Міністерство освіти і науки України

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ
УНІВЕРСИТЕТ
БУДІВНИЦТВА ТА АРХІТЕКТУРИ

О.О. Аршава, А.П. Харченко, Л.І. Щелкунова

МАТЕМАТИЧНА СТАТИСТИКА

Навчально-методичний посібник

Харків 2018

Міністерство освіти і науки України
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
БУДІВНИЦТВА ТА АРХІТЕКТУРИ

О.О. Аршава, А.П. Харченко, Л.І Щелкунова

МАТЕМАТИЧНА СТАТИСТИКА

Рекомендовано
науково-методичною радою університету
як навчально-методичний посібник
для студентів технічних вищів

Харків 2018

Щ – 45
УДК 517.22.17

Рецензенти:
Л.Б. Коваленко, канд. фіз.-мат. наук, доцент каф. вищої математики Харківського національного університету міського господарства ім.. О.О. Бєкєтова (ХНУМГ).
Є.В. Поклонський, канд. фіз.-мат. наук, доцент каф. вищої математики Харківського національного університету будівництва та архітектури (ХНУБА)

Рекомендовано кафедрою вищої математики ХНУБА,
протокол № 3 від 23.11.2017
Затверджено науково-методичною радою університету,
протокол № 3 від 30.11.2017

Щ – 45 Аршава О.О., А.П. Харченко, Щелкунова Л.І. Математична статистика: навчально-методичний посібник. – Х.: ХНУБА, 2018. – 100 с.

Викладено теоретичні та практичні основи таких розділів математичної статистики, як варіаційні ряди, статистичні оцінки, перевірка статистичних гіпотез, кореляційно-регресійний аналіз.

Велика кількість докладно розібраних прикладів та вправ дозволяє використовувати посібник як на практичних заняттях так і в процесі самостійної роботи студента.

Видання містить питання для самоперевірки та перелік завдань для самостійної роботи для кожного розділу.

Призначено для студентів технічних вищів.

Іл. 14: ; табл. 57 ; бібліогр.: 9 назв.

©О.О. Аршава, А.П. Харченко, Л.І.Щелкунова, 2018

ВСТУП

Математична статистика – розділ математики, в якому вивчаються методи збору, систематизації і обробки результатів спостережень з метою виявлення статистичних закономірностей ознаки або ознак певної сукупності елементів.

Математична статистика спирається на теорію ймовірностей. Але якщо теорія ймовірностей вивчає закономірності випадкових подій на підставі абстрактного опису дійсності, то математична статистика оперує результатами спостережень над випадковим явищем, які подають вибірку із деякої скінченої або гіпотетично нескінченої генеральної сукупності.

Математична статистика розв'язує два типа задач:

- статистичне оцінювання (точкове, інтервальне) параметрів генеральної сукупності;
- перевірка правдивості статистичних гіпотез про значення параметрів генеральної сукупності або про закон розподілу ознаки генеральної сукупності на підставі обробки результатів вибірки.

При великої кількості спостережень випадкові впливи у значній мірі погашаються і здобутий результат виявляється практично невипадковим, передбаченим. Цей принцип є базою для практичного використання ймовірнісних і математико-статистичних методів дослідження. Мета вказаних методів полягає у вивченні закономірності масових випадкових подій, прогнозуванні їх характеристик і впливу на хід цих подій, контролюючи та обмежуючи область дії випадковості.

Широкому впровадженню математико-статистичних методів дослідження сприяла поява електронних обчислювальних машин, зокрема, персональних комп'ютерів. Серед множини статистичних пакетів прикладних програм можна виділити наступні: STATGRAPHICS, STATISTICA, SPSS, STADIA тощо.

Розділ 1 ВИБІРКОВИЙ МЕТОД

Нехай маємо деяку сукупність однорідних об’єктів. Наприклад: а) партія деталей; б) партія ламп.

Як вивчити кількісні та якісні ознаки, які характеризують дану сукупність?

В а) і б) за якісну ознаку можна прийняти стандартність деталі або лампи. За кількісну ознаку у випадку а) можна прийняти контролюючий розмір деталі, у випадку б) – час роботи лампи. Іноді для вивчення кількісної або якісної ознаки проводять так зване суцільне обстеження, тобто обстежують кожний об’єкт, що на практиці не завжди можливо. Тому замість перевірки всіх об’єктів з даної партії перевіряють тільки невелику її частину, вибрану випадково.

Означення 1. Вся сукупність однорідних об’єктів, яка підлягає вивченню, називається генеральною сукупністю.

Означення 2. Частина випадково відібраних об’єктів з генеральної сукупності називається вибіркової сукупністю або вибіркою.

Означення 3. Число об’єктів у генеральній сукупності або вибірки називається об’ємом генеральної сукупності або вибірки і позначається відповідно через N і n.

Означення 4. Вибірка є представницькою (репрезентативною), якщо вона утворена випадково і достатньо добре відтворює генеральну сукупність.

Із означення 4 випливає, що вибірка може бути репрезентативною, якщо правильно організувати спосіб відбору з генеральної сукупності.

Способи відбору:

1. Випадкова вибірка з поверненням (повторний відбір);
2. Випадкова відбірка без повернення (без повторний відбір);
3. Механічний відбір;
4. Типовий відбір;
5. Серійний відбір.

Механічний відбір – вся генеральна сукупність механічно розділяється на стільки частин, скільки елементів повинно бути у вибірки. Потім з кожної частини випадково по одному елементу складається вибірка.

Типовий відбір використовується, коли у виготовлені елементів вибірки приймають участь різні станки, цеха. Вся продукція розбивається на частини і далі як у механічному відбору формують випадкову вибірку.

Серійний відбір – об'єкти вибираються з генеральної сукупності не по одному, а «серіями».

1.1 Статистичний розподіл вибірки

Нехай з генеральної сукупності кількісної ознаки X об'ємом N елементів виконана вибірка об'ємом n елементів, причому ознака X прийняла значення $x_1, x_2, \ldots, x_к$, при цьому значення x_1 повторилось n_1 разів, x_2 – n_2 разів,…, $x_к$ – n_k разів, де $n_1 + n_2 + \ldots + n_k = \sum_{i=1}^{k} n_i = n$.

Означення. Якщо ознака X приймає тільки ізольовані одне від одного значення, то вона називається дискретною, а якщо вона може приймати значення, які скільки завгодно мало відрізняються одне від одного, то вона називається неперервною.

Означення. Значення $x_1, x_2, \ldots, x_к$, які приймає ознака X, називаються її варіантами.

Означення. Послідовність варіант, виписаних у порядку зростання, називається варіаційним рядом.

Означення. Кількість спостережень $n_1, n_2, \ldots, n_k$ разів появи відповідно варіант $x_1, x_2, \ldots, x_к$ називаються частотами варіант, а відношення $\frac{n_i}{n} = W_i$ $(i = 1, 2, \ldots, k)$ – відносними частотами (частостями).

Означення. Таблиця, яка встановлює відповідність між: а) варіантами та їх частотами; б) варіантами та їх відносними частотами, називається статистичним розподілом вибірки, тобто

Таблиця 1

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k

або

Таблиця 2

x_i	x_1	x_2	…	x_k
W_i	$\frac{n_1}{n}$	$\frac{n_2}{n}$	…	$\frac{n_k}{n}$

$$\sum_{i=1}^{k} W_i = 1.$$

Статистичний розподіл вибірки є аналогом (реалізацією) розподілу ознаки (випадкової величини) X.

Зауваження. У випадку, коли ознака X є неперервною, будувати варіаційний ряд не дуже зручно за великою кількістю вихідних даних. Тому число інтервалів k за формулою Стерджеса рекомендовано вибирати

$k = 1+3{,}322\lg n$, а величина інтервала $h=\frac{x_{\max} - x_{\min}}{1+3{,}322\lg n}$, де $x_{мах}$ – найбільше значення ознаки, x_{min} – найменше значення ознаки, n – число вихідних даних (об’єм вибірки).

За початок першого інтервалу вибирають величину $x_{поч.} = x_{min} - \frac{h}{2}$, а кінець останнього інтервалу ($x_{кінц.}$) повинен задовольняти умові

$x_{кінц.}-h \le x_{max} < x_{кінц.}$.

Приклад 1. 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 (1.1)

Знайти об’єм вибірки та її статистичний розподіл частоти та відносних частот.

Розв’язання. Оскільки вибірка складається з 20 значень, то об’єм вибірки $n = 20$. Вважаючи означення дискретної ознаки, вона задовольняє її умовам.

Побудуємо варіаційний ряд вибірки, тобто розташуємо варіанти в порядку їх зростання.

1, 1, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 5, 5, 5, 6, 6, 7**.** (1.2)

У даній вибірці всього сім різних значень, тобто варіант 1, 2, 3, 4, 5, 6, 7. Знайдемо їх частоти $n_1=2;\ n_2=2;\ n_3=4;\ n_4=6;\ n_5=3;\ n_6=2;\ n_7=1$; та їх відносні частоти за формулою $W_i = \frac{n_i}{n}$; $W_1=\frac{2}{20}=0,1;$ $W_2=\frac{2}{20}=0,1;$ $W_3=\frac{4}{20}=0,2;$ $W_4=\frac{6}{20}=0,3;$ $W_5=\frac{3}{20}=0,15;$ $W_6=\frac{2}{20}=0,1;$ $W_7=\frac{1}{20}=0,05.$

Запишемо шукані статистичні розподіли.

Таблиця 3

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

$n=\sum n_i=20$

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

Таблиця 4

Контроль $\sum W_i=1$

Приклад 2. При вимірюванні діаметрів валиків після шліфовки одержані наступні результати: 6,75; 6,77; 6,77; 6,73; 6,76; 6,74; 6,70; 6,75; 6,71;

6,72; 6,72; 6,79; 6,71; 6,78; 6,73; 6,70; 6,73; 6,77; 6,75; 6,74. Знайти об'єм вибірки та побудувати її статистичні розподіли.

Розв'язання. Об'єм вибірки дорівнює $n=20$. Так як варіанти мало відрізняються одне від одного, то маємо інтервальний ряд. Для його побудови треба знайти кількість інтервалів $k=1+3,322lgn$, то довжину кожного інтервалу $h=\frac{x_{\max}-x_{\min}}{k}$. $k=1+3,322\cdot lg20=1+3,322\cdot 1,3=1+4,32=5,32;$

$x_{max}=6,79;\ x_{min}=6,70;\ h=\frac{6,79-6,70}{5,32}\approx 0,02.$ За початок першого інтервалу вибираємо величину $x_{поч.}=x_{min}-\frac{h}{2}=6,70-0,01=6,69.$

Будуємо статистичні розподіли інтервального варіаційного ряду, для якого кінцеві значення інтервалів не входять в кількість значень варіантів, які попали в даний інтервал.

Таблиця 5

$(x_i;x_{i+1})$	6,69-6,71	6,71-6,73	6,73-6,75	6,75-6,77	6,77-6,79	6,79-6,81
n_i	2	4	5	4	4	1

Контроль: $n=2+4+5+4+4+1=20$

Таблиця 6

$(x_i;x_{i+1})$	6,69-6,71	6,71-6,73	6,73-6,75	6,75-6,77	6,77-6,79	6,79-6,81
W_i	0,1	0,2	0,25	0,2	0,2	0,05

Контроль: $W=0,1+0,2+0,25+0,2+0,2+0,05=1$

1.2 Дискретний статистичний розподіл вибірки

При вивченні варіаційних рядів поряд з поняттям частоти використовується поняття накопиченої частоти, позначається $n_i^{нак.}$. Накопичена частота показує, скільки спостерігалось варіант із значенням ознаки, меншим *X*. Відношення накопиченої частоти $n_i^{нак.}$ до загального числа спостережень *n* називається накопиченою частотою $W_i^{нак.}$. Накопичені частоти (частості) для кожного інтервалу знаходяться послідовним додаванням частот (частостей) всіх попередніх інтервалів, включно даний.

Для графічного зображення варіаційних рядів найбільш часто використовуються емпірична функція розподілу (комулята), полігон, гістограма.

