки генерального середнього та генеральної дисперсії.

Розв'язання. Об'єм вибірки $n = 50$. Побудуємо статистичний розподіл вибірки:

x_i	44	45	46	47	48
n_i	6	11	23	7	3

Перевірка: $n = 6 + 11 + 23 + 7 + 3 = 50$.

Незміщеною оцінкою генерального середнього є вибіркове середнє.

Оскільки варіанти вибірки є великими числами, перейдемо до умовних варіант

$$u_i = x_i - 46.$$

Як C вибрано число 46:

$$u_1 = x_1 - 46 = 44 - 46 = -2; \quad u_2 = x_2 - 46 = 45 - 46 = -1;$$

$$u_3 = x_3 - 46 = 46 - 46 = 0; \quad u_4 = x_4 - 46 = 47 - 46 = 1;$$

$$u_5 = x_5 - 46 = 48 - 46 = 5.$$

Тепер можна знайти вибіркове середнє:

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^m n_i u_i + C = \frac{-2 \cdot 6 + (-1) \cdot 11 + 0 \cdot 23 + 1 \cdot 7 + 2 \cdot 3}{50} + 46 = \\ &= \frac{-10}{50} + 46 = 45,8.\end{aligned}$$

Щоб знайти незміщену оцінку генеральної дисперсії (виправлену вибіркoву дисперсію), визначимо вибіркoву дисперсію й помножимо її на поправку Бесселя:

$$\begin{aligned}D_B &= \frac{1}{n} \sum_{i=1}^m n_i x_i^2 - \left(\frac{1}{n} \sum_{i=1}^m n_i x_i \right)^2 = \\ &= \frac{(-2)^2 \cdot 6 + (-1)^2 \cdot 11 + 0^2 \cdot 23 + 1^2 \cdot 7 + 2^2 \cdot 3}{50} - \\ &\quad - \left[\frac{-2 \cdot 6 + (-1) \cdot 11 + 0 \cdot 23 + 1 \cdot 7 + 2 \cdot 3}{50} \right]^2 = \\ &= \frac{54}{50} - \left[\frac{-10}{50} \right]^2 = 1,04. \\ S^2 &= \frac{n}{n-1} D_B = \frac{50}{49} \cdot 1,04 \approx 1,06.\end{aligned}$$

Отже, незміщеними оцінками генерального середнього та генеральної дисперсії є

$$\bar{x}_T^* = 45,8; \quad \bar{D}_T^* \approx 1,06.$$

9.2. Методи визначення точкових статистичних оцінок

9.2.1. Метод моментів

Методом моментів знаходження точкових оцінок називають метод, за яким для обчислення невідомих параметрів заданого розподілу прирівнюють теоретичні та емпіричні моменти.

Якщо розподіл визначається одним параметром, то для знаходження оцінки прирівнюють математичне сподівання до вибіркового середнього

$$M(X) = \bar{x},$$

а потім із цього рівняння визначають шукану точкову оцінку невідомого параметра.

Якщо розподіл визначається двома параметрами, то їх точкові оцінки знаходять із системи рівнянь:

$$\begin{cases} M(X) = \bar{x}; \\ D(X) = D_B. \end{cases} \quad (9.5)$$

Лівими частинами цих рівнянь є математичне сподівання та дисперсія, які дорівнюють відповідно вибіркового середнього та вибірковій дисперсії.

Приклад 9.3. Випадкова величина X (зріст дорослої людини) розподілена за нормальним розподілом з параметрами a, σ . У результаті статистичних досліджень отримано такий статистичний розподіл зросту дорослих людей для $n = 1000$ осіб:

Зріст, см	(145; 155]	(155; 165]	(165; 175]	(175; 185]	(185; 195]	(195; 205]	(205; 215]
Кількість осіб	24	112	263	322	202	66	11

Знайти методом моментів точкову оцінку невідомих параметрів a, σ нормального розподілу.

Розв'язання. Перетворимо інтервальний статистичний розподіл на точковий, вибравши як варіанти середини частинних інтервалів:

Зріст, см	150	160	170	180	190	200	210
Кількість осіб	24	112	263	322	202	66	11

Оскільки нормальний закон розподілу залежить від двох параметрів, треба знайти вибіркве середнє та вибіркву дисперсію, а потім прирівняти їх відповідно до математичного сподівання та дисперсії.

Перейдемо до умовних варіант

$$u_i = \frac{x_i - 180}{10}$$

(x_i – зріст людини):

$$u_1 = \frac{150-180}{10} = -3; \quad u_2 = \frac{160-180}{10} = -2; \quad u_3 = \frac{170-180}{10} = -1;$$

$$u_4 = \frac{180-180}{10} = 0; \quad u_5 = \frac{190-180}{10} = 1; \quad u_6 = \frac{200-180}{10} = 2;$$

$$u_7 = \frac{210-180}{10} = 3.$$

Знайдемо вибіркве середнє та вибіркву дисперсію:

$$\begin{aligned} \bar{x} &= \frac{10}{n} \sum_{i=1}^m n_i u_i + 180 = \\ &= \frac{10 \cdot (-3 \cdot 24 + (-2) \cdot 112 + (-1) \cdot 263 + 0 \cdot 322 + 1 \cdot 202 + 2 \cdot 66 + 3 \cdot 11)}{1000} + 180 = \\ &= 0,01 \cdot (-72 - 224 - 263 + 0 + 202 + 132 + 33) + 180 = \\ &= 0,01 \cdot (-192) + 180 = 178,08; \end{aligned}$$

$$\begin{aligned} D_B &= b^2 \left(\frac{1}{n} \sum_{i=1}^m n_i u_i^2 - \left(\frac{1}{n} \sum_{i=1}^m n_i u_i \right)^2 \right) = \\ &= 10^2 \cdot \left(\frac{(-3)^2 \cdot 24 + (-2)^2 \cdot 112 + (-1)^2 \cdot 263 + 0^2 \cdot 322 + 1^2 \cdot 202 + 2^2 \cdot 66 + 3^2 \cdot 11}{1000} \right) - \\ &\quad - \left[\frac{(-3) \cdot 24 + (-2) \cdot 112 + (-1) \cdot 263 + 0 \cdot 322 + 1 \cdot 202 + 2 \cdot 66 + 3 \cdot 11}{1000} \right]^2 = \\ &= 100 \cdot \left(\frac{1492}{1000} - \left[\frac{-192}{1000} \right]^2 \right) = 100 * 1,455136 = 145,5136. \end{aligned}$$

Оскільки параметр a нормального закону розподілу є математичним сподіванням, а параметр σ – середнім квадратичним відхиленням, то оцінками цих параметрів є

$$a^* = \bar{x} = 178,08; \quad \sigma^* = \sqrt{D_B} = \sqrt{145,5136} \approx 12,0629.$$

9.2.2. Метод найменших квадратів

Згідно з цим методом статистичні оцінки визначаються з умови мінімізації суми квадратів відхилень варіант вибірки від статистичної оцінки θ^* .

Отже, використовуючи метод найменших квадратів, можна, наприклад, визначити статистичну оцінку для $\bar{X}_T = M(X)$. Для цього скористаємося функцією $u = \sum_{i=1}^n (x_i - \theta^*)^2 n_i$. Використовуючи умову екстремуму, отримаємо:

$$\frac{\partial u}{\partial \theta^*} = -2 \sum_{i=1}^n (x_i - \theta^*) n_i = 0 \rightarrow$$

$$\rightarrow \sum_{i=1}^n x_i n_i - \sum_{i=1}^n \theta^* n_i = 0 \rightarrow$$

$$\theta^* = \frac{\sum_{i=1}^n x_i n_i}{n} = \bar{x}_B.$$

Звідси для $\theta = \bar{X}_T$ точковою статистичною оцінкою буде $\theta^* = \bar{x}$ – вибіркова середня.

9.2.3. Метод максимальної правдоподібності

Метод максимальної правдоподібності займає головне місце в теорії статистичного оцінювання параметрів θ . На нього свого часу зауважував К. Гаусс, а розробив його Р. Фішер. Цей метод розглянемо детальніше.

Метод максимальної правдоподібності полягає у знаходженні максимуму функції одного або кількох оцінюваних параметрів.

Припустимо, що X – дискретна випадкова величина з відомим законом розподілу, який визначається невідомим параметром θ . За даними вибірки x_1, x_2, \dots, x_n , отриманої в результаті спостережень над випадковою величиною X , необхідно знайти точкову оцінку $\theta^* = \theta^*(x_1, x_2, \dots, x_n)$ параметра θ .

Функцією правдоподібності L дискретної випадкової величини називають функцію аргумента θ

$$L(x_1, x_2, \dots, x_n; \theta) = p(x_1; \theta) \cdot p(x_2; \theta) \cdot \dots \cdot p(x_n; \theta), \quad (9.6)$$

де $p(x_i; \theta)$ – ймовірність того, що в результаті випробування випадкова величина X набуде значення x_i .

Оцінкою найбільшої або максимальної правдоподібності параметра θ називають таке його значення θ^* , за якого функція правдоподібності досягає свого максимуму.

Логарифмічною функцією правдоподібності називають функцію $\ln L$.

Оскільки функції L і $\ln L$ досягають свого максимуму при одному і тому самому значенні θ , здебільшого зручніше знаходити максимум функції $\ln L$, а не L .

Якщо випадкова величина X неперервна, то відомою вважається щільність розподілу ймовірностей $f(x)$, а невідомим – параметр, від якого залежить ця щільність.

Функцією правдоподібності L неперервної випадкової величини називають функцію аргумента θ

$$L(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdot \dots \cdot f(x_n; \theta). \quad (9.7)$$

Оцінку найбільшої правдоподібності невідомого параметра розподілу неперервної випадкової величини шукають так само, як і у випадку дискретної випадкової величини, а саме:

1) визначають похідну $\frac{\partial \ln L}{\partial \theta}$ (або $\frac{\partial L}{\partial \theta}$);

2) знаходять корені θ_i^* рівняння $\frac{\partial \ln L}{\partial \theta} = 0$ (або $\frac{\partial L}{\partial \theta} = 0$). Ці рівняння називають **рівняннями правдоподібності**;

3) визначають другу похідну $\frac{\partial^2 \ln L}{\partial \theta^2}$ (або $\frac{\partial^2 L}{\partial \theta^2}$). Корінь θ_i^* рівняння правдоподібності, для якого друга похідна від'ємна, беруть за оцінку θ^* найбільшої правдоподібності параметра θ .

Якщо параметр θ двовимірний, тобто $\theta = (\theta_1, \theta_2)$, то для знаходження максимуму функції правдоподібності складають і розв'язують систему рівнянь:

$$\begin{cases} \frac{\partial \ln L}{\partial \theta_1} = 0; \\ \frac{\partial \ln L}{\partial \theta_2} = 0. \end{cases} \quad \text{або} \quad \begin{cases} \frac{\partial L}{\partial \theta_1} = 0; \\ \frac{\partial L}{\partial \theta_2} = 0. \end{cases} \quad (9.8)$$

Наприклад, коли ознака генеральної сукупності X має нормальний закон розподілу, то функція максимальної правдоподібності набере такого вигляду:

$$L(x_1, x_2, \dots, x_n; \theta_1^*, \theta_2^*) = \frac{1}{(2\pi\theta_2^*)^{\frac{n}{2}}} \cdot e^{-\frac{\sum_{i=1}^n (x_i - \theta_1^*)^2}{2\theta_2^*}}. \quad (9.9)$$

Водночас за статистичні оцінки θ_1^*, θ_2^* вибирають ті їх значення, за яких задана вибірка буде найімовірнішою, тобто функція $L(x_1, x_2, \dots, x_n; \theta_1^*, \theta_2^*)$ досягає максимуму.

На практиці зручно від функції (9.9) перейти до її логарифма, а саме:

$$\ln L(x_1, x_2, \dots, x_n; \theta_1^*, \theta_2^*) = -\frac{n}{2} (\ln \pi + \ln \theta_2^*) - \frac{\sum_{i=1}^n (x_i - \theta_1^*)^2}{2\theta_2^*}. \quad (9.10)$$

Згідно з необхідною умовою екстремуму для цієї функції дістанемо:

$$\begin{cases} \frac{\partial \ln L}{\partial \theta_1^*} = -\frac{\sum_{i=1}^n (x_i - \theta_1^*)}{\theta_1^*} = 0; \\ \frac{\partial \ln L}{\partial \theta_2^*} = -\frac{n}{2\theta_2^*} + \frac{\sum_{i=1}^n (x_i - \theta_1^*)^2}{2(\theta_2^*)^2} = 0. \end{cases} \quad (9.11)$$

З першого рівняння системи дістанемо:

$$\theta_1^* = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}.$$

З другого рівняння системи маємо:

$$\theta_2^* = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = D_B.$$

Отже, для $\bar{X}_\Gamma = M(X)$ точковою статистичною оцінкою є \bar{x} для $D_\Gamma - D_B$.

Приклад 9.4. Випадкова величина X (кількість битого скляного посуду в одній упаковці) розподілена за законом Пуассона з невідомим параметром λ . У результаті статистичних досліджень отримано такий емпіричний розподіл кількості битого скляного посуду в $n = 1000$ упаковках:

x_i	0	1	2	3	4	5	6
n_i	554	324	98	19	3	1	1

Знайти методом максимальної правдоподібності точкову оцінку невідомого параметра λ розподілу Пуассона.

Розв'язання. Оскільки випадкова величина X розподілена за законом Пуассона, то функція правдоподібності має вигляд:

$$L = \left(\frac{\lambda^0 e^{-\lambda}}{0!}\right)^{554} \cdot \left(\frac{\lambda^1 e^{-\lambda}}{1!}\right)^{324} \cdot \left(\frac{\lambda^2 e^{-\lambda}}{2!}\right)^{98} \cdot \left(\frac{\lambda^3 e^{-\lambda}}{3!}\right)^{19} \cdot \left(\frac{\lambda^4 e^{-\lambda}}{4!}\right)^3 \cdot \left(\frac{\lambda^5 e^{-\lambda}}{5!}\right)^1 \cdot \left(\frac{\lambda^6 e^{-\lambda}}{6!}\right)^1.$$

Запишемо логарифмічну функцію правдоподібності:

$$\ln L = 600 \ln \lambda - 1000\lambda - \ln(2^{133} \cdot 3^{25} \cdot 5^2).$$

Визначимо похідну логарифмічної функції розподілу за λ :

$$\frac{\partial \ln L}{\partial \lambda} = \frac{600}{\lambda} - 1000.$$

Прирівнявши її до нуля, знайдемо єдиний корінь рівняння правдоподібності:

$$\lambda^* = 0,6.$$

Оскільки друга похідна логарифмічної функції правдоподібності

$$\frac{\partial^2 \ln L}{\partial \lambda^2} = -\frac{600}{\lambda^2} < 0$$

завжди від'ємна, точковою оцінкою максимальної правдоподібності параметра λ розподілу Пуассона буде

$$\lambda^* = 0,6.$$

9.3. Інтервальні оцінки

Якщо об'єм вибірки малий, то точкові оцінки можуть давати значні похибки, тому питання точності оцінок у цьому випадку дуже важливе і використовують інтервальні оцінки.

Інтервальною називають статистичну оцінку, що визначається двома числами, кінцями інтервалів.

Нехай знайдена за даними вибірки статистична оцінка θ^* буде оцінкою невідомого параметра θ .

Різниця між статистичною оцінкою θ^* та її оцінювальним параметром θ , взята за абсолютним значенням, називається *точністю оцінки*, а саме:

$$|\theta^* - \theta| < \delta, \tag{9.12}$$

де δ є точністю оцінки.

Оскільки θ^* є випадковою величиною, то і δ буде випадковою, тому нерівність (9.12) справджуватиметься з певною ймовірністю.

Ймовірність, з якою береться нерівність (9.12), тобто

$$P(|\theta^* - \theta| < \delta) = \gamma \tag{9.13}$$

називають *надійністю*.

Рівність (9.13) можна записати так:

$$P(\theta^* - \delta < \theta < \theta^* + \delta) = \gamma \quad (9.14)$$

Найчастіше число γ задається наперед і, залежно від обставин, воно дорівнює 0,95 або 0,99 або 0,999.

Інтервал $[\theta^* - \delta; \theta^* + \delta]$, що покриває оцінюваний параметр θ генеральної сукупності зі заданою надійністю γ , називають *довірчим*.

Зауваження. Кінці довірчого інтервалу є випадковими величинами.

9.3.1. Побудова довірчого інтервалу для \bar{X}_Γ при відомому значенні σ_Γ із заданою надійністю γ

Нехай ознака X генеральної сукупності має нормальний закон розподілу. Побудуємо довірчий інтервал для \bar{X}_Γ , знаючи числові значення середнього квадратичного відхилення генеральної сукупності σ_Γ ; із заданою надійністю γ . Оскільки \bar{x} як точкова незміщена статистична оцінка для $\bar{X}_\Gamma = M(X)$ має нормальний закон розподілу з числовими характеристиками $M(\bar{x}) = \alpha$, $\sigma(\bar{x}) = \frac{\sigma_\Gamma}{\sqrt{n}}$, то, скориставшись (9.14), дістанемо

$$P(|\bar{x} - \alpha| < \delta) = \gamma. \quad (9.15)$$

Випадкова величина $\bar{x} - \alpha$ має нормальний закон розподілу з числовими характеристиками

$$M(\bar{x} - \alpha) = M(\bar{x}) - \alpha = \alpha - \alpha = 0,$$

$$D(\bar{x} - \alpha) = D(\bar{x}) = \frac{D_\Gamma}{n},$$

$$\sigma(\bar{x}) = \frac{\sigma_\Gamma}{\sqrt{n}}.$$

Тому $\frac{\bar{x} - \alpha}{\frac{\sigma_\Gamma}{\sqrt{n}}}$ матиме нормований нормальний закон розподілу $N(0; 1)$.

Звідси рівність (9.15) можна записати, назначивши $\frac{\delta}{\frac{\sigma_\Gamma}{\sqrt{n}}} = x$ так:

$$P\left(\left|\frac{\bar{x} - \alpha}{\frac{\sigma_\Gamma}{\sqrt{n}}}\right| < x\right) = \gamma \quad (9.16)$$

або

$$P\left(\bar{x} - \frac{x \cdot \sigma_{\Gamma}}{\sqrt{n}} < a < \bar{x} + \frac{x \cdot \sigma_{\Gamma}}{\sqrt{n}}\right) = \gamma.$$

Згідно з формулою нормованого нормального закону

$$P(|X - a| < \delta) = 2\Phi(\delta)$$

для (9.16) вона набирає такого вигляду:

$$P\left(\left|\frac{\bar{x} - a}{\frac{\sigma_{\Gamma}}{\sqrt{n}}}\right| < x\right) = 2\Phi(x) = \gamma. \quad (9.17)$$

З рівності (9.17) знаходимо аргументи x , а саме:

$$2\Phi(x) = \gamma \rightarrow \Phi(x) = \frac{\gamma}{2}.$$

Аргумент x знаходимо за значенням функції Лапласа, яка дорівнює $\frac{1}{2}\gamma$ за таблицею.

Отже, довірчий інтервал дорівнюватиме:

$$\bar{x} - \frac{x \cdot \sigma_{\Gamma}}{\sqrt{n}} < a < \bar{x} + \frac{x \cdot \sigma_{\Gamma}}{\sqrt{n}}. \quad (9.18)$$

Величина $\frac{x \cdot \sigma_{\Gamma}}{\sqrt{n}}$ називається *точністю оцінки* або *похибкою вибірки*.

Приклад 9.5. Вимірявши 40 випадково відібраних після виготовлення деталей, знайшли вибіркочну середню, що дорівнює 15 см. Із надійністю $\gamma = 0,99$ побудувати довірчий інтервал для середньої величини всієї партії деталей, якщо генеральна дисперсія дорівнює $0,99 \text{ см}^2$.

Розв'язання. Для побудови довірчого інтервалу необхідно знати: \bar{x} , σ_{Γ} , n , x .

З умови задачі маємо: $\bar{x} = 15 \text{ см}$, $\sigma_{\Gamma} = \sqrt{D_{\Gamma}} = \sqrt{0,99} = 0,3 \text{ см}$, $n = 40$. Величина x обчислюється з рівняння:

$$\Phi(x) = \frac{\gamma}{2} = \frac{0,99}{2} = 0,495;$$

$\Phi(x) = 0,495 \rightarrow x = 2,58$ (за таблицею значень функції Лапласа).

Знайдемо числові значення кінців довірчого інтервалу:

$$\bar{x} - \frac{x \cdot \sigma_{\Gamma}}{\sqrt{n}} = 15 - \frac{2,58 \cdot 0,3}{\sqrt{40}} = 15 - 0,12 = 14,88 \text{ см};$$

$$\bar{x} + \frac{x \cdot \sigma_{\Gamma}}{\sqrt{n}} = 15 + \frac{2,58 \cdot 0,3}{\sqrt{40}} = 15 + 0,12 = 15,12 \text{ см.}$$

Відтак маємо:

$$14,88 < \bar{X}_{\Gamma} < 15,12.$$

Отже, з надійністю 0,99 (99% гарантії) оцінюваний параметр \bar{X}_{Γ} перебуває всередині інтервалу [14,87; 15,13].

Приклад 9.6. Маємо такі дані про розміри основних фондів (у млн грн) на 30-ти випадково вибраних підприємствах:

4,2; 2,4; 4,9; 6,7; 4,5; 2,7; 3,9; 2,1; 5,8; 4,0;
 2,8; 7,8; 4,4; 6,6; 2,0; 6,2; 7,0; 8,1; 0,7; 6,8;
 9,4; 7,6; 6,3; 8,8; 6,5; 1,4; 4,6; 2,0; 7,2; 9,1.

Побудувати інтервальний статистичний розподіл із довжиною кроку $h = 2$ млн грн.

З надійністю $\gamma = 0,999$ знайти довірчий інтервал для \bar{X}_{Γ} , якщо $\sigma_{\Gamma} = 5$ млн грн.

Розв'язання. Інтервальний статистичний розподіл буде таким:

$h = 2$ млн грн.	2 – 4	4 – 6	6 – 8	8 – 10
n_i	9	7	10	4

Для визначення \bar{x} необхідно побудувати дискретний статистичний розподіл, що має такий вигляд:

x_i^*	3	5	7	9
n_i	9	7	10	4

$$n = \sum n_i = 30.$$

Тоді

$$\bar{x} = \frac{\sum n_i x_i^*}{n} = \frac{3 \cdot 9 + 5 \cdot 7 + 7 \cdot 10 + 9 \cdot 4}{30} = \frac{168}{30} = 5,6 \text{ млн грн.}$$

Для побудови довірчого інтервалу із заданою надійністю $\gamma = 0,999$ необхідно знайти x :

$$\Phi(x) = \frac{\gamma}{2} = \frac{0,999}{2} = 0,4995;$$

$\Phi(x) = 0,4995 \rightarrow x = 3,4$ (за таблицею значень функції Лапласа).

Обчислюємо кінці інтервалу:

$$\bar{x} - \frac{x \cdot \sigma_{\Gamma}}{\sqrt{n}} = 5,6 - \frac{3,4 \cdot 5}{\sqrt{30}} = 5,6 - 3,1 = 2,5 \text{ млн грн.}$$

$$\bar{x} + \frac{x \cdot \sigma_{\Gamma}}{\sqrt{n}} = 5,6 + \frac{3,4 \cdot 5}{\sqrt{30}} = 5,6 + 3,1 = 8,7 \text{ млн грн.}$$

Отже, довірчий інтервал для \bar{X}_{Γ} буде $2,5 < \bar{X}_{\Gamma} < 8,7$.

9.3.2. Побудова довірчого інтервалу для \bar{X}_{Γ} при невідомому значенні σ_{Γ} із заданою надійністю γ

Для малих вибірок, з якими стикаємося, досліджуючи різні ознаки в економіці, техніці чи сільському господарстві, для оцінювання $\bar{X}_{\Gamma} = a$ при невідомому значенні σ_{Γ} неможливо скористатися нормальним законом розподілу. Тому для побудови довірчого інтервалу застосовується випадкова величина

$$t = \frac{\bar{x} - a}{\frac{s}{\sqrt{n}}}, \quad (9.19)$$

що має розподіл Стюдента з $k = n - 1$ ступенями свободи.

Тоді (9.19) набуває такого вигляду:

$$P\left(\left|\frac{\bar{x} - a}{\frac{s}{\sqrt{n}}}\right| < t_{\gamma}\right) = P\left(\bar{x} - \frac{t_{\gamma} \cdot s}{\sqrt{n}} < a < \bar{x} + \frac{t_{\gamma} \cdot s}{\sqrt{n}}\right) = 2 \int_0^{t_{\gamma}} f(t) dt = \gamma,$$

оскільки $f(t)$ для розподілу Стюдента є функцією парною.

Обчисливши за цим статистичним розподілом \bar{x} , S і визначивши за таблицею розподілу Стюдента значення t_γ , будемо довірчий інтервал

$$\bar{x} - \frac{t_\gamma \cdot S}{\sqrt{n}} < a < \bar{x} + \frac{t_\gamma \cdot S}{\sqrt{n}}. \quad (9.20)$$

Тут $t_\gamma(\gamma; k = n - 1)$ обчислюємо за заданою надійністю γ і числом ступенів свободи $k = n - 1$ за таблицею.

Приклад 9.7. Випадково вибрана партія з двадцяти приладів була випробувана щодо терміну безвідказної роботи кожного з них t_i . Результати випробувань наведено у вигляді дискретного статистичного розподілу:

t_i	100	170	240	310	380
n_i	2	5	10	2	1

З надійністю $\gamma = 0,99$ побудувати довірчий інтервал для «а» (середнього часу безвідказної роботи приладу).

Розв'язання. Для побудови довірчого інтервалу необхідно знайти середнє вибіркоче і виправлене середнє квадратичне відхилення.

Обчислимо \bar{x} :

$$\bar{x} = \frac{\sum n_i t_i}{n} = \frac{100 \cdot 2 + 170 \cdot 5 + 240 \cdot 10 + 310 \cdot 2 + 380 \cdot 1}{20} = \frac{4450}{20} = 222,5.$$

Визначимо D_B :

$$D_B = \frac{100^2 \cdot 2 + 170^2 \cdot 5 + 240^2 \cdot 10 + 310^2 \cdot 2 + 380^2 \cdot 1}{20} - 222,5^2 = \frac{1077100}{20} - 49506,25 = 53855 - 49506,25 = 4348,75.$$

Виправлене середнє квадратичне відхилення дорівнюватиме:

$$S = \sqrt{\frac{n}{n-1} D_B} = \sqrt{\frac{20}{20-1} \cdot 4348,75} \approx 67,66.$$

За таблицею значень $\int_0^t f(t) dt = \gamma = 0,99$ розподілу Стюдента за заданою надійністю $\gamma = 0,99$ і числом ступенів свободи $k = n - 1 = 20 - 1 = 19$ знаходимо значення $t_\gamma(\gamma = 0,99; k = 19) = 2,86$.

Обчислимо кінці довірчого інтервалу:

$$\bar{x} - \frac{t_{\gamma} \cdot S}{\sqrt{n}} = 222,5 - \frac{2,86 \cdot 67,66}{\sqrt{0}} = 222,5 - 43,27 = 179,23 \text{ год.};$$

$$\bar{x} + \frac{t_{\gamma} \cdot S}{\sqrt{n}} = 222,5 + \frac{2,86 \cdot 67,66}{\sqrt{0}} = 222,5 + 43,27 = 265,77 \text{ год.}$$

Отже, з надійністю $\gamma = 0,99$ можна стверджувати, що $\bar{X}_{\Gamma} = a$ буде міститися в інтервалі

$$179,23 < a < 265,77.$$

При великих обсягах вибірки, а саме: $n > 30$ на підставі центральної граничної теореми теорії ймовірностей (теореми Ляпунова) розподіл Стюдента наближається до нормального закону. У цьому разі t_{γ} знаходиться за таблицею значень функції Лапласа.

9.3.3. Побудова довірчих інтервалів із заданою надійністю γ для D_{Γ} , σ_{Γ}

У разі, коли ознака X має нормальний закон розподілу, для побудови довірчого інтервалу із заданою надійністю γ для D_{Γ} , σ_{Γ} застосовуємо випадкову величину

$$\chi^2 = \frac{n-1}{\sigma_{\Gamma}^2} S^2, \quad (9.21)$$

що має розподіл χ^2 із $n-1$ ступенями свободи.