Означення. Емпіричною функцією розподілу $F^*(x)$ називається відносна частота (частость) того, що ознака (випадкова величина) *X* прийме значення, менше заданого аргументу *x*, тобто $F^*(x) = W_x^{нак.} = \dfrac{n_x^{нак.}}{n}$ (1.3)

Властивості $F^*(x)$:

1) $0 \le F^*(x) \le 1$;

2) $F(x_{\min})=0$, де $x_{\min}$ є найменшою варіантою варіаційного ряду;

3) $F\ (x)\,|x > x_{\max} = 1$, де $x_{\max}$ є найбільшою варіантою варіаційного ряду;

4) $F\ (x)$ є неспадною функцією аргументу *x*, тобто $F\ (x_2) \ge F\ (x_1)$ при $x_2 \ge x_1$.

1.3 Полігон частот і відносних частот (частостей)

Дискретний статистичний розподіл вибірки можна зобразити графічно у вигляді ламаної лінії, відрізки якої з'єднують точки $(x_1; n_1), (x_2; n_2), \ldots (x_k; n_k)$, або $(x_1; w_1), (x_2; w_2), \ldots, (x_k; w_k)$.

Перша ламана лінія зветься полігоном частот, друга – полігоном відносних частот.

Полігон, як правило, застосовується для зображення дискретного варіаційного ряду.

Приклад 3. За даним дискретним статистичним розподілом вибірки

Таблиця 7

x_i	-6	-4	-2	2	4	6
n_i	5	10	15	20	40	10
w_i	0,05	0,1	0,25	0,2	0,4	0,1

Необхідно:

1. Побудувати $F^*(x)$ і графічно зобразити;
2. Графічно зобразити полігони частот і відносних частот.

Розв'язання.

1. Згідно з означенням та властивостями $F^*(x)$ має такий вигляд:

$$F^*(x) = \frac{n_x^{нак.}}{n} = \begin{cases} 0 & x \le -6 \\ 0{,}05 & -6 < x \le -4 \\ 0{,}15 & -4 < x \le -2 \\ 0{,}3 & -2 < x \le 2 \\ 0{,}5 & 2 < x \le 4 \\ 0{,}9 & 4 < x \le 6 \\ 1{,}0 & x > 6 \end{cases}$$

Зобразимо функцію $F^*(x)$ графічно (рис.1.1):

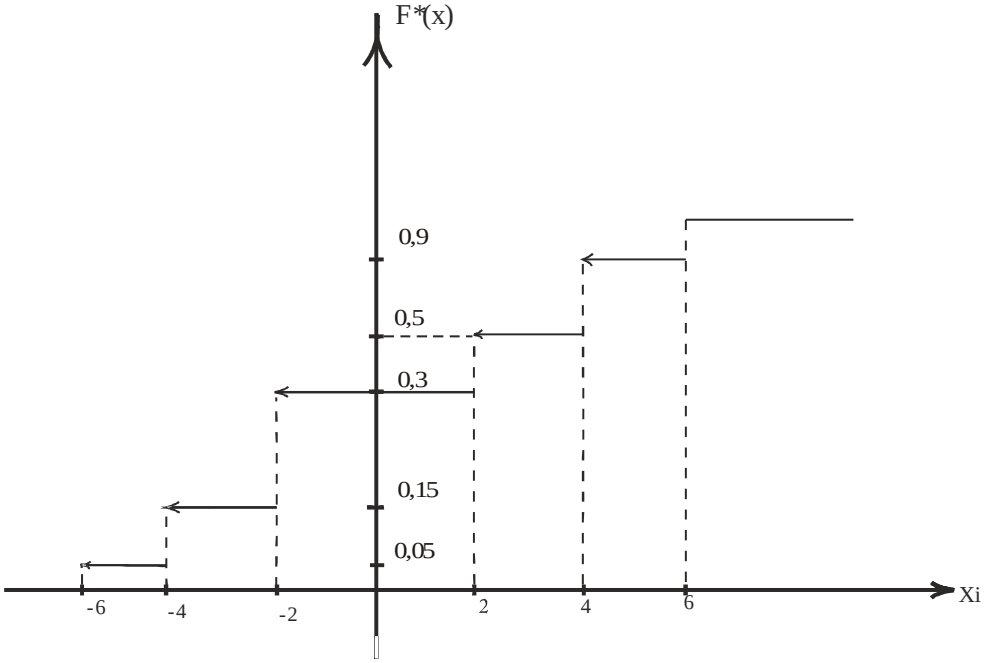


Рис.1.1

2) Зобразимо графічно полігон частот і відносних частот (рис.1.2, 1.3):

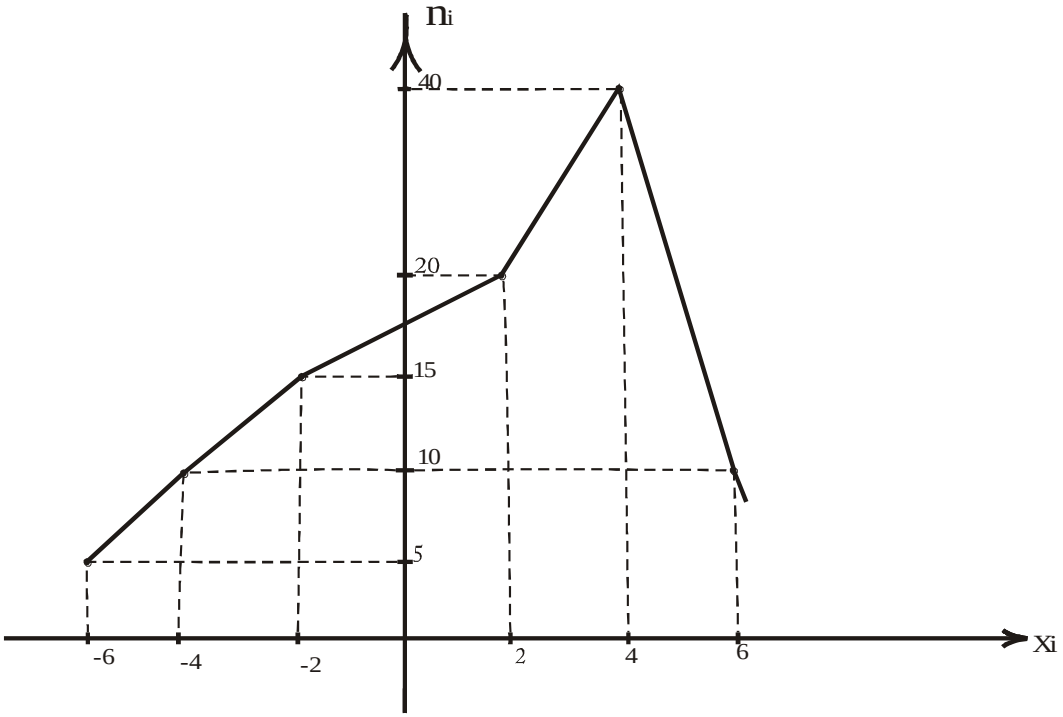


Рис.1.2

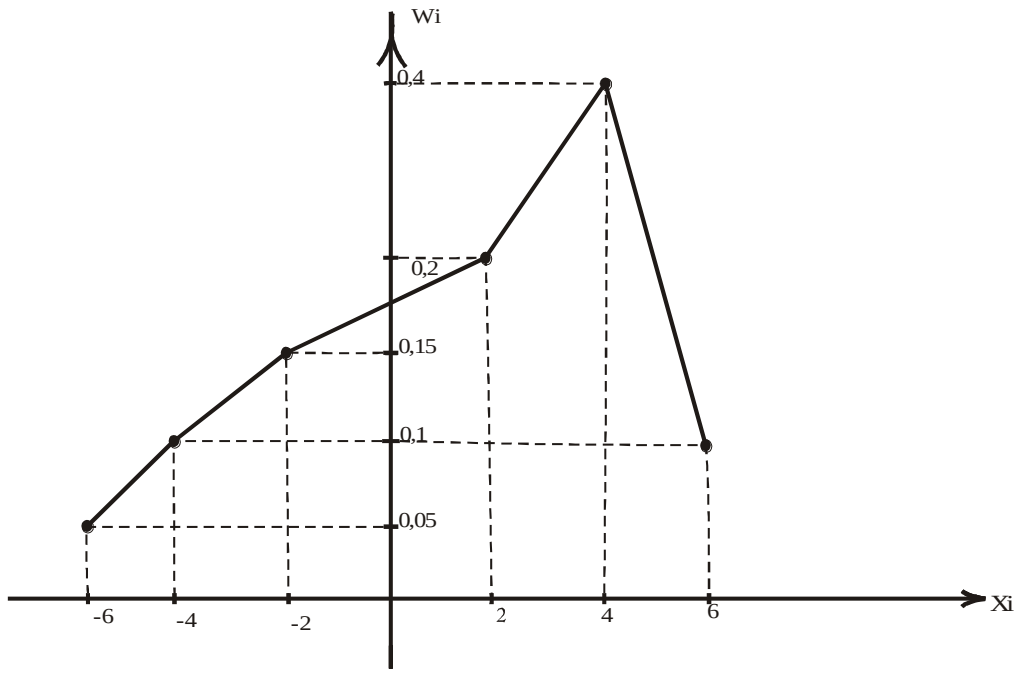


Рис.1.3

Зауваження. При побудові полігона інтервального варіаційного ряду за значення варіанти x_i вибирається середина частинних інтервалів, а в якості частот n_i та відносних частот w_i вибирається їх значення для відповідного інтервалу.

Приклад 4. Вибірка задана інтервальним розподілом частот:

Таблиця 8

$(x_i\,;x_{i+1})$	(1;2)	(2;3)	(3;4)	(4;5)	(5;6)	(6;7)	(7;8)
n_i	19	9	12	14	7	22	17

$\sum n_i = 100$

Побудувати полігон відносних частот.

Розв’язання. Запишемо інтервальний розподіл відносних частот

Таблиця 9

$(x_i\,;x_{i+1})$	(1;2)	(2;3)	(3;4)	(4;5)	(5;6)	(6;7)	(7;8)
w_i	0,19	0,09	0,12	0,14	0,07	0,22	0,17

$\sum w_i = 1$

Знайдемо середини частинних інтервалів

$$x_1^{сер} = \frac{x_1 + x_2}{2} = \frac{1+2}{2} = 1{,}5; \quad x_2^{сер} = \frac{x_2 + x_3}{2} = \frac{2+3}{2} = 2{,}5;$$

$$x_3^{сер} = \frac{x_3 + x_4}{2} = \frac{3+4}{2} = 3{,}5; \quad x_4^{сер} = \frac{x_4 + x_5}{2} = \frac{4+5}{2} = 4{,}5;$$

$$x_5^{сер} = \frac{x_5 + x_6}{2} = \frac{5+6}{2} = 5{,}5; \quad x_6^{сер} = \frac{x_6 + x_7}{2} = \frac{6+7}{2} = 6{,}5;$$

$$x_7^{сер} = \frac{x_7 + x_8}{2} = \frac{7+8}{2} = 7{,}5.$$

Відкладемо на осі абсцис одержані середини частинних інтервалів $x_i^{сер}$, а на осі ординат – значення відповідних їм відносних частот w_i. Послідовно з'єднуючи між собою точки ($x_i^{сер}$; w_i) – відрізками, отримуємо полігон відносних частот.