Оскільки випадкові події

$$A(\chi_1^2 < \chi^2 < \chi_2^2) \text{ і } B\left(\frac{1}{\chi_2^2} < \frac{1}{\chi^2} < \frac{1}{\chi_1^2}\right)$$

є рівноймовірними, тобто їх імовірності рівні ($P(A) = P(B)$), то маємо:

$$P(\chi_1^2 < \chi^2 < \chi_2^2) = P\left(\frac{1}{\chi_2^2} < \frac{1}{\chi^2} < \frac{1}{\chi_1^2}\right). \quad (9.22)$$

Підставляючи в (9.22) $\chi^2 = \frac{n-1}{\sigma_\Gamma^2} S^2$ дістанемо

$$\begin{aligned} P\left(\frac{1}{\chi_2^2} < \frac{1}{\chi^2} < \frac{1}{\chi_1^2}\right) &= P\left(\frac{1}{\chi_2^2} < \frac{1}{\frac{(n-1)S^2}{\sigma_\Gamma^2}} < \frac{1}{\chi_1^2}\right) = P\left(\frac{1}{\chi_2^2} < \frac{\sigma_\Gamma^2}{(n-1)S^2} < \frac{1}{\chi_1^2}\right) = \\ &= P\left(\frac{(n-1)S^2}{\chi_2^2} < \sigma_\Gamma^2 < \frac{(n-1)S^2}{\chi_1^2}\right) = \gamma. \end{aligned}$$

Отже, довірчий інтервал для $\sigma_\Gamma^2 = D_\Gamma$ матиме вигляд:

$$\frac{(n-1)S^2}{\chi_2^2} < D_\Gamma < \frac{(n-1)S^2}{\chi_1^2}. \quad (9.23)$$

Тоді довірчий інтервал для σ_Γ випливає із (9.23) і буде таким:

$$\frac{S\sqrt{n-1}}{\chi_2} < \sigma_\Gamma < \frac{S\sqrt{n-1}}{\chi_1}. \quad (9.24)$$

Значення χ_1^2, χ_2^2 знаходимо за таблицею згідно з рівностями:

$$P(\chi^2 > \chi_1^2) = 1 - \frac{\alpha}{2}; \quad (9.25)$$

$$P(\chi^2 > \chi_2^2) = \frac{\alpha}{2}; \quad (9.26)$$

$$\alpha = 1 - \gamma.$$

Приклад 9.8. У результаті статистичних досліджень випадкової величини X отримано вибірку об'єму $n = 25$ із таким статистичним розподілом:

x_i	0	1	2	3	4	5	6
n_i	1	3	4	6	5	4	2

З надійністю $\gamma = 0,99$ побудувати довірчі інтервали для D_Γ, σ_Γ .

Розв'язання. Для побудови довірчих інтервалів необхідно знайти значення S^2, S .

Обчислимо значення \bar{x} :

$$\bar{x} = \frac{\sum x_i n_i}{n} = \frac{0 \cdot 1 + 1 \cdot 3 + 2 \cdot 4 + 3 \cdot 6 + 4 \cdot 5 + 5 \cdot 4 + 6 \cdot 2}{25} = \frac{81}{25} = 3,24.$$

Обчислимо D_B :

$$D_B = \frac{0^2 \cdot 1 + 1^2 \cdot 3 + 2^2 \cdot 4 + 3^2 \cdot 6 + 4^2 \cdot 5 + 5^2 \cdot 4 + 6^2 \cdot 2}{25} - 3,24^2 = \frac{325}{25} - 10,498 = \\ = 13 - 10,498 = 2,502.$$

Виправлена дисперсія і виправлене середнє квадратичне відхилення дорівнюватимуть:

$$S^2 = \frac{n}{n-1} \cdot D_B = \frac{25}{25-1} \cdot 2,502 = 2,634;$$

$$S = \sqrt{S^2} = \sqrt{2,634} \approx 1,623.$$

Оскільки $\alpha = 1 - \gamma = 1 - 0,99 = 0,01$, то згідно з (9.25), (9.26) знаходимо значення χ_1^2 , χ_2^2 , а саме:

$$P(\chi^2 > \chi_1^2) = 1 - \frac{0,01}{2} = 0,995;$$

$$P(\chi^2 > \chi_2^2) = \frac{0,01}{2} = 0,005.$$

За таблицею знаходимо:

$$\chi_1^2(0,995; 24) = 10,9;$$

$$\chi_2^2(0,005; 24) = 45,5.$$

Обчислимо кінці довірчого інтервалу для D_Γ :

$$\frac{(n-1)S^2}{\chi_2^2} = \frac{25-1}{45,5} \cdot 2,634 = 1,389;$$

$$\frac{(n-1)S^2}{\chi_1^2} = \frac{25-1}{10,9} \cdot 2,634 = 5,799.$$

Отже, довірчий інтервал для D_Γ буде таким:

$$1,389 < D_\Gamma < 5,799.$$

Довірчий інтервал для σ_Γ становить

$$1,179 < \sigma_\Gamma < 2,408.$$

Запитання для самоконтролю

1. Що називається статистичною оцінкою?
2. Що називається точковою статистичною оцінкою?
3. Що таке незміщена точкова статистична оцінка?
4. Що таке зміщена точкова статистична оцінка?
5. Що називають ефективною точковою статистичною оцінкою?
6. Що називають обґрунтованою точковою статистичною оцінкою?
7. Яку статистичну оцінку називають спроможною?
8. Яку статистичну оцінку називають сильно спроможною?
9. У чому сутність методу моментів?
10. У чому сутність методу найменших квадратів?
11. У чому сутність методу максимальної правдоподібності?
12. Що називається функцією правдоподібності дискретної випадкової?
13. Що називається логарифмічною функцією правдоподібності?
14. Що називається функцією правдоподібності неперервної випадкової?
15. Який вигляд мають рівняння правдоподібності.
16. Що є точковою незміщеною статистичною оцінкою для \bar{X}_T ?
17. Що означає точкова незміщена статистична оцінка для D_T ?
18. Що називається виправленою дисперсією?
19. Що називається виправленим середнім квадратичним відхиленням?
20. Який закон розподілу ймовірностей має випадкова величина \bar{x} ?
21. Який закон розподілу має випадкова величина $\frac{n-1}{\sigma_T^2} S^2$?
22. Який закон розподілу ймовірностей має випадкова величина $\frac{S\sqrt{n-1}}{\sigma_T}$?
23. Визначення інтервальної статистичної оцінки для параметрів генеральної сукупності.
24. Що називають точністю оцінки?
25. Що називають надійністю оцінки?
26. Що називають довірчим інтервалом?
27. Який закон розподілу ймовірностей має випадкова величина $\frac{\bar{x}-\alpha}{\frac{\sigma_T}{\sqrt{n}}}$?
28. Як побудувати довірчий інтервал із заданою надійністю γ при відомому значенні σ_T ?
29. Який закон розподілу ймовірностей має випадкова величина $\frac{\bar{x}-\alpha}{\frac{s}{\sqrt{n}}}$?
30. Як побудувати довірчий інтервал для \bar{X}_T із заданою надійністю γ при невідомому значенні σ_T ?
31. Як побудувати довірчий інтервал для \bar{X}_T із заданою надійністю γ при невідомому значенні σ_T і при обсягах вибірки $n > 30$?
32. Як побудувати довірчий інтервал із надійністю γ для D_T , σ_T , використовуючи розподіл χ^2 ?

Задачі для самостійного розв'язування

1. Із нормально розподіленої сукупності з $D(X) = \sigma^2$ зроблено вибірку обсягом n . Знайти оцінку для $M(X)$. Перевірити її на незміщеність, ефективність та обґрунтованість.

2. Із нормально розподіленої сукупності з $M(X) = a$ зроблено вибірку обсягом n . Знайти оцінку для $D(X)$. Перевірити її на незміщеність, ефективність та обґрунтованість.

3. З генеральної сукупності отримано деяку вибірку:

x_i	0,001	0,003	0,005	0,007	0,009	0,011	0,013	0,015
n_i	5	30	34	13	8	5	3	2

Знайти зміщену оцінку генеральної дисперсії.

4. Випадкова величина X розподілена за нормальним розподілом з параметрами a, σ . Таблиця містить дані про зріст 1000 учнів старших класів:

Зріст, см	143-146	146-149	149-152	152-155	155-158	158-161	161-164	164-167	167-170	170-173	173-176	176-179	179-182	182-185	185-188
Число учнів	1	2	8	26	65	120	181	201	170	120	64	27	10	3	1

Знайти методом моментів точкову оцінку невідомих параметрів a, σ нормального розподілу.

5. Випадкова величина X розподілена за нормальним розподілом. Знайти методом моментів точкову оцінку невідомих параметрів.

$[x_i; x_{i+1})$	$[-1; 0)$	$[0; 1)$	$[1; 2)$	$[2; 3)$	$[3; 4)$	$[4; 5)$	$[5; 6)$
n_i	1	2	5	11	6	3	2

6. Випадкова величина X розподілена за розподілом Пуассона. Знайти методом моментів точкову оцінку невідомих параметрів.

x_i	0	1	2	3	4	5	6	7	8	9
n_i	4	11	17	27	16	15	5	2	2	1

7. Випадкова величина X розподілена за біноміальним розподілом. Знайти методом моментів точкову оцінку невідомих параметрів.

x_i	0	1	2	3	4	5	6	7	8
n_i	3	10	16	34	17	11	5	3	1

8. Визначити мінімальний обсяг вибірки n для того, щоб із надійністю $\gamma = 0,98$ можна було дістати оцінку математичного сподівання нормально розподіленої сукупності $M(X) = 0,2$, якщо середнє квадратичне відхилення в генеральній сукупності $\sigma = 1,5$ і оцінка знаходиться за допомогою вибіркової середньої величини.

9. Знайти з надійністю $0,95$ довірчий інтервал оцінки математичного сподівання a нормально розподіленої ознаки X генеральної сукупності, якщо відомі вибіркова середня \bar{x} , об'єм вибірки n та середнє квадратичне відхилення σ генеральної сукупності.

а) $\bar{x} = 14$; $n = 25$, $\sigma = 5$; б) $\bar{x} = 10,2$; $n = 16$, $\sigma = 4$;

в) $\bar{x} = 16,8$; $n = 25$, $\sigma = 5$; г) $\bar{x} = 2000$; $n = 1600$, $\sigma = 40$.

10. У будинку відпочинку випадковим способом було відібрано 20 осіб і виміряно їх зріст x_i . Здобуті результати наведено у вигляді інтервального статистичного розподілу:

x_i , см	165,5 – 170,5	170,5 – 175,5	175,5 – 180,5	180,5 – 185,5
n_i	4	6	8	2

Із надійністю $\gamma = 0,99$ побудувати довірчий інтервал для \bar{X}_Γ , якщо $\sigma_\Gamma = 2$.

11. 28 однотипних приладів були випробувані щодо їх безвідмовної роботи x_i . Результати вимірювання наведено в таблиці:

x_i , год	100	110	120	130	140	150
n_i	10	6	5	4	2	1

З надійністю $\gamma = 0,99$ побудувати довірчий інтервал для \bar{X}_Γ .

12. За вибіркою з нормального розподілу з надійністю $\gamma = 0,95$ знайти інтервальні оцінки для \bar{X}_Γ , якщо

а) $\sigma_\Gamma = 4$;

б) σ – невідоме.

x_i	-2	1	2	3	4	5
n_i	4	3	4	6	2	1

Знайти також довірчий інтервал для D_Γ, σ_Γ якщо $\gamma = 0,9$.

13. Під час вибіркового аналізу максимального завантаження літака при певній кількості пального, дістали такі дані, т: 78; 80; 85; 88; 89; 90; 92; 93; 92; 85; 93; 89; 80; 90; 85; 89.

Припускаючи, що максимальне завантаження літака має нормальний розподіл, знайти довірчий інтервал для дисперсії та середнього квадратичного відхилення з надійністю $\gamma = 0,95$.

14. Побудувати довірчий інтервал з надійністю $\gamma = 0,96$ для невідомої дисперсії генеральної сукупності спостережень випадкової величини X , розподіленої нормально, за вибіркою із обсягом $n = 20$ та відомою виправленою вибірковою дисперсією $S^2 = 10$.

Розділ 10

ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ

Інформація, яку здобувають на підставі вибірки, реалізованої із генеральної сукупності, може бути використана для формулювання певних суджень про всю генеральну сукупність. Наприклад, розпочавши виготовляти покриття нового типу для автомобілів, відбирають певну кількість цих покриттів і піддають їх певним тестам.

За результатами тестів можна зробити висновок про те, чи кращі нові покриття від покриттів старого типу, чи ні. А це, своєю чергою, дає підставу для ухвалення рішення: виготовляти їх чи ні.

Такі рішення називають *статистичними*.

Статистичні рішення мають імовірнісний характер, тобто завжди існує ймовірність того, що ці рішення будуть помилковими.

Головна цінність ухвалення статистичних рішень полягає в тому, що в межах імовірнісних категорій можна об'єктивно виміряти ступінь ризику, що відповідає тому чи іншому рішення.

Будь-які статистичні висновки, здобуті на підставі обробки вибірки, називають *статистичними гіпотезами*.

Статистичні гіпотези про значення параметрів ознак генеральної сукупності називають *параметричними*.

Наприклад, висувається статистична гіпотеза про числові значення генеральної середньої \bar{x}_G , генеральної дисперсії D_G , генерального середнього квадратичного відхилення σ_G та ін.

Статистичні гіпотези, що висуваються на підставі обробки вибірки про закон розподілу ознаки генеральної сукупності, називаються *непараметричними*. Наприклад, на підставі обробки вибірки може бути висунута гіпотеза, що ознака генеральної сукупності має нормальний закон розподілу, експоненціальний закон та ін.

10.1. Статистична перевірка гіпотез

Статистичною називається гіпотеза про вигляд невідомого розподілу випадкової величини або про параметри відомих розподілів.

Нульовою (основною) називають висунуту гіпотезу H_0 .

Зміст нульової гіпотези записується так:

$$H_0 : \bar{x}_T = a;$$

Конкуруючою (альтернативною) називають гіпотезу H_a , яка суперечить нульовій.

Кожній нульовій гіпотезі можна протиставити кілька альтернативних (конкуруючих) гіпотез.

Приміром, нульова гіпотеза стверджує: $H_0: \bar{x}_T = a$, а альтернативна гіпотеза – $H_a: \bar{x}_T > a$, тобто заперечує твердження нульової.

Простою називають гіпотезу, яка містить лише одне твердження.

Складною називають гіпотезу, яка складається зі скінченного або нескінченного числа простих гіпотез.

Складна статистична гіпотеза є неоднозначною. Вона може стверджувати, що значення параметра генеральної сукупності належить певній області ймовірних значень, яка може бути дискретною і неперервною.

Наприклад:

$$H_0 : \bar{x}_T \in [2; 2,1; 2,2].$$

Перевірка гіпотези здійснюється за даними вибірки, тобто статистичними методами. Тому перевірку гіпотези за даними вибірки називають статистичною.

У результаті статистичної перевірки правильності основної гіпотези H_0 за результатами вибірки у двох випадках може бути ухвалено неправильне рішення, тобто можуть бути допущені помилки двох видів.

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

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

Ймовірність здійснити помилку першого роду позначають α і називають **рівнем значущості**.

Найчастіше рівень значущості приймають рівним 0,05 або 0,01. Якщо прийнято рівень значущості рівним 0,05, то це означає, що в п'яти випадках із 100 ми ризикуємо одержати похибку першого роду (відкинути правильну гіпотезу).

Ймовірність здійснити помилку другого роду позначають β .

Для перевірки нульової гіпотези вибирається деяка випадкова величина K , розподіл якої відомий, і яка називається **статистичним критерієм** (або просто критерієм) перевірки нульової гіпотези.

Спостережним значенням критерію узгодження називають значення відповідного критерію, обчислене за даними вибірки.

Критичною областю називають множину значень критерію, при яких нульову гіпотезу відхиляють.

Областю прийняття гіпотези (областю допустимих значень) називають множину значень критерію, при яких нульову гіпотезу приймають.

Основний принцип перевірки статистичних гіпотез такий: якщо спостережуване значення критерію належить критичній області, нульову гіпотезу відхиляють; якщо спостережуване значення критерію належить області прийняття гіпотези, гіпотезу приймають.

Критичними точками $k_{кр}$ називають точки, які відокремлюють критичну область від області прийняття гіпотези.

Правосторонньою називають критичну область, яка визначається нерівністю

$$K > k_{кр} (k_{кр} > 0). \quad (10.1)$$

Лівосторонньою називають критичну область, яка визначається нерівністю

$$K < k_{кр} (k_{кр} < 0). \quad (10.2)$$

Двосторонньою називають критичну область, яка визначається нерівностями

$$K < k_1, K > k_2 (k_2 > k_1). \quad (10.3)$$

Зокрема, якщо критичні точки симетричні відносно нуля, то двостороння критична область визначається нерівностями

$$K < -k_{кр}, K > k_{кр} (k_{кр} > 0), \quad (10.4)$$

або рівносильною їй нерівністю

$$|K| > k_{кр}. \quad (10.5)$$

Лівобічна і правобічна області визначаються однією критичною точкою, двобічна критична область – двома критичними точками, симетричними щодо нуля.

Для знаходження правосторонньої критичної області достатньо знайти критичну точку $k_{кр}$. Для цього задають малу ймовірність α ; число α називається-

ся рівнем значущості критерію. Далі знаходять критичну точку, виходячи з умови, що при справедливості нульової гіпотези виконується рівність

$$P\{K > k_{кр}\} = \alpha.$$

Для кожного критерію є відповідні таблиці, з яких знаходять критичну точку, що задовольняє цю формулу.

Аналогічно знаходять лівосторонню критичну область.

Двосторонню критичну область знаходять за умови справедливості нульової гіпотези з рівності

$$P\{K < k_1\} + P\{K > k_2\} = \alpha. \quad (10.6)$$

Якщо розподіл критерію симетричний щодо нуля, то

$$P\{K < k_1\} = P\{K > k_2\}.$$

Врахувавши рівність (10.6), одержимо

$$P\{K > k_{кр}\} = \frac{\alpha}{2}. \quad (10.7)$$

Тоді двосторонню критичну область знаходять з рівності (10.7). Критичні точки знаходять з відповідних таблиць.

При знаходженні критичної області доцільно враховувати потужність критерію.

Потужністю критерію називають імовірність належності критерію критичній області за умови, що правильна альтернативна гіпотеза.

Тобто потужність критерію є імовірність того, що основна гіпотеза буде відхилена, якщо альтернативна гіпотеза правильна.

Якщо рівень значущості вже обрано, то критичну область доцільно будувати так, щоб потужність критерію була максимальною. Виконання цієї вимоги забезпечує мінімальну імовірність похибки другого роду.

Зауваження. Єдиний спосіб одночасного зменшення ймовірностей похибок першого та другого роду це є збільшення об'єму вибірки.

Загальний алгоритм перевірки правильності нульової гіпотези

Для перевірки правильності H_0 задається так званий рівень значущості α .

Пропонується такий алгоритм перевірки правильності H_0 :

1. Сформулювати H_0 й одночасно альтернативну гіпотезу H_α .

2. Вибрати статистичний критерій, який відповідав би сформульованій нульовій гіпотезі.

3. Визначити допустиму імовірність похибки першого роду, тобто рівень значущості α ;

4. Знайти за відповідною таблицею критичну область (критичну точку) для обраної статистичної характеристики.

До критичної області належать такі значення статистичної характеристики, за яких гіпотеза H_0 відхиляється на користь альтернативної гіпотези H_α .

Наголосимо, що між рівнем значущості (та критичною областю існує такий зв'язок: якщо гіпотеза H_0 правильна, то з імовірністю значення вибіркової функції будуть належати критичній області.

10.2. Перевірка рівності вибіркового середнього гіпотетичному генеральному середньому

Нехай із генеральної сукупності з відомою дисперсією σ^2 отримано вибірку об'єму n .

Правило 1. Для того, щоб при заданому рівні значущості α перевірити нульову гіпотезу $H_0: a = a_0$ (про рівність генерального середнього a нормальної сукупності з відомою дисперсією σ^2 гіпотетичному (прогнозованому) значенню a_0) при конкуруючій гіпотезі $H_\alpha: a \neq a_0$, потрібно обчислити спостережуване значення критерію

$$U_{\text{спост}} = \frac{(\bar{x} - a_0)\sqrt{n}}{\sigma} \quad (10.8)$$

і за таблицею функції Лапласа знайти критичну точку $u_{\text{кр}}$ двосторонньої критичної області з рівності

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

Якщо $|U_{\text{спост}}| < u_{\text{кр}}$, немає підстав відхилити нульову гіпотезу. Якщо $|U_{\text{спост}}| > u_{\text{кр}}$, нульову гіпотезу відхиляють.

Правило 2. При конкуруючій гіпотезі $H_\alpha: a > a_0$ критичну точку правосторонньої критичної області знаходять з рівності

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

Якщо $U_{\text{спост}} < u_{\text{кр}}$, немає підстав відхиляти нульову гіпотезу. Якщо $U_{\text{спост}} > u_{\text{кр}}$, то нульову гіпотезу відхиляють.

Правило 3. При конкуруючій гіпотезі $H_a: a < a_0$ спочатку знаходять допоміжну критичну точку $u_{\text{кр}}$ за правилом 2, а потім вважають межею лівосторонньої критичної області

$$u'_{\text{кр}} = -u_{\text{кр}}.$$

Якщо $U_{\text{спост}} > -u_{\text{кр}}$, то немає підстав відхиляти нульову гіпотезу. Якщо $U_{\text{спост}} < -u_{\text{кр}}$, то нульову гіпотезу відхиляють.

Потужність критерію перевірки основної гіпотези $H_0: a = a_0$ про рівність генерального середнього a гіпотетичному значенню a_0 при відомому середньому квадратичному відхиленні σ знаходять залежно від вигляду альтернативної гіпотези.

При альтернативній гіпотезі $H_a: a > a_0$ для гіпотетичного значення генерального середнього $a = a_1 > a_0$ потужність правостороннього критерію

$$1 - \beta = \frac{1}{2} - \Phi(u_{\text{кр}} - \lambda),$$

де $u_{\text{кр}}$ знаходять з рівності $\Phi(u_{\text{кр}}) = \frac{1}{2} - \alpha$, $\lambda = \frac{(a_1 - a_0)\sqrt{n}}{\sigma}$. При різних значеннях a_1 функція потужності одностороннього критерію

$$\pi_1(a_1) = \frac{1}{2} - \Phi(u_{\text{кр}} - \lambda).$$

При альтернативній гіпотезі $H_a: a \neq a_0$ для гіпотетичного значення генерального середнього $a = a_1$ потужність двостороннього критерію

$$1 - \beta = 1 - [\Phi(u_{\text{кр}} - \lambda) + \Phi(u_{\text{кр}} + \lambda)],$$

де $u_{\text{кр}}$ знаходять з рівності $\Phi(u_{\text{кр}}) = \frac{1 - \alpha}{2}$, $\lambda = \frac{(a_1 - a_0)\sqrt{n}}{\sigma}$. При різних значеннях a_1 функція потужності одностороннього критерію

$$\pi_1(a_1) = 1 - [\Phi(u_{\text{кр}} - \lambda) + \Phi(u_{\text{кр}} + \lambda)].$$

Приклад 10.1. Із нормальної генеральної сукупності з відомим середнім квадратичним відхиленням $\sigma = 4,8$ отримано вибірку об'єму $n = 144$, за якою

знайдено вибіркоче середнє $\bar{x} = 16$. Потрібно при рівні значущості $\alpha = 0,05$ перевірити нульову гіпотезу $H_0 : a = a_0 = 15$ при конкуруючій гіпотезі:

А. $H_a : a \neq 15$.

Б. $H_a : a > 15$.

В. $H_a : a < 15$.

Крім того, необхідно знайти потужність правостороннього та двохстороннього критеріїв.

Розв'язання. Обчислимо спочатку спостережуване значення критерію

$$U_{\text{спост}} = \frac{(\bar{x} - a_0)\sqrt{n}}{\sigma} = \frac{(16 - 15)\sqrt{144}}{4.8} = 2,5.$$

А. Skorистаємося правилом 1. Знайдемо критичну точку з рівності

$$\Phi(u_{\text{кр}}) = \frac{1 - \alpha}{2} = \frac{1 - 0,05}{2} = 0,475.$$

За таблицею функції Лапласа визначимо критичну точку:

$$u_{\text{кр}} \approx 1,96.$$

Оскільки $|U_{\text{спост}}| > u_{\text{кр}}$, основна гіпотеза відхиляється. Тобто, вибіркоче та генеральні середні суттєво відрізняються.

Б. Skorистаємося правилом 2. Знайдемо критичну точку з рівності:

$$\Phi(u_{\text{кр}}) = \frac{1}{2} - \alpha = \frac{1}{2} - 0,05 = 0,45.$$

За таблицею функції Лапласа визначимо критичну точку:

$$u_{\text{кр}} \approx 1,64.$$

Оскільки $U_{\text{спост}} > u_{\text{кр}}$, основна гіпотеза відхиляється. Інакше кажучи, вибіркоче та гіпотетичне генеральне середні суттєво відрізняються.

В. Skorистаємося правилом 3. Критична точка буде такою ж, як і в пункті б), але з протилежним знаком:

$$u_{\text{кр}} \approx -1,64.$$

Оскільки $U_{\text{спост}} > u_{\text{кр}}$, то немає підстав відхиляти нульову гіпотезу. Тобто вибіркоче та гіпотетичне генеральне середні несуттєво відрізняються.

Тепер знайдемо потужності правостороннього та двохстороннього критеріїв. Нагадаємо, що критичні точки у цих випадках різні та дорівнюють 1,64 і 1,96 відповідно.

Знайдемо параметр λ , який входить в обидва рівняння для визначення потужності критеріїв:

$$\lambda = \frac{(\bar{x}-a_0)\sqrt{n}}{\sigma} = \frac{(16-15)\sqrt{144}}{4,8} = 2,5.$$

Отже, потужності відповідно правостороннього та двохстороннього критеріїв будуть такі:

$$\begin{aligned} \pi_1^{\text{правост}}(16) &= \frac{1}{2} - \Phi(u_{\text{кр}} - \lambda) = \frac{1}{2} - \Phi(1,64 - 2,5) = \frac{1}{2} + \Phi(0,86) = \\ &= 0,5 + 0,3051 = 0,8051. \end{aligned}$$

$$\begin{aligned} \pi_1^{\text{двост}}(16) &= 1 - [\Phi(u_{\text{кр}} - \lambda) + \Phi(u_{\text{кр}} + \lambda)] = \\ &= 1 - [\Phi(1,96 - 2,5) + \Phi(1,96 + 2,5)] = \\ &= 1 - [-\Phi(0,54) + \Phi(4,46)] = 1 + 0,2054 - 0,5 = 0,7054. \end{aligned}$$

Тобто ймовірності того, що нульова гіпотеза буде відхилена, якщо правильною є конкуруюча гіпотеза, дорівнюють 0,8051 і 0,7054 відповідно для правостороннього та двохстороннього критеріїв.

Нехай із генеральної сукупності отримано вибірку об'єму n з невідомою дисперсією.

Якщо дисперсія генеральної сукупності невідома, то як критерій перевірки основної гіпотези беруть випадкову величину

$$T = \frac{(\bar{x}-a_0)\sqrt{n}}{s}, \quad (10.9)$$

де $s = \sqrt{\frac{n \sum_i n_i x_i^2 - (\sum_i n_i x_i)^2}{n(n-1)}}$ – виправлене вибіркове середнє квадратичне відхилення. Величина T має розподіл Стюдента з $k = n - 1$ ступенями вільності.

Правило 1. Для того, щоб при заданому рівні значущості α перевірити нульову гіпотезу $H_0 : a = a_0$ про рівність невідомого генерального середнього a нормальної сукупності з невідомою дисперсією гіпотетичному (прогно-

зованому) значенню a_0 при конкуруючій гіпотезі $H_a: a \neq a_0$ потрібно обчислити спостережуване значення критерію

$$T_{\text{спост}} = \frac{(\bar{x} - a_0)\sqrt{n}}{s}$$

і за таблицею критичних точок розподілу Стьюдента при заданому рівні значущості α , розміщеному у верхній частині таблиці, і кількості ступенів вільності $k = n - 1$ знайти критичну точку $t_{\text{двост кр}}(\alpha; k)$.