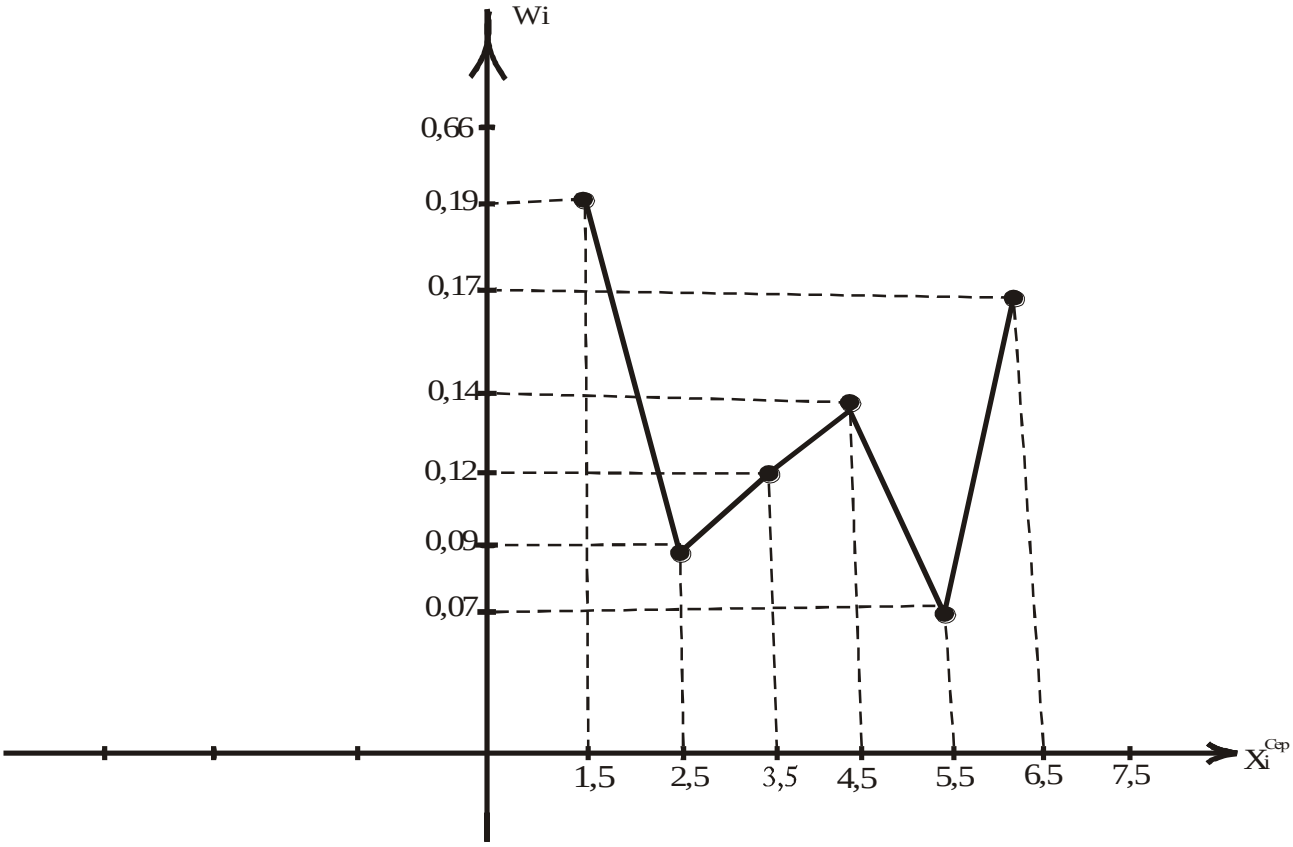


Рис.1.4

Означення. Гістограмою частот називають ступінчасту фігуру, що складається з прямокутників, основами яких є частинні інтервали довжиною h, а висоти дорівнюють відношенню $\frac{n_i}{h}$ (щільність частоти). Площа i-го частинного прямокутника $h\frac{n_i}{h} = n_i$, тобто сумі частот варіант, що потрапили в

i-й частинний інтервал. Площа гістограми частот дорівнює сумі всіх частот, тобто обсягу вибірки, тобто площа гістограми частот дорівнює $\sum n_i = n$.

Аналогічно, гістограмою відносних частот називають ступінчасту фігуру, що складається з прямокутників, основами яких є частинні інтервали довжиною h, а висоти дорівнюють відношення $\frac{w_i}{n}$ (щільність відносної частоти). Площа і-го частинного прямокутника дорівнює $h\frac{w_i}{h} = w_i$, тобто $S = \sum_{i=1}^{k} w_i = 1$, площа гістограми відносних частот дорівнює одиниці.

Зауваження 1. Гістограму можна побудувати тільки для інтервального статистичного розподілу.

Зауваження 2. Очевидно, що при збільшенні n можна вибрати все більш малі інтервали (h), при цьому гістограма буде наближатися до деякої кривої, яка обмежує площу близьку до 1. Ця крива є графіком щільності розподілу випадкової величини X.

Зауваження 3. Комулятивна крива (комулята) – крива накопичених частот або відносних частот. Для дискретного статистичного розподілу комулятою є ламана лінія, яка з'єднує точки $(x_i; n_i^{нак.}), (x_i; w_i^{нак.})$ або ($i=1,2,\ldots,k$). Для інтервального статистичного розподілу є ламана лінія, яка починається з точки, абсциса яка дорівнює початку першого інтервалу, а ордината – накопиченій частоті, рівній нулю.

Зауваження 4. Полігон і гістограма аналогічні криві розподілу ознаки X, а емпірична функція розподілу $F^*(x)$ - функція розподілу випадкової величини X.

Розглянемо приклад побудови гістограми, полігона частот для інтервального статистичного розподілу.

Приклад 4. Вибірку задано інтервальним розподілом частот

Таблиця 9

(x_i, x_{i+1})	(1;2)	(2;3)	(3;4)	(4;5)	(5;6)
n_i	13	9	5	16	7

Розв'язання.

1) Побудуємо гістограму частот.

Знайдемо об'єм вибірки *n=13+9+5+16+7=50* і довжину інтервалів $h = x_{i+1} - x_i = 1$, а також висоти прямокутників $\frac{n_i}{h}$.

$\frac{n_1}{h} = \frac{13}{1} = 13$; $\frac{n_2}{h} = \frac{9}{1} = 9$; $\frac{n_3}{h} = \frac{5}{1} = 5$; $\frac{n_4}{h} = \frac{16}{1} = 16$; $\frac{n_5}{h} = \frac{7}{1} = 7$. Відповідно до заданих даних побудуємо прямокутники з основою $h = 2$ і висотами $\frac{n_i}{h}$.

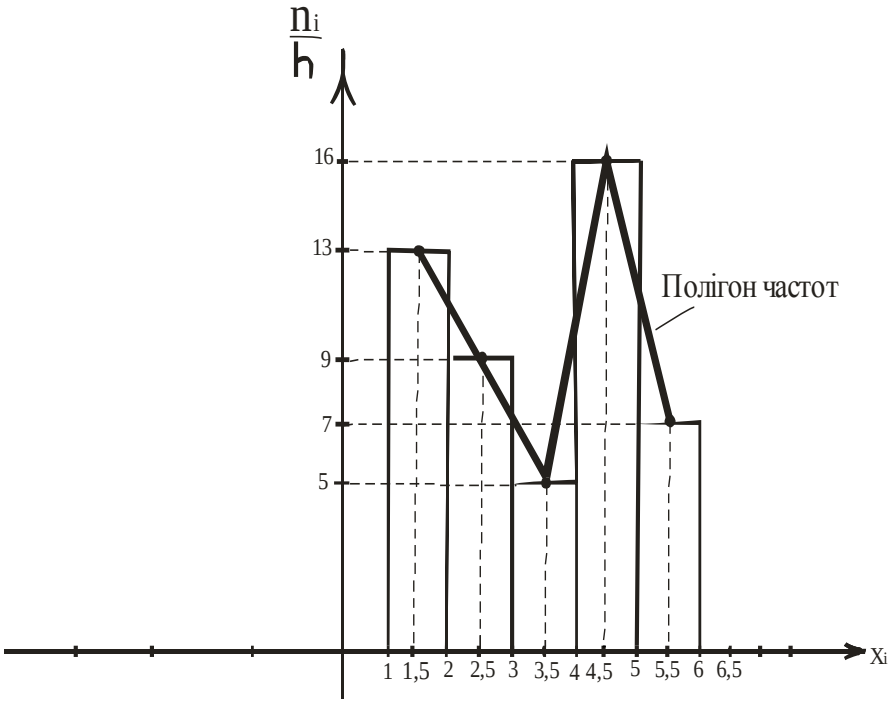


Рис.1.5

Прямокутники утворюють гістограму частот (рис.1.5).

2) Для побудови полігона частот треба з'єднати середини верхньої основи кожного прямокутника ламаною лінією.

1.4 Числові характеристики вибірки

Побудувавши статистичний розподіл вибірки і зобразивши його графічно, можна одержати первісне уявлення про закономірності, які мають місце у ряду спостережень. Але на практиці часто цього не достатньо. Така ситуація виникає, коли необхідно уточнити ті або інші відомості про ряд розподілу, або коли виникає необхідність порівняти два ряди і більше. При цьому необхідно порівняти однотипні варіаційні ряди, тобто ряди, які одержані при обробці зрівнювальних статистичних даних. Зрівнювальні розподіли можуть суттєво відрізнятися одне від одного. Тому для подальшого вивчення зміни значень випадкової величини використовують числові характеристики варіаційних рядів, які зображають деякі сталі величини, які подають варіаційний ряд в цілому і відображають властивості, сукупності закономірностей, що вивчаються. До таких числових характеристик відносяться середня величина ряду розподілу, величини, які відображають варіацію змін – розмах, дисперсія, середнє квадратичне відхилення та інші.

1. Вибіркова середня величина $\overline{x_в}$ вибірки. Дана величина визначається формулами

$$\overline{x_в} = \frac{\sum_{i=1}^{k} x_i n_i}{n} \text{ (для упорядкованої вибірки) або } \overline{x_в} = \sum_{i=1}^{n} x_i w_i \qquad (1.4)$$

де x_i - значення *i*-ої варіанти, n_i - частота *i*-ої варіанти, n - об'єм вибірки, k – кількість варіант у вибірках, w_i - відносна частота *i*-ої варіанти. Для інтервального варіаційного ряду x_i є серединою інтервалів варіаційного ряда.

$$\overline{x_в} = \frac{\sum_{i=1}^{n} x_i}{n} \text{ (для невпорядкованої вибірки)} \qquad (1.5)$$

Статистичний зміст середньої величини заклечається у тому, що навколо неї концентруються інші результати спостережень.

2. Відхил варіант визначають формулою

$$(x_i - \overline{x_в})n_i,\ i = \overline{1,k} \qquad (1.6)$$

При цьому очевидно, що $\sum_{i=1}^{k}(x_i - \overline{x_в})n_i = 0$

3. Модою $Mo*$ вибірки варіант називається варіанта, якій відповідає найбільша частота. Якщо варіаційний ряд має одну таку частоту, то ряд зветься одномодальним, дві – двомодальним і т.д. Для визначення моди інтервального статистичного розподілу необхідно знайти такий частинний інтервал, що має найбільшу частоту появи.

Мода обчислюється за формулою

$$Mo* = x_{i-1} + \frac{n_{Mo} - n_{Mo-1}}{2n_{Mo} - n_{Mo-1} - n_{Mo+1}} h, \qquad (1.7)$$

де x_{i-1} - початок модального інтервалу; h - довжина часткового інтервалу; n_{Mo} - частота модального інтервалу; n_{Mo-1} - частота домодального інтервалу; n_{Mo+1} - частота післямодального інтервалу.

4. Медіана $Me*$. Медіаною дискретного статистичного розподілу називають варіанту, яка поділяє варіаційний ряд на дві рівні частини за кількістю варіант. Для визначення медіани інтервального статистичного розподілу вибірки необхідно визначити медіанний частковий інтервал. Якщо, наприклад, на і-му інтервалі $[x_{i-1}, x_i]$ $F*(x_{i-1}) < 0{,}5$ і $F*(x_i) > 0{,}5$, то беручи до уваги те, що досліджувана ознака X є неперервною і при цьому $F*(x)$ є не спадною функцією, всередині інтервалу $[x_{i-1}, x_i]$неодмінно існує таке значення $x = Me*$, де $F*(Me*) = 0{,}5$. Медіана обчислюється за формулою

$$Me* = x_{i-1} + \frac{0{,}5 - F*(x_{i-1})}{F*(x_i) - F*(x_{i-1})} h, \qquad (1.8)$$

де $h = x_i - x_{i-1}$ - довжина інтервала

Приклад 5. За даними інтервального статистичного розподілу вибірки

Таблиця 11

h =4	0-4	4-8	8-12	12-16	16-20	20-24
n_i	6	14	20	25	30	5

побудувати гістограму частот і $F*(x)$. Визначити Мо*, Ме*.

Розв’язання. Гістограму частот зобразимо на рис. 6

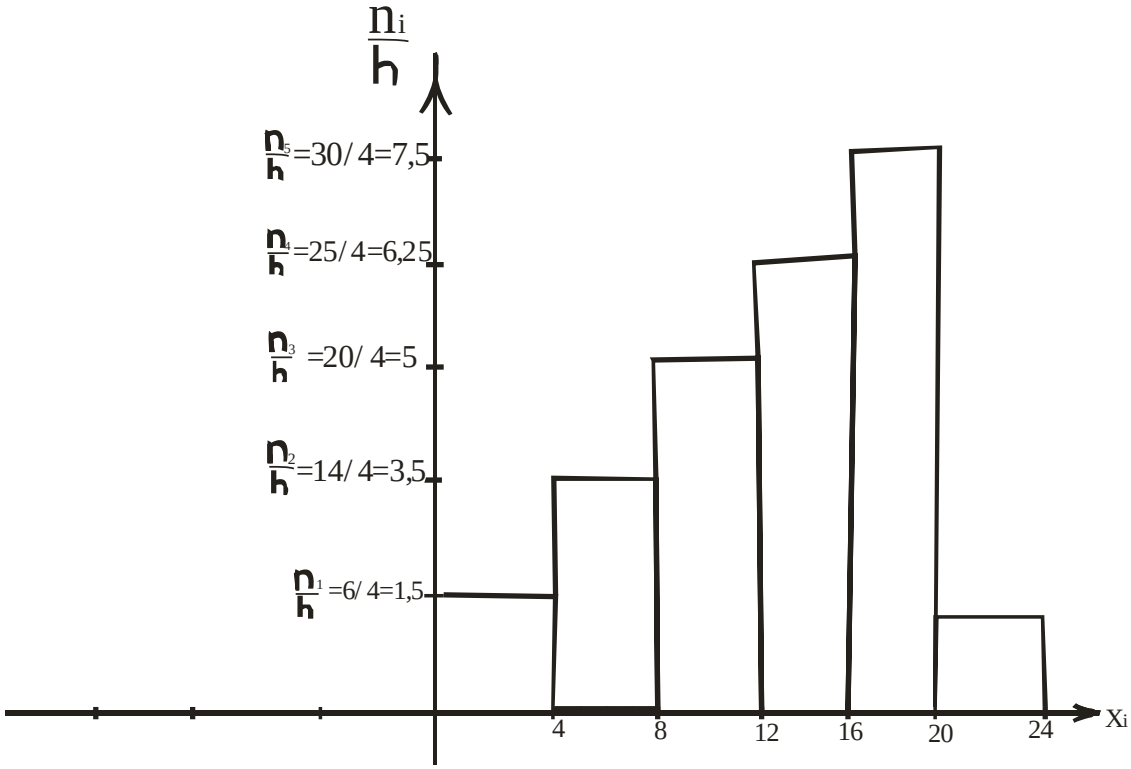


Рис.1.6

Побудуємо графік $F*(x)$, для цього спочатку розрахуємо її значення на кожному з інтервалів, скориставшись формулою $F*(x)=\frac{n_x^{нак}}{n}$.

$$F*(x)=\begin{cases} 0 & x\le 0 \\ \frac{6}{100}=0{,}06 & 0<x\le 4 \\ \frac{20}{100}=0{,}2 & 4<x\le 8 \\ \frac{40}{100}=0{,}4 & 8<x\le 12 \\ \frac{65}{100}=0{,}65 & 12<x\le 16 \\ \frac{95}{100}=0{,}95 & 16<x\le 20 \\ \frac{100}{100}=1 & 20<x\le 24 \end{cases}$$

Графік $F*(x)$ зображено на рис. 1.7.

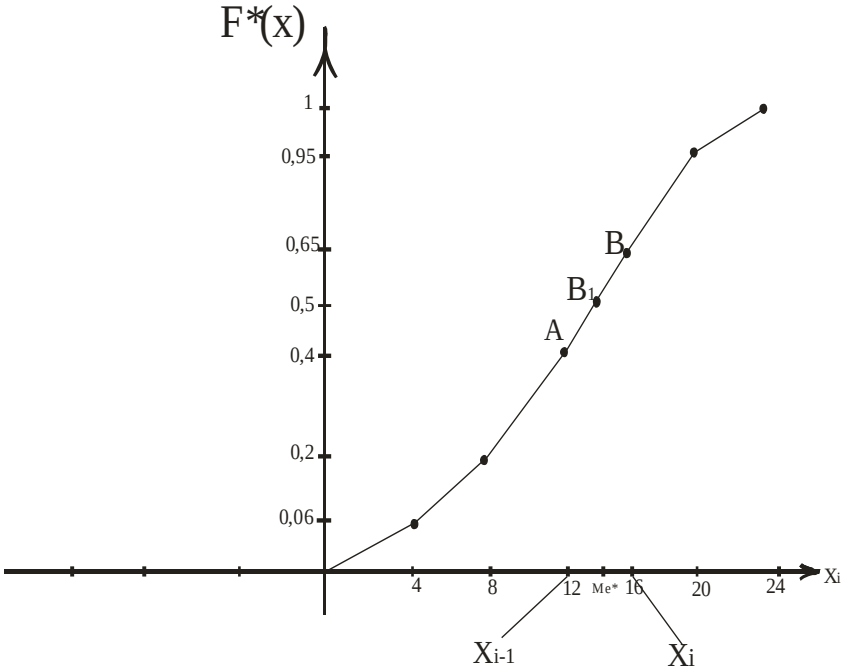


Рис.1.7

Використовуючи рис.7 і формулу (1.8) обчислимо значення *Me**.

$x_{i-1} = 12$, $x_i = 16$, $h = x_i - x_{i-1} = 4$, $F*(x_{i-1}) = 0{,}4$, $F*(x_i) = 0{,}65$,

$$Me* = 12 + \frac{0{,}5 - 0{,}4}{0{,}65 - 0{,}4} 4 = 13,$$

Відповідно до таблиці 11 найбільшу частоту має інтервал (16-20), значення якої дорівнює 30, тому за формулою (1.7) обчислимо воду *Mo** інтервального розподілу, використовуючи значення $x_{i-1} = 16$, $h = x_i - x_{i-1} = 4$, $h_{Mo} = 30$, $h_{Mo-1} = 25$, $h_{Mo+1} = 5$

$$Mo* = x_{i-1} + \frac{h_{Mo} - h_{Mo-1}}{2n_{Mo} - n_{Mo-1} - n_{Mo+1}} h = 16 + \frac{30 - 25}{2 \cdot 30 - 25 - 5} \cdot 4 = 16{,}7$$

Таким чином, для вихідного інтервального статистичного розподілу, заданого таблицею 11 медіана *Me*=13,6; Mo*=16,7.*

5. Розмах R.

Для грубого розходження варіант відносно середнього значення $\overline{x_в}$ застосовується величина, яка дорівнює різниці між найбільшою і найменшою варіантами варіаційного ряду $R = x_{\max} - x_{\min}$ (1.9)

6. Дисперсія вибірки.

Для більш точного оцінювання розходження варіант вибірки відносно $\overline{x_в}$ вибирається дисперсія.

Вибірковою дисперсією $D_в$ називається середній квадратичний відхил варіант від їх середньої арифметичної, яку можна обчислити за формулами.

$$D_в = \frac{\sum_{i=1}^{k}(x_i - \overline{x_в})^2}{n} \cdot n_i \text{ (для упорядкованої вибірки)} \qquad (1.10)$$

$$D_в^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x_в})^2}{n} \text{(для невпорядкованої вибірки)} \qquad (1.11)$$

Дисперсію $D_в$, як правило, обчислюють за формулою

$$D_в = \frac{\sum_{i=1}^{k} x_i^2 n_i}{n} - (\overline{x_в})^2 \qquad (1.12)$$

7. Середнє квадратичне відхилення $D_в$ вибірки.

В якості міри розходження (варіації) бажано мати характеристику, яка вимірюється в тих же одиницях, що і значення ознаки. Такою характеристикою є вибіркове середнє квадратичне відхилення $\sigma_в$, яке обчислюється за формулою

$$\sigma_в = \sqrt{D_в} = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \overline{x_в})^2 \cdot n_i}{n}} \qquad (1.13)$$

8. Коефіцієнт варіації $\overline{\nu}$.

Для порівняння оцінок варіацій статистичних рядів з різними значеннями $\overline{x_в} = 0$ вводиться поняття коефіцієнта варіації, який дорівнює процентному відношення вибіркового середнє-квадратичного відхилу $\sigma_в$ до середньої арифметичної $\overline{x_в}$, тобто:

$$\overline{\nu} = \frac{\sigma_в}{\overline{x_в}} \cdot 100\% \quad (\overline{x_в} \neq 0) \qquad (1.14)$$

Зауваження. Для визначення $\overline{x_в}, D_в, \sigma_в$ для інтервального статистичного розподілу вибірки треба перейти до дискретного статистичного розподілу, варіантами якого є середина часткових інтервалів

$x_i^* = x_{i-1} + \frac{h}{2} = x_i - \frac{h}{2}$, який має вигляд:

Таблиця 12

$x_i^* = x_{i-1} + \frac{h}{2} = x_i - \frac{h}{2}$	x_1^*	x_2^*	x_3^*	…	x_k^*
n_i	n_1	n_2	n_3	…	n_k

Приклад 6. За даними інтервального статистичного розподілу вибірки, в якому наведено розподіл маси новонароджених x_i

Таблиця 13

	1-1,2	1,2-1,4	1,4-1,6	1,6-1,8	1,8-2,0	2,0-2,2	2,2-2,4	2,4-2,6	2,6-2,8	2,8-3,0	3,0-3,2
n_i	5	12	18	22	36	24	19	15	11		2

Обчислити $\overline{x_в}$, $D_в$, $\sigma_в$.

Розв’язання. Побудуємо дискретний статистичний розподіл за заданим інтервальним. Оскільки, $h = 0{,}2$, то одержимо

Таблиця 14

$x_i^* = x_{i-1} + \frac{h}{2} = x_i - \frac{h}{2}$	1,1	1,3	1,5	1,7	1,9	2,1	2,3	2,5	2,7	2,9	3,1
n_i	5	12	18	22	36	24	19	15	11	9	2

Підрахуємо число варіант $n = 5+12+18+22+36+24+19+15+11+9+2=173$

За формулами (1.4), (1.12), і (1.13) обчислимо $\overline{x_в}$, $D_в$, $\sigma_в$.

$$\overline{x_в} = \frac{\sum_{i=1}^{k} x_i^* n_i}{n} = \frac{1{,}1\cdot 5 + 1{,}3\cdot 12 + 1{,}5\cdot 18 + 1{,}7\cdot 22 + 1{,}9\cdot 36 + 2{,}1\cdot 24 + 2{,}3\cdot 19 + 2{,}5\cdot 15 + 2{,}7\cdot 11 + 2{,}8\cdot 9 + 3{,}1\cdot 2}{173} =$$

$$= \frac{347{,}5}{173} = 2{,}008671 \text{ кг.}$$

Отже, $\overline{x_в} \approx 2{,}009$ кг.

$$D_в = \frac{\sum_{i=1}^{k} x_i^2 n_i}{n} - (\overline{x_в})^2 = \frac{(1{,}1)^2\cdot 5 + (1{,}5)^2\cdot 1{,}8 + (1{,}7)^2\cdot 22 + (1{,}9)^2\cdot 36}{173} +$$

$$+ \frac{(2{,}1)^2\cdot 24 + (2{,}3)^3\cdot 19 + (2{,}5)^2\cdot 15 + (2{,}7)^2\cdot 11 + (2{,}9)^2\cdot 9 + (3{,}1)^2\cdot 2}{173} - (2{,}009)^2 \approx 0{,}217$$

$$D_в \approx 0{,}217,\ \sigma_в = \sqrt{0{,}217} \approx 0{,}466$$

1.5 Емпіричні моменти

1. Початкові емпіричні моменти.