Якщо $|T_{\text{спост}}| < t_{\text{двост кр}}$, то немає підстав відхиляти нульову гіпотезу. Якщо $|T_{\text{спост}}| > t_{\text{двост кр}}$ – нульову гіпотезу відхиляють.

Правило 2. При конкуруючій гіпотезі $H_a: a > a_0$ за рівнем значущості α , розміщеним у нижній частині таблиці розподілу Стьюдента, і при кількості ступенів вільності $k = n - 1$ знаходять критичну точку правосторонньої критичної області $t_{\text{правост кр}}(\alpha; k)$.

Якщо $T_{\text{спост}} < t_{\text{правост кр}}$, немає підстав відхиляти нульову гіпотезу. Якщо $T_{\text{спост}} > t_{\text{правост кр}}$, нульову гіпотезу відхиляють.

Правило 3. При конкуруючій гіпотезі $H_a: a < a_0$ спочатку знаходять допоміжну критичну точку (за правилом 2) $t_{\text{правост кр}}(\alpha; k)$ і вважають межею лівосторонньої критичної області $t_{\text{лівост кр}}(\alpha; k) = -t_{\text{правост кр}}(\alpha; k)$.

Якщо $T_{\text{спост}} > -t_{\text{правост кр}}$, немає підстав відхиляти нульову гіпотезу. Якщо $T_{\text{спост}} < -t_{\text{правост кр}}$, нульову гіпотезу відхиляють.

Приклад 10.2. Проведено 10 незалежних експериментів над випадковою величиною X , що має нормальний закон розподілу з невідомими значеннями a , σ . Наслідки експериментів подано у вигляді статистичного ряду:

x_i	2,5	2	-2,3	1,9	-2,1	2,4	2,3	-2,5	1,5	-1,7
n_i	1	1	1	1	1	1	1	1	1	1

При рівні значущості $\alpha = 0,001$ перевірити правильність нульової гіпотези $H_0: a = 0,9$, при альтернативній гіпотезі $H_a: a < 0,9$.

Розв'язання. Запишемо статистичний ряд у вигляді статистичного розподілу й обчислимо \bar{x} , s :

x_i	-2,5	-2,3	-2,1	-1,7	1,5	1,9	2	2,3	2,4	2,5
n_i	1	1	1	1	1	1	1	1	1	1

$$n = 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 = 10;$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^m n_i x_i = \frac{1}{10} (-2,5 \cdot 1 + (-2,3) \cdot 1 + (-2,1) \cdot 1 + (-1,7) \cdot 1 + 1,5 \cdot 1 + 1,9 \cdot 1 + 2 \cdot 1 + 2,3 \cdot 1 + 2,4 \cdot 1 + 2,5 \cdot 1) = 0,4;$$

$$D_B = \frac{1}{n} \sum_{i=1}^m n_i x_i^2 - \left(\frac{1}{n} \sum_{i=1}^m n_i x_i \right)^2 = \frac{1}{10} \left((-2,5)^2 \cdot 1 + (-2,3)^2 \cdot 1 + (-2,1)^2 \cdot 1 + (-1,7)^2 \cdot 1 + 1,5^2 \cdot 1 + 1,9^2 \cdot 1 + 2^2 \cdot 1 + 2,3^2 \cdot 1 + 2,4^2 \cdot 1 + 2,5^2 \cdot 1 \right) - 0,4^2 = 4,6 - 0,16 = 4,44;$$

$$S^2 = \frac{n}{n-1} D_B = \frac{10}{9} \cdot 4,44 = 4,933;$$

$$s = \sqrt{S^2} = \sqrt{4,933} \approx 2,22.$$

При альтернативній гіпотезі $H_a: a < 0,9$ будується лівобічна критична область. Для цього необхідно знайти критичну точку

$$t_{\text{лівост кр}}(\alpha; k) = - t_{\text{правост кр}}(\alpha; k);$$

$$t_{\text{правост кр}}(\alpha; k) = t(0,001; 9) = 4,78;$$

$$t_{\text{лівост кр}}(\alpha; k) = -4,78.$$

Обчислимо спостережуване значення критерію:

$$T_{\text{спост}} = \frac{(\bar{x} - a_0) \sqrt{n}}{s} = \frac{(0,4 - 0,9) \cdot \sqrt{10}}{2,22} = - \frac{0,5 \cdot \sqrt{10}}{2,22} \approx -0,712.$$

Оскільки $T_{\text{спост}} > - t_{\text{правост кр}}$, то немає підстав відхилити $H_0: a = 0,9$.

Приклад 10.3. Реалізувавши вибірку з генеральної сукупності, ознака якої X має нормальний закон розподілу, дістали статистичний розподіл:

x_i	6	7	8	9	10	11	12	13	14
n_i	1	3	6	8	6	6	5	3	2

При рівні значущості $\alpha = 0,01$ перевірити правильність нульової гіпотези

$H_0: a = 8$, якщо альтернативна гіпотеза $H_a: a \neq 8$.

Розв'язання. Обчислимо значення \bar{x} , s :

$$n = 1 + 3 + 6 + 8 + 6 + 6 + 5 + 3 + 2 = 40;$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^m n_i x_i = \frac{1}{40} (6 \cdot 1 + 7 \cdot 3 + 8 \cdot 6 + 9 \cdot 8 + 10 \cdot 6 + 11 \cdot 6 + 12 \cdot 5 + 13 \cdot 3 + 14 \cdot 2) = \frac{6+21+48+72+60+60+60+39+28}{40} = 10;$$

$$D_B = \frac{1}{n} \sum_{i=1}^m n_i x_i^2 - \left(\frac{1}{n} \sum_{i=1}^m n_i x_i \right)^2 = \frac{1}{40} (6^2 \cdot 1 + 7^2 \cdot 3 + 8^2 \cdot 6 + 9^2 \cdot 8 + 10^2 \cdot 6 + 11^2 \cdot 6 + 12^2 \cdot 5 + 13^2 \cdot 3 + 14^2 \cdot 2) - 10^2 = \frac{36+147+384+648+600+726+720+507+392}{100} - 100 = 104 - 100 = 4;$$

$$S^2 = \frac{n}{n-1} D_B = \frac{40}{39} \cdot 4 = 4,103;$$

$$s = \sqrt{S^2} = \sqrt{4,103} \approx 2,03.$$

При альтернативній гіпотезі $H_a: a \neq 8$ будемо двобічну критичну область. Враховуючи, що σ_T є невідомою величиною, для побудови цієї області беремо статистичний критерій.

$$t_{\text{двост кр}}(\alpha; k) = t_{\text{двост кр}}(0,01; 39) = 2,7.$$

Обчислимо спостережуване значення критерію:

$$T_{\text{спост}} = \frac{(\bar{x} - a_0) \sqrt{n}}{s} = \frac{(10-8) \sqrt{40}}{2,03} = \frac{2 \cdot 2 \sqrt{10}}{2,03} \approx 6,23.$$

Оскільки $|T_{\text{спост}}| > t_{\text{двост кр}}$, то немає підстав приймати нульову гіпотезу, тобто гіпотеза $H_0: a = 8$ відхиляється.

10.3. Перевірка рівності виправленої вибіркової дисперсії генеральній дисперсії

Нехай з нормальної генеральної сукупності отримано вибірку об'єму n , для якої знайдено виправлену вибірку дисперсію S^2 .

Правило 1. Для того, щоб при заданому рівні значущості α перевірити нульову гіпотезу $H_0: \sigma^2 = \sigma_0^2$ про рівність невідомої генеральної дисперсії σ^2

гіпотетичному (прогнозованому) значенню σ_0^2 при конкуруючій гіпотезі $H_a: \sigma^2 > \sigma_0^2$, потрібно обчислити спостережуване значення критерію

$$\chi_{\text{спост}}^2 = \frac{(n-1)S^2}{\sigma_0^2} \quad (10.10)$$

і за таблицею критичних точок розподілу χ^2 при заданому рівні значущості α і кількості ступенів вільності $k = n - 1$ знайти критичну точку $\chi_{\text{крит}}^2(\alpha; k)$.

Якщо $\chi_{\text{спост}}^2 < \chi_{\text{крит}}^2$, то немає підстав відхиляти нульову гіпотезу. Якщо $\chi_{\text{спост}}^2 > \chi_{\text{крит}}^2$ – нульову гіпотезу відхиляють.

Правило 2. При конкуруючій гіпотезі $H_a: \sigma^2 \neq \sigma_0^2$ знаходять ліву $\chi_{\text{лів крит}}^2(1 - \frac{\alpha}{2}; k)$ і праву $\chi_{\text{прав крит}}^2(\frac{\alpha}{2}; k)$ критичні точки.

Якщо $\chi_{\text{лів крит}}^2 < \chi_{\text{спост}}^2 < \chi_{\text{прав крит}}^2$, то немає підстав відхиляти нульову гіпотезу.

Якщо $\chi_{\text{спост}}^2 < \chi_{\text{лів крит}}^2$ або $\chi_{\text{спост}}^2 > \chi_{\text{прав крит}}^2$ – нульову гіпотезу відхиляють.

Правило 3. При конкуруючій гіпотезі $H_a: \sigma^2 < \sigma_0^2$ знаходять критичну точку $\chi_{\text{крит}}^2(1 - \alpha; k)$.

Якщо $\chi_{\text{спост}}^2 > \chi_{\text{крит}}^2$, то немає підстав відхиляти нульову гіпотезу. Якщо $\chi_{\text{спост}}^2 < \chi_{\text{крит}}^2$ – нульову гіпотезу відхиляють.

Приклад 10.4. Із нормальної генеральної сукупності отримано вибірку об'єму $n = 20$, для якої знайдено вибіркочну виправлену дисперсію $S^2 = 2,7$. Перевірити при рівні значущості $\alpha = 0,05$ основну гіпотезу $H_0: \sigma^2 = \sigma_0^2 = 3$, якщо альтернативна гіпотеза:

А. $H_a: \sigma^2 > 3$;

Б. $H_a: \sigma^2 \neq 3$;

В. $H_a: \sigma^2 < 3$.

Розв'язання. Знайдемо спостережуване значення критерію:

$$\chi_{\text{спост}}^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(20-1) \cdot 2,7}{3} = 17,1.$$

А. Скористаємося правилом 1. За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0,05$ і кількості ступенів вільності

$$k = n - 1 = 20 - 1 = 19$$

знаходимо критичну точку:

$$\chi_{\text{крит}}^2(0,05; 19) = 30,14.$$

Оскільки $\chi_{\text{спост}}^2 < \chi_{\text{крит}}^2$, то немає підстав відхилити основну гіпотезу про рівність генеральної дисперсії гіпотетичному значенню. Інакше кажучи, різниця між виправленою вибірковою дисперсією $S^2 = 2,7$ і гіпотетичною генеральною дисперсією $\sigma_0^2 \leq 3$ незначуща.

Б. Скористаємося правилом 2. За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0,05$ і кількості ступенів вільності

$$k = n - 1 = 20 - 1 = 19$$

знаходимо критичні точки:

$$\chi_{\text{лів крит}}^2\left(1 - \frac{\alpha}{2}; k\right) = \chi_{\text{лів крит}}^2(0,975; 19) = 8,91;$$

$$\chi_{\text{прав крит}}^2\left(\frac{\alpha}{2}; k\right) = \chi_{\text{прав крит}}^2(0,025; 19) = 32,85.$$

Оскільки $\chi_{\text{лів крит}}^2 < \chi_{\text{спост}}^2 < \chi_{\text{прав крит}}^2$, то немає підстав відхилити основну гіпотезу про рівність генеральної дисперсії гіпотетичному значенню.

В. Скористаємося правилом 3. За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0,05$ і кількості ступенів вільності

$$k = n - 1 = 20 - 1 = 19$$

знаходимо критичну точку:

$$\chi_{\text{крит}}^2(1 - \alpha; k) = \chi_{\text{крит}}^2(0,95; 19) = 10,12.$$

Оскільки $\chi_{\text{спост}}^2 > \chi_{\text{крит}}^2$, то немає підстав відхилити основну гіпотезу про рівність генеральної дисперсії гіпотетичному значенню.

10.4. Перевірка правильності нульової гіпотези про рівність двох генеральних середніх ($M(X) = M(Y)$)

Нехай задано дві генеральні сукупності, ознаки яких X і Y мають нормальний закон розподілу і водночас незалежні одна від одної. Необхідно перевірити правдивість $H_0: M(X) = M(Y)$ ($\bar{x}_T = \bar{y}_T$) при відомих значеннях D_X, D_Y ознак генеральних сукупностей.

З кожної генеральної сукупності здійснюють вибірку відповідно з обсягами n' і n'' і будують статистичні розподіли:

x_i	x_1	x_2	x_3	x_m
n'_i	n'_1	n'_2	n'_3	n'_m

y_j	y_1	y_2	y_3	y_m
n''_j	n''_1	n''_2	n''_3	n''_m

Тут $n' = \sum n'_i$, $n'' = \sum n''_j$.

Обчислюються значення

$$\bar{x} = \frac{\sum_{i=1}^m x_i n'_i}{n'}, \quad \bar{y} = \frac{\sum_{j=1}^m y_j n''_j}{n''}.$$

За статистичний критерій береться випадкова величина

$$U_{\text{спост}} = \frac{(\bar{x} - \bar{y})}{\sigma(\bar{x} - \bar{y})}, \tag{10.11}$$

що має нормальний закон розподілу.

Оскільки $D(\bar{x} - \bar{y}) = \frac{D_X}{n'} + \frac{D_Y}{n''}$, то статистичний критерій набере такого вигляду:

$$U_{\text{спост}} = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{D_X}{n'} + \frac{D_Y}{n''}}}. \tag{10.12}$$

Залежно від формулювання альтернативної гіпотези H_a будуються відповідно правобічна, лівобічна та двобічна критичні області.

Спостережуване значення критерію, відповідно, обчислюється:

$$U_{\text{спост}} = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{D_X}{n'} + \frac{D_Y}{n''}}}.$$

Приклад 10.5. За заданими статистичними розподілами двох вибірок, реалізованих із двох генеральних сукупностей, ознаки яких мають нормальний закон розподілу зі значенням дисперсій генеральних сукупностей $D_X = 10, D_Y = 15$,

x_i	12,2	13,2	14,2	15,2	16,2
n'_i	5	15	40	30	10

y_j	8,4	12,4	16,4	20,4	24,4
n''_j	10	15	35	20	20

при рівні значущості $\alpha = 0,01$ перевірити правдивість нульової гіпотези

$H_0: M(X) = M(Y)$, якщо альтернативна гіпотеза $H_a: M(X) > M(Y)$.

Розв'язання. Оскільки $n' = \sum n'_i = 100$, $n'' = \sum n''_j = 100$, обчислимо \bar{x} та \bar{y} :

$$\begin{aligned}\bar{x} &= \frac{\sum_{i=1}^m x_i n'_i}{n'} = \frac{12,2 \cdot 5 + 13,2 \cdot 15 + 14,2 \cdot 40 + 15,2 \cdot 30 + 16,2 \cdot 10}{100} = \\ &= \frac{62,5 + 198 + 568 + 456 + 162}{100} = \frac{1446,5}{100} = 14,465;\end{aligned}$$

$$\begin{aligned}\bar{y} &= \frac{\sum_{j=1}^m y_j n''_j}{n''} = \frac{8,4 \cdot 10 + 12,4 \cdot 15 + 16,4 \cdot 35 + 20,4 \cdot 20 + 24,4 \cdot 20}{100} = \\ &= \frac{84 + 186 + 574 + 408 + 488}{100} = \frac{1740}{100} = 17,4.\end{aligned}$$

Для альтернативної гіпотези $H_a: M(X) > M(Y)$ будується правобічна критична область. Критичну точку $u_{кр}$ знаходимо з рівності

$$\Phi(u_{кр}) = \frac{1-2\alpha}{2} = \frac{1-2 \cdot 0,01}{2} = \frac{0,98}{2} = 0,49.$$

За таблицею визначимо критичну точку:

$$u_{кр} \approx 2,34.$$

Обчислимо спостережуване значення критерію

$$U_{\text{спост}} = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{D_X + D_Y}{n' + n''}}} = \frac{14,465 - 17,4}{\sqrt{\frac{10 + 15}{100 + 100}}} = -\frac{2,935}{\sqrt{0,1 + 0,15}} = -\frac{2,935}{0,5} = -5,87.$$

Оскільки $U_{\text{спост}} < u_{кр}$, то $H_0: M(X) = M(Y)$ не відхиляється.

10.5. Перевірка правильності нульової гіпотези про рівність двох дисперсій

Одним із важливих завдань математичної статистики є порівняння двох або кількох вибірових дисперсій. Таке порівняння дає можливість визначити, чи можна вважати вибірові дисперсії статистичними оцінками однієї і тієї ж дисперсії генеральної сукупності. Воно застосовується насамперед під час обчислення дисперсій за результатами технологічних вимірювань.

Порівняння дисперсій D_X, D_Y здійснюється зіставленням виправлених дисперсій S_x^2, S_y^2 , які, відповідно, мають закон розподілу χ^2 із $k_1 = n' - 1, k_2 = n'' - 1$ ступенями свободи, де n' і n'' є обсяги першої і другої вибірок.

Нехай перша вибірка здійснена з генеральної сукупності з ознакою Y , дисперсія якої дорівнює D_Y , друга – з генеральної сукупності з ознакою X , дисперсія якої дорівнює D_X . Необхідно перевірити правильність нульової гіпотези

$$H_0: D_X = D_Y.$$

За статистичний критерій береться випадкова величина

$$F = \frac{S_x^2}{S_y^2}, \quad (10.13)$$

яка має розподіл Фішера–Снедекора із k_1 і k_2 ступенями свободи, де S_x^2 є більшою з виправлених дисперсій, одержаною внаслідок обробки результатів вибірок, S_y^2 є меншою з виправлених дисперсій.

Приклад 10.6. Під час дослідження стабільності температури в термостаті дістали такі результати: 21,2; 21,8; 21,3; 21,0; 21,4; 21,3.

З метою стабілізації температури було використано удосконалений пристрій, після цього заміри температури показали такі результати: 37,7; 37,6; 37,6; 37,4. Чи можна за рівня значущості $\alpha = 0,01$ вважати використання удосконаленого пристрою до стабілізатора температури ефективним?

Розв'язання. Очевидно, що ефективність стабілізаторів без удосконаленого пристрою і з ним залежить від дисперсій вимірюваних ними температур. Отже, задача звелась до порівняння двох дисперсій.

Обчислимо виправлені вибірові дисперсії:

$$\bar{y} = \frac{\sum_{j=1}^m y_j n_j''}{n''} = \frac{21,2+21,8+21,3+21,0+21,4+21,3}{6} = 21,333;$$

$$\begin{aligned}
D_B(Y) &= \frac{1}{n''} \sum_{j=1}^m n_j'' y_j^2 - \left(\frac{1}{n''} \sum_{j=1}^m n_j'' y_j \right)^2 = \\
&= \frac{21,2^2 + 21,8^2 + 21,3^2 + 21,0^2 + 21,4^2 + 21,3^2}{6} - 21,333^2 = \\
&= \frac{2731,02}{6} - 455,097 = 455,17 - 455,097 = 0,073;
\end{aligned}$$

$$S_y^2 = \frac{n''}{n''-1} D_B(Y) = \frac{6}{6-1} \cdot 0,073 = 0,0876;$$

$$\bar{x} = \frac{\sum_{i=1}^m x_i n_i'}{n'} = \frac{37,7 + 37,6 + 37,6 + 37,4}{4} = \frac{150,3}{4} = 37,575;$$

$$\begin{aligned}
D_B(X) &= \frac{1}{n'} \sum_{i=1}^m n_i' x_i^2 - \left(\frac{1}{n'} \sum_{i=1}^m n_i' x_i \right)^2 = \\
&= \frac{37,7^2 + 37,6^2 + 37,6^2 + 37,4^2}{4} - 37,575^2 = \\
&= \frac{5647,57}{4} - 1411,880625 = 1411,8925 - 1411,880625 = 0,011875;
\end{aligned}$$

$$S_x^2 = \frac{n'}{n'-1} D_B(X) = \frac{4}{4-1} \cdot 0,011875 = 0,01583.$$

Обчислимо спостережуване значення критерію:

$$F^* = \frac{S_\delta^2}{S_m^2} = \frac{0,0876}{0,01583} = 5,534.$$

Число ступенів свободи:

для більшої виправленої дисперсії $S_\delta^2 = S_y^2$, $k_1 = n'' - 1 = 6 - 1 = 5$,

для меншої $S_m^2 = S_x^2$, $k_2 = n' - 1 = 4 - 1 = 3$.

Оскільки удосконалення стабілізатора температур може тільки зменшити дисперсію, то будуємо правобічну критичну область. Отже,

$$H_a: S_y^2 > S_x^2.$$

Критичну точку знаходимо за таблицею відповідно до заданого рівня значущості $\alpha = 0,01$ і числа ступенів свободи $k_1 = 5$, $k_2 = 3$,

$$F_{кр}(\alpha; k_1, k_2) = F_{кр}(0,01; 5, 3) = 28,2.$$

Оскільки $F^* < F_{кр}$, то дані спостережень не дають підстав відхилити нульову гіпотезу, тобто вдосконалення термостабілізатора є ефективним.

Запитання для самоконтролю

1. Статистична гіпотеза – це ... ?
2. Які статистичні гіпотези називають параметричними?
3. Які статистичні гіпотези називають непараметричними?
4. Нульовою (основною) називають гіпотезу ...?
5. Конкуруючою (альтернативною) називають гіпотезу ...?
6. Простою називають гіпотезу ...?
7. Складною називають гіпотезу...?
8. Коли виникає помилка першого роду?
9. Коли виникає помилка другого роду?
10. Що таке рівень значущості?
11. Що називають статистичним критерієм перевірки?
12. Що називають спостережним значенням критерію узгодження?
13. Що називають критичною областю?
14. Областю прийняття гіпотези (областю допустимих значень) називають ...?
15. Основний принцип перевірки статистичних гіпотез.
16. Критичними точками $k_{кр}$ називають ...
17. Правосторонньою називають критичну область ...
18. Лівосторонньою називають критичну область ...
19. Двосторонньою називають критичну область ...
20. Потужністю критерію називають ...
21. Загальний алгоритм перевірки правильності нульової гіпотези.
22. Алгоритм перевірки правильності нульової гіпотези про рівність генерального середнього a нормальної сукупності з відомою дисперсією σ^2 гіпотетичному значенню a_0 при конкуруючій гіпотезі $H_a: a \neq a_0$; $H_a: a < a_0$; $H_a: a > a_0$.
23. Алгоритм перевірки правильності нульової гіпотези про рівність генерального середнього a нормальної сукупності з невідомою дисперсією гіпотетичному значенню a_0 при конкуруючій гіпотезі $H_a: a \neq a_0$; $H_a: a < a_0$; $H_a: a > a_0$.
24. Який статистичний критерій вибирається для перевірки основної гіпотези $H_0: a = a_0$ при невідомій дисперсії?
25. Який закон розподілу ймовірностей має статистичний критерій
$$T = \frac{(\bar{x} - a_0)\sqrt{n}}{s}?$$
26. Алгоритм перевірки правильності нульової гіпотези про рівність невідомої генеральної дисперсії σ^2 гіпотетичному (прогнозованому) значенню σ_0^2 при конкуруючій гіпотезі $H_a: \sigma^2 > \sigma_0^2$; $H_a: \sigma^2 \neq \sigma_0^2$; $H_a: \sigma^2 < \sigma_0^2$.

27. Алгоритм перевірки правильності гіпотези про рівність двох генеральних середніх при відомих значеннях D_X, D_Y .

28. Який статистичний критерій вибирається для перевірки правильності: $M(X) = M(Y)$?

29. Який закон розподілу ймовірностей має статистичний критерій $U_{\text{спост}} = \frac{(\bar{x} - \bar{y})}{\sigma(\bar{x} - \bar{y})}$?

30. Алгоритм перевірки правильності нульової гіпотези про рівність двох дисперсій.

31. Який статистичний критерій вибирається для перевірки правильності: $H_0: D_X = D_Y$;

32. Який закон розподілу ймовірностей $F = \frac{S_{\delta}^2}{S_m^2}$?

Задачі для самостійного розв'язування

1. За заданим статистичним розподілом вибірки, реалізованим із генеральної сукупності, ознака X якої має нормальний закон розподілу

x_i	4,2	6,2	8,2	10,2	12,2
n_i	6	8	12	8	2

При рівні значущості $\alpha = 0,01$ перевірити правильність нульової гіпотези $H_0: a = 10$, якщо альтернативна гіпотеза $H_a: a > 10$, $\sigma_{\Gamma}^2 = 16$.

2. Проведено 25 незалежних вимірювань випадкової величини X , що має нормальний закон розподілу зі значенням $\sigma_{\Gamma}^2 = 4$:

x_i	2,4	5,4	8,4	11,4	14,4	17,4
n_i	2	3	10	6	3	1

При рівні значущості $\alpha = 0,001$ перевірити правильність нульової гіпотези $H_0: a = 10,5$, якщо альтернативна гіпотеза $H_a: a < 10,5$.

3. Маємо дані про розподіл підприємств певної області за зростанням виробітку на одного працівника у відсотках до наступного року:

$x_i, \%$	75	85	95	105	115	125
n_i	5	8	10	5	2	1

Враховуючи, що ознака має нормальний закон розподілу зі значенням $\sigma_1^2 = 36$, перевірити правильність нульової гіпотези при $\alpha = 0,01$: $H_0: a = 90$, якщо $H_a: a \neq 90$.

4. У результаті двадцяти незалежних вимірювань певної величини X дістали статистичний розподіл:

x_i	3,4	6,4	9,4	12,4	15,4	18,4
n_i	2	4	8	3	2	1

Припускаючи, що випадкова величина X має нормальний закон розподілу, при рівні значущості $\alpha = 0,01$ перевірити правильність $H_0: a = 10$, якщо $H_a: a > 10$.

5. Результати вимірювання зросту дівчаток віком 16 років дали такі показники:

$h = 4, \text{ см}$	160–164	164–168	168–172	172–176	176–180
n_i	4	6	20	4	2

Вважаючи, що випадкова величина X – зріст дівчаток – має нормальний закон розподілу, при рівні значущості $\alpha = 0,001$ перевірити правильність нульової гіпотези $H_0: a = 180$, якщо альтернативна гіпотеза $H_a: a \neq 180$.

6. Вимірювання барометром атмосферного тиску протягом 100 діб дали такі результати:

$x_i, \text{ мм рт. ст.}$	744,4	746,4	748,4	750,4	752,4	754,4
n_i	10	20	30	20	15	5

Вважаючи, що X – атмосферний тиск – є випадковою величиною, яка має нормальний закон розподілу, при рівні значущості $\alpha = 0,01$ перевірити правильність нульової гіпотези $H_0: a = 749,2$, якщо альтернативна гіпотеза $H_a: a < 749,2$.

7. Із нормальної генеральної сукупності отримано вибірку об'єму $n = 15$, для якої знайдено вибіркочну виправлену дисперсію $S^2 = 14,9$. Переві-

рити при рівні значущості $\alpha = 0,01$ основну гіпотезу $H_0: \sigma^2 = \sigma_0^2 = 14,1$, якщо альтернативна гіпотеза:

а) $H_a: \sigma^2 > 14,1$;

б) $H_a: \sigma^2 \neq 14,1$;

в) $H_a: \sigma^2 < 14,1$.

8. Із нормальної генеральної сукупності отримано вибірку об'єму $n = 17$, для якої знайдено вибіркочну виправлену дисперсію $S^2 = 0,24$. Перевірити при рівні значущості $\alpha = 0,05$ основну гіпотезу $H_0: \sigma^2 = \sigma_0^2 = 0,18$, якщо альтернативна гіпотеза:

а) $H_a: \sigma^2 > 0,18$;

б) $H_a: \sigma^2 \neq 0,18$;

в) $H_a: \sigma^2 < 0,18$.

9. Точність роботи верстата-автомата перевіряється по дисперсії контрольованого розміру виробів, яка не повинна перевищувати 0,15. Вибірковому контролю було піддано 25 виробів і за результатами визначена оцінка дисперсії $S^2 = 0,25$. Припускається, що розмір виробів – нормально розподілена випадкова величина. Перевірити гіпотезу, що верстат забезпечує необхідну точність.

10. Із нормальної генеральної сукупності отримано вибірку об'єму $n = 31$.

x_i	10,1	10,3	10,6	11,2	11,5	11,8	12,0
n_i	1	3	7	10	6	3	1

Перевірити при рівні значущості $\alpha = 0,05$ основну гіпотезу $H_0: \sigma^2 = \sigma_0^2 = 0,18$, якщо альтернативна гіпотеза:

а) $H_a: \sigma^2 > 0,18$;

б) $H_a: \sigma^2 \neq 0,18$;

в) $H_a: \sigma^2 < 0,18$.

11. Електролампочки на 220 В виготовлялися двома заводами. З першої партії, виготовленої заводом № 1, здійснили вибірку обсягом $n_j'' = 25$, а з другої партії – обсягом $n_i' = 36$. Першу і другу партії електролампочок перевірили на тривалість роботи. Результати перевірки наведено у вигляді статистичних розподілів такого вигляду:

y_j	48	50	52	54	56	x_i	53	56	59	62	65
n_j''	2	3	14	5	1	n_i'	4	6	10	12	4

Відомо, що ознаки Y – тривалість роботи електролампочки першого заводу і X – тривалість роботи електролампочки другого заводу є випадковими величинами, які незалежні між собою і мають нормальний закон розподілу зі значеннями $\sigma_Y = 50$, $\sigma_X = 72$. При рівні значущості $\alpha = 0,01$ перевірити правильність $H_0: M(X) = M(Y)$, якщо альтернативна гіпотеза $H_a: M(X) > M(Y)$.

12. Досліджувався місячний прибуток робітників у гривнях двох заводів однієї і тієї ж галузі виробництва. Результати досліджень подано двома статистичними розподілами:

y_j	150,6	160,6	170,6	180,6	190,6
n_j''	12	28	40	18	2

x_i	140,8	160,8	180,8	200,8	220,8
n_i'	2	6	32	8	2

Ознаки X і Y є незалежними і мають нормальний закон розподілу. При рівні значущості $\alpha = 0,01$ перевірити правильність нульової гіпотези $H_0: M(X) = M(Y)$, якщо альтернативна гіпотеза $H_a: M(X) < M(Y)$.

13. Визначався обсяг валової продукції на підприємствах однієї і тієї ж галузі у двох районах України. Результати розрахунків подано двома статистичними розподілами:

y_j , млн грн	380	400	420	440	460
n_j''	5	15	30	40	10

x_i , млн грн	360	400	440	480	500	540
n_i'	10	20	30	20	15	5

Враховуючи, що ознаки X і Y є незалежними і мають нормальний закон розподілу, при рівні значущості $\alpha = 0,01$ перевірити правильність $H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза $H_a : M(X) \neq M(Y)$.

14. Вимірювалось споживання масла за одну добу одним мешканцем у двох регіонах країни. Результати вимірювання подано двома статистичними розподілами:

y_j , мг	15,99	18,99	21,99	21,99	24,99
n_j''	4	6	20	10	5

x_i , мг	14,55	20,55	26,55	30,55	36,55
n_i'	6	14	16	6	4

Ознаки X і Y (добове споживання масла в мг) є незалежними випадковими величинами, які мають нормальний закон розподілу ймовірностей. При рівні значущості $\alpha = 0,01$ перевірити правильність нульової гіпотези $H_0 : D_X = D_Y$, якщо альтернативна гіпотеза $H_a : D_X > D_Y$.

15. Визначалися річні середні витрати електроенергії на комунально-побутові вимоги для одного мешканця у двох містах. Результати розрахунків подано двома статистичними розподілами для першого і другого міст:

y_j , Вт/м.	700	708	716	724	732	740
n_j''	5	6	9	6	3	1

x_i , Вт/м.	706	710	714	718	722	726	730
n_i'	8	10	12	5	2	2	1

Ознаки X і Y (річні витрати в кВт/особу) є незалежними між собою і мають нормальний закон розподілу. При рівні значущості $\alpha = 0,001$ перевірити правильність нульової гіпотези $H_0 : D_X = D_Y$, якщо альтернативна гіпотеза $H_a : D_X < D_Y$.

Розділ 11

СТАТИСТИЧНА ПЕРЕВІРКА ГІПОТЕЗИ ПРО ЗАКОН РОЗПОДІЛУ ГЕНЕРАЛЬНОЇ СУКУПНОСТІ

11.1. Перевірка гіпотези про нормальний розподіл

Нехай статистичний розподіл вибірки задано у вигляді послідовності рівновіддалених варіант і відповідних їм частот:

x_i	x_1	x_2	...	x_m
n_i	n_1	n_2	...	n_m

За допомогою критерію Пірсона перевірити гіпотезу про нормальний розподіл генеральної сукупності.

Правило 1. Для того, щоб при заданому рівні значущості α перевірити гіпотезу про нормальний розподіл генеральної сукупності, необхідно:

- 1) обчислити вибіркове середнє \bar{x} і вибіркове середнє квадратичне відхилення σ_B ;
- 2) визначити теоретичні частоти

$$n'_i = \frac{n \cdot h}{\sigma_B} \cdot \varphi(u_i), \quad (11.1)$$

де n – об'єм вибірки; h – крок (різниця між двома сусідніми варіантами); $\varphi(u)$ – диференціальна функція Лапласа; $u_i = \frac{x_i - \bar{x}}{\sigma_B}$;

- 3) порівняти емпіричні та теоретичні частоти за допомогою критерію Пірсона. Для цього:

а) складають розрахункову таблицю

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$
1.					

і знаходять спостережуване значення критерію Пірсона

$$\chi^2_{\text{спост}} = \sum_{i=1}^m \frac{(n_i - n'_i)^2}{n'_i}; \quad (11.2)$$

б) за таблицею критичних точок розподілу χ^2 при заданому рівні значущості α і кількості ступенів вільності $k = s - 3$ (s – кількість варіант вибірки) знаходять критичну точку $\chi^2_{\text{крит}}(\alpha; k)$ правосторонньої критичної області. Якщо $\chi^2_{\text{спост}} < \chi^2_{\text{крит}}$, немає підстав відхиляти гіпотезу про нормальний розподіл генеральної сукупності. Іншими словами, емпіричні та теоретичні частоти відрізняються несуттєво (випадково). Якщо $\chi^2_{\text{спост}} > \chi^2_{\text{крит}}$, гіпотезу відхиляють. Іншими словами, емпіричні та теоретичні частоти відрізняються суттєво.

Нехай статистичний розподіл вибірки задано у вигляді послідовності інтервалів і відповідних їм частот

$(x_i; x_{i+1}]$	$(x_1; x_2]$	$(x_2; x_3]$.	$(x_m; x_{m+1}]$
n_i	n_1	n_2	n_m

За допомогою критерію Пірсона перевірити гіпотезу про нормальний розподіл генеральної сукупності.

Правило 2. Для того, щоб при заданому рівні значущості α перевірити гіпотезу про нормальний розподіл генеральної сукупності, необхідно:

1) обчислити вибіркове середнє \bar{x}^* і вибіркове середнє квадратичне відхилення σ^* , причому як варіанти x_i^* беруть середнє арифметичне кінців інтервалу:

$$x_i^* = \frac{x_i + x_{i+1}}{2};$$

2) пронумерувати досліджувану випадкову величину X , тобто перейти до випадкової величини $Z = \frac{X - \bar{x}^*}{\sigma^*}$ і обчислити кінці інтервалів: $z_i = \frac{x_i - \bar{x}^*}{\sigma^*}$,

причому найменше значення Z , тобто z_1 , вважають рівним $-\infty$, а найбільше, тобто z_{m+1} – рівним ∞ ;

3) обчислити теоретичні частоти

$$n'_i = n \cdot P_i, \quad (11.3)$$

де n – об'єм вибірки; $P_i = \Phi(z_{i+1}) - \Phi(z_i)$ – ймовірності потрапляння X в інтервали $(x_i; x_{i+1}]$; $\Phi(z)$ – функція Лапласа;

4) порівняти емпіричні та теоретичні частоти за допомогою критерію Пірсона. Для цього:

А. Складають розрахункову таблицю, за якою знаходять спостережуване значення критерію Пірсона

$$\chi^2_{\text{спост}} = \sum_{i=1}^m \frac{(n_i - n'_i)^2}{n'_i}.$$

Б. За таблицею критичних точок розподілу χ^2 при заданому рівні значущості α і кількості ступенів вільності $k = s - 3$ (s – кількість інтервалів вибірки) знаходять критичну точку $\chi^2_{\text{крит}}(\alpha; k)$ правосторонньої критичної області. Якщо $\chi^2_{\text{спост}} < \chi^2_{\text{крит}}$, немає підстав відхилити гіпотезу про нормальний розподіл генеральної сукупності. Тобто, емпіричні та теоретичні частоти відрізняються несуттєво (випадково). Якщо $\chi^2_{\text{спост}} > \chi^2_{\text{крит}}$, гіпотезу відхиляють. Інакше кажучи, емпіричні та теоретичні частоти відрізняються суттєво.

Приклад 11.1. Використовуючи критерій Пірсона, при рівні значущості $\alpha = 0,05$ перевірити, чи справджується гіпотеза про нормальний розподіл генеральної сукупності X , якщо з цієї сукупності отримано таку вибірку об'ємом $n = 100$:

$(x_i; x_{i+1}]$	(3; 7]	(7; 11]	(11; 15]	(15; 19]	(19; 23]	(23; 27]	(27; 31]
n_i	6	16	19	17	15	14	13

Розв'язання. Спочатку обчислимо вибіркове середнє \bar{x}^* і середнє квадратичне відхилення вибірки σ^* . Для цього перетворимо інтервальний статистичний розподіл вибірки на точковий:

x_i	5	9	13	17	21	25	29
n_i	6	16	19	17	15	14	13

Отже,

$$\bar{x}^* = \frac{5 \cdot 6 + 9 \cdot 16 + 13 \cdot 19 + 17 \cdot 17 + 21 \cdot 15 + 25 \cdot 14 + 29 \cdot 13}{100} = 17,52;$$

$$\begin{aligned} \sigma^* &= \sqrt{\frac{5^2 \cdot 6 + 9^2 \cdot 16 + 13^2 \cdot 19 + 17^2 \cdot 17 + 21^2 \cdot 15 + 25^2 \cdot 14 + 29^2 \cdot 13}{100} - 17,52^2} = \\ &= \sqrt{\frac{35868}{100} - 306,95} = \sqrt{51,73} \approx 7,19; \end{aligned}$$

Знайдемо нормовані інтервали $(z_i; z_{i+1}]$, враховуючи, що вибіркове середнє $\bar{x}^* = 17,52$ і середнє квадратичне відхилення вибірки $\sigma^* = 7,19$. Для цього складемо розрахункову таблицю (лівий кінець першого інтервалу вважаємо рівним $-\infty$, а правий кінець останнього інтервалу – рівним ∞).

i	x_i	x_{i+1}	$x_i - \bar{x}^*$	$x_{i+1} - \bar{x}^*$	$z_i = \frac{x_i - \bar{x}^*}{\sigma^*}$	$z_{i+1} = \frac{x_{i+1} - \bar{x}^*}{\sigma^*}$
1	3	7	-14,52	-10,52	$-\infty$	-1,46
2	7	11	-10,52	-6,52	-1,46	-0,91
3	11	15	-6,52	-2,5	-0,91	-0,35
4	15	19	-2,5	1,48	-0,35	0,21
5	19	23	1,48	5,48	0,21	0,76
6	23	27	5,48	9,48	0,76	1,32
7	27	31	9,48	13,48	1,32	∞

Визначимо теоретичні ймовірності P_i і теоретичні частоти n'_i :

$$n'_i = n \cdot P_i$$

Для цього складемо розрахункову таблицю:

i	z_i	z_{i+1}	$\Phi(z_i)$	$\Phi(z_{i+1})$	$P_i = \Phi(z_{i+1}) - \Phi(z_i)$	$n'_i = n \cdot P_i$
1	$-\infty$	-1,46	-0,5000	-0,4279	0,0721	7,21
2	-1,46	-0,91	-0,4279	-0,3186	0,1093	10,93
3	-0,91	-0,35	-0,3186	-0,1368	0,1818	18,18
4	-0,35	0,21	-0,1368	0,0832	0,22	22
5	0,21	0,76	0,0832	0,2764	0,1932	19,32
6	0,76	1,32	0,2764	0,4066	0,1302	13,02
7	1,32	∞	0,4066	0,5000	0,0934	9,34
Σ	–	–	–	–	1	100

Порівняємо емпіричні та теоретичні частоти, використовуючи критерій Пірсона. Для цього обчислимо спочатку спостережуване значення критерію Пірсона за допомогою розрахункової таблиці.

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$	n_i^2	$\frac{n_i^2}{n'_i}$
1	6	7,21	-1,21	1,4641	0,2031	36	4,9931
2	16	10,93	5,07	25,7049	2,3518	256	23,4218
3	19	18,18	0,82	0,6724	0,0370	361	19,8570
4	17	22	-5	25	1,1364	289	13,1364
5	15	19,32	-4,32	18,6624	0,9660	225	11,6460
6	14	13,02	0,98	0,9604	0,0738	196	15,0538
7	13	9,34	3,66	13,3956	1,4342	169	18,0942
Σ	100	100			$\chi_{\text{спост}}^2 = 6,2023$	6,2023	106,2023

Перевірка:

$$\chi_{\text{спост}}^2 = \sum_{i=1}^m \frac{n_i^2}{n'_i} - 100.$$

Оскільки $\sum_{i=1}^m \frac{n_i^2}{n'_i} - 100 = 106,2023 - 100 = 6,2023 = \chi_{\text{спост}}^2$, то обчислення проведено правильно.

За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0,05$ і кількості ступенів вільності

$$k = s - 3 = 7 - 3 = 4$$

знайдемо критичну точку правосторонньої критичної області:

$$\chi_{\text{крит}}^2(0,05; 4) = 9,49.$$

Оскільки $\chi_{\text{спост}}^2 < \chi_{\text{крит}}^2$, гіпотезу про нормальний розподіл генеральної сукупності не відхиляємо. Тобто, емпіричні та теоретичні частоти відрізняються несуттєво.

11.2. Перевірка гіпотези про рівномірний розподіл

Нехай задано інтервальний статистичний розподіл вибірки:

$(x_i; x_{i+1}]$	$(x_1; x_2]$	$(x_2; x_3]$		$(x_m; x_{m+1}]$
n_i	n_1	n_2	...	n_m

n – об'єм вибірки.

За допомогою критерію Пірсона перевірити гіпотезу про рівномірний розподіл генеральної сукупності.

Правило 1. Для того, щоб перевірити гіпотезу про рівномірний розподіл досліджуваної випадкової величини X , тобто про щільність X :

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a; b]; \\ 0, & x \notin [a; b], \end{cases} \quad (11.4)$$

необхідно:

1) оцінити параметри a і b – кінці інтервалу, у якому спостерігалися можливі значення X , за формулами:

$$a^* = \bar{x} - \sqrt{3}\sigma_B, \quad b^* = \bar{x} + \sqrt{3}\sigma_B, \quad (11.5)$$

де a^* і b^* – оцінки параметрів);

2) знайти щільність ймовірності передбачуваного розподілу:

$$f(x) = \begin{cases} \frac{1}{b^*-a^*}, & x \in [a^*; b^*]; \\ 0, & x \notin [a^*; b^*], \end{cases} \quad (11.6)$$

3) визначити теоретичні частоти:

$$n'_1 = n \cdot P_1 = n \cdot \frac{1}{b^*-a^*} (x_1 - a^*); \quad (11.7)$$

$$n'_i = n \cdot P_i = n \cdot \frac{1}{b^*-a^*} (x_i - x_{i-1});$$

$$n'_m = n \cdot P_m = n \cdot \frac{1}{b^*-a^*} (b^* - x_m).$$

4) порівняти емпіричні та теоретичні частоти за допомогою критерію Пірсона, взявши кількість ступенів вільності $k = s - 3$, де s – кількість інтервалів вибірки.

Приклад 11.2. Передбачалося, що про стабільність економічної ситуації в країні (відсутність воєн, стихійних лих і т. д.) за останні 50 років можна судити за характером розподілу населення за віком (n_i – млн чол.): при спокійній обстановці розподіл повинен бути рівномірним. В результаті проведеного дослідження, для однієї з країн були отримані такі дані.

$(x_i; x_{i+1}]$	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
n_i	14	9	10	8	16	13	12	18

Чи є підстави вважати, що в країні була нестабільна ситуація? При рівні значущості $\alpha = 0,01$ перевірити гіпотезу про рівномірний розподіл та зробити відповідні висновки.

Розв'язання. Спочатку обчислимо вибіркове середнє \bar{x}^* і середнє квадратичне відхилення вибірки σ^* . Для цього перетворимо інтервальний статистичний розподіл вибірки на точковий:

x_i	5	15	25	35	45	55	65	75
n_i	14	9	10	8	16	13	12	18

Отже,

$$\bar{x}^* = \frac{5 \cdot 14 + 15 \cdot 9 + 25 \cdot 10 + 35 \cdot 8 + 45 \cdot 16 + 55 \cdot 13 + 65 \cdot 12 + 75 \cdot 18}{100} = \frac{4300}{100} = 43;$$

$$\sigma^* = \sqrt{\frac{5^2 \cdot 14 + 15^2 \cdot 9 + 25^2 \cdot 10 + 35^2 \cdot 8 + 45^2 \cdot 16 + 55^2 \cdot 13 + 65^2 \cdot 12 + 75^2 \cdot 18}{100} - 43^2} =$$

$$= \sqrt{\frac{242100}{100} - 1849} = \sqrt{2421 - 1849} = \sqrt{572} \approx 23,92.$$

Оцінимо параметри a і b – кінців інтервалу, у якому спостерігалися можливі значення X :

$$a^* = \bar{x} - \sqrt{3}\sigma_B = 43 - \sqrt{3} \cdot 23,92 = 1,57;$$

$$b^* = \bar{x} + \sqrt{3}\sigma_B = 43 + \sqrt{3} \cdot 23,92 = 84,43.$$

Знайдемо щільність ймовірності передбачуваного розподілу:

$$f(x) = \begin{cases} 0,0121, & x \in [a^*; b^*]; \\ 0, & x \notin [a^*; b^*]. \end{cases}$$

Визначимо теоретичні частоти:

$$n'_1 = n \cdot P_1 = 100 \cdot 0,0121 \cdot (10 - 1,57) = 10,1;$$

$$n'_2 = n \cdot P_2 = 100 \cdot 0,0121(20 - 10) = 12,1;$$

$$n'_3 = n \cdot P_3 = 100 \cdot 0,0121(30 - 20) = 12,1;$$

$$n'_4 = n \cdot P_4 = 100 \cdot 0,0121(40 - 30) = 12,1;$$

$$n'_5 = n \cdot P_5 = 100 \cdot 0,0121(50 - 40) = 12,1;$$

$$n'_6 = n \cdot P_6 = 100 \cdot 0,0121(60 - 50) = 12,1;$$

$$n'_7 = n \cdot P_7 = 100 \cdot 0,0121(70 - 60) = 12,1;$$

$$n'_8 = n \cdot P_8 = 100 \cdot 0,0121(84,43 - 70) = 17,3.$$

Порівняємо емпіричні та теоретичні частоти за допомогою критерію Пірсона, взявши кількість ступенів вільності $k = s - 3$, де s – кількість інтервалів вибірки.

Для цього обчислимо спочатку спостережуване значення критерію Пірсона за допомогою розрахункової таблиці.

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$	n_i^2	$\frac{n_i^2}{n'_i}$
1	14	10,1	3,9	15,21	1,5059	196	19,4059
2	9	12,1	-3,1	9,61	0,7942	81	6,6942
3	10	12,1	-2,1	4,41	0,3645	100	8,2645
4	8	12,1	-4,1	16,81	1,3893	64	5,2893
5	16	12,1	3,9	15,21	1,2570	256	21,157
6	13	12,1	0,9	0,81	0,0669	169	13,9669
7	12	12,1	-0,1	0,01	0,0008	144	11,9008
8	18	17,3	0,7	0,49	0,0283	324	18,7283
Σ	100	100			$\chi^2_{\text{спост}} = 5,4069$		105,4069

Перевірка:

$$\chi^2_{\text{спост}} = \sum_{i=1}^m \frac{n_i^2}{n_i'} - 100.$$

Оскільки $\sum_{i=1}^m \frac{n_i^2}{n_i'} - 100 = 105,4069 - 100 = 5,4069 = \chi^2_{\text{спост}}$, то обчислення проведено правильно.

За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0,01$ і кількості ступенів вільності

$$k = s - 3 = 8 - 3 = 5$$

знайдемо критичну точку правосторонньої критичної області:

$$\chi^2_{\text{крит}}(0,01; 5) = 15,09.$$

Оскільки $\chi^2_{\text{спост}} < \chi^2_{\text{крит}}$, гіпотезу про рівномірний розподіл генеральної сукупності не відхиляємо. Тобто, емпіричні та теоретичні частоти відрізняються несуттєво. Отже, підстави вважати, що в країні була нестабільна ситуація, немає.

11.3. Перевірка гіпотези про показниковий розподіл

Нехай задано інтервальний статистичний розподіл вибірки

$(x_i; x_{i+1}]$	$(x_1; x_2]$	$(x_2; x_3]$		$(x_m; x_{m+1}]$
n_i	n_1	n_2	...	n_m

n – об'єм вибірки.

За допомогою критерію Пірсона перевірити гіпотезу про показниковий розподіл генеральної сукупності.

Правило 1. Для того, щоб при рівні значущості α перевірити гіпотезу про показниковий розподіл досліджуваної випадкової величини X , необхідно:

1) знайти вибіркове середнє \bar{x} , узявши як варіанти середини інтервалів:

$$x_i^* = \frac{x_i + x_{i+1}}{2};$$

2) за оцінку параметра λ показникового розподілу взяти величину, обернену до вибіркового середнього,

$$\lambda^* = \frac{1}{\bar{x}}, \quad (11.8)$$

3) знайти ймовірності потрапляння випадкової величини X у частинні інтервали $(x_i; x_{i+1}]$ за формулою

$$P_i = P(x_i < X < x_{i+1}) = e^{-\lambda^* x_i} - e^{-\lambda^* x_{i+1}}, \quad (11.9)$$

4) обчислити теоретичні частоти

$$n'_i = n \cdot P_i;$$

5) порівняти емпіричні та теоретичні частоти за допомогою критерію Пірсона, узявши кількість ступенів вільності $k = s - 2$, де s – кількість інтервалів вибірки.

Приклад 11.3. В результаті випробування 200 елементів на тривалість роботи отримано емпіричне розподіл, наведене в таблиці (у першому рядку вказані інтервали часу в годинах, в другому рядку – частоти, тобто кількість елементів, які працювали в межах відповідного інтервалу). Використовуючи критерій Пірсона, при рівні значущості $\alpha = 0,05$ перевірити, чи узгоджується гіпотеза про показниковий розподіл генеральної сукупності X з емпіричним розподілом вибірки.

$(x_i; x_{i+1}]$	(0; 5]	(5; 10]	(10; 15]	(15; 20]	(20; 25]	(25; 30]
n_i	133	45	15	4	2	1

Розв'язання. Спочатку обчислимо вибірконе середнє \bar{x} . Для цього перетворимо інтервальний статистичний розподіл вибірки на точковий:

x_i	2,5	7,5	12,5	17,5	22,5	27,5
n_i	133	45	15	4	2	1

Отже,

$$\bar{x} = \frac{2,5 \cdot 133 + 7,5 \cdot 45 + 12,5 \cdot 15 + 17,5 \cdot 4 + 22,5 \cdot 2 + 27,5 \cdot 1}{200} = \frac{1000}{200} = 5.$$

Оцінимо параметр λ показникового розподілу взявши величину, обернену до вибіркового середнього,

$$\lambda^* = \frac{1}{\bar{x}} = \frac{1}{5} = 0,2.$$

Знайдемо ймовірності потрапляння випадкової величини X у частинні інтервали $(x_i; x_{i+1}]$ за формулою

$$\begin{aligned} P_1 &= P(0 < X < 5) = e^{-0,2 \cdot 0} - e^{-0,2 \cdot 5} = \\ &= 1 - e^{-1} = 1 - 0,3679 = 0,6321; \end{aligned}$$

$$\begin{aligned} P_2 &= P(5 < X < 10) = e^{-0,2 \cdot 5} - e^{-0,2 \cdot 10} = \\ &= e^{-1} - e^{-2} = 0,3679 - 0,1353 = 0,2326; \end{aligned}$$

$$\begin{aligned} P_3 &= P(10 < X < 15) = e^{-0,2 \cdot 10} - e^{-0,2 \cdot 15} = \\ &= e^{-2} - e^{-3} = 0,1353 - 0,0498 = 0,0855; \end{aligned}$$

$$\begin{aligned} P_4 &= P(15 < X < 20) = e^{-0,2 \cdot 15} - e^{-0,2 \cdot 20} = \\ &= e^{-3} - e^{-4} = 0,0498 - 0,0183 = 0,0315; \end{aligned}$$

$$\begin{aligned} P_5 &= P(20 < X < 25) = e^{-0,2 \cdot 20} - e^{-0,2 \cdot 25} = \\ &= e^{-4} - e^{-5} = 0,0183 - 0,0067 = 0,0116; \end{aligned}$$

$$\begin{aligned} P_6 &= P(25 < X < 30) = e^{-0,2 \cdot 25} - e^{-0,2 \cdot 30} = \\ &= e^{-5} - e^{-6} = 0,0067 - 0,0025 = 0,0042. \end{aligned}$$

Визначимо теоретичні частоти:

$$n'_1 = n \cdot P_1 = 200 \cdot 0,6321 = 126,42;$$

$$n'_2 = n \cdot P_2 = 200 \cdot 0,2326 = 46,52;$$

$$n'_3 = n \cdot P_3 = 200 \cdot 0,0855 = 17,1;$$

$$n'_4 = n \cdot P_4 = 200 \cdot 0,0315 = 6,3;$$

$$n'_5 = n \cdot P_5 = 200 \cdot 0,0116 = 2,32;$$

$$n'_6 = n \cdot P_6 = 200 \cdot 0,0042 = 0,84.$$

Порівняємо емпіричні та теоретичні частоти за допомогою критерію Пірсона, взявши кількість ступенів вільності $k = s - 2$, де s – кількість інтервалів вибірки.

Для цього обчислимо спочатку спостережуване значення критерію Пірсона за допомогою розрахункової таблиці.

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$
1	133	126,42	6,58	43,2964	0,3425
2	45	46,52	-1,52	2,3104	0,0497
3	15	17,1	-2,1	4,41	0,2579
4	4	6,3	-2,3	5,29	0,8397
5	2	2,32	-0,32	0,1024	0,0441
6	1	0,84	0,16	0,0256	0,0305
Σ					$\chi^2_{\text{спост}} = 1,5644$

За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0,05$ і кількості ступенів вільності

$$k = s - 3 = 6 - 2 = 1$$

знайдемо критичну точку правосторонньої критичної області:

$$\chi^2_{\text{крит}}(0,05; 3) = 9,49.$$

Оскільки $\chi^2_{\text{спост}} < \chi^2_{\text{крит}}$, гіпотезу про показниковий розподіл генеральної сукупності не відхиляємо. Інакше кажучи, емпіричні та теоретичні частоти відрізняються несуттєво.

11.4. Перевірка гіпотези про біноміальний розподіл

Нехай проведено n дослідів. Кожний дослід складається із N незалежних випробувань, для яких ймовірність появи деякої події A одна й та ж. Реєструється кількість появ події A в кожному досліді. У підсумку отримуємо статистичний розподіл дискретної випадкової величини X , яка характеризує кількість появ події A (у першому рядку наведено кількість появ події A в одному досліді, а в другому рядку – частоту n_i , тобто кількість дослідів, у яких зареєстровано x_i появ події A):

x_i	0	1	N
n_i	n_1	n_2		n_N

За допомогою критерію Пірсона перевірити гіпотезу про біноміальний закон розподілу дискретної випадкової величини X .