Середнє арифметичне значення зважених варіант у степені *k* (*k=1,2,3,…)* називають початковим емпіричним моментом k-го порядку ν_k^*, який обчислюється за формулою $\nu_k^* = \frac{\sum_{i=1}^{m} x_i^k \cdot n_i}{n}$ $(n_1 + n_2 + \ldots + n_m = n)$ (1.15)

Очевидно, $\nu_1^* = \frac{\sum_{i=1}^{m} x_i \cdot n_i}{n} = \overline{x_в}$ є початковим моментом першого порядку варіаційного ряду. При *k*=2 одержимо початковий момент другого порядку, тобто

$$\nu_2^* = \frac{\sum_{i=1}^{m} x_i^2 \cdot n_i}{n} \qquad (1.16)$$

Очевидно, вибіркова дисперсія $D_в$ дорівнює $D_в^2 = \nu_2^* - (\nu_1^*)^2$ (1.17)

2. Центральні емпіричні моменти k-го порядку. (*k=1,2,3,…)* називають центральним емпіричним моментом k-го порядку $\mu_k^* = \frac{\sum_{i=1}^{m} (x_i - \overline{x_в})^k \cdot n_i}{n}$ (1.18)

Знайдемо центральні емпіричні моменти при *k=1*

$$\mu_k^* = \frac{\sum_{i=1}^{m}(x_i - \overline{x_в})^k \cdot n_i}{n} = \frac{\sum_{i=1}^{m} x_i n_i}{n} - \overline{x_в}\frac{\sum_{i=1}^{m} n_1}{n} = \overline{x_в} - \overline{x_в} = 0$$

При k=2 одержимо

$$\mu_2^* = \frac{\sum_{i=1}^{m}(x_i - \overline{x_в})^2}{n} = \frac{\sum_{i=1}^{m} x_i^2 n_i}{n} - (\overline{x_в})^2 = \overline{x_в^2} - (\overline{x})^2 = D_в$$ вибіркова дисперсія.

На практиці найчастіше застосовуються центральні емпіричні моменти третього та четвертого порядків, що обчислюють за формулами

$$\mu_3^* = \frac{\sum_{i=1}^{m}(x_i - \overline{x_в})^3 \cdot n_i}{n} = \nu_3^* - 3\nu_2^* \cdot \nu_1^* + 2(\nu_1^*)^2 \qquad (1.19)$$

$$\mu_4^* = \frac{\sum_{i=1}^{m}(x_i - \overline{x_в})^4 \cdot n_i}{n} = \nu_4^* - 4\nu_3^* \cdot \nu_1^* + 6\nu_2^*(\nu_1^*)^2 - 3(\nu_1^*)^4 \qquad (1.20)$$

3. Коефіцієнт асиметрії A_s^*.

Центральний емпіричний момент третього порядку застосовується для обчислення асиметрії:

$$A_s^* = \frac{M_3^*}{D_в} = \frac{\sum_{i=1}^{m}(x_i - \overline{x_в})^3 \cdot n_i}{D_в} \qquad (1.21)$$

Якщо $A_s^* = 0$, то розподіл має симетричну форму, тобто варіанти, рівновіддалені від $\overline{x_в}$, мають однакову частоту. При $A_s^* > 0$ ($A_s^* < 0$) говорять про додатну (правосторонню) – варіанти $x_i > \overline{x_в}$ переважають варіанти $x_i \le \overline{x_в}$ або від'ємну (лівосторонню) асиметрію, варіанти статистичного розподілу $x_i < \overline{x_в}$ переважають $x_i > \overline{x_в}$.

4. Ексцес варіаційного ряду.

Ексцесом варіаційного ряду називають число

$$E_s^* = \frac{M_4^*}{\sigma_в^4} - 3 = \frac{\sum_{i=1}^{m}(x_i n_i - \overline{x_в})^4}{\sigma_в^4} - 3 = \frac{\nu_4^* - 4\nu_3^* + 6\nu_2^* \cdot (\nu_1^*)^2 - 3(\nu_1^*)^4}{\sigma_в^4} - 3 \qquad (1.22)$$

Ексцес, як правило, використовується при дослідженні неперервних ознак генеральних сукупностей, оскільки він оцінює крутизну закону розподілу випадкової величини порівняно з нормальним. Для нормального закону $E_s^* = 0$.

Приклад 7. Оцінки в балах x_i, одержані абітурієнтами на вступних іспитах з математики, наведені у вигляді дискретного розподілу:

Таблиця 15

x_i	15	25	35	45	55	65	75	85
n_i	5	10	15	20	25	15	8	2

Обчислити A_s^*, E_s^*.

Розв’язання.

1. Використовуючи (1.21) та (1.19) і враховуючи, що $n = \sum n_i = 100$, обчислимо

$$A_s^* = \frac{\mu_3^*}{S_в^3} = \frac{\nu_3^* - 3\nu_2^* \cdot \nu_1^* + 2(\nu_1^*)^3}{S_в^3}$$

$$\nu_1^* = \overline{x_в} = \frac{15\cdot 5 + 25\cdot 10 + 35\cdot 15 + 45\cdot 20 + 55\cdot 25 + 65\cdot 15 + 75\cdot 8 + 85\cdot 2}{100} = 48{,}7$$

$$\nu_1^* = 48{,}7$$

$$\nu_2^* = \frac{\sum_{i=1}^{8} x_i^2 n_i}{n} = \frac{(15)^2 + (25)^2\cdot 10 + (35)^2\cdot 15 + (45)^2\cdot 20 + (55)^2\cdot 25 + (65)^2\cdot 15 + (75)^2\cdot 8 + (85)^2\cdot 2}{100} = 2647$$

$$\nu_2^* = 2647$$

$$\nu_3^* = \frac{\sum_{i=1}^{8} x_i^3 n_i}{n} = \frac{(15)^3\cdot 5 + (25)^3\cdot 10 + (35)^3 15 + (45)^3\cdot 20 + (55)^3\cdot 25 + (65)^3\cdot 15 + (75)^3\cdot 8 + (85)^3\cdot 2}{100} =$$

$$= \frac{16875 + 156250 + 643125 + 1822500 + 4159375 + 4119375 + 3375000 + 1228250}{100} = \frac{15520750}{100} = 155207{,}5$$

$$\nu_3^* = 155207{,}5$$

$$D_в = \sum_{i=1}^{8} x_i^2 n_i - (\overline{x_в})^2 = \nu_2^* - (\nu_1^*)^2 = 2647 - (48{,}7)^2 = 275{,}31$$

$$S_в = \sqrt{S_в^2} = 16{,}59$$

$$A_s^* = \frac{155207{,}5 - 3 \cdot 2647 \cdot 48{,}7 + 2(48{,}7)^3}{4567{,}39} = \frac{155207{,}7 - 386726{,}7 + 231002{,}6}{4567{,}4} = -\frac{516{,}6}{4567{,}4} - 0{,}113$$

Відповідь: A_s^*=*-0,11*. Висновок: оскільки A_3 порівняно малий, то статистичний розподіл близький до симетричного.

2. Для знаходження ексцеса E_s^* скористуємося формулою (1.22)

$$E_s^* = \frac{\nu_4^* - 4\nu_3^* \cdot \nu_3^* + 6\nu_2^* \cdot (\nu_1^*)^2 - 3(\nu_1^*)^4}{\delta_в^4} - 3$$

$\nu_1^* = 48{,}7$; $\nu_2^* = 2647$; $\nu_3^* = 155207{,}5$; $\sigma_в$ =16, 59 знайдені в п.1.

Визначимо початковий момент

$$\nu_4^* = \frac{\sum_{i=1}^{8} x_i n_i}{n} = \frac{(15)^4 \cdot 5 + (25)^4 \cdot 10 + (35)^4 \cdot 15 + (45)^4 \cdot 20 + (55)^4 \cdot 25 + (65)^4 \cdot 15 + (75)^4 \cdot 8 + (85)^4 \cdot 2}{100} =$$

$$= \frac{253125 + 3906250 + 22509375 + 82012500 + 228765625 + 267759375 + 253125000 + 104401250}{100} =$$

$$= \frac{962732500}{100} = 9627325$$

$\nu_4^* = 9627325$, $\delta_в^4 = (16{,}59)^4 = 75795{,}6$

$$E_s^* = \frac{9627324 - 4 \cdot 155207{,}5 \cdot 48{,}7 + 6 \cdot 2647 \cdot (48{,}7)^2 - 3(48{,}7)^4}{75795{,}6} =$$

E_s^*=*15,77.* Оскільки $E_s^* > 0$, то вершина закону розподілу випадкової величини, заданого щільністю ймовірностей, буде гостровершинною.

Питання для самоперевірки

1)Дати визначення генеральної та вибіркової сукупності.

2)Яка вибірка є представницькою (репрезентативною)?

3)Способи відбору вибірки, дати означення кожного з них.

4)Дати визначення дискретної та неперервної ознаки *X*.

5) Що називається варіантою, варіаційним рядом?

6)Що таке частота, відносна частота варіант?

7)Дати визначення дискретного статистичного розподілу вибірки.

8) Аналогом чого є статистичний розподіл вибірки?

9)Як будується статистичний розподіл вибірки для неперервної ознаки *X*?

10)Дати визначення емпіричної функції розподілу $F^*(x)$.

11)Властивості $F^*(x)$.

12)Що таке полігон частот та відносних частот і для чого він використовується?

13) Як будується полігон частот та відносних частот для інтервального варіаційного ряду?

14) Дати означення гістограми частот та відносних частот.

15) Дати означення кумулятивної кривої (помуляти) для дискретного і інтервального розподілу ознаки *X*.

16) $\overline{x_в}$, $D_в$, $\sigma_в$ для дискретного статистичного розподілу вибірки.

17) $\overline{x_в}$, $D_в$, $\sigma_в$ для інтервального статистичного розподілу вибірки.

18)Як визначається Me^* і Mo^* для інтервального статистичного розподілу?

19)Що називається розмахом, коефіцієнтом варіації?

20)Що таке початковий момент *k*-го порядку?

21)Що таке центральний момент *k*-го порядку

22)Асиметрія і ексцес статистичного розподілу вибірки.

Індивідуальні завдання

1. При вивченні випадкової величини X у результаті 40 незалежних спостережень дістали вибірку: 10, 13, 10, 9, 9, 12, 12, 6, 7, 9, 8, 9, 11, 9, 14, 13, 9, 8, 8, 7, 10, 10, 11, 11, 11, 12, 8, 7, 9, 10, 14, 13, 8, 8, 9, 10, 11, 11, 12, 12.

Потрібно: 1) Побудувати дискретній статистичний розподіл для цієї вибірки, а також полігон частот і $F^*(x)$; 2)Обчислити $\overline{x_в}$, $S_в$, R, V; 3) Знайти M_o^*, Me^*.

Відповідь: $\overline{x_в}$=10, $S_в$=2.

2. П'ятдесят абітурієнтів на вступних іспитах з інформатики дістали 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.

Потрібно: 1) Побудувати дискретний статистичний розподіл. Полігон частот, $F^*(x)$; 2) Обчислити $\overline{x_в}$, $S_в$, R, V; 3) Знайти M_o^*, Me^*.

Відповідь: $\overline{x_в}$=15,78, $S_в$=1,93.

3. Через кожну 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.

Потрібно: 1) Побудувати дискретний статистичний розподіл, полігон частот і $F^*(x)$; 2) Обчислити $\overline{x_в}$, $S_в$, R, V; 3) Знайти M_o^*, Me^*.

Відповідь: $\overline{x_в}$=220,25, $S_в$=2,66.

4. За допомогою радіодальноміра було здійснено 16 вимірювань однієї і тієї ж самої відстані. Результати вимірювання в метрах наведені у вигляді статистичного ряду: 201, 195, 207, 203, 191, 208, 198, 210, 204, 192, 195, 211, 206, 196, 208, 197.

Потрібно: 1) Побудувати дискретний статистичний розподіл, полігон частот і $F^*(x)$; 2) Обчислити $\overline{x_в}$, $S_в$, R, V; 3) Знайти M_o^*, Me^*.

Відповідь: $\overline{x_в}$=210 м, $S_в$=13,85.

5. Для обчислення середньої врожайності озимої пшениці x_i кооперативне поле площею 2000 га було поділено на 20 рівних ділянок n_i. Фактичний урожай на кожній ділянці наведено в таблиці:

x_i, ц/га	25	30	35	40	45
n_i	2	3	8	4	3

Потрібно: 1) Побудувати дискретний статистичний розподіл, полігон частот і $F^*(x)$; 2) Обчислити $\overline{x_в}$, $S_в$, R, V; 3) Знайти M_o^*, Me^*.