Правило 1. Для того, щоб при рівні значущості α перевірити гіпотезу про біноміальний розподіл дискретної випадкової величини X , необхідно:

1) знайти за формулою Бернуллі

$$P_N(i) = C_N^i p^i (1-p)^{N-i} \quad (11.10)$$

ймовірності $P_N(i)$ появи події A рівно i разів у N випробуваннях ($i = 0, 1, 2, \dots, s$), де s – максимальна кількість спостережуваних появ події A в одному досліді, тобто $s \leq N$);

2) знайти теоретичні частоти

$$n'_i = n \cdot P_N(i);$$

де n – кількість дослідів;

3) порівняти емпіричні та теоретичні частоти за допомогою критерію Пірсона, поклавши кількість ступенів вільності $k = s - 1$ (вважається, що ймовірність p появи події A задана, тобто вона не оцінювалась за вибіркою і не об'єднувались малочисельні частоти).

Якщо ймовірність p була оцінена за вибіркою, то $k = s - 2$. Якщо, крім того, було об'єднано малочисельні частоти, то s – кількість варіант вибірки, які залишилися після об'єднання частот.

Приклад 11.4. Чотири монети були підкинуті 20160 разів, водночас комбінації: чотири «герби», три «герби» і «цифра», два «герби» і дві «цифри», один «герб» і три «цифри», чотири «цифри» з'явилися таку відповідну кількість разів:

1181, 4909, 7583, 5085, 1402.

Чи свідчать ці дані про те, що кількість «гербів», яка з'явилась на чотирьох монетах має біноміальний розподіл. Застосувати критерій χ^2 , при рівні значущості $\alpha = 0,05$.

Розв'язання. Спочатку знайдемо за формулою Бернуллі (11.10) ймовірності $P_N(i)$ появи події A – кількість випавши гербів. Дані запишемо у таблицю.

$$P_4(4) = C_4^4 \cdot 0,5^4 \cdot (1-0,5)^{4-4} = 1 \cdot 0,0625 \cdot 1 = 0,0625;$$

$$P_4(3) = C_4^3 \cdot 0,5^3 \cdot (1-0,5)^{4-3} = 4 \cdot 0,125 \cdot 0,5 = 0,25;$$

$$P_4(2) = C_4^2 \cdot 0,5^2 \cdot (1-0,5)^{4-2} = 6 \cdot 0,25 \cdot 0,25 = 0,375;$$

$$P_4(1) = C_4^1 \cdot 0,5^1 \cdot (1-0,5)^{4-1} = 4 \cdot 0,5 \cdot 0,125 = 0,25;$$

$$P_4(0) = C_4^0 \cdot 0,5^0 \cdot (1-0,5)^{4-0} = 1 \cdot 1 \cdot 0,0625 = 0,0625.$$

X	4	3	2	1	0
$P(X)$	0,0625	0,25	0,375	0,25	0,0625

Знайдемо теоретичні частоти, використовуючи формулу:

$$n'_i = n \cdot P_N(i).$$

$$n'_1 = 20160 \cdot 0,0625 = 1260;$$

$$n'_2 = 20160 \cdot 0,25 = 5040;$$

$$n'_3 = 20160 \cdot 0,375 = 7560;$$

$$n'_4 = 20160 \cdot 0,25 = 5040;$$

$$n'_5 = 20160 \cdot 0,0625 = 1260.$$

Порівняємо емпіричні та теоретичні частоти за допомогою критерію Пірсона. Для цього складемо розрахункову таблицю:

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$
0	1181	1260	-79	6241	4,9532
1	4909	5040	-131	17161	3,405
2	7583	7560	23	529	0,0699
3	5085	5040	45	2025	0,4018

4	1402	1260	142	20164	16,0032
5	1181	1260	-79	6241	4,9532
Σ					$\chi^2_{\text{спост}} = 24,8331$

Із розрахункової таблиці отримаємо

$$\chi^2_{\text{спост}} = 24,8331.$$

За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0,05$ і кількості ступенів вільності

$$k = s - 2 = 5 - 2 = 3$$

знаходимо критичну точку правосторонньої критичної області:

$$\chi^2_{\text{крит}}(0,05; 3) = 7,82.$$

Оскільки $\chi^2_{\text{спост}} > \chi^2_{\text{крит}}$, є підстави відхилити гіпотезу про розподіл випадкової величини X (кількості «гербів» за біноміальним законом).

11.5. Перевірка гіпотези про розподіл Пуассона

Нехай задано точковий статистичний розподіл вибірки. За допомогою критерію Пірсона перевірити гіпотезу про розподіл генеральної сукупності за законом Пуассона.

Правило 1. Для того, щоб при рівні значущості α перевірити гіпотезу про те, що досліджувана випадкова величина X розподілена за законом Пуассона, необхідно:

- 1) знайти за цим статистичним розподілом вибіркове середнє \bar{x} ;
- 2) взяти за оцінку параметра λ розподілу Пуассона вибіркове середнє

$$\lambda^* = \bar{x}; \quad (11.11)$$

3) знайти за формулою Пуассона (або за таблицею значень функції Пуассона) ймовірності P_i появи рівно i подій у n випробуваннях

($i = 1, 2, 3, \dots, r$), де r – максимальна кількість спостережуваних подій; n – об'єм вибірки;

4) визначити теоретичні частоти

$$n'_i = n \cdot P_i;$$

5) порівняти емпіричні та теоретичні частоти за допомогою критерію Пірсона, узявши кількість ступенів вільності $k = s - 2$, де s – кількість варіант вибірки (якщо проводилося об'єднання малочисельних частот в одну групу, то s – кількість варіант, які залишилися після об'єднання).

Приклад 11.4. У $n = 1000$ перевірках партій товару реєструвалася кількість x_i неякісної продукції, внаслідок чого було отримано такий статистичний розподіл кількості браку x_i в n_i партіях товару:

x_i	0	1	2	3	4	5	6	7
n_i	242	349	234	107	43	21	3	1

При рівні значущості $\alpha = 0,05$ перевірити гіпотезу про те, що кількість неякісної продукції X розподілена за законом Пуассона.

Розв'язання. Спочатку знайдемо вибіркове середнє:

$$\bar{x} = \frac{0 \cdot 242 + 1 \cdot 349 + 2 \cdot 234 + 3 \cdot 107 + 4 \cdot 43 + 5 \cdot 21 + 6 \cdot 3 + 7 \cdot 1}{1000} = \frac{1440}{1000} = 1,44.$$

Візьмемо за оцінку параметра λ розподілу Пуассона вибіркове середнє

$$\lambda^* = \bar{x} = 1,44.$$

Припускаємо, що закон Пуассона має вигляд:

$$P_{1000}(i) = 1,44^i \cdot \frac{e^{-1,44}}{i!}.$$

Поклавши $i = 0, 1, 2, \dots, 7$, обчислимо ймовірності $P_i = P_{1000}(i)$:

$$P_0 = 1,44^0 \cdot \frac{e^{-1,44}}{0!} \approx 0,2369; \quad P_1 = 1,44^1 \cdot \frac{e^{-1,44}}{1!} \approx 0,3412;$$

$$P_2 = 1,44^2 \cdot \frac{e^{-1,44}}{2!} \approx 0,2456; \quad P_3 = 1,44^3 \cdot \frac{e^{-1,44}}{3!} \approx 0,1179;$$

$$P_4 = 1,44^4 \cdot \frac{e^{-1,44}}{4!} \approx 0,0424;$$

$$P_5 = 1,44^5 \cdot \frac{e^{-1,44}}{5!} \approx 0,0122;$$

$$P_6 = 1,44^6 \cdot \frac{e^{-1,44}}{6!} \approx 0,0029;$$

$$P_7 = 1,44^7 \cdot \frac{e^{-1,44}}{7!} \approx 0,0006;$$

Знайдемо теоретичні частоти $n'_i = n \cdot P_i$:

$$n'_0 = 1000 \cdot 0,2369 = 236,9;$$

$$n'_1 = 1000 \cdot 0,3112 = 311,2;$$

$$n'_2 = 1000 \cdot 0,2369 = 236,9;$$

$$n'_3 = 1000 \cdot 0,1179 = 117,9;$$

$$n'_4 = 1000 \cdot 0,0424 = 42,4;$$

$$n'_5 = 1000 \cdot 0,0122 = 12,2;$$

$$n'_6 = 1000 \cdot 0,0029 = 2,9;$$

$$n'_7 = 1000 \cdot 0,0006 = 0,6.$$

Оскільки частоти $n_6 = 3$ і $n_7 = 1$ – малочисельні (менше п'яти), об'єднаємо їх із частотою $n_5 = 21$, а саме

$$n_5 = 21 + 3 + 1 = 25.$$

Як теоретичну частоту, що відповідає об'єднаній частоті, візьмемо суму відповідних теоретичних частот:

$$n'_5 = 12,2 + 2,9 + 0,6 = 15,7.$$

Порівняємо емпіричні та теоретичні частоти за допомогою критерію Пірсона. Для цього складемо розрахункову таблицю:

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$
0	242	236,9	5,1	26,01	0,1098
1	349	341,2	7,8	60,84	0,1784
2	234	245,6	-11,6	134,56	0,5479
3	107	117,9	-10,9	118,81	1,0077
4	43	42,4	0,6	0,36	0,0085
5	25	15,7	9,3	86,49	5,5089
Σ	1000				$\chi^2_{\text{спост}} = 7,3612$

Із розрахункової таблиці отримаємо

$$\chi^2_{\text{спост}} = 7,3612.$$

За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0,05$ і кількості ступенів вільності

$$k = s - 1 = 6 - 2 = 4$$

знаходимо критичну точку правосторонньої критичної області:

$$\chi_{\text{крит}}^2(0,05; 4) = 9,49.$$

Оскільки $\chi_{\text{спост}}^2 < \chi_{\text{крит}}^2$, то немає підстав відхилити гіпотезу про розподіл випадкової величини X (кількості неякісного товару в партії) за законом Пуассона.

Запитання для самоконтролю

1. Диференціальна функція Лапласа має вигляд ...?
2. Як знайти спостережуване значення критерію Пірсона?
3. Як знайти критичне значення критерію Пірсона $\chi_{\text{крит}}^2$?
4. Якщо $\chi_{\text{спост}}^2 < \chi_{\text{крит}}^2$, то гіпотеза про нормальний розподіл ... ?
5. Якщо $\chi_{\text{спост}}^2 > \chi_{\text{крит}}^2$, то гіпотеза про нормальний розподіл ... ?
6. $\Phi(z)$ – функція Лапласа має вигляд ... ?
7. Як за допомогою критерію Пірсона перевірити гіпотезу про нормальний розподіл генеральної сукупності?
8. Як визначити теоретичні частоти при перевірці гіпотези про нормальний розподіл генеральної сукупності?
9. Як за допомогою критерію Пірсона перевірити гіпотезу про нормальний розподіл, якщо вибірка задана інтервальним розподілом?
10. Як визначити кількість ступенів вільності при перевірці гіпотези про нормальний розподіл за критерієм Пірсона?
11. Як за допомогою критерію перевірити гіпотезу про рівномірний розподіл генеральної сукупності?
12. Як оцінити параметри a і b – кінці інтервалу, у якому спостерігалися можливі значення X (рівномірний розподіл)?
13. Як визначити теоретичні частоти при перевірці гіпотези про рівномірний розподіл?
14. Як визначити кількості ступенів вільності при перевірці гіпотези про рівномірний розподіл за критерієм Пірсона?
15. Як за допомогою критерію Пірсона перевірити гіпотезу про показниковий розподіл генеральної сукупності?
16. Яку величину беруть за оцінку параметра λ показникового розподілу?

17. Як знайти ймовірності потрапляння випадкової величини X у частинні інтервали $(x_i; x_{i+1}]$ (показників розподіл)?
18. Як визначити теоретичні частоти при перевірці гіпотези про показниковий розподіл?
19. Як визначити кількість ступенів вільності при перевірці гіпотези про показниковий розподіл за критерієм Пірсона?
20. Як знайти ймовірності $P_N(i)$ появи події A рівно i разів у N випробуваннях?
21. Як визначити теоретичні частоти при перевірці гіпотези про біноміальний розподіл?
22. Як визначити кількість ступенів вільності при перевірці гіпотези про біноміальний розподіл за критерієм Пірсона?
23. Як за допомогою критерію Пірсона перевірити гіпотезу про біноміальний розподіл генеральної сукупності?
24. Яку величину беруть за оцінку параметра λ розподілу Пуассона?
25. Як знайти ймовірності P_i появи рівно i подій у n випробуваннях (розподіл Пуассона)?
26. Як визначити кількість ступенів вільності під час перевірки гіпотези про розподіл Пуассона за критерієм Пірсона?
27. Як за допомогою критерію Пірсона перевірити гіпотезу про розподіл Пуассона генеральної сукупності?

Задачі для самостійного розв'язування

1. Використовуючи критерій Пірсона (χ^2 – квадрат) з рівнем значущості α , перевірити гіпотезу про нормальний розподіл генеральної сукупності X , якщо відомі емпіричні n_i та теоретичні n'_i частоти.

а)

n_i	15	26	25	30	26	21	24	20	13
n'_i	9,1	16,5	25,3	32	33,9	29,8	22	13,5	7

$\alpha = 0,05$

б)

n_i	15	26	25	30	26	21	24	20	13
n'_i	9,1	16,5	25,3	32	33,9	29,8	22	13,5	7

$\alpha = 0,01$

в) $\alpha = 0,05$

n_i	8	16	40	72	36	18	10
n'_i	6	18	36	76	39	18	7

г) $\alpha = 0,05$

n_i	5	10	20	8	7
n'_i	6	14	18	7	5

2. Застосовуючи критерій Пірсона, при рівні значимості 0,05 перевірити узгодженість з нормальним розподілом генеральної сукупності X емпіричного розподілу вибірки об'єму $n = 200$:

x_i	0,3	0,5	0,7	0,9	1,1	1,3	1,5	1,7	1,9	2,1	2,3
n_i	6	9	26	25	30	26	21	24	20	8	5

3. За даним інтервальним розподілом вибірки об'єму $n = 100$ при рівні значущості $\alpha = 0,05$ за критерієм згоди Пірсона перевірити гіпотезу про нормальний розподіл генеральної сукупності.

$(x_i; x_{i+1}]$	(3,8]	(8,13]	(13,18]	(18,23]	(23,28]	(28,33]	(33,38]
n_i	6	8	15	40	16	8	7

4. Таблиця містить дані про зріст 1000 учнів старших класів. Перевірити, чи не протирічають статистичні дані нормальному закону розподілу ($\alpha = 0,01$).

Зріст, см	143-146	146-149	149-152	152-155	155-158	158-161	161-164	164-167	167-170	170-173	173-176	176-179	179-182	182-185	185-188
Число учнів	1	2	8	26	65	120	181	201	170	120	64	27	10	3	1

5. За допомогою критерію χ^2 – квадрат перевірити гіпотезу про нормальний закон розподілу у сукупності – розміри деталей після шліфування, на підставі даних, які наводяться в таблиці ($\alpha = 0,01$).

$(x_i - x_{i+1}]$	3,6–3,7	3,7–3,8	3,8–3,9	3,9–4,0	4,0–4,1	4,1–4,2	4,2–4,3	4,3–4,4
n_i	1	22	40	79	27	26	4	1

6. Перевірити з допомогою критерію χ^2 – квадрат узгодженість статистичного розподілу з нормальним розподілом при рівні значимості $\alpha = 0,01$:

$x_i - x_{i+1}$	0–3	3–6	6–9	9–12	12–15	15–18	18–21	21–24	24–27	27–30
n_i	1	3	4	6	11	10	7	5	2	1

7. Застосовуючи критерій Пірсона χ^2 при рівні значимості $\alpha = 0,05$ перевірити, чи узгоджується гіпотеза про нормальний розподіл генеральної сукупності з цим емпіричним розподілом.

$x_i - x_{i+1}$	-20 – -10	-10 – 0	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
n_i	20	47	80	89	40	16	8

8. Перевірити з допомогою критерію χ^2 узгодженість статистичних даних з рівномірним законом розподілу, рівень значимості $\alpha = 0,05$:

$(x_i - x_{i+1}]$	(-1 ; 1]	(1 ; 3]	(3 ; 5]	(5 ; 7]	(7 ; 9]
n_i	10	11	8	9	12

9. Проведено $n = 200$ випробувань, в результаті яких подія A виникла в різні моменти. Отримали емпіричний розподіл. За допомогою критерію Пірсона при $\alpha = 0,05$ перевірити гіпотезу про те, що час появи події розподілений рівномірно.

$(x_i - x_{i+1}]$	2 – 4	4 – 6	6 – 8	8 – 10	10 – 12	12 – 14	14 – 16	16 – 18	18 – 20	20 – 22
n_i	21	16	15	26	22	14	21	22	18	25

10. За спостереженнями, наведеними в таблиці, за допомогою критерію χ^2 перевірити згоду з рівномірним розподілом. У першому рядку таблиці вказано ліву границю інтервалу (i – номер інтервалу ($i; i + 1$)).

a)	i	0	1	2	3	4	5	6	7	8	9	10	11
$\alpha = 0,05$	n_i	45	41	34	54	39	43	41	33	37	41	47	39

б) $\alpha = 0,1$

i	0	1	2	3	4	5	6	7	8	9
n_i	74	92	83	79	80	73	77	75	76	91

в) $\alpha = 0,05$

i	0	1	2	3	4	5	6	7	8	9
n_i	16	15	19	13	14	19	14	11	13	16

11. За спостереженнями, наведеними в таблиці, за допомогою критерію χ^2 перевірити гіпотезу, що випадкова величина має пуассонівський розподіл:

а) $\alpha = 0,05$

x_i	0	1	2	3	4
n_i	110	65	21	3	1

б) $\alpha = 0,05$

x_i	0	1	2	3	4	5	6	7
n_i	112	168	130	68	32	5	1	1

в) $\alpha = 0,01$

x_i	0	1	2	3	4	5	6	7
n_i	8	17	16	10	6	2	0	1

12. В результаті випробування 200 елементів на тривалість роботи отримали емпіричний розподіл. Треба при рівні значимості $\alpha = 0,05$ перевірити гіпотезу про те, що час роботи елементів розподілено за показниковим законом.

$x_i - x_{i+1}$	0 – 5	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30
n_i	133	45	15	4	2	1

13. В результаті реєстрації часу приходу 800 відвідувачів виставки отримали статистичний розподіл. При рівні значимості $\alpha = 0,01$ перевірити за критерієм χ^2 гіпотезу про те, що час приходу відвідувачів виставки розподілений за показниковим законом.

$x_i - x_{i+1}$	0 – 1	1 – 2	2 – 3	3 – 4	4 – 5	5 – 6	6 – 7	7 – 8
n_i	259	167	109	74	70	47	40	34

14. Дослід, який полягає в одночасному підкиданні чотирьох монет, повторили 100 раз. Емпіричний розподіл дискретної випадкової величини X – число випавших «гербів» вийшов таким:

x_i	0	1	2	3	4
n_i	8	20	42	22	8

При рівні значимості $\alpha = 0,05$ перевірити гіпотезу про біноміальний закон за допомогою критерію χ^2 .

15. За заданими статистичними розподілами вибірки висунути H_0 про закон розподілу ознаки генеральної сукупності і при рівні значущості $\alpha = 0,01$ перевірити її правильність:

А. Вимірювалась маса новонароджених дітей у пологовому будинку. Результати вимірювань подано інтервальним статистичним розподілом:

x_i , кг, $h = 0,2$	1 – 1,2	1,2 – 1,4	1,4 – 1,6	1,6 – 1,8	1,8 – 2	2 – 2,2	2,2 – 2,4	2,4 – 2,6	2,6 – 2,8	2,8 – 3	3 – 3,2
n_i	5	12	18	22	36	24	19	15	11	9	2

Б. Протягом 50 років вимірювався рівень води навесні під час повені відносно умовного нуля. Результати вимірювання наведено у вигляді інтервального статистичного розподілу:

x_i , см, $h = 24$	0–24	24–48	48–72	72–96	96–120	120–146	146–170	170–196	196–220
n_i	1	2	4	6	12	16	6	3	1

В. Вимірювання показників виконання річного плану підприємствами певної галузі наведено у вигляді інтервального статистичного розподілу:

x_i , %, $h = 10$	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90	90–100	100–120
n_i	2	5	13	16	25	12	10	5	3	1

Г. Вимірювався місячний дохід робітника певної галузі. Результати вимірювання подано інтервальним статистичним розподілом:

x_i , грн., $h = 20$	100–120	120–140	140–160	160–180	180–200	200–220	220–240	240–260	260–280
n_i	10	15	20	25	30	40	10	4	2

16. За наведеними результатами 50-ти вимірів значень деякої неперервної випадкової величини потрібно:

- 1) згрупувати результати спостережень (побудувати інтервальний статистичний ряд);
- 2) побудувати гістограму частот та емпіричну функцію розподілу;
- 3) знайти точкові оцінки числових характеристик;
- 4) вважаючи, що досліджувана величина має нормальний закон розподілу, обчислити теоретичні частоти та побудувати теоретичну криву розподілу;
- 5) за критеріями Пірсона та Колмогорова при заданому рівні значущості перевірити гіпотезу про узгодженість емпіричних даних з теоретичними;
- 6) побудувати довірчі інтервали з певною надійністю для невідомих числових характеристик розподілу.

24	23	20	28	32	28	12	19	18	40
17	39	38	22	24	23	31	22	30	23
15	14	24	30	23	12	34	27	28	22
13	11	24	8	24	26	19	7	24	27
22	29	24	18	26	25	24	15	20	25

$$\gamma = 0,95; \alpha = 0,05.$$

Розділ 12

ЕЛЕМЕНТИ КОРЕЛЯЦІЙНОГО ТА РЕГРЕСІЙНОГО АНАЛІЗУ

Кожній величині, яку отримують у результаті проведення експерименту, притаманний елемент випадковості, що виявляється більшою чи меншою мірою залежно від її природи.

У разі сумісної появи двох і більше величин у результаті проведення експерименту дослідник має підстави для встановлення певної залежності між ними, зв'язку.

Строгій функціональній залежності між змінними, у буквальному розумінні цього слова, у реальному світі не існує, бо вони перебувають під впливом випадкових факторів, наслідки якого передбачити практично неможливо. Тому між змінними існує особлива форма зв'язку, яку називають стохастичною і яка в математичній статистиці трансформується, не змінюючи своєї сутності, у статистичну залежність.

Наприклад, при дослідженні двох змінних X та Y зміна значень $X = x_i$ призводить до такої зміни значень Y , яку можна розбити на два компоненти: систематичну, що пов'язана із залежністю, котра існує між X та Y , і випадкову, яка зазнає впливу випадкових факторів.

Показником, що вимірює стохастичний зв'язок між змінними, є *коефіцієнт кореляції*, який свідчить, з певною мірою ймовірності, наскільки зв'язок між змінними близький.

За наявності кореляційного зв'язку між змінними необхідно виявити його форму функціональної залежності (лінійна чи нелінійна), а саме:

$$y = a_0 + a_1x; \quad (12.1)$$

$$y = a_0 + a_1x + a_2x^2; \quad (12.2)$$

$$y = a_0 + \frac{a_1}{x}. \quad (12.3)$$

Наведені можливі залежності між змінними X і Y (12.1), (12.2), (12.3) називають *функціями регресії*. Форму зв'язку між змінними X і Y можна

встановити, застосовуючи кореляційні поля. Кожній точці з координатами x_i, y_i відповідає певне числове значення ознак X та Y .

Отже, на основі розміщення точок кореляційного поля дослідник має підстави для гіпотетичного припущення про лінійні чи нелінійні залежності між ознаками X і Y .

Припустимо, що нам відома функціональна залежність між випадковими величинами Y та X вигляду

$$Y = f(X; a_1; a_2; \dots; a_m) \quad (12.4)$$

з невідомими параметрами $a_1; a_2; \dots; a_m$.

Нехай внаслідок n незалежних випробувань одержані варіанти ознак Y та X , які оформлені у статистичній таблиці вигляду:

X	x_1	x_2	...	x_k	...	x_n
Y	y_1	y_2	...	y_k	...	y_n

Для знаходження оцінок параметрів функціональної залежності $a_1; a_2; \dots; a_m$ за даними вибірки застосуємо метод найменших квадратів. Цей метод ґрунтується на тому, що найімовірніші значення параметрів $a_1; a_2; \dots; a_m$ повинні давати мінімум функції

$$S = \sum_{k=1}^n [y_k - f(x_k; a_1; a_2; \dots; a_m)]^2 \quad (12.5)$$

Якщо функція $f(x_k; a_1; a_2; \dots; a_m)$ має неперервні частинні похідні відносно невідомих параметрів $a_1; a_2; \dots; a_m$, то необхідною умовою існування мінімуму функції S буде система m рівнянь з m невідомими

$$\frac{\partial S}{\partial a_k} = 0, \quad k = 1, 2, \dots, m. \quad (12.6)$$

Знаходження функціональної залежності між випадковими величинами X та Y з використанням даних випробувань (або вибірки) називають вирівнюванням емпіричних даних вздовж кривої $y = f(x; a_1; a_2; \dots; a_m)$.

12.1. Рівняння лінійної парної регресії

Нехай між змінними X та Y теоретично існує певна лінійна залежність. Це твердження може ґрунтуватися на тій підставі, наприклад, що кореляційне поле для пар $(x_i; y_i)$ має такий вигляд (рис. 12.1).

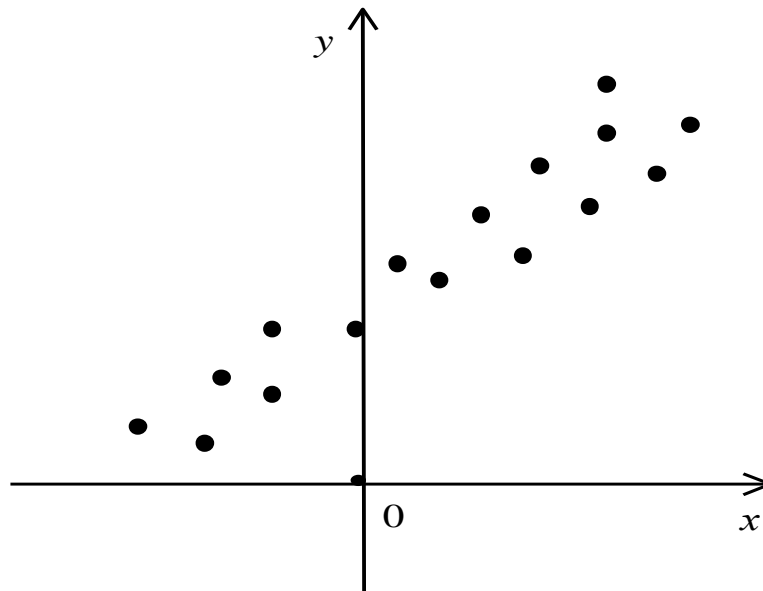


Рис. 12.1. Кореляційне поле для пар

Як бачимо, насправді між ознаками X та Y спостерігається не такий тісний зв'язок, як це передбачає функціональна залежність.

Окремі спостережувані значення y , як правило, відхилятимуться від передбаченої лінійної залежності під впливом випадкових збудників, які здебільшого є невідомими. Відхилення від передбаченої лінійної форми зв'язку можуть статися внаслідок неправильної специфікації рівняння, тобто ще з початку неправильно вибране рівняння, що описує залежність між X і Y .

Будемо вважати, що специфікація рівняння вибрана правильно. Ураховуючи вплив на значення Y збурювальних випадкових факторів, лінійне рівняння зв'язку X і Y можна подати в такому вигляді:

$$y = a_0 + a_1x + \varepsilon, \quad (12.7)$$

де a_0 , a_1 є невідомі параметри регресії, ε є випадковою змінною, що характеризує відхилення y від гіпотетичної теоретичної регресії.

Отже, в рівнянні (12.7) значення « y » подається у вигляді суми двох частин: систематичної $a_0 + a_1x$ і випадкової ε . Параметри a_0 , a_1 є невідомими величинами, а ε є випадковою величиною, що має нормальний закон розподілу з числовими характеристиками: $M(\varepsilon) = 0$; $D(\varepsilon) = const$.

У результаті статистичних спостережень дослідник дістає характеристики для незалежної змінної x і відповідні значення залежної змінної y .

Отже, необхідно визначити параметри a_0 , a_1 . Але істинні значення цих параметрів дістати неможливо, оскільки ми користуємося інформацією, здобутою від вибірки обмеженого обсягу. Тому знайдені значення параметрів будуть лише статистичними оцінками істинних (невдомих нам) параметрів a_0 , a_1 .

На практиці найчастіше оцінки невідомих параметрів a_0, a_1 визначаються за методом найменших квадратів, розробка якого належить К. Гауссу і П. Лапласу. Цей метод почали широко застосовувати в економіко-статистичних обчисленнях, відколи була створена теорія регресії.

В основі методу найменших квадратів є принцип мінімізації суми квадратів залишків моделі.

Згідно з формулою (12.5) маємо

$$S = \sum_{k=1}^n [y_k - (a_0 + a_1 x_k)]^2 \quad (12.8)$$

Ця функція S неперервно диференційовна, тому згідно з необхідними умовами існування мінімуму S повинні виконуватись рівності $\frac{\partial S}{\partial a_0} = 0$ та $\frac{\partial S}{\partial a_1} = 0$.

У нашому випадку ці рівності мають вигляд:

$$\begin{cases} \sum_{k=1}^n [y_k - (a_0 + a_1 x_k)] = 0; \\ \sum_{k=1}^n [y_k - (a_0 + a_1 x_k)] \cdot x_k = 0; \end{cases}$$

Реалізація цього принципу дає можливість отримати систему нормальних рівнянь:

$$\begin{cases} n a_0 + a_1 \sum_{k=1}^n x_k = \sum_{k=1}^n y_k; \\ a_0 \sum_{k=1}^n x_k + a_1 \sum_{k=1}^n x_k^2 = \sum_{k=1}^n x_k y_k. \end{cases} \quad (12.9)$$

У цій системі n – кількість спостережень, $\sum_{k=1}^n x_k$, $\sum_{k=1}^n y_k$, $\sum_{k=1}^n x_k^2$, $\sum_{k=1}^n x_k y_k$ – величини, які можна розрахувати на основі вихідних спостережень над змінними Y і X .

Розв'язавши систему нормальних рівнянь, одержимо оцінки невідомих параметрів моделі \hat{a}_0 і \hat{a}_1 :

$$\hat{a}_1 = \frac{(\sum_{k=1}^n x_k) \cdot (\sum_{k=1}^n y_k) - n \cdot \sum_{k=1}^n x_k y_k}{(\sum_{k=1}^n x_k)^2 - n \cdot \sum_{k=1}^n x_k^2}, \quad (12.10)$$

$$\hat{a}_0 = \frac{(\sum_{k=1}^n x_k) \cdot (\sum_{k=1}^n x_k y_k) - (\sum_{k=1}^n x_k^2) \cdot (\sum_{k=1}^n y_k)}{(\sum_{k=1}^n x_k)^2 - n \cdot \sum_{k=1}^n x_k^2}. \quad (12.11)$$

Тоді рівняння прямої регресії матиме вигляд:

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 x. \quad (12.12)$$

Якщо кількість значень x_k та y_k велика, то обчислення параметрів \hat{a}_0 та \hat{a}_1 за формулами (12.10; 12.11) ускладнюється. Для спрощення обчислень поділимо два рівняння системи (12.9) на n . Отримаємо:

$$\begin{cases} a_0 + a_1 \frac{\sum_{k=1}^n x_k}{n} = \frac{\sum_{k=1}^n y_k}{n}; \\ a_0 \frac{\sum_{k=1}^n x_k}{n} + a_1 \frac{\sum_{k=1}^n x_k^2}{n} = \frac{\sum_{k=1}^n x_k y_k}{n}; \end{cases}$$

або

$$\begin{cases} a_0 + a_1 \bar{x} = \bar{y}; \\ a_0 \bar{x} + a_1 \bar{x}^2 = \overline{xy}. \end{cases} \quad (12.13)$$

Розв'язавши систему рівнянь (12.13), отримали:

$$\hat{a}_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\bar{x}^2 - (\bar{x})^2}; \quad (12.14)$$

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}. \quad (12.15)$$

Коваріацією вибірки називається величина:

$$K(X; Y) = \overline{xy} - \bar{x} \cdot \bar{y}. \quad (12.16)$$

Коефіцієнтом кореляції називають величину

$$r = \frac{K(X; Y)}{\sigma_x \cdot \sigma_y}, \quad (12.17)$$

де $\sigma_x = \sqrt{\bar{x}^2 - (\bar{x})^2}$ – середньоквадратичне відхилення змінної X , $\sigma_y = \sqrt{\bar{y}^2 - (\bar{y})^2}$ – середньоквадратичне відхилення змінної Y .

Коефіцієнт кореляції оцінює залежність між X та Y і має такі властивості:

1. $-1 \leq r \leq 1$ – залежність між X та Y тим сильніша, чим значення r ближче до 1. Якщо $r > 0$, то із зростанням X зростає і Y (кореляція додатна). Якщо $r < 0$, то при зростанні X величина Y у середньому спадає (кореляція від'ємна).

2. $r = \pm 1$ тоді і лише тоді, коли Y є лінійною функцією від X і навпаки.

3. Якщо статистичні змінні X та Y незалежні, то $r = 0$ ($K(X; Y) = 0$). Такі змінні X та Y , для яких $r = 0$, називають некорельованими, а для яких $r \neq 0$ – корельованими.

Коефіцієнт кореляції служить для оцінки тісноти лінійного зв'язку між випадковими змінними X та Y : чим $|r|$ ближче до 1, тим зв'язок сильніший, чим ближче $|r|$ до нуля – слабший. Від'ємний знак свідчить про обернений зв'язок, додатній – про прямиий.

Коефіцієнтом детермінації називається величина R^2 , яка визначається за формулою:

$$R^2 = r^2. \quad (12.18)$$

На основі коефіцієнта детермінації R^2 можна зробити висновок про ступінь значущості вимірюваного зв'язку на основі лінійної регресії. $R^2 \in [0; 1]$.

Оскільки коефіцієнт детермінації R^2 характеризує, якою мірою варіація залежної змінної визначається варіацією незалежної змінної, то що ближче R^2 до одиниці, то суттєвішим є зв'язок між цими змінними.

Приклад 12.1. Залежність обсягу отриманого прибутку деяким умовним підприємством регіону від вартості основних виробничих фондів наведено парним статистичним розподілом вибірки:

Основні фонди, млн грн, x_k	2,5	2,8	3	3,2	3,5	4,2	4,5	5	5,3	6
Прибуток, млн грн, y_k	1,2	1,5	1,7	2,2	2,6	3,1	3,4	4,2	4,7	5,4

Методом найменших квадратів визначити оцінки невідомих параметрів лінійної парної регресії. Обчислити коефіцієнт кореляції та детермінації, зробити висновки.

Розв'язання. З таблиці бачимо, що зі збільшенням значень ознаки X залежна змінна Y має тенденцію до збільшення.

Тому припускаємо, що між ознаками X та Y існує лінійна функціональна залежність

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 x.$$

Для визначення параметрів \hat{a}_0 та \hat{a}_1 скористаємося розрахунковою таблицею, що має такий вигляд:

№ з/п	x_k	y_k	x_k^2	$x_k y_k$	y_k^2
1.	2,5	1,2	6,25	3,0	1,44
2.	2,8	1,5	7,84	4,2	2,25
3.	3	1,7	9	5,1	2,89
4.	3,2	2,2	10,24	7,0	4,84
5.	3,5	2,6	12,25	9,1	6,76
6.	4,2	3,1	17,64	13,0	9,61
7.	4,5	3,4	20,25	15,3	11,56
8.	5	4,2	25	21,0	17,64
9.	5,3	4,7	28,09	24,9	22,09
10	6	5,4	36	32,4	29,16
Σ	40	30	172,56	135,07	108,24

Скориставшись формулами (12.10; 12.11), де $n = 10$, отримаємо:

$$\hat{a}_1 = \frac{(\sum_{k=1}^n x_k) \cdot (\sum_{k=1}^n y_k) - n \cdot \sum_{k=1}^n x_k y_k}{(\sum_{k=1}^n x_k)^2 - n \cdot \sum_{k=1}^n x_k^2} = \frac{40 \cdot 30 - 10 \cdot 135,07}{40^2 - 10 \cdot 172,56} = 1,12;$$

$$\hat{a}_0 = \frac{(\sum_{k=1}^n x_k) \cdot (\sum_{k=1}^n x_k y_k) - (\sum_{k=1}^n x_k^2) \cdot (\sum_{k=1}^n y_k)}{(\sum_{k=1}^n x_k)^2 - n \cdot \sum_{k=1}^n x_k^2} = \frac{40 \cdot 135,07 - 172,56 \cdot 30}{40^2 - 10 \cdot 172,56} = -1,799.$$

Отже, рівняння регресії буде таким:

$$\hat{Y} = -1,799 + 1,12x.$$

Для обчислення r необхідно знайти $K(X; Y)$, σ_x , σ_y :

$$K(X; Y) = \bar{x}\bar{y} - \bar{x} \cdot \bar{y} = 13,507 - 4 \cdot 3 = 1,507;$$

$$\sigma_x = \sqrt{\overline{x^2} - (\bar{x})^2} = \sqrt{17,256 - 4^2} = \sqrt{1,256} = 1,12;$$

$$\sigma_y = \sqrt{\overline{y^2} - (\bar{y})^2} = \sqrt{10,824 - 3^2} = \sqrt{1,824} = 1,35;$$

$$r = \frac{K(X; Y)}{\sigma_x \cdot \sigma_y} = \frac{1,507}{1,12 \cdot 1,35} = 0,996;$$

$$R^2 = r^2 = 0,996^2 = 0,992.$$

Як бачимо, коефіцієнт кореляції близький за своїм значенням до одиниці, що свідчить про те, що залежність між X та Y є практично лінійною.

Коефіцієнт детермінації $R^2 = 0,992$. Це означає, що зміна обсягу прибутку підприємства на 99,2% визначається варіацією вартості основних фондів, і 0,8% – іншими випадковими факторами.

12.2. Парна нелінійна регресія

Якщо відображені на площині XOY групи точок $(x_i; y_i)$, розміщуються, нагадуючи деякі криві, то доцільно вважати, що між досліджуваними величинами існує нелінійна залежність. Тепер знову виникло завдання підібрати таку криву, яка б на основі методу найменших квадратів мала найменші відхилення від точок, здобутих при спостереженні, знайти її рівняння і визначити тісноту зв'язку.

Розглянемо деякі найпростіші види нелінійної кореляційної залежності.

Нехай зі зростанням однієї випадкової величини умовні середні другої зростають (спадають), досягають максимуму (мінімуму), а потім спадають (зростають). Тоді можна вважати, що між ними існує *параболічна залежність* виду:

$$y = a_0 + a_1x + a_2x^2 + \varepsilon. \quad (12.19)$$

Методом найменших квадратів на основі даних випробувань необхідно оцінити невідомі параметри a_0, a_1, a_2 . Для цієї залежності формула (12.5) матиме вигляд:

$$S = \sum_{k=1}^n [y_k - (a_0 + a_1x_k + a_2x_k^2)]^2. \quad (12.20)$$

Необхідні умови існування мінімуму функції S є рівності нулю частинних похідних першого порядку: $\frac{\partial S}{\partial a_0} = 0$; $\frac{\partial S}{\partial a_1} = 0$; $\frac{\partial S}{\partial a_2} = 0$.

У нашому випадку:

$$\begin{cases} -2 \sum_{k=1}^n [y_k - (a_0 + a_1x_k + a_2x_k^2)]x_k^2 = 0; \\ -2 \sum_{k=1}^n [y_k - (a_0 + a_1x_k + a_2x_k^2)]x_k = 0; \\ -2 \sum_{k=1}^n [y_k - (a_0 + a_1x_k + a_2x_k^2)] = 0. \end{cases}$$

Реалізація цього принципу дає можливість отримати систему нормальних рівнянь:

$$\begin{cases} a_2 \sum_{k=1}^n x_k^4 + a_1 \sum_{k=1}^n x_k^3 + a_0 \sum_{k=1}^n x_k^2 = \sum_{k=1}^n x_k^2 y_k; \\ a_2 \sum_{k=1}^n x_k^3 + a_1 \sum_{k=1}^n x_k^2 + a_0 \sum_{k=1}^n x_k = \sum_{k=1}^n x_k y_k; \\ a_2 \sum_{k=1}^n x_k^2 + a_1 \sum_{k=1}^n x_k + n \cdot a_0 = \sum_{k=1}^n y_k. \end{cases} \quad (12.21)$$

Система (12.21) є неоднорідною лінійною системою трьох рівнянь з невідомими a_0, a_1, a_2 . Розв'язок цієї системи можна знайти різними методами (матричним, за правилом Крамера, методом Гаусса, а його вигляд буде громіздкий при доволі великій кількості випробувань n).

Розв'язавши систему нормальних рівнянь, одержимо оцінки невідомих параметрів моделі $\hat{a}_0; \hat{a}_1; \hat{a}_2$.

Тоді рівняння регресії матиме вигляд:

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 x + \hat{a}_2 x^2. \quad (12.22)$$

Приклад 12.2. За даними (див. табл.) про господарську діяльність десяти підприємств побудувати парну кореляційно-регресійну модель залежності обсягів виробництва (y – показник) від основних виробничих фондів (x – фактор). Встановити форму зв'язку та математичне рівняння зв'язку шляхом побудови графіка кореляційної залежності. Знайти оцінки параметрів рівняння парної параболічної регресії $\hat{a}_0; \hat{a}_1; \hat{a}_2$.

x	2,8	4,2	5	5,4	5,8	7	8,2	9	10	10,2
y	7	9	9,8	11	10	13	12	11,4	8	6,6

Розв'язання. З таблиці бачимо, що зі збільшенням значень ознаки X залежна змінна Y має тенденцію до зростання, досягає максимуму, а потім спадає.

Тому припускаємо, що між ознаками X та Y існує параболічна функціональна залежність:

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 x + \hat{a}_2 x^2.$$

Для визначення параметрів \hat{a}_0 , \hat{a}_1 та \hat{a}_2 скористаємося розрахунковою таблицею, що має такий вигляд:

№ з/п	x_k	y_k	x_k^2	$x_k y_k$	y_k^2	x_k^3	x_k^4	$x_k^2 y_k$
1.	2,8	7	7,84	19,6	49	21,952	61,4656	54,88
2.	4,2	9	17,64	37,8	81	74,088	311,17	158,76
3.	5	9,8	25	49	96,04	125	625	245
4.	5,4	11	29,16	59,4	121	157,464	850,306	320,76
5.	5,8	10	33,64	58	100	195,112	1131,65	336,4
6.	7	13	49	91	169	343	2401	637
7.	8,2	12	67,24	98,4	144	551,368	4521,22	806,88
8.	9	11,4	81	102,6	129,96	729	6561	923,4
9.	10	8	100	80	64	1000	10000	800
10.	10,2	6,6	104,04	67,32	43,56	1061,21	10824,3	686,664
Σ	67,6	97,8	514,56	663,12	997,56	4258,19	37287,1	4969,74

Підставивши у систему (12.21) дані з розрахункової таблиці, отримаємо:

$$\begin{cases} 37287,1a_2 + 4258,19a_1 + 514,56a_0 = 4969,74; \\ 4258,19a_2 + 514,56a_1 + 67,6a_0 = 663,12; \\ 514,56a_2 + 67,6a_1 + 10a_0 = 97,8. \end{cases}$$

Розв'язавши систему нормальних рівнянь, одержимо оцінки невідомих параметрів моделі \hat{a}_0 ; \hat{a}_1 ; \hat{a}_2 .

$$\hat{a}_0 = -4,78;$$

$$\hat{a}_1 = 4,87;$$

$$\hat{a}_2 = -0,36.$$

Отже, рівняння регресії буде таким:

$$\hat{Y} = -4,78 + 4,87x - 0,36x^2.$$

Нехай зі зростанням однієї випадкової величини умовні середні другої спадають, але не на ту ж величину, як це буває в разі лінійної залежності, а розмір зміни ніби згасає. У такому разі можна вважати, що залежність *гіперболічна*:

$$y = a_0 + \frac{a_1}{x} + \varepsilon. \quad (12.23)$$

Методом найменших квадратів на основі даних випробувань необхідно оцінити невідомі параметри a_0, a_1 . Для цієї залежності формула (12.5) матиме вигляд:

$$S = \sum_{k=1}^n \left[y_k - \left(a_0 + \frac{a_1}{x_k} \right) \right]^2. \quad (12.24)$$

Необхідні умови існування мінімуму функції S є рівності нулю частинних похідних першого порядку: $\frac{\partial S}{\partial a_0} = 0; \frac{\partial S}{\partial a_1} = 0$.

У нашому випадку:

$$\begin{cases} \sum_{k=1}^n \left[y_k - \left(a_0 + \frac{a_1}{x_k} \right) \right] = 0; \\ \sum_{k=1}^n \left[y_k - \left(a_0 + a_1 \frac{1}{x_k} \right) \right] \cdot \frac{1}{x_k} = 0. \end{cases}$$

Реалізація цього принципу дає можливість отримати систему нормальних рівнянь:

$$\begin{cases} na_0 + a_1 \sum_{k=1}^n \frac{1}{x_k} = \sum_{k=1}^n y_k; \\ a_0 \sum_{k=1}^n \frac{1}{x_k} + a_1 \sum_{k=1}^n \frac{1}{x_k^2} = \sum_{k=1}^n \frac{y_k}{x_k}. \end{cases} \quad (12.25)$$

Система (12.25) є неоднорідною лінійною системою двох рівнянь з невідомими a_0, a_1 .

Розв'язавши систему нормальних рівнянь, одержимо оцінки невідомих параметрів моделі $\hat{a}_0; \hat{a}_1$.

Тоді рівняння регресії матиме вигляд:

$$\hat{Y} = \hat{a}_0 + \frac{\hat{a}_1}{x}. \quad (12.26)$$

Приклад 12.3. Підприємства легкої промисловості регіону отримали інформацію, що характеризує залежність обсягу випуску продукції (Y , млн грн) від обсягу капіталовкладень (X , млн грн). Установити форму залежності між X і Y , знайти рівняння регресії і оцінити тісноту зв'язку.

x_k	3	7	7	10	12	14	17	20	21	22
y_k	13	15	19	22	21	20	26	30	26	27

Розв'язання. Припускаємо, що між ознаками X та Y існує гіперболічна функціональна залежність

$$\hat{Y} = \hat{a}_0 + \frac{\hat{a}_1}{x}$$

Для визначення параметрів \hat{a}_0 та \hat{a}_1 скористаємося розрахунковою таблицею, що має такий вигляд:

№ з/п	x_k	y_k	$\frac{1}{x_k}$	$\frac{1}{x_k^2}$	$\frac{y_k}{x_k}$	y_k^2
1.	3	13	0,333333	0,111111	4,333333	169
2.	7	15	0,142857	0,020408	2,142857	225
3.	7	19	0,142857	0,020408	2,714286	361
4.	10	22	0,1	0,01	2,2	484
5.	12	21	0,083333	0,006944	1,75	441
6.	14	20	0,071429	0,005102	1,428571	400
7.	17	26	0,058824	0,00346	1,529412	676
8.	20	30	0,05	0,0025	1,5	900
9.	21	26	0,047619	0,002268	1,238095	676
10.	22	27	0,045455	0,002066	1,227273	729
Σ	133	219	1,075707	0,184268	20,06383	5061

Скориставшись формулами (12.25), отримаємо систему нормальних рівнянь:

$$\begin{cases} 10a_0 + 1,075707a_1 = 219; \\ 1,075707a_0 + 0,184268a_1 = 20,06383. \end{cases}$$

Розв'язавши цю систему нормальних рівнянь, одержимо оцінки невідомих параметрів моделі $\hat{a}_0; \hat{a}_1$.

$$\hat{a}_0 = 27,38;$$

$$\hat{a}_1 = -50,97.$$

Отже, рівняння регресії буде таким:

$$\hat{Y} = 27,38 - \frac{50,97}{x}.$$

Для обчислення r необхідно знайти $K(X; Y)$, σ_x , σ_y :

$$K(X; Y) = \left(\frac{y}{x}\right) - \frac{1}{\bar{x}} \cdot \bar{y} = 2,006383 - 0,1075707 \cdot 21,9 = -0,3494;$$

$$\sigma_x = \sqrt{\left(\frac{1}{x^2}\right) - \left(\frac{1}{x}\right)^2} = \sqrt{0,184268 - 0,1075707^2} = \sqrt{0,0068553} = 0,082797;$$

$$\sigma_y = \sqrt{y^2 - (\bar{y})^2} = \sqrt{506,1 - 21,9^2} = \sqrt{26,49} = 5,146844;$$

$$r = \frac{K(X; Y)}{\sigma_x \cdot \sigma_y} = \frac{-0,3494}{0,082797 \cdot 5,146844} \approx -0,82.$$

Як бачимо, коефіцієнт кореляції $|r|$ близький до одиниці, що свідчить про те, що зв'язок між X та Y є тісним.

Нехай зі зростанням однієї випадкової величини умовні середні другої зростають (спадають), але не на ту ж величину, як це буває в разі лінійної залежності, а розмір зміни ніби прискорюється (згасає). У такому разі можна вважати, що залежність **показникова**:

$$y = a_0 \cdot a_1^x + \varepsilon. \quad (12.27)$$

Для оцінки невідомих параметрів a_0 , a_1 рівняння показникової регресії, насамперед зведемо її до лінійного вигляду прологарифмувавши ліву і праву частину рівняння (12.27).

$$\ln y = \ln(a_0 \cdot a_1^x) \Rightarrow \ln y = \ln a_0 + x \ln a_1.$$

Введемо нові змінні $y' = \ln y$, $b_0 = \ln a_0$; $b_1 = \ln a_1$. Тоді рівняння регресії матиме вигляд:

$$y' = b_0 + b_1 \cdot x. \quad (12.28)$$

Методом найменших квадратів оцінимо невідомі параметри b_0, b_1 , розв'язавши систему нормальних рівнянь:

$$\begin{cases} nb_0 + b_1 \sum_{k=1}^n x_k = \sum_{k=1}^n y'_k; \\ b_0 \sum_{k=1}^n x_k + b_1 \sum_{k=1}^n x_k^2 = \sum_{k=1}^n x_k y'_k. \end{cases} \quad (12.29)$$

Розв'язавши систему нормальних рівнянь (12.29), одержимо оцінки невідомих параметрів моделі \hat{b}_0 ; \hat{b}_1 .

Оцінки \hat{a}_0 та \hat{a}_1 невідомих параметрів здійснимо за формулами:

$$\hat{a}_0 = e^{\hat{b}_0}; \quad \hat{a}_1 = e^{\hat{b}_1}.$$

Тоді рівняння регресії матиме вигляд:

$$\hat{Y} = \hat{a}_0 \cdot \hat{a}_1^x. \quad (12.30)$$

Приклад 12.4. Дослідити виробничий процес у регіоні (табл.) за допомогою класичної виробничої моделі, що описує залежність між обсягом валової продукції (y) та обсягом основного капіталу (x). Установити форму залежності між X і Y , знайти рівняння регресії і оцінити тісноту зв'язку.

x_k	60	64	72	76	78	82	84	92	94	98
y_k	6,4	6,8	7,8	10,4	15,4	25,4	30,4	40,8	50,4	56,2

Розв'язання. Припускаємо, що між ознаками X та Y існує показникові функціональна залежність

$$\hat{Y} = \hat{a}_0 \cdot \hat{a}_1^x.$$

Прологарифмуємо ліву і праву частину рівняння та введемо додаткові позначення.

$$\ln y = \ln(a_0 \cdot a_1^x) \Rightarrow \ln y = \ln a_0 + x \ln a_1$$

$$y' = \ln y, \quad b_0 = \ln a_0; \quad b_1 = \ln a_1.$$

Тоді рівняння регресії матиме вигляд:

$$y' = b_0 + b_1 \cdot x.$$

Для визначення параметрів b_0 та b_1 скористаємося розрахунковою таблицею, що має такий вигляд:

№ з/п	x_k	y_k	$y'_k = \ln y_k$	x_k^2	$y'_k \cdot x_k$	$(y'_k)^2$
1.	60	6,4	1,856298	3600	111,3779	3,445842
2.	64	6,8	1,916923	4096	122,683	3,674592
3.	72	7,8	2,054124	5184	147,8969	4,219424
4.	76	10,4	2,341806	5776	177,9772	5,484054
5.	78	15,4	2,734368	6084	213,2807	7,476766
6.	82	25,4	3,234749	6724	265,2494	10,4636
7.	84	30,4	3,414443	7056	286,8132	11,65842
8.	92	40,8	3,708682	8464	341,1988	13,75432
9.	94	50,4	3,919991	8836	368,4792	15,36633
10.	98	56,2	4,028917	9604	394,8338	16,23217
Σ	800	250	29,2103	65424	2429,79	91,77552

$$b_1 = \frac{(\sum_{k=1}^n x_k) \cdot (\sum_{k=1}^n y'_k) - n \cdot \sum_{k=1}^n x_k y'_k}{(\sum_{k=1}^n x_k)^2 - n \cdot \sum_{k=1}^n x_k^2} = \frac{800 \cdot 29,2103 - 10 \cdot 2429,79}{800^2 - 10 \cdot 65424} = 0,065285;$$

$$b_0 = \frac{(\sum_{k=1}^n x_k) \cdot (\sum_{k=1}^n x_k y'_k) - (\sum_{k=1}^n x_k^2) \cdot (\sum_{k=1}^n y'_k)}{(\sum_{k=1}^n x_k)^2 - n \cdot \sum_{k=1}^n x_k^2} = \frac{800 \cdot 2429,79 - 65424 \cdot 29,2103}{800^2 - 10 \cdot 65424} = -2,301779.$$

Знайдемо оцінки \hat{a}_0 та \hat{a}_1 невідомих параметрів, використовуючи такі формули:

$$\hat{a}_0 = e^{b_0} = e^{-2,301779} = 0,1001;$$

$$\hat{a}_1 = e^{b_1} = e^{0,065285} = 1,0675.$$

Тоді рівняння регресії матиме вигляд:

$$\hat{Y} = 0,1001 \cdot 1,0675^x.$$

Для оцінки тісноти зв'язку використовуємо коефіцієнт кореляції:

$$r = \frac{K(X;Y)}{\sigma_x \cdot \sigma_{y'}} = \frac{9,2966}{11,93314711 \cdot 0,803203423} = 0,97;$$

$$K(X; Y') = \overline{xy'} - \bar{x} \cdot \bar{y}' = 242,979 - 80 \cdot 2,92103 = 9,2966;$$

$$\sigma_x = \sqrt{\overline{x^2} - (\bar{x})^2} = \sqrt{6542,4 - 80^2} = 11,93314711;$$

$$\sigma_y = \sqrt{\overline{(y')^2} - (\bar{y}')^2} = \sqrt{9,177552 - 2,92103^2} = 0,803203423.$$

Як бачимо, коефіцієнт кореляції близький за своїм значенням до одиниці, що свідчить про те, що зв'язок між X та Y є тісним.

12.3. Множинна лінійна регресія

На практиці здебільшого залежна змінна Y пов'язана з впливом не одного, а кількох аргументів. У цьому разі регресію називають множинною. Водночас якщо аргументи в функції регресії в першій степені, то множинна регресія називається *лінійною*, в іншому разі – *множинною нелінійною регресією*.

Лінійна множинна регресія

Визначення статистичних точкових оцінок

Розглянемо лінійну залежність Y від m аргументів (X_1, X_2, \dots, X_m) .

Лінійна модель у цьому разі набуває такого вигляду:

$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + \dots + a_m x_{im}. \quad (12.31)$$

Для вибірки обсягу n матимемо систему лінійних рівнянь

$$\begin{aligned}
 y_1 &= a_0 + a_1x_{11} + a_2x_{12} + \dots + a_mx_{1m} + \varepsilon_1; \\
 y_2 &= a_0 + a_1x_{21} + a_2x_{22} + \dots + a_mx_{2m} + \varepsilon_2; \\
 y_3 &= a_0 + a_1x_{31} + a_2x_{32} + \dots + a_mx_{3m} + \varepsilon_3; \\
 &\dots \\
 y_n &= a_0 + a_1x_{n1} + a_2x_{n2} + \dots + a_mx_{nm} + \varepsilon_n.
 \end{aligned}
 \tag{12.32}$$

Параметри рівняння (12.31) є величинами сталими, але невідомими. Ці параметри оцінюють статистичними точковими оцінками $\hat{a}_0; \hat{a}_1; \hat{a}_2; \dots; \hat{a}_m$, які отримують шляхом обробки результатів вибірки, i є величинами випадковими.