Відповідь: $\overline{x_в}$ =35,75 ц/га , $S_в$ =5,76 ц/га.

6. На кожну сотню деталей, що їх виготовляє цех, у середньому припадає дві браковані. Було перевірена 10 партій по 100 деталей у кожній. Відхилення кількості виявлених бракованих деталей від середнього x_i наведено в таблиці:

Номер партії	1	2	3	4	5	6	7	8	9	10
x_i	-1	0	1	1	-1	1	0	-2	2	1

Потрібно: 1) Побудувати дискретний статистичний розподіл, полігон частот і $F^*(x)$; 2) Обчислити $\overline{x_в}$, $S_в$, R, V; 3) Знайти M_o^*, Me^*.

Відповідь: $\overline{x_в}$ =0,2 , $S_в$ =1,233.

7. Результати вимірювання максимальної місткості x_i 20-ти конденсаторів у пікофарадах наведено в таблиці:

Номер	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
x_i, пФ	4,4	4,31	4,4	4,4	4,65	4,56	4,71	4,54	4,36	4,56	4,31	4,42	4,6	4,35	4,5	4,4	4,43	4,48	4,42	4,45

Потрібно: 1) Побудувати дискретний статистичний розподіл, полігон частот і $F^*(x)$; 2) Обчислити $\overline{x_в}$, $S_в$, R, V; 3) Знайти M_o^*, Me^*.

Відповідь: $\overline{x_в}$=4,47 пФ , $S_в$=1,108 пФ.

8. Із партії однотипних амперметрів N для контролю відібрано 10 штук. Вимірювання показали такі відхилення показали такі відхилення x_i від номіналу в міліамперметрах:

Номер партії	1	2	3	4	5	6	7	8	9	10
x_i, мА	1	3	-2	2	4	2	5	3	-2	4

Потрібно: 1) Побудувати дискретний статистичний розподіл, полігон частот і $F^*(x)$; 2) Обчислити $\overline{x_в}$, $S_в$, R, V; 3) Знайти M_o^*, Me^*.

Відповідь: $\overline{x_в}$=2 мА, $S_в$=2,23 мА.

9. 200 однотипних деталей n_i після шліфування були піддані контрольним вимірюванням, результати яких наведено в таблиці x_i:

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

Потрібно: 1) Побудувати полігон частот, комуляту, $F^*(x)$; 2) Обчислити $\overline{x_в}$, $S_в$, R, V; 3) Знайдемо M_o^*, Me^*.

Відповідь: $\overline{x_в}$=14,34, $S_в$=0,039.

10. Із партії сталевих кілець, виготовлених заводом, була здійснена вибірка обсягом 200, і результати вимірювання товщини цих кілець x_i наведено у вигляді дискретного статистичного розподілу:

x_i, Мм	14,41	14,43	14,45	14,47	14,49	14,51	14,53	14,55	14,57	14,59	14,61	14,63
n_i	2	2	8	9	9	14	41	76	21	11	4	3

Потрібно: 1) Побудувати полігон відносних частот і комуляту; 2) Обчислити $\overline{x_в}$, $S_в$, R, V; 3) Знайти M_o^*, Me^*.

Відповідь: $\overline{x_в}$=14,34, $S_в$=0,039.

11. У відділі технічного контролю було виміряно діаметри 200 валиків із партії, виготовленої одним верстатом-автоматом. Відхилення виміряних діаметрів від номіналу наведено як інтервальний статистичний розподіл, де X= x_i вимірюється в мікронах:

h=5, Мк	-20…-15	-15…-10	-10…-5	-5…0	0…5	5…10	10…15	15…20	20…25	25…30
n_i	7	11	15	24	49	41	26	17	7	3

Потрібно: 1) Побудувати гістограму частот і $F^*(x)$; 2) Обчислити $\overline{x_в}$, $S_в$, A_S^*, E_S^*, M_o^*, Me^*.

Відповідь: $\overline{x_в}$=4,3 мк; $S_в$=9,79 мк; A_S^*=-0,128 мк; E_S^*=-0,16; M_o^*=3,79 мк; Me^*= -1,46 мк.

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

x ,мм h=2мм	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

Потрібно: 1) Побудувати гістограму частот і $F^*(x)$; 2) Обчислити $\overline{x_в}$, $S_в$, A_S^*, E_S^*, M_o^*, Me^*.

Відповідь: $\overline{x_в}$=14,34 мм; $S_в$=0,039; A_S^*=0,311 мм; E_S^*=1,549 мм; M_o^*=15,34 мм; Me^*= 15,39 мм.

13. Після шліфування 200 однотипних деталей були піддані контрольним вимірюванням, результати яких наведено у формі інтервального статистичного розподілу:

h=1	3,65 – 3,75	3,75 – 3,85	3,85 – 3,95	3,95 – 4,05	4,05 – 4,15	4,15 – 4,25	4,25 – 4,35	4,35 – 4,45
n_i	1	22	40	79	27	26	4	1

Потрібно: 1) Побудувати гістограму частот і $F^*(x)$; 2) Обчислити $\overline{x_в}$, $S_в$, A_S^*, E_S^*, M_o^*, Me^*.

Відповідь: $\overline{x_в}$=4,004 мм; $S_в$=0,126; A_S^*=0,311 мм; E_S^*=-0,117 мм; M_o^*=4,38мм; Me^*= 4,875 мм.

14. У 100 осіб було виміряно зріст x_i. Результати вимірювання наведено як інтервальний статистичний розподіл:

x_i ,см h=4см м	168- 172	172 - 176	176 - 180	180 – 184	184 - 188	192 - 196	196 - 166
n_i	10	20	30	25	10	3	2

Потрібно: 1) Побудувати гістограму частот і $F^*(x)$; 2) Обчислити $\overline{x_в}$, $S_в$, A_S^*, E_S^*, M_o^*, Me^*.

Відповідь: $\overline{x_в}$=178,88 см; $S_в$=98,87 см; A_S^*=0,0063 см; E_S^*=-2,9999 см; M_o^*=178,7 см; Me^*= 178,6 см.

15. 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 ,кг h=0,5	1,0 – 1,5	1,5 – 2,0	2,0 – 2,5	2,5 – 3,0	3,0 – 3,5	3,5 – 4,0	4,0 – 4,5	4,5 – 5,0
n_i	2	8	10	30	40	6	3	1

Потрібно: 1) Побудувати гістограму відносних частот і $F^*(x)$; 2) Обчислити $\overline{x_в}$, $S_в$, A_S^*, E_S^*, M_o^*, Me^*.

Відповідь: $\overline{x_в}$=2,915 кг; $S_в$=0,625 кг; A_S^*=-0,26; E_S^*=0,73; M_o^*=3,11 кг; Me^*=3 кг.

16. Число розладнань у роботі верстатів-автоматів заводу протягом року наведено у вигляді інтервального статистичного розподілу:

x_i h=2	0 – 2	2 - 4	4 - 6	6 - 8	8 - 10	10 - 12	12 - 14	14 - 16	16 - 18	18 - 20	20 - 22	22 - 24
n_i	2	5	7	11	15	18	26	20	14	10	6	3

Потрібно: 1) Побудувати гістограму відносних частот і комуляту $F^*(x)$; 2) Обчислити $\overline{x_в}$, $S_в$, A_S^*, E_S^*, M_o^*, Me^*

Відповідь: $\overline{x_в}$=12,58; $S_в$=4,88; A_S^*=-0,13; E_S^*=-0,39; M_o^*=13; Me^*=11.

17. 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

Потрібно: 1) Побудувати гістограму відносних частот і комуляту $F^*(x)$; 2) Обчислити $\overline{x_в}$, $S_в$, A_S^*, E_S^*, M_o^*, Me^*.

Відповідь: $\overline{x_в}$=58,2%; $S_в$=21,21%; A_S^*=0,0043%; E_S^*=-0,22%; M_o^*=54,1%; Me^*=57,2%.

18. Кількість виготовленого полотна за зміну x_i ткачами наведено у формі інтервального статистичного розподілу:

x_i, м h=10	52,8 – 62,8	62,8 – 72,8	72,8 – 82,8	82,8 – 92,8	92,8 – 102,8	102,8 – 112,8	112,8 – 122,8	122,8 – 132,8	132,8 – 142,8
n_i	8	12	25	39	26	18	12	9	4

Потрібно: 1) Побудувати гістограму відносних частот і комуляту $F^*(x)$; 2) Обчислити $\overline{x_в}$, $S_в$, A_S^*, E_S^*, M_o^*, Me^*.

Відповідь: $\overline{x_в}$=93,13 м; $S_в$=19,1 м; A_S^*=0,31; E_S^*=0,38; M_o^*=88 м;

Me^*=84,12 м.

19. Рівень води в річці x_i відносно номінально вимірюється протягом 50-ти років навесні. Результати вимірювання наведено у формі інтервальго 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

Потрібно: 1) Побудувати гістограму частот і комуляту $F^*(x)$;

2) Обчислити $\overline{x_в}$, $S_в$, E_S^*, Δ_S^*, M_o^*.

Відповідь: $\overline{x_в}$=118,1 см; $S_в$=36,2 см; Δ_S^*=0,38; E_S^*=-0,12; M_o^*=126,86

20. Залежність урожайності x_i ярової пшениці від кількості опадів протягом весни зображено у вигляді інтервального статистичного розподілу:

x_i, ц/га h=2	4,2 – 6,2	6,2 – 8,2	8,2 – 10,2	10,2 – 12,2	12,2 – 14,2	14,2 – 16,2	16,2 – 18,2	18,2 – 20,2	20,2 – 22,2
n_i	5	15	20	25	30	18	8	2	1

Потрібно: 1) Побудувати гістограму частот і комуляту $F^*(x)$;

2) Обчислити $\overline{x_в}$, $S_в$, A_S^*, E_S^*, M_o^*, Me^*.

Відповідь: $\overline{x_в}$=11,9 ц/га; $S_в$=40,8 ц/га; A_S^*=0,0002; E_S^*=-2,9999; M_o^*=12,79 ц/га; Me^*=11 ц/га.

Розділ 2 СТАТИСТИЧНІ ОЦІНКИ ПАРАМЕТРІВ ГЕНЕРАЛЬНОЇ СУКУПНОСТІ

Сформулюємо задачу оцінки параметрів генеральної сукупності у загальному вигляді. Нехай розподіл ознаки *X* генеральної сукупності задається функцією ймовірності $\varphi(x_i,\theta)=P(X=x_i)$ (для дискретної випадкової величини) або щільністю ймовірності $\varphi(x,\theta)$ (для неперервної випадкової величини), яка містить невідомий параметр θ.

Для обчислення параметра θ неможливо дослідити усі елементи генеральної сукупності. Тому параметр θ оцінюють за допомогою вибірки, яка складається із значень $x_1, x_2, ..., x_n$. Ці значення можна розглядати як часткові значення (реалізації) *n* незалежних випадкових величин $X_1, X_2, ..., X_n$, кожна з яких має той же самий закон розподілу, що і випадкова величина *X*.

Означення. Оцінкою θ^* параметра θ називають будь-яку функцію результатів спостережень над випадковою величиною *X* (інакше – статистикою), за допомогою якої судять про значення параметра θ:

$$\theta^* = \theta_n^*(X_1, X_2, ..., X_n).$$

Оскільки $X_1, X_2, ..., X_n$ – випадкові величини, то й оцінка θ^* також є випадковою величиною, яка залежить від закону розподілу випадкової величини *X* і числа *n*.

Параметри генеральної сукупності $M(X)=\bar{X}_Г, D_Г, \sigma_Г, \text{Mo}, \text{Me}, r_{xy}$ є величинами сталими, але їх числове значення невідоме. Ці параметри оцінюються параметрами вибірки: $\bar{x}_В, D_В, \sigma_В, \text{Mo}^*, \text{Me}^*, r_В$, які дістають у процесі обробці вибірки. Вони є величинами непередбачуваними, тобто випадковими.

2. 1 Точкові статистичні оцінки параметрів генеральної сукупності

Статистична оцінка θ^* параметра θ, яка визначається одним числом (точкою), називається *точковою*.