Для визначення компонентів $\hat{a}_0; \hat{a}_1; \hat{a}_2; \dots; \hat{a}_m$ (статистичних точкових оцінок) застосовується метод найменших квадратів. Необхідно, щоб сума квадратів відхилень фактичних даних від теоретичних була мінімальною. Цю вимогу можна представити так:

$$S = \sum_{i=1}^n [y_i - (a_0 + a_1x_{i1} + a_2x_{i2} + \dots + a_mx_{im})]^2. \tag{12.33}$$

Ця функція S неперервно диференційовна, тому згідно з необхідними умовами існування мінімуму S повинні виконуватись рівності частинних похідних нулю.

У нашому випадку ці рівності мають вигляд:

$$\left\{ \begin{aligned}
 &\sum_{i=1}^n [y_i - (a_0 + a_1x_{i1} + a_2x_{i2} + \dots + a_mx_{im})] = 0; \\
 &\sum_{i=1}^n [y_i - (a_0 + a_1x_{i1} + a_2x_{i2} + \dots + a_mx_{im})] \cdot x_{i1} = 0; \\
 &\sum_{i=1}^n [y_i - (a_0 + a_1x_{i1} + a_2x_{i2} + \dots + a_mx_{im})] \cdot x_{i2} = 0; \\
 &\dots \\
 &\sum_{i=1}^n [y_i - (a_0 + a_1x_{i1} + a_2x_{i2} + \dots + a_mx_{im})] \cdot x_{im} = 0.
 \end{aligned} \right.$$

Знайдемо частинну похідну окресленого виразу за компонентами вектора A і прирівняємо до нуля:

$$\frac{\partial S}{\partial A} = -2X'Y + 2X'XA = 0.$$

Звідси, отримаємо систему рівнянь у матричній формі, якій повинен задовольняти вектор A при дотриманні вимоги:

$$X'XA = X'Y. \quad (12.36)$$

Якщо до матриці $X'X$ існує обернена матриця $(X'X)^{-1}$, то отримаємо розв'язком системи нормальних рівнянь вектор-стовпець шуканих оцінок параметрів регресії:

$$A = (X'X)^{-1} \cdot (X'Y). \quad (12.37)$$

На відміну від простої моделі регресії алгоритм визначення параметрів багатофакторної моделі є більш складним та трудомістким.

Приклад 12.5. На основі наведених у таблиці даних залежність обсягу отриманого прибутку підприємствами регіону від розміру основних виробничих фондів та затрат праці побудувати множинну лінійну регресійну модель.

Таблиця

Номер підприємства	Прибуток, млн грн, y	Основні фонди, млн грн, x_1	Затрати праці, млн днів, x_2
1	1,2	2,5	4,0
2	1,5	2,8	4,2
3	1,9	3,0	3,6
4	2,2	3,6	4,6
5	2,8	3,9	4,3
6	3,1	4,2	5,1
7	3,4	4,5	5,3
8	4,5	5,0	4,8
9	4,8	5,6	5,4
10	5,4	6,0	5,8

Розв'язування. Попередній аналіз вхідної інформації дає можливість зробити висновок про наявність лінійної форми зв'язку між вибраними економічними показниками:

$$\hat{y} = a_0 + a_1x_1 + a_2x_2,$$

де y – прибуток, млн грн; x_1 – вартість основних виробничих фондів, млн грн; x_2 – затрати праці, млн днів.

Для знаходження оцінок параметрів моделі використаємо математичний апарат матричної алгебри.

Введемо позначення:

$$A = \begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{a}_2 \end{pmatrix}; \quad Y = \begin{pmatrix} 1,2 \\ 1,5 \\ 1,9 \\ 2,2 \\ 2,8 \\ 3,1 \\ 3,4 \\ 4,5 \\ 4,8 \\ 5,4 \end{pmatrix}; \quad X = \begin{pmatrix} 1 & 2,5 & 4,0 \\ 1 & 2,8 & 4,2 \\ 1 & 3,0 & 3,6 \\ 1 & 3,6 & 4,6 \\ 1 & 3,9 & 4,3 \\ 1 & 4,2 & 5,1 \\ 1 & 4,5 & 5,3 \\ 1 & 5,0 & 4,8 \\ 1 & 5,6 & 5,4 \\ 1 & 6,0 & 5,8 \end{pmatrix}.$$

Вектор-стовпець шуканих оцінок параметрів регресії знайдемо, користуючись формулою:

$$A = (X'X)^{-1} \cdot (X'Y),$$

де X' – матриця, транспонована до матриці X .

1. Знаходимо добуток двох матриць:

$$X'X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2,5 & 2,8 & 3,0 & 3,6 & 3,9 & 4,2 & 4,5 & 5,0 & 5,6 & 6,0 \\ 4,0 & 4,2 & 3,6 & 4,6 & 4,3 & 5,1 & 5,3 & 4,8 & 5,4 & 5,8 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2,5 & 4,0 \\ 1 & 2,8 & 4,2 \\ 1 & 3,0 & 3,6 \\ 1 & 3,6 & 4,6 \\ 1 & 3,9 & 4,3 \\ 1 & 4,2 & 5,1 \\ 1 & 4,5 & 5,3 \\ 1 & 5,0 & 4,8 \\ 1 & 5,6 & 5,4 \\ 1 & 6,0 & 5,8 \end{pmatrix} = \begin{pmatrix} 10 & 14,1 & 47,1 \\ 41,1 & 181,5 & 200,2 \\ 47,1 & 200,2 & 226,2 \end{pmatrix}.$$

2. Знаходимо обернену матрицю до матриці $(X'X)$:

$$(X'X)^{-1} = \begin{pmatrix} 8,92 & 1,22 & -2,93 \\ 1,22 & 0,4 & -0,61 \\ -2,93 & -0,61 & 1,51 \end{pmatrix}.$$

3. Знаходимо добуток матриць X' та Y :

$$X'Y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2,5 & 2,8 & 3,0 & 3,6 & 3,9 & 4,2 & 4,5 & 5,0 & 5,6 & 6,0 \\ 4,0 & 4,2 & 3,6 & 4,6 & 4,3 & 5,1 & 5,3 & 4,8 & 5,4 & 5,8 \end{pmatrix} \cdot \begin{pmatrix} 1,2 \\ 1,5 \\ 1,9 \\ 2,2 \\ 2,8 \\ 3,1 \\ 3,4 \\ 4,5 \\ 4,8 \\ 5,4 \end{pmatrix} = \begin{pmatrix} 30,8 \\ 141,84 \\ 152,77 \end{pmatrix}.$$

4. Знаходимо вектор-стовпець шуканих оцінок параметрів регресії:

$$A = (X'X)^{-1} \cdot (X'Y) = \begin{pmatrix} 8,92 & 1,22 & -2,93 \\ 1,22 & 0,4 & -0,61 \\ -2,93 & -0,61 & 1,51 \end{pmatrix} \cdot \begin{pmatrix} 30,8 \\ 141,84 \\ 152,77 \end{pmatrix} = \begin{pmatrix} -0,97 \\ 1,4 \\ -0,37 \end{pmatrix}.$$

Отже, нами отримано таку множинну лінійну регресійну модель:

$$\hat{y} = -0,97 + 1,4x_1 - 0,37x_2.$$

Запитання для самоконтролю

1. Дати визначення статистичної залежності між ознаками X та Y .
2. Що означає кореляційна залежність між ознаками X та Y ?
3. Що називають функціями регресії?
4. Лінійне рівняння зв'язку X і Y .
5. Яким методом найчастіше визначають оцінки невідомих параметрів?
6. Який принцип лежить в основі методу найменших квадратів?
7. Який вигляд має функція S у випадку парної лінійної регресії?
8. Необхідними умовами існування мінімуму S ?
9. У якому випадку можна вважати, що між X та Y існує лінійна залежність?
 10. Записати модель парної лінійної регресії.
 11. Коваріацією вибірки називається?
 12. Коефіцієнтом кореляції називають?
 13. Залежність між X та Y тим сильніша, чим значення ... ?
 14. Який кореляційний зв'язок називається прямим; оберненим?
 15. У яких межах знаходиться коефіцієнт кореляції?
 16. Якщо статистичні змінні X та Y незалежні, то r дорівнює?
 17. Коефіцієнтом детермінації називається величина R^2 ...
 18. У яких межах знаходиться коефіцієнт детермінації?
 19. У якому випадку можна вважати, що між X та Y існує параболічна залежність?
 20. Записати модель парної регресії при параболічній залежності.
 21. Який вигляд має функція S у випадку параболічної залежності?
 22. У якому випадку можна вважати, що між X та Y існує гіперболічна залежність?
 23. Записати модель парної регресії при гіперболічній залежності.
 24. Який вигляд має функція S у випадку гіперболічної залежності?
 25. У якому випадку можна вважати, що між X та Y існує показникова залежність?
 26. Записати модель парної регресії при показниковій залежності.
 27. Який вигляд має функція S у випадку показникової залежності?
 28. Якщо залежна змінна Y пов'язана з впливом не одного, а кількох аргументів, то регресію називають ...?
 29. Записати модель множинної лінійної регресії.
 30. Записати у векторно-матричній формі модель лінійної множинної регресії.
 31. Який вигляд має функція S у випадку множинної лінійної регресії?
 32. Вектор-стовпець оцінок параметрів множинної лінійної регресії знаходять за формулою.

Задачі для самостійного розв'язування

1. За дослідними даними методом найменших квадратів визначити оцінки невідомих параметрів лінійної парної регресії. Обчислити коефіцієнт кореляції та детермінації, зробити висновки.

а)

x_k	0,5	1,5	2,5	3,5
y_k	1,32	0,81	0,18	-0,46

б)

x_k	1	2	3	4
y_k	1,1	0,48	-0,1	-0,8

в)

x_k	2	4	6	8
y_k	0,81	2,3	3,4	4,54

г)

x_k	0	0,5	1	1,5	2	2,5	3	3,5	4	4,5
y_k	1,67	1,32	1,1	0,81	0,48	0,18	-0,1	-0,46	-0,8	-1,15

2. За дослідними даними методом найменших квадратів визначити оцінки невідомих параметрів рівняння регресії при параболічній залежності.

а)

x_k	0	2	4	6	8	10
y_k	5	-1	0,5	1,5	1,5	8,5

б)

x_k	-4	-3	-2	-1	0	1	2	3	4
y_k	6	3	1	0,3	-0,1	-0,2	0	0,2	1

в)

x_k	-3	-2	-1	0	1	2	3
y_k	-0,7	0	0,5	0,4	0,9	0,8	0,5

г)

x_k	0	1	2	3	4	5	6	7	8
y_k	0,1	0,48	0,81	1,26	2,3	2,85	3,4	3,96	4,54

3. За дослідними даними методом найменших квадратів визначити оцінки невідомих параметрів рівняння регресії при гіперболічній залежності.

а)

x_k	0,25	0,5	1	2
y_k	6	4	3	3

б)

x_k	1	2	5	10
y_k	10	5	2	1

в)

x_k	8	9	10	11	12	15	18	20	22	24
y_k	50	45	42	36	30	15	12	8	7	5

4. За дослідними даними методом найменших квадратів визначити оцінки невідомих параметрів рівняння регресії при показниковій залежності.

а)

x_k	48	50	54	58	60	62	64	66	68	70
y_k	18	18	19	26	25	26	30	28	26	34

б)

x_k	1	1,2	1,3	1,7	2,5	2,8	3	3,5	3,8	4,2
y_k	10	15	20	25	30	36	40	42	40	42

5. Маючи статистичні дані про результати господарської діяльності підприємств побудувати багатofакторну лінійну модель залежності показника y (чистий прибуток підприємств, млн грн) від двох факторів: x_1 (основні фонди, млн грн), x_2 (оборотні фонди, млн грн).

y	13	8	15	7	18	21	25	26	28	30	33	38
x_1	21	19	18	15	24	23	21	24	25	26	25	29
x_2	16	7	11	9	15	17	19	21	20	22	21	20

6. Підприємство, що складається з багатьох філій, досліджує залежність свого річного товарообігу y (млн у. о.) від торгової площі своїх філій x_1 (тис. кв. м) і середньоденної інтенсивності потоку покупців x_2 (тис. чол./день). Побудувати багатofакторну лінійну модель залежності. Дані за філіями наведено в таблиці.

Номер філії	Річний товарообіг, (млн у.о.)	Торгова площа філії, тис. кв. м	Середньоденна інтенсивність потоку покупців, тис. чол./день
1.	2,93	0,31	10,24
2.	5,27	0,98	7,15
3.	6,85	1,21	10,81
4.	7,01	1,29	9,89
5.	7,02	1,12	13,72
6.	8,35	1,49	13,92
7.	4,33	0,78	8,54
8.	5,77	0,94	12,36
9.	7,68	1,29	12,27
10.	3,16	0,48	11,01
11.	1,52	0,24	8,25
12.	3,15	0,55	9,31

7. Маючи статистичні дані про результати господарської діяльності підприємств, побудувати багатofакторну лінійну модель залежності прибутку підприємства у від інвестицій x_1 , витрат на рекламу x_2 та витрат на заробітну плату x_3 .

Номер підприємства	Прибуток підприємства	Розмір інвестицій	Витрати на рекламу	Витрати на заробітну плату
1.	15,70	17,37	5,28	1,42
2.	17,34	18,24	6,47	1,58
3.	21,57	22,47	6,98	1,98
4.	33,50	18,47	7,05	2,04
5.	32,30	16,82	7,94	2,38
6.	37,90	17,60	8,12	3,48
7.	40,78	17,12	8,69	3,07
8.	48,02	19,81	9,31	3,84
9.	43,30	18,67	10,45	4,28

ДОДАТКИ

Таблиця значень функції $\varphi(x) = \frac{1}{2\pi} e^{-\frac{x^2}{2}}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3988	3986	3989	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3952	3845	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	3725	3712	3697
0,4	3683	3668	3653	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	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	0794	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	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	0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	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	0011	0011	0010	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
4,0	0001	0001	0001	0001	0001	0001	0001	0001	0001	0001
4,1	0001	0001	0001	0001	0001	0001	0001	0001	0001	0001
4,2	0001	0001	0001	0001	0000	0000	0000	0000	0000	0000

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

x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)
0,00	0,0000	0,50	0,1915	1,00	0,3413	1,50	0,4332	2,00	0,4772	2,50	0,4938
0,01	0,0040	0,51	0,1950	1,01	0,3438	1,51	0,4345	2,01	0,4778	2,52	0,4941
0,02	0,0080	0,52	0,1985	1,02	0,3461	1,52	0,4357	2,02	0,4783	2,54	0,4945
0,03	0,0120	0,53	0,2019	1,03	0,3485	1,53	0,4370	2,03	0,4788	2,56	0,4948
0,04	0,0160	0,54	0,2054	1,04	0,3508	1,54	0,4382	2,04	0,4793	2,58	0,4951
0,05	0,0199	0,55	0,2088	1,05	0,3531	1,55	0,4394	2,05	0,4798	2,60	0,4953
0,06	0,0239	0,56	0,2123	1,06	0,3554	1,56	0,4406	2,06	0,4803	2,62	0,4956
0,07	0,0279	0,57	0,2157	1,07	0,3577	1,57	0,4418	2,07	0,4808	2,64	0,4959
0,08	0,0319	0,58	0,2190	1,08	0,3599	1,58	0,4429	2,08	0,4812	2,66	0,4961
0,09	0,0359	0,59	0,2224	1,09	0,3621	1,59	0,4441	2,09	0,4817	2,68	0,4963
0,10	0,0398	0,60	0,2257	1,10	0,3643	1,60	0,4452	2,10	0,4821	2,70	0,4965
0,11	0,0438	0,61	0,2291	1,11	0,3665	1,61	0,4463	2,11	0,4826	2,72	0,4967
0,12	0,0478	0,62	0,2324	1,12	0,3686	1,62	0,4474	2,12	0,4830	2,74	0,4969
0,13	0,0517	0,63	0,2357	1,13	0,3708	1,63	0,4484	2,13	0,4834	2,76	0,4971
0,14	0,0557	0,64	0,2389	1,14	0,3729	1,64	0,4495	2,14	0,4838	2,78	0,4973
0,15	0,0596	0,65	0,2422	1,15	0,3749	1,65	0,4505	2,15	0,4842	2,80	0,4974
0,16	0,0636	0,66	0,2454	1,16	0,3770	1,66	0,4515	2,16	0,4846	2,82	0,4976
0,17	0,0675	0,67	0,2486	1,17	0,3790	1,67	0,4525	2,17	0,4850	2,84	0,4977
0,18	0,0714	0,68	0,2517	1,18	0,3810	1,68	0,4535	2,18	0,4854	2,86	0,4979
0,19	0,0753	0,69	0,2549	1,19	0,3830	1,69	0,4545	2,19	0,4857	2,88	0,4980
0,20	0,0793	0,70	0,2580	1,20	0,3849	1,70	0,4554	2,20	0,4861	2,90	0,4981
0,21	0,0832	0,71	0,2611	1,21	0,3869	1,71	0,4564	2,21	0,4864	2,92	0,4982
0,22	0,0871	0,72	0,2642	1,22	0,3888	1,72	0,4573	2,22	0,4868	2,94	0,4984
0,23	0,0910	0,73	0,2673	1,23	0,3907	1,73	0,4582	2,23	0,4871	2,96	0,4985
0,24	0,0948	0,74	0,2704	1,24	0,3925	1,74	0,4591	2,24	0,4875	2,98	0,4986
0,25	0,0987	0,75	0,2734	1,25	0,3944	1,75	0,4599	2,25	0,4878	3,00	0,4987
0,26	0,1026	0,76	0,2764	1,26	0,3962	1,76	0,4608	2,26	0,4881	3,05	0,4989
0,27	0,1064	0,77	0,2794	1,27	0,3980	1,77	0,4616	2,27	0,4884	3,10	0,49903
0,28	0,1103	0,78	0,2823	1,28	0,3997	1,78	0,4625	2,28	0,4887	3,15	0,49918
0,29	0,1141	0,79	0,2852	1,29	0,4015	1,79	0,4633	2,29	0,4890	3,20	0,49931
0,30	0,1179	0,80	0,2881	1,30	0,4032	1,80	0,4641	2,30	0,4893	3,25	0,49942
0,31	0,1217	0,81	0,2910	1,31	0,4049	1,81	0,4649	2,31	0,4896	3,30	0,49952
0,32	0,1255	0,82	0,2939	1,32	0,4066	1,82	0,4656	2,32	0,4898	3,35	0,49960
0,33	0,1293	0,83	0,2967	1,33	0,4082	1,83	0,4664	2,33	0,4901	3,40	0,49966
0,34	0,1331	0,84	0,2995	1,34	0,4099	1,84	0,4671	2,34	0,4904	3,50	0,49977
0,35	0,1368	0,85	0,3023	1,35	0,4115	1,85	0,4678	2,35	0,4906	3,60	0,49984
0,36	0,1406	0,86	0,3051	1,36	0,4131	1,86	0,4686	2,36	0,4909	3,70	0,49989
0,37	0,1443	0,87	0,3078	1,37	0,4147	1,87	0,4693	2,37	0,4911	3,80	0,499928
0,38	0,1480	0,88	0,3106	1,38	0,4162	1,88	0,4699	2,38	0,4913	3,90	0,499952
0,39	0,1517	0,89	0,3133	1,39	0,4177	1,89	0,4706	2,39	0,4916	4,00	0,499968
0,40	0,1554	0,90	0,3159	1,40	0,4192	1,90	0,4713	2,40	0,4918	4,20	0,499987
0,41	0,1591	0,91	0,3186	1,41	0,4207	1,91	0,4719	2,41	0,4920	4,40	0,4999946
0,42	0,1628	0,92	0,3212	1,42	0,4222	1,92	0,4726	2,42	0,4922	4,60	0,4999979
0,43	0,1664	0,93	0,3238	1,43	0,4236	1,93	0,4732	2,43	0,4925	4,80	0,4999992
0,44	0,1700	0,94	0,3264	1,44	0,4251	1,94	0,4738	2,44	0,4927	5,00	0,4999997
0,45	0,1736	0,95	0,3289	1,45	0,4265	1,95	0,4744	2,45	0,4929		
0,46	0,1772	0,96	0,3315	1,46	0,4279	1,96	0,4750	2,46	0,4931		
0,47	0,1808	0,97	0,3340	1,47	0,4292	1,97	0,4756	2,47	0,4932		
0,48	0,1844	0,98	0,3365	1,48	0,4306	1,98	0,4761	2,48	0,4934		
0,49	0,1879	0,99	0,3389	1,49	0,4319	1,99	0,4767	2,49	0,4936		

Таблиця значень функції Пуассона $P_n(m) = \frac{\lambda^m}{m!} e^{-\lambda}$

$m \backslash \lambda$	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
0	0,9048	8187	7408	6703	6065	5488	4966	4493	4066	3679
1	0905	1637	2222	2681	3033	3293	3476	3595	3659	3679
2	0045	0164	0333	0536	0758	0988	1217	1438	1647	1839
3	0002	0011	0033	0072	0126	0198	0284	0383	0494	0613
4	0000	0001	0003	0007	0016	0030	0050	0077	0111	0153
5	0000	0000	0000	0001	0002	0004	0007	0012	0020	0031
6	0000	0000	0000	0000	0000	0000	0001	0002	0003	0005
$m \backslash \lambda$	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2
0	0,3329	3012	2725	2466	2231	2019	1827	1653	1496	1353
1	3662	3614	3543	3452	3347	3230	3106	2975	2842	2707
2	2014	2169	2303	2417	2510	2584	2640	2678	2700	2707
3	0738	0867	0998	1128	1255	1378	1496	1607	1710	1804
4	0203	0260	0324	0395	0471	0551	0636	0723	0812	0902
5	0045	0062	0084	0111	0141	0176	0216	0260	0309	0361
6	0008	0012	0018	0026	0035	0047	0061	0078	0098	0120
7	0001	0002	0003	0005	0008	0011	0015	0020	0027	0034
8	0000	0000	0001	0001	0001	0002	0003	0005	0006	0009
9	0000	0000	0000	0000	0000	0000	0001	0001	0001	0002
$m \backslash \lambda$	2,1	2,2	2,3	2,4	2,5	2,6	2,7	2,8	2,9	3
0	0,1225	1108	1003	0907	0821	0743	0672	0608	0550	0498
1	2572	2438	2306	2177	2052	1931	1815	1703	1596	1494
2	2700	2681	2652	2613	2565	2510	2450	2384	2314	2240
3	1890	1966	2033	2090	2138	2176	2205	2225	2237	2240
4	0992	1082	1169	1254	1336	1414	1488	1557	1622	1680
5	0417	0476	0538	0602	0668	0735	0804	0872	0940	1008
6	0146	0174	0206	0241	0278	0319	0362	0407	0455	0504
7	0044	0055	0068	0083	0099	0118	0139	0163	0188	0216
8	0011	0015	0019	0025	0031	0038	0047	0057	0068	0081
9	0003	0004	0005	0007	0009	0011	0014	0018	0022	0027
10	0001	0001	0001	0002	0002	0003	0004	0005	0006	0008
11	0000	0000	0000	0000	0000	0001	0001	0001	0002	0002
$m \backslash \lambda$	3	3,5	4	4,5	5	6	7	8	9	10
0	0,0498	0302	0183	0111	0067	0025	0009	0003	0001	0000
1	1494	1057	0733	0500	0337	0149	0064	0027	0011	0005
2	2240	1850	1465	1125	0842	0446	0223	0107	0050	0023
3	2240	2158	1954	1687	1404	0892	0521	0286	0150	0076
4	1680	1888	1954	1898	1755	1339	0912	0573	0337	0189
5	1008	1322	1563	1708	1755	1606	1277	0916	0607	0378
6	0504	0771	1042	1281	1462	1606	1490	1221	0911	0631
7	0216	0385	0595	0824	1044	1377	1490	1396	1171	0901
8	0081	0169	0298	0463	0653	1033	1304	1396	1318	1126
9	0027	0066	0132	0232	0363	0688	1014	1241	1318	1251
10	0008	0023	0053	0104	0181	0413	0710	0993	1186	1251
11	0002	0007	0019	0043	0082	0225	0452	0722	0970	1137
12	0001	0002	0006	0016	0034	0113	0263	0481	0728	0948
13	0000	0001	0002	0006	0013	0052	0142	0296	0504	0729
14	0000	0000	0001	0002	0005	0022	0071	0169	0324	0521
15	0000	0000	0000	0001	0002	0009	0033	0090	0194	0347
16	0000	0000	0000	0000	0000	0003	0014	0045	0109	0217
17	0000	0000	0000	0000	0000	0001	0006	0021	0058	0128
18	0000	0000	0000	0000	0000	0000	0002	0009	0029	0071
19	0000	0000	0000	0000	0000	0000	0001	0004	0014	0037
20	0000	0000	0000	0000	0000	0000	0000	0002	0006	0019
21	0000	0000	0000	0000	0000	0000	0000	0001	0003	0009
22	0000	0000	0000	0000	0000	0000	0000	0000	0001	0004

Таблиця значень функції $t_\alpha = t(\alpha; n)$

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

Таблиця значень функції $q = q(\alpha; n)$

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

Критичні точки розподілу Стьюдента

Кількість ступенів вільності	Рівень значущості α (двостороння критична область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,31	12,71	31,82	63,66	318,29	636,58
2	2,92	4,30	6,96	9,92	22,33	31,60
3	2,35	3,18	4,54	5,84	10,21	12,92
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,02	2,57	3,36	4,03	5,89	6,87
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,41
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	4,02	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	3,97
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	1,72	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	3,50	3,79
24	1,71	2,06	2,49	2,80	3,47	3,75
25	1,71	2,06	2,49	2,79	3,45	3,73
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
∞	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 α (одностороння критична область)					

Критичні точки розподілу χ^2

Кількість ступенів вільності k	Рівень значущості α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,63	5,02	3,84	0,0039	0,00098	0,00016
2	9,21	7,38	5,99	0,103	0,051	0,020
3	11,34	9,35	7,81	0,352	0,216	0,115
4	13,28	11,14	9,49	0,711	0,484	0,297
5	15,09	12,83	11,07	1,15	0,831	0,554
6	16,81	14,45	12,59	1,64	1,24	0,872
7	18,48	16,01	14,07	2,17	1,69	1,24
8	20,09	17,53	15,51	2,73	2,18	1,65
9	21,67	19,02	16,92	3,33	2,70	2,09
10	23,21	20,48	18,31	3,94	3,25	2,56
11	24,73	21,92	19,68	4,57	3,82	3,05
12	26,22	23,34	21,03	5,23	4,40	3,57
13	27,69	24,74	22,36	5,89	5,01	4,11
14	29,14	26,12	23,68	6,57	5,63	4,66
15	30,58	27,49	25,00	7,26	6,26	5,23
16	32,00	28,85	26,30	7,96	6,91	5,81
17	33,41	30,19	27,59	8,67	7,56	6,41
18	34,81	31,53	28,87	9,39	8,23	7,01
19	36,19	32,85	30,14	10,12	8,91	7,63
20	37,57	34,17	31,41	10,85	9,59	8,26
21	38,93	35,48	32,67	11,59	10,28	8,90
22	40,29	36,78	33,92	12,34	10,98	9,54
23	41,64	38,08	35,17	13,09	11,69	10,20
24	42,98	39,36	36,42	13,85	12,40	10,86
25	44,31	40,65	37,65	14,61	13,12	11,52
26	45,64	41,92	38,89	15,38	13,84	12,20
27	46,96	43,19	40,11	16,15	14,57	12,88
28	48,28	44,46	41,34	16,93	15,31	13,56
29	49,59	45,72	42,56	17,71	16,05	14,26
30	50,89	46,98	43,77	18,49	16,79	14,95

Критичні точки розподілу F Фішера–Снедекора

Рівень значущості $\alpha = 0,01$												
$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5404	5624	5764	5859	5928	5981	6022	6056	6083	6107
2	98,50	99,00	99,16	99,25	99,30	99,33	99,36	99,38	99,39	99,40	99,41	99,42
3	34,12	30,82	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	14,55	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,96	9,89
6	13,75	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	6,99	6,84	6,72	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,73	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,77	4,71
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,04	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	3,62	3,55
17	8,40	6,11	5,19	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,46
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,43	3,37
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36	3,30
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,29	3,23
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,24	3,17
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,18	3,12
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,14	3,07
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,09	3,03
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	3,06	2,99
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	3,02	2,96
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,99	2,93
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,96	2,90
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,93	2,87
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,91	2,84
31	7,53	5,36	4,48	3,99	3,67	3,45	3,28	3,15	3,04	2,96	2,88	2,82
32	7,50	5,34	4,46	3,97	3,65	3,43	3,26	3,13	3,02	2,93	2,86	2,80
33	7,47	5,31	4,44	3,95	3,63	3,41	3,24	3,11	3,00	2,91	2,84	2,78
34	7,44	5,29	4,42	3,93	3,61	3,39	3,22	3,09	2,98	2,89	2,82	2,76
35	7,42	5,27	4,40	3,91	3,59	3,37	3,20	3,07	2,96	2,88	2,80	2,74
36	7,40	5,25	4,38	3,89	3,57	3,35	3,18	3,05	2,95	2,86	2,79	2,72
37	7,37	5,23	4,36	3,87	3,56	3,33	3,17	3,04	2,93	2,84	2,77	2,71
38	7,35	5,21	4,34	3,86	3,54	3,32	3,15	3,02	2,92	2,83	2,75	2,69
39	7,33	5,19	4,33	3,84	3,53	3,30	3,14	3,01	2,90	2,81	2,74	2,68
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,73	2,66
41	7,30	5,16	4,30	3,81	3,50	3,28	3,11	2,98	2,87	2,79	2,71	2,65

Рівень значущості $\alpha = 0,01$												
$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9	10	11	12
42	7,28	5,15	4,29	3,80	3,49	3,27	3,10	2,97	2,86	2,78	2,70	2,64
43	7,26	5,14	4,27	3,79	3,48	3,25	3,09	2,96	2,85	2,76	2,69	2,63
44	7,25	5,12	4,26	3,78	3,47	3,24	3,08	2,95	2,84	2,75	2,68	2,62
45	7,23	5,11	4,25	3,77	3,45	3,23	3,07	2,94	2,83	2,74	2,67	2,61
46	7,22	5,10	4,24	3,76	3,44	3,22	3,06	2,93	2,82	2,73	2,66	2,60
47	7,21	5,09	4,23	3,75	3,43	3,21	3,05	2,92	2,81	2,72	2,65	2,59
48	7,19	5,08	4,22	3,74	3,43	3,20	3,04	2,91	2,80	2,71	2,64	2,58
49	7,18	5,07	4,21	3,73	3,42	3,19	3,03	2,90	2,79	2,71	2,63	2,57
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,78	2,70	2,63	2,56
55	7,12	5,01	4,16	3,68	3,37	3,15	2,98	2,85	2,75	2,66	2,59	2,53
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,56	2,50
65	7,04	4,95	4,10	3,62	3,31	3,09	2,93	2,80	2,69	2,61	3,53	2,47
70	7,01	4,92	4,07	3,60	3,29	3,07	2,91	2,78	2,67	2,59	2,51	2,45
75	6,99	4,90	4,05	3,58	3,27	3,05	2,89	2,76	2,65	2,57	2,49	2,43
80	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,64	2,55	2,48	2,42
85	6,94	4,86	4,02	3,55	3,24	3,02	2,86	2,73	2,62	2,54	2,46	2,40
90	6,93	4,85	4,01	3,53	3,23	3,01	2,84	2,72	2,61	2,52	2,45	2,39
95	6,91	4,84	3,99	3,52	3,22	3,00	2,83	2,70	2,60	2,51	2,44	2,38
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50	2,43	2,37
125	6,84	4,78	3,94	3,47	3,17	2,95	2,79	2,66	2,55	2,47	2,39	2,33
150	6,81	4,75	3,91	3,45	3,14	2,92	2,76	2,63	2,53	2,44	2,37	2,31
175	6,78	4,73	3,90	3,43	3,12	2,91	2,74	2,61	2,51	2,42	2,35	2,29
200	6,76	4,71	3,88	3,41	3,11	2,89	2,73	2,60	2,50	2,41	2,34	2,27
300	6,72	4,68	3,85	3,38	3,08	2,86	2,70	2,57	2,47	2,38	2,31	2,24
400	6,70	4,66	3,83	3,37	3,06	2,85	2,68	2,56	2,45	2,37	2,29	2,23
500	6,69	4,65	3,82	3,36	3,05	2,84	2,68	2,55	2,44	2,36	2,28	2,22
700	6,67	4,64	3,81	3,35	3,04	2,83	2,66	2,54	2,43	2,35	2,27	2,21
1000	6,66	4,63	3,80	3,34	3,04	2,82	2,66	2,53	2,43	2,34	2,27	2,20
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,25	2,18

Рівень значущості $\alpha = 0,01$												
$k_2 \backslash k_1$	14	16	20	24	30	40	50	75	100	200	500	∞
1	6143	6170	6209	6234	6260	6286	6302	6324	6334	6350	6360	6366
2	99,43	99,44	99,45	99,46	99,47	99,48	99,48	99,48	99,49	99,49	99,50	99,50
3	26,92	26,83	26,69	26,60	26,50	26,41	26,35	26,28	26,24	26,18	26,15	26,13
4	14,25	14,15	14,02	13,93	13,84	13,75	13,69	13,61	13,58	13,52	13,49	13,46
5	9,77	9,68	9,55	9,47	9,38	9,24	9,29	9,17	9,13	9,08	9,04	9,02
6	7,60	7,52	7,40	7,31	7,23	7,14	7,09	7,02	6,99	6,93	6,90	6,88
7	6,36	6,28	6,16	6,07	5,99	5,91	5,86	5,79	5,75	5,70	5,67	5,65
8	5,56	5,48	5,36	5,28	5,20	5,12	5,07	5,00	4,96	4,91	4,88	4,86
9	5,01	4,92	4,81	4,73	4,65	4,57	4,52	4,45	4,41	4,36	4,33	4,31
10	4,60	4,52	4,41	4,33	4,25	4,17	4,12	4,05	4,01	3,96	3,93	3,91
11	4,29	4,21	4,10	4,02	3,94	3,86	3,81	3,74	3,71	3,66	3,62	3,60
12	4,05	3,97	3,86	3,78	3,70	3,62	3,57	3,50	3,47	3,41	3,38	3,36
13	3,86	3,78	3,66	3,59	3,51	3,43	3,38	3,31	3,27	3,22	3,19	3,17
14	3,70	3,62	3,51	3,43	3,35	3,27	3,22	3,14	3,11	3,06	3,03	3,00
15	3,56	3,49	3,37	3,29	3,21	3,13	3,08	3,01	2,98	2,92	2,89	2,87
16	3,45	3,37	3,26	3,18	3,10	3,02	2,97	2,90	2,86	2,81	2,78	2,75
17	3,35	3,27	3,16	3,08	3,00	2,92	2,87	2,80	2,76	2,71	2,68	2,65
18	3,27	3,19	3,08	3,00	2,92	2,84	2,78	2,71	2,68	2,62	2,59	2,57
19	3,19	3,12	3,00	2,92	2,84	2,76	2,71	2,64	2,60	2,55	2,51	2,49
20	3,13	3,05	2,94	2,86	2,78	2,69	2,64	2,57	2,54	2,48	2,44	2,42
21	3,07	2,99	2,88	2,80	2,72	2,64	2,58	2,51	2,48	2,42	2,38	2,36
22	3,02	2,94	2,83	2,75	2,67	2,58	2,53	2,46	2,42	2,36	2,33	2,31
23	2,97	2,89	2,78	2,70	2,62	2,54	2,48	2,41	2,37	2,32	2,28	2,26
24	2,93	2,85	2,74	2,66	2,58	2,49	2,44	2,37	2,33	2,27	2,24	2,21
25	2,89	2,81	2,70	2,62	2,54	2,45	2,40	2,33	2,29	2,23	2,19	2,17
26	2,86	2,78	2,66	2,58	2,50	2,42	2,36	2,29	2,25	2,19	2,16	2,13
27	2,82	2,75	2,63	2,55	2,47	2,38	2,33	2,26	2,22	2,16	2,12	2,10
28	2,79	2,72	2,60	2,52	2,44	2,35	2,30	2,23	2,19	2,13	2,09	2,06
29	2,77	2,69	2,57	2,49	2,41	2,33	2,27	2,20	2,16	2,10	2,06	2,03
30	2,74	2,66	2,55	2,47	2,39	2,30	2,25	2,17	2,13	2,07	2,03	2,01
31	2,72	2,64	2,52	2,45	2,36	2,27	2,22	2,14	2,11	2,04	2,01	1,98
32	2,70	2,62	2,50	2,42	2,34	2,25	2,20	2,12	2,08	2,02	1,98	1,96
33	2,68	2,60	2,48	2,40	2,32	2,23	2,18	2,10	2,06	2,00	1,96	1,93
34	2,66	2,58	2,46	2,38	2,30	2,21	2,16	2,08	2,04	1,98	1,94	1,91
35	2,64	2,56	2,44	2,36	2,28	2,19	2,14	2,06	2,02	1,96	1,92	1,89
36	2,62	2,54	2,43	2,35	2,26	2,18	2,12	2,04	2,00	1,94	1,90	1,87
37	2,61	2,53	2,41	2,33	2,25	2,16	2,10	2,03	1,98	1,92	1,88	1,85
38	2,59	2,51	2,40	2,32	2,23	2,14	2,09	2,01	1,97	1,90	1,86	1,84
39	2,58	2,50	2,38	2,30	2,22	2,13	2,07	1,99	1,95	1,89	1,85	1,82
40	2,56	2,48	2,37	2,29	2,20	2,11	2,06	1,98	1,94	1,87	1,83	1,80
41	2,55	2,47	2,36	2,28	2,19	2,10	2,04	1,97	1,92	1,86	1,82	1,79

Рівень значущості $\alpha = 0,01$												
$k_2 \backslash k_1$	14	16	20	24	30	40	50	75	100	200	500	∞
42	2,54	2,46	2,34	2,26	2,18	2,09	2,03	1,95	1,91	1,85	1,80	1,78
43	2,53	2,45	2,33	2,25	2,17	2,08	2,02	1,94	1,90	1,83	1,79	1,76
44	2,52	2,44	2,32	2,24	2,15	2,07	2,01	1,93	1,89	1,82	1,78	1,75
45	2,51	2,43	2,31	2,23	2,14	2,05	2,00	1,92	1,88	1,81	1,77	1,74
46	2,50	2,42	2,30	2,22	2,13	2,04	1,99	1,91	1,86	1,80	1,76	1,73
47	2,49	2,41	2,29	2,21	2,12	2,03	1,98	1,90	1,85	1,79	1,74	1,71
48	2,48	2,40	2,28	2,20	2,12	2,02	1,97	1,89	1,84	1,78	1,73	1,70
49	2,47	2,39	2,27	2,19	2,11	2,02	1,96	1,88	1,83	1,77	1,72	1,69
50	2,46	2,38	2,27	2,18	2,10	2,01	1,95	1,87	1,82	1,76	1,71	1,68
55	2,42	2,34	2,23	2,15	2,06	1,97	1,91	1,83	1,78	1,71	1,67	1,64
60	2,39	2,31	2,20	2,12	2,03	1,94	1,88	1,79	1,75	1,68	1,63	1,60
65	2,37	2,29	2,17	2,09	2,00	1,91	1,85	1,77	1,72	1,65	1,60	1,57
70	2,35	2,27	2,15	2,07	1,98	1,89	1,83	1,74	1,70	1,62	1,57	1,54
75	2,33	2,25	2,13	2,05	1,96	1,87	1,81	1,72	1,67	1,60	1,55	1,52
80	2,31	2,23	2,12	2,03	1,94	1,85	1,79	1,70	1,65	1,58	1,53	1,49
85	2,30	2,22	2,10	2,02	1,93	1,83	1,77	1,69	1,64	1,56	1,51	1,47
90	2,29	2,21	2,09	2,00	1,92	1,82	1,76	1,67	1,62	1,55	1,49	1,46
95	2,28	2,20	2,08	1,99	1,90	1,81	1,75	1,66	1,61	1,53	1,48	1,44
100	2,27	2,19	2,07	1,98	1,89	1,80	1,74	1,65	1,60	1,52	1,47	1,43
125	2,23	2,15	2,03	1,94	1,85	1,76	1,69	1,60	1,55	1,47	1,41	1,37
150	2,20	2,12	2,00	1,92	1,83	1,73	1,66	1,57	1,52	1,43	1,38	1,33
175	2,19	2,10	1,98	1,90	1,81	1,71	1,64	1,55	1,50	1,41	1,35	1,30
200	2,17	2,09	1,97	1,89	1,79	1,69	1,63	1,53	1,48	1,39	1,33	1,28
300	2,14	2,06	1,94	1,85	1,76	1,66	1,59	1,50	1,44	1,35	1,28	1,22
400	2,13	2,05	1,92	1,84	1,75	1,64	1,58	1,48	1,42	1,32	1,25	1,19
500	2,12	2,04	1,92	1,83	1,74	1,63	1,57	1,47	1,41	1,31	1,23	1,16
600	2,11	2,03	1,91	1,82	1,73	1,63	1,56	1,46	1,40	1,30	1,22	1,15
700	2,11	2,03	1,90	1,82	1,72	1,62	1,55	1,45	1,39	1,29	1,21	1,14
800	2,10	2,02	1,90	1,81	1,72	1,62	1,55	1,45	1,39	1,29	1,20	1,13
1000	2,10	2,02	1,90	1,81	1,72	1,61	1,54	1,44	1,38	1,28	1,19	1,11
∞	2,08	2,00	1,88	1,79	1,70	1,59	1,52	1,42	1,36	1,25	1,15	1,00

Рівень значущості $\alpha = 0,05$

$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9	10	11	12
1	161	199	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	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	5,94	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	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	3,64	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	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	2,46	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26	2,23
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,24	2,20
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,22	2,18
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,20	2,16
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,17	2,13
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,15	2,12
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,14	2,10
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09
31	4,16	3,30	2,91	2,68	2,52	2,41	2,32	2,25	2,20	2,15	2,11	2,08
32	4,15	3,29	2,90	2,67	2,51	2,40	2,31	2,24	2,19	2,14	2,10	2,07
33	4,14	3,28	2,89	2,66	2,50	2,39	2,30	2,23	2,18	2,13	2,09	2,06
34	4,13	3,28	2,88	2,65	2,49	2,38	2,29	2,23	2,17	2,12	2,08	2,05
35	4,12	3,27	2,87	2,64	2,49	2,37	2,29	2,22	2,16	2,11	2,07	2,04
36	4,11	3,26	2,87	2,63	2,48	2,36	2,28	2,21	2,15	2,11	2,07	2,03
37	4,11	3,25	2,86	2,63	2,47	2,36	2,27	2,20	2,14	2,10	2,06	2,02
38	4,10	3,24	2,85	2,62	2,46	2,35	2,26	2,19	2,14	2,09	2,05	2,02
39	4,09	3,24	2,85	2,61	2,46	2,34	2,26	2,19	2,13	2,08	2,04	2,01
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00
41	4,08	3,23	2,83	2,60	2,44	2,33	2,24	2,17	2,12	2,07	2,03	2,00

Рівень значущості $\alpha = 0,05$

$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9	10	11	12
42	4,07	3,22	2,83	2,59	2,44	2,32	2,24	2,17	2,11	2,06	2,03	1,99
43	4,07	3,21	2,82	2,59	2,43	2,32	2,23	2,16	2,11	2,06	2,02	1,99
44	4,06	3,21	2,82	2,58	2,43	2,31	2,23	2,16	2,10	2,05	2,01	1,98
45	4,06	3,20	2,81	2,58	2,42	2,31	2,22	2,15	2,10	2,05	2,01	1,97
46	4,05	3,20	2,81	2,57	2,42	2,30	2,22	2,15	2,09	2,04	2,00	1,97
47	4,05	3,20	2,80	2,57	2,41	2,30	2,21	2,14	2,09	2,04	2,00	1,96
48	4,04	3,19	2,80	2,57	2,41	2,29	2,21	2,14	2,08	2,03	1,99	1,96
49	4,04	3,19	2,79	2,56	2,40	2,29	2,20	2,13	2,08	2,03	1,99	1,96
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,99	1,95
55	4,02	3,16	2,77	2,54	2,38	2,27	2,18	2,11	2,06	2,01	1,97	1,93
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92
65	3,99	3,14	2,75	2,51	2,36	2,24	2,15	2,08	2,03	1,98	1,94	1,90
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97	1,93	1,89
75	3,97	3,12	2,73	2,49	2,34	2,22	2,13	2,06	2,01	1,96	1,92	1,88
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,91	1,88
85	3,95	3,10	2,71	2,48	2,32	2,21	2,12	2,05	1,99	1,94	1,90	1,87
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,98	1,93	1,89	1,86
95	3,94	3,09	2,70	2,47	2,31	2,20	2,11	2,04	1,98	1,93	1,89	1,86
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,89	1,85
125	3,92	3,07	2,68	2,44	2,29	2,17	2,08	2,01	1,96	1,91	1,87	1,83
150	3,90	3,06	2,66	2,43	2,27	2,16	2,07	2,00	1,94	1,89	1,85	1,82
175	3,90	3,05	2,66	2,42	2,27	2,15	2,06	1,99	1,93	1,89	1,84	1,81
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,84	1,80
300	3,87	3,03	2,63	2,40	2,24	2,13	2,04	1,97	1,91	1,86	1,82	1,78
400	3,86	3,02	2,63	2,39	2,24	2,12	2,03	1,96	1,90	1,85	1,81	1,78
500	3,86	3,01	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,81	1,77
600	3,86	3,01	2,62	2,39	2,23	2,11	2,02	1,95	1,90	1,85	1,80	1,77
700	3,85	3,01	2,62	2,38	2,23	2,11	2,02	1,95	1,89	1,84	1,80	1,77
800	3,85	3,01	2,62	2,38	2,23	2,11	2,02	1,95	1,89	1,84	1,80	1,76
1000	3,85	3,00	2,61	2,38	2,22	2,11	2,02	1,95	1,89	1,84	1,80	1,76
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,79	1,75

Рівень значущості $\alpha = 0,05$												
$k_2 \backslash k_1$	14	16	20	24	30	40	50	75	100	200	500	∞
1	245	246	248	249	250	251	252	253	253	254	254	254
2	19,42	19,43	19,45	19,45	19,46	19,47	19,48	19,48	19,49	19,49	19,49	19,50
3	8,71	8,69	8,66	8,64	8,62	8,59	8,58	8,56	8,55	8,54	8,53	8,53
4	5,87	5,84	5,80	5,77	5,75	5,72	5,70	5,68	5,66	5,65	5,64	5,63
5	4,64	4,60	4,56	4,53	4,50	4,46	4,44	4,42	4,41	4,39	4,37	4,37
6	3,96	3,92	3,87	3,84	3,81	3,77	3,75	3,73	3,71	3,69	3,68	3,67
7	3,53	3,49	3,44	3,41	3,38	3,34	3,32	3,29	3,27	3,25	3,24	3,23
8	3,24	3,20	3,15	3,12	3,08	3,04	3,02	2,99	2,97	2,95	2,94	2,93
9	3,03	2,99	2,94	2,90	2,86	2,83	2,80	2,77	2,76	2,73	2,72	2,71
10	2,86	2,83	2,77	2,74	2,70	2,66	2,64	2,60	2,59	2,56	2,55	2,54
11	2,74	2,70	2,65	2,61	2,57	2,53	2,51	2,47	2,46	2,43	2,42	2,40
12	2,64	2,60	2,54	2,51	2,47	2,43	2,40	2,37	2,35	2,32	2,31	2,30
13	2,55	2,51	2,46	2,42	2,38	2,34	2,31	2,28	2,26	2,23	2,22	2,21
14	2,48	2,44	2,39	2,35	2,31	2,27	2,24	2,21	2,19	2,16	2,14	2,13
15	2,42	2,38	2,33	2,29	2,25	2,20	2,18	2,14	2,12	2,10	2,08	2,07
16	2,37	2,33	2,28	2,24	2,19	2,15	2,12	2,09	2,07	2,04	2,02	2,01
17	2,33	2,29	2,23	2,19	2,15	2,10	2,08	2,04	2,02	1,99	1,97	1,96
18	2,29	2,25	2,19	2,15	2,11	2,06	2,04	2,00	1,98	1,95	1,93	1,92
19	2,26	2,21	2,16	2,11	2,07	2,03	2,00	1,96	1,94	1,91	1,89	1,88
20	2,22	2,18	2,12	2,08	2,04	1,99	1,97	1,93	1,91	1,88	1,86	1,84
21	2,20	2,16	2,10	2,05	2,01	1,96	1,94	1,90	1,88	1,84	1,83	1,81
22	2,17	2,13	2,07	2,03	1,98	1,94	1,91	1,87	1,85	1,82	1,80	1,78
23	2,15	2,11	2,05	2,01	1,96	1,91	1,88	1,84	1,82	1,79	1,77	1,76
24	2,13	2,09	2,03	1,98	1,94	1,89	1,86	1,82	1,80	1,77	1,75	1,73
25	2,11	2,07	2,01	1,96	1,92	1,87	1,84	1,80	1,78	1,75	1,73	1,71
26	2,09	2,05	1,99	1,95	1,90	1,85	1,82	1,78	1,76	1,73	1,71	1,69
27	2,08	2,04	1,97	1,93	1,88	1,84	1,81	1,76	1,74	1,71	1,69	1,67
28	2,06	2,02	1,96	1,91	1,87	1,82	1,79	1,75	1,73	1,69	1,67	1,65
29	2,05	2,01	1,94	1,90	1,85	1,81	1,77	1,73	1,71	1,67	1,65	1,64
30	2,04	1,99	1,93	1,89	1,84	1,79	1,76	1,72	1,70	1,66	1,64	1,62
31	2,03	1,98	1,92	1,88	1,83	1,78	1,75	1,70	1,68	1,65	1,62	1,61
32	2,01	1,97	1,91	1,86	1,82	1,77	1,74	1,69	1,67	1,63	1,61	1,59
33	2,00	1,96	1,90	1,85	1,81	1,76	1,72	1,68	1,66	1,62	1,60	1,58
34	1,99	1,95	1,89	1,84	1,80	1,75	1,71	1,67	1,65	1,61	1,59	1,57
35	1,99	1,94	1,88	1,83	1,79	1,74	1,70	1,66	1,63	1,60	1,57	1,56
36	1,98	1,93	1,87	1,82	1,78	1,73	1,69	1,65	1,62	1,59	1,56	1,55
37	1,97	1,93	1,86	1,82	1,77	1,72	1,68	1,64	1,62	1,58	1,55	1,54
38	1,96	1,92	1,85	1,81	1,76	1,71	1,68	1,63	1,61	1,57	1,54	1,53
39	1,95	1,91	1,85	1,80	1,75	1,70	1,67	1,62	1,60	1,56	1,53	1,52
40	1,95	1,90	1,84	1,79	1,74	1,69	1,66	1,61	1,59	1,55	1,53	1,51
41	1,94	1,90	1,83	1,79	1,74	1,69	1,65	1,61	1,58	1,54	1,52	1,50

Рівень значущості $\alpha = 0,05$												
$k_2 \backslash k_1$	14	16	20	24	30	40	50	75	100	200	500	∞
42	1,94	1,89	1,83	1,78	1,73	1,68	1,65	1,60	1,57	1,53	1,51	1,49
43	1,93	1,89	1,82	1,77	1,72	1,67	1,64	1,59	1,57	1,53	1,50	1,48
44	1,92	1,88	1,81	1,77	1,72	1,67	1,63	1,59	1,56	1,52	1,49	1,48
45	1,92	1,87	1,81	1,76	1,71	1,66	1,63	1,58	1,55	1,51	1,49	1,47
46	1,91	1,87	1,80	1,76	1,71	1,65	1,62	1,57	1,55	1,51	1,48	1,46
47	1,91	1,86	1,80	1,75	1,70	1,65	1,61	1,57	1,54	1,50	1,47	1,46
48	1,90	1,86	1,79	1,75	1,70	1,64	1,61	1,56	1,54	1,49	1,47	1,45
49	1,90	1,85	1,79	1,74	1,69	1,64	1,60	1,56	1,53	1,49	1,46	1,44
50	1,89	1,85	1,78	1,74	1,69	1,63	1,60	1,55	1,52	1,48	1,46	1,44
55	1,88	1,83	1,76	1,72	1,67	1,61	1,58	1,53	1,50	1,46	1,43	1,41
60	1,86	1,82	1,75	1,70	1,65	1,59	1,56	1,51	1,48	1,44	1,41	1,39
65	1,85	1,80	1,73	1,69	1,63	1,58	1,54	1,49	1,46	1,42	1,39	1,37
70	1,84	1,79	1,72	1,67	1,62	1,57	1,53	1,48	1,45	1,40	1,37	1,35
75	1,83	1,78	1,71	1,66	1,61	1,55	1,52	1,47	1,44	1,39	1,36	1,34
80	1,82	1,77	1,70	1,65	1,60	1,54	1,51	1,45	1,43	1,38	1,35	1,32
85	1,81	1,76	1,70	1,65	1,59	1,54	1,50	1,45	1,42	1,37	1,34	1,31
90	1,80	1,75	1,68	1,64	1,59	1,53	1,49	1,44	1,41	1,36	1,33	1,30
95	1,80	1,75	1,68	1,63	1,58	1,52	1,48	1,43	1,40	1,35	1,32	1,29
100	1,79	1,75	1,68	1,63	1,57	1,52	1,48	1,42	1,39	1,34	1,31	1,28
125	1,77	1,73	1,66	1,60	1,55	1,49	1,45	1,40	1,36	1,31	1,27	1,25
150	1,76	1,71	1,64	1,59	1,54	1,48	1,4	1,38	1,34	1,29	1,25	1,22
175	1,75	1,70	1,63	1,58	1,52	1,46	1,42	1,36	1,33	1,27	1,23	1,20
200	1,74	1,69	1,62	1,57	1,52	1,46	1,41	1,35	1,32	1,26	1,22	1,19
300	1,72	1,68	1,61	1,5	1,50	1,43	1,39	1,33	1,30	1,23	1,19	1,15
400	1,72	1,67	1,60	1,54	1,49	1,42	1,38	1,32	1,28	1,22	1,17	1,13
500	1,71	1,6	1,59	1,54	1,48	1,42	1,38	1,31	1,28	1,21	1,16	1,11
600	1,71	1,6	1,59	1,54	1,48	1,41	1,37	1,31	1,27	1,20	1,15	1,10
700	1,71	1,66	1,59	1,53	1,48	1,41	1,37	1,30	1,27	1,20	1,15	1,09
800	1,70	1,6	1,58	1,53	1,47	1,41	1,37	1,30	1,26	1,20	1,14	1,09
1000	1,70	1,65	1,58	1,53	1,47	1,41	1,36	1,30	1,26	1,19	1,13	1,08
∞	1,69	1,64	1,57	1,52	1,46	1,39	1,35	1,28	1,24	1,17	1,11	1,00

ОГІРКО Ольга Ігорівна,
кандидат технічних наук
ГАЛАЙКО Наталія Володимирівна

ТЕОРІЯ
ЙМОВІРНОСТЕЙ
ТА МАТЕМАТИЧНА
СТАТИСТИКА

Навчальний посібник

Редагування *І. Б. Попик*

Макетування *Н. М. Лесь*

Друк *А. М. Радченко*

Підписано до друку 28.12.2017.
Формат 60×84/8. Папір офсетний. Умовн. друк. арк. 33,94.
Тираж 100 прим. Зам. № 145-17.

Львівський державний університет внутрішніх справ
Україна, 79007, м. Львів, вул. Городоцька, 26.

Свідоцтво про внесення суб'єкта видавничої справи до державного реєстру
видавців, виготівників і розповсюджувачів видавничої продукції
ДК № 2541 від 26 червня 2006 р.