Завжди існує множина функцій від результатів спостережень $X_1, X_2, ..., X_n$, які можна розглядати у якості оцінки параметра θ. Тому виникає питання: якими властивостями має володіти оцінка, щоб вважати її якісною у деякому сенсі? Серед найбільш важливих властивостей розглядають наступні: незміщеність, ґрунтовність, ефективність.

Означення. Точкова статистична оцінка називається незміщеною, коли математичне сподівання цієї оцінки точно дорівнює оцінювальному параметру θ, а саме:

$$M\left(\theta^*\right)=\theta. \qquad (2.1)$$

Якщо

$$M(\theta^*) \neq \theta, \tag{2.2}$$

точкова статистична оцінка θ^* називається зміщеною відносно параметра генеральної сукупності θ.

Різниця

$$\theta^* - \theta = \delta \tag{2.3}$$

називається *зміщенням статистичної оцінки* θ^*.

Оцінювальний параметр може мати кілька точкових незміщених статистичних оцінок. Тому порівнюють їх дисперсії як міру розсіювання і виявляють ту, яка має меншу дисперсію.

Означення. Точкова статистична оцінка називається ефективною, коли при заданому обсязі вибірки вона має мінімальну дисперсію.

Означення. Точкова статистична оцінка називається ґрунтовною, якщо у разі необмеженого збільшення обсягу вибірки θ^* наближається до оцінювального параметра θ, а саме:

$$\lim_{n\to\infty} P\left(\left|\theta^* - \theta\right| < \delta\right) = 1. \tag{2.4}$$

2.2 Методи визначення точкових статистичних оцінок

Існують три методи визначення точкових статистичних оцінок для параметрів генеральної сукупності.

Метод аналогій. Цей метод базується на тому, що для параметрів генеральної сукупності вибирають такі самі параметри вибірки, тобто для оцінки $\overline{X}_{\Gamma} = M(X)$, D_{Γ} вибирають аналогічні статистики — $\overline{x}_{\text{В}}$, $D_{\text{В}}$.

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

Отож, використовуючи метод найменших квадратів, можна, наприклад, визначити статистичну оцінку для $\overline{X}_{\Gamma} = M(X)$. Для цього скористаємося функцією $u = \sum_{i=0}^{n} \left(x_i - \theta^*\right)^2 n_i$. Використовуючи необхідну умову існування екстремуму, дістанемо

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

звідки

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

$$\theta^* = \frac{\sum_{i=1}^{n} x_i n_i}{n} = \overline{x}_{\text{В}}.$$

Звідси для $\theta = \overline{X}_{\Gamma}$ точковою статистичною оцінкою буде $\theta^* = \overline{x}_{\text{В}}$ — вибіркова середня.

Метод максимальної правдоподібності. Цей метод посідає центральне місце в теорії статистичного оцінювання параметрів θ.

Нехай ознака генеральної сукупності X визначається лише одним параметром θ і має щільність ймовірностей $f(x;\ \theta)$. У разі реалізації вибірки з варіантами $x_1, x_2, ..., x_n$ щільність ймовірностей вибірки буде такою:

$$f\left(x_1, x_2, ..., x_n, \theta^*\right) = f\left(x_1, \theta^*\right) \cdot f\left(x_2, \theta^*\right) \cdot ... \cdot f\left(x_n, \theta^*\right). \tag{2.5}$$

При цьому варіанти розглядаються як незалежні випадкові величини, які мають один і той самий закон розподілу, що й ознака генеральної сукупності X.

Суть цього методу полягає в тому, що, фіксуючи значення варіант $x_1, x_2, ..., x_n$, визначають таке значення параметра θ^*, при якому функція (2.5) максимізується. Функція (2.5) називається функцією максимальної правдоподібності і позначається так: $L = L(\theta^*)$.

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

$$f\left(x_1, x_2, ..., x_n, \theta_1^*, \theta_2^*\right) = \frac{1}{(2\pi\theta_2)^{\frac{n}{2}}} \cdot e^{-\frac{\sum_{i=1}^{n}\left(x_i - \theta_1^*\right)^2}{2\theta_2^*}}. \tag{2.6}$$

При цьому за статистичні оцінки θ_1^*, θ_2^* вибирають ті їх значення, за яких задана вибірка буде найімовірнішою, тобто функція (2.6) досягає максимуму.

На практиці зручно від функції (2.6) перейти до її логарифма, а саме:

$$\ln f(x_1, x_2, ..., x_n, \theta_1^*, \theta_2^*) =$$

$$= L\left(x_1, x_2, ..., x_n, \theta_1^*, \theta_2^*\right) = -\frac{n}{2}\left(\ln\pi + \ln\theta_2^*\right) - \frac{\sum\left(x_i - \theta_1^*\right)^2}{2\theta_2^*}.$$

Згідно з необхідною умовою екстремуму для цієї функції дістанемо:

$$\begin{cases} \dfrac{\partial L}{\partial \theta_1^*} = -\dfrac{1}{\theta_2^*}\sum_{i=1}^{n}\left(x_i - \theta_1^*\right) = 0, \\ \dfrac{\partial L}{\partial \theta_2^*} = -\dfrac{n}{2\theta_2^*} + \dfrac{1}{2\left(\theta_2^*\right)^2} \cdot \sum_{i=1}^{n}\left(x_i - \theta_1^*\right)^2 = 0. \end{cases} \tag{2.7}$$

З першого рівняння системи (2.7) дістанемо

$$\theta_1^* = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \overline{x}_{\text{В}}; \tag{2.8}$$

з другого рівняння системи (2.7) маємо:

$$\theta_2^* = \frac{1}{n} \cdot \sum_{i=1}^{n}\left(x_i - \overline{x}_{\text{В}}\right)^2 = D_{\text{В}}. \tag{2.9}$$

Отже, для генеральної середньої $\overline{X}_{\Gamma} = M(X)$ точковою статистичною оцінкою є $\overline{x}_{\text{В}}$, для генеральної дисперсії D_{Γ} — вибіркова дисперсія $D_{\text{В}}$.

Можна показати, що точковою незміщеною статистичною оцінкою для генеральної середньої $\overline{X}_{\Gamma} = M(X)$ є вибіркова середня $\overline{x}_{\text{В}}$. Дійсно,

$$M(\bar{x}_{\text{В}}) = M\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{\sum M(x_i)}{n} = \left|\begin{matrix}\text{Ураховуючите, що} \\ M(x_i) = \bar{X}_{\Gamma} = a\end{matrix}\right| = \frac{\sum_{i=1}^{n} a}{n} = \frac{na}{n} = a.$$

Отже, $M(\bar{x}_{\text{В}}) = \bar{X}_{\Gamma}$.

Вибіркова дисперсія $D_{\text{В}}$ є точковою зміщеною статистичною оцінкою для генеральної дисперсії D_{Γ}, де $\frac{n-1}{n}$ — коефіцієнт зміщення, який зменшується зі збільшенням обсягу вибірки *n*.

Якщо розглянути величину $\frac{n}{n-1} D_{\text{В}}$, то її математичне сподівання

$$M\left(\frac{n}{n-1} D_{\text{В}}\right) = \frac{n}{n-1} M(D_{\text{В}}) = \frac{n}{n-1} \cdot \frac{n-1}{n} D_{\Gamma} = D_{\Gamma}.$$

Величина $\frac{n}{n-1} D_{\text{В}}$ є точковою незміщеною статистичною оцінкою для генеральної дисперсії D_{Γ}, називається виправленою дисперсією і позначається через S^2:

$$S^2 = \frac{n}{n-1} D_{\text{В}}$$

або

$$S^2 = \frac{n}{n-1} \cdot \frac{\sum_{i=1}^{n} (x_i - \bar{x}_{\text{В}})^2}{n} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_{\text{В}})^2}{n-1}. \qquad (2.10)$$

Величину

$$S = \sqrt{\frac{n}{n-1} D_{\text{В}}} \qquad (2.11)$$

називають виправленим середнім квадратичним відхиленням.

Виправлене середнє квадратичне відхилення є зміщеною точковою статистичною оцінкою для σ_{Γ}, оскільки

$$M(S) = \sqrt{\frac{2}{k}} \cdot \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \sigma_{\Gamma}, \qquad (2.12)$$

де $k = n - 1$ є кількістю ступенів свободи;

$$\sqrt{\frac{2}{k}} \cdot \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)}$$ — коефіцієнт зміщення.

Приклад. 200 однотипних деталей були піддані шліфуванню. Результати вимірювання наведені як дискретний статистичний розподіл, поданий у табличній формі:

x_i, мм	3,7	3,8	3,9	4,0	4,1	4,2	4,3	4,4
n_i	1	22	40	79	27	26	4	1

Знайти точкові незміщені статистичні оцінки для $\overline{X}_{\text{Г}} = M(x),\ D_{\text{Г}}$.

Розв'язання. Оскільки точковою незміщеною оцінкою для $\overline{X}_{\text{Г}}$ є $\overline{x}_{\text{В}}$, то обчислимо

$$\overline{x}_{\text{В}} = \frac{\sum x_i n_i}{n} =$$

$$= \frac{3{,}7\cdot 1 + 3{,}8\cdot 22 + 3{,}9\cdot 40 + 4{,}0\cdot 79 + 4{,}1\cdot 27 + 4{,}2\cdot 26 + 4{,}3\cdot 4 + 4{,}4\cdot 1}{200} =$$

$$= \frac{3{,}7 + 83{,}6 + 156 + 316 + 110{,}7 + 109{,}2 + 17{,}2 + 4{,}4}{200} = \frac{808{,}8}{200} = 4{,}004\ \text{мм}.$$

Для визначення точкової незміщеної статистичної оцінки для $D_{\text{Г}}$ обчислимо $D_{\text{В}}$:

$$\frac{\sum x_i^2 n_i}{n} = \frac{(3{,}7)^2\cdot 1 + (3{,}8)^2\cdot 22 + (3{,}9)^2\cdot 40 + (4{,}0)^2\cdot 79 +}{200} =$$

$$\frac{+(4{,}1)^2\cdot 27 + (4{,}2)^2\cdot 26 + (4{,}3)^2\cdot 4 + (4{,}4)^2\cdot 1}{200} =$$

$$= \frac{13{,}69 + 317{,}68 + 608{,}4 + 1264 + 453{,}87 + 458{,}64 + 73{,}96 + 19{,}36}{200} =$$

$$= \frac{3209{,}6}{200} = 16{,}048$$

$$D_{\text{В}} = \frac{\sum x_i^2 n_i}{n} - (\overline{x}_{\text{В}})^2 = 16{,}048 - (4{,}004)^2 =$$

$$= 16{,}048 - 16{,}032016 = 0{,}015984$$

Тоді точкова незміщена статистична оцінка для $D_{\text{Г}}$ становитиме:

$$S^2 = \frac{n}{n-1} D_{\text{В}} = \frac{200}{200-1}\cdot 0{,}015984 = \frac{200}{199}\cdot 0{,}015984 = 0{,}01606\ \text{мм}^2.$$

Приклад. Граничне навантаження на сталевий болт x_i, що вимірювалось в лабораторних умовах, задано як інтервальний статистичний розподіл:

x_i, км/мм^2	4,5—5,5	5,5—6,5	6,5—7,5	7,5—8,5	8,5—9,5	9,5—10,5	10,5—11,5	11,5—12,5	12,5—13,5	13,5—14,5
n_i	40	32	28	24	20	18	16	12	8	4

Визначити точкові незміщені статистичні оцінки для $\overline{X}_{\text{Г}} = M(x),\ D_{\text{Г}}$.

Розв'язання. Для визначення точкових незміщених статистичних оцінок $\overline{x}_{\text{В}}$, S^2 перейдемо від інтервального статистичного розподілу до дискретного, який набирає такого вигляду:

$x_i^* = x_{i-1} + \frac{h}{2}$	5	6	7	8	9	10	11	12	13	14
n_i	40	32	28	24	20	18	16	12	8	4

Обчислимо $\overline{x}_{\text{В}}$: $n = \sum n_i = 202$,

$$\bar{x}_{\text{В}} = \frac{\sum x_i^* n_i}{n} = \frac{5\cdot 40 + 6\cdot 32 + 7\cdot 28 + 8\cdot 24 + 9\cdot 20 + 10\cdot 18 +}{202}$$

$$\frac{+11\cdot 16 + 12\cdot 12 + 13\cdot 8 + 14\cdot 4}{202} = \frac{1620}{202} = 8{,}02 \text{ кг/мм}^2.$$

Отже, точкова незміщена статистична оцінка для $\bar{X}_{\text{Г}} = M(x)$, $\bar{x}_{\text{В}} = 8{,}02$ кг/мм^2.
Для визначення S^2 обчислимо $D_{\text{В}}$:

$$\frac{\sum (x_i^*)^2 n_i}{n} = \frac{(5)^2\cdot 40 + (6)^2\cdot 32 + (7)^2\cdot 28 + (8)^2\cdot 24 + (9)^2\cdot 20 +}{202}$$

$$\frac{+(10)^2\cdot 18 + (11)^2\cdot 16 + (12)^2\cdot 12 + (13)^2\cdot 8 + (14)^2\cdot 4}{202} = \frac{14280}{202} \approx 70{,}69.$$

$$D_{\text{В}} = \frac{\sum (x_i^*)^2 n_i}{n} - (\bar{x}_{\text{В}})^2 = 70{,}69 - (8{,}02)^2 = 70{,}69 - 64{,}32 \approx 6{,}37 \text{ кг/мм}^2.$$

$$S^2 = \frac{n}{n-1} D_{\text{В}} = \frac{202}{202-1}\cdot 6{,}37 = \frac{202}{201}\cdot 6{,}37 \approx 6{,}4.$$

Звідси точкова незміщена статистична оцінка для $D_{\text{Г}}$ є $S^2 = 6{,}4$ кг/мм^2.

2.3 Закони розподілу ймовірностей для $\bar{x}_{\text{В}}$, S^2, S

Числові характеристики вибірки є випадковими величинами, що мають певні закони розподілу ймовірностей. Так, $\bar{x}_{\text{В}}$ (вибіркова середня) на підставі центральної граничної теореми теорії ймовірностей (теореми Ляпунова) матиме нормальний закон розподілу з числовими характеристиками:

$$M(\bar{x}_{\text{В}}) = M\left(\frac{\sum x_i n_i}{n}\right) = \frac{1}{n}\sum M(x_i)\cdot n_i = \frac{1}{n}\sum a\cdot n_i = a\frac{\sum n_i}{n} = a;$$

$$\left(a = M(x) = \bar{X}_{\text{Г}}\right);$$

$$D(\bar{x}_{\text{В}}) = D\left(\frac{\sum x_i n_i}{n}\right) = \frac{D_{\text{Г}}}{n};$$

$$\sigma(\bar{x}_{\text{В}}) = \frac{\sigma_{\text{Г}}}{\sqrt{n}}.$$

Отже, випадкова величина $\bar{x}_{\text{В}}$ має закон розподілу $N\left(a; \frac{\sigma_{\text{Г}}}{\sqrt{n}}\right)$.

Випадкова величина

$$\frac{n-1}{\sigma^2}\cdot S^2 = \sum_{i=1}^{n-1}\left(\frac{y_i}{\sigma}\right)^2 \qquad (2.13)$$

матиме розподіл χ^2 із $k = n - 1$ ступенями свободи.

Звідси випливає, що випадкова величина $\frac{\sqrt{n-1}}{\sigma}S$ матиме розподіл χ із $k = n - 1$ ступенями свободи.
Таким чином, можна стверджувати,що:

випадкова величина $\bar{x}_B \sim N(a;\, b)$, тут символ ~ потрібно читати «розподілена як»;

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

випадкова величина $S \sim \frac{\chi(n-1)}{\sqrt{n-1}}\sigma$.

2.4 Інтервальні статистичні оцінки для параметрів генеральної сукупності

Точкові статистичні оцінки θ^* є випадковими величинами, а тому наближена заміна θ на θ^* часто призводить до істотних похибок, особливо коли обсяг вибірки малий. У цьому разі застосовують інтервальні статистичні оцінки.

Статистична оцінка, що визначається двома числами, кінцями інтервалів, називається інтервальною.

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

$$\left|\theta^* - \theta\right| < \delta, \tag{2.14}$$

де δ є точністю оцінки.

Оскільки θ^* є випадковою величиною, то і δ буде випадковою, тому нерівність (2.14) справджуватиметься з певною ймовірністю.

Імовірність, з якою береться нерівність (2.14), тобто

$$P\left(\left|\theta^* - \theta\right| < \delta\right) = \gamma, \tag{2.15}$$

називають надійністю.

Рівність (2.15) можна записати так:

$$P\left(\theta^* - \delta < \theta < \theta^2 + \delta\right) = \gamma. \tag{2.16}$$

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

2.5 Побудова довірчого інтервалу для $\bar{X}_Г$ при відомому значенні $\sigma_Г$ із заданою надійністю γ

Нехай ознака *X* генеральної сукупності має нормальний закон розподілу. Побудуємо довірчий інтервал для $\bar{X}_Г$, знаючи числове значення середнього квадратичного відхилення генеральної сукупності $\sigma_Г$, із заданою надійністю γ. Оскільки $\bar{x}_В$ як точкова незміщена статистична оцінка для $\bar{X}_Г = M(x)$ має нормальний закон розподілу з числовими характеристиками $M(\bar{x}_В) = \bar{X}_Г = a$, $\sigma(\bar{x}_В) = \frac{\sigma_Г}{\sqrt{n}}$, то, скориставшись (2.16), дістанемо

$$P\left(\left|\bar{x}_{\text{В}} - a\right| < \delta\right) = \gamma\,. \qquad (2.17)$$

Випадкова величина $\bar{x}_{\text{В}} - a$ має нормальний закон розподілу з числовими характеристиками

$$M\left(\bar{x}_{\text{В}} - a\right) = M\left(\bar{x}_{\text{В}}\right) - a = a - a = 0;$$

$$D\left(\bar{x}_{\text{В}} - a\right) = D\left(\bar{x}_{\text{В}}\right) = \frac{D_{\text{Г}}}{n};$$

$$\sigma\left(\bar{x}_{\text{В}}\right) = \frac{\sigma_{\text{Г}}}{\sqrt{n}}.$$

Тому $\dfrac{\bar{x}_{\text{В}} - a}{\dfrac{\sigma_{\text{Г}}}{\sqrt{n}}}$ матиме нормований нормальний закон розподілу $N(0;\ 1)$.

Звідси рівність (2.17) можна записати, назначивши $\dfrac{\delta}{\dfrac{\sigma_{\text{Г}}}{\sqrt{n}}} = x,$ так:

$$P\left(\left|\frac{\bar{x}_{\text{В}} - a}{\dfrac{\sigma_{\text{Г}}}{\sqrt{n}}}\right| < x\right) = \gamma \qquad (2.18)$$

або

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

Згідно з формулою нормованого нормального закону

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

для (2.18) вона набирає такого вигляду:

$$P\left(\left|\frac{\bar{x}_{\text{В}} - a}{\dfrac{\sigma_{\text{Г}}}{\sqrt{n}}}\right| < x\right) = 2\Phi(x) = \gamma. \qquad (2.19)$$

З рівності (2.19) знаходимо аргументи x, а саме:

$$2\Phi\left(x\right) = \gamma \rightarrow \Phi\left(x\right) = 0{,}5\gamma.$$

Аргумент x знаходимо за значенням функції Лапласа, яка дорівнює 0,5 γ за таблицею (додаток Б).

Отже, довірчий інтервал для $\bar{x}_{\text{Г}}$ при відомому значенні $\sigma_{\text{Г}}$ має вигляд

$$\bar{x}_{\text{В}} - \frac{x \cdot \sigma_{\text{Г}}}{\sqrt{n}} < a < \bar{x}_{\text{В}} + \frac{x \cdot \sigma_{\text{Г}}}{\sqrt{n}}\,, \qquad (2.20)$$

що можна зобразити умовно на рис. 2.1.

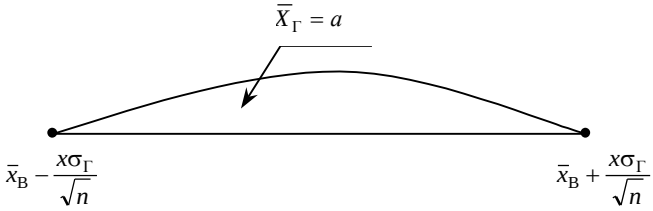


Рис. 2.1

Величина $\frac{x \cdot \sigma_Г}{\sqrt{n}}$ називається точністю оцінки, або похибкою вибірки.

Приклад. Вимірявши 40 випадково відібраних після виготовлення деталей, знайшли вибіркову середню, що дорівнює 15 см. Із надійністю $\gamma = 0{,}99$ побудувати довірчий інтервал для середньої величини всієї партії деталей, якщо генеральна дисперсія дорівнює $0{,}09\,\text{см}^2$.

Розв'язання. Для побудови довірчого інтервалу необхідно знати:

$$\bar{x}_В,\ \ \sigma_Г,\ n,\ x.$$

З умови задачі маємо: $\bar{x}_В = 15\,\text{см}$, $\sigma_Г = \sqrt{D_Г} = \sqrt{0{,}09\,\text{см}^2} = 0{,}3\,\text{см}$, $n = 40$, тобто $\sqrt{n} = \sqrt{40} = 6{,}32$. Величина x обчислюється з рівняння

$$\Phi(x) = 0{,}5\gamma = 0{,}5 \cdot 0{,}99 = 0{,}495.$$

$$\Phi(x) = 0{,}495, \quad x = 2{,}58\ [\text{за таблицею значень функції Лапласа}].$$

Знайдемо числові значення кінців довірчого інтервалу:

$$\bar{x}_В - \frac{\sigma_Г \cdot x}{\sqrt{n}} = 15 - \frac{0{,}3 \cdot 2{,}58}{6{,}32} = 15 - 0{,}12 = 14{,}88\ \text{см}.$$

$$\bar{x}_В + \frac{\sigma_Г \cdot x}{\sqrt{n}} = 15 + \frac{0{,}3 \cdot 2{,}58}{6{,}32} = 15 + 0{,}12 = 15{,}12\ \text{см}.$$

Таким чином, маємо:

$$14{,}88 < \bar{X}_Г < 15{,}12.$$

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

Приклад. 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 $\bar{X}_Г$, якщо $\sigma_Г$ = 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 $\bar{x}_{\text{В}}$ 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

$$\bar{x}_{\text{В}} = \frac{\sum x_i^* n_i}{n} = \frac{3\cdot 9 + 5\cdot 7 + 7\cdot 10 + 9\cdot 4}{30} = \frac{27+35+70+36}{30} =$$
$$= \frac{168}{30} = 5{,}6 \text{ млн грн.}$$

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

$$\Phi(x) = 0{,}5\gamma = 0{,}5\cdot 0{,}999 = 0{,}4995, \quad x \approx 3{,}4.$$

Обчислюємо кінці інтервалу:

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

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

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

Приклад. Якого значення має набувати надійність оцінки γ, щоб за обсягу вибірки n = 100 похибка її не перевищувала 0,01 при $\sigma_{\text{Г}} = 5$.

Розв'язання. Позначимо похибку вибірки

$$\frac{x\cdot\sigma_{\text{Г}}}{\sqrt{n}} = \varepsilon, \quad x = \frac{\varepsilon\sqrt{n}}{\sigma_{\text{Г}}} = \frac{0{,}01\sqrt{100}}{5} = \frac{0{,}01\cdot 10}{5} = 0{,}02.$$

Далі маємо

$$P\left(\left|\frac{\bar{x}_{\text{В}} - a}{\frac{\sigma_{\text{Г}}}{\sqrt{n}}}\right| < x\right) = 2\Phi(x) = 2\Phi(0{,}02) = 2\cdot 0{,}008 = 0{,}016.$$

Як бачимо, надійність мала.

Приклад. Визначити обсяг вибірки n, за якого похибка $\varepsilon = 0{,}01$ гарантується з імовірністю 0,999, якщо $\sigma_{\text{Г}} = 5$.

Розв'язання. За умовою задачі $P\left(\left|\frac{\bar{x}_{\text{В}}-a}{\frac{\sigma_{\text{Г}}}{\sqrt{n}}}\right|<x\right)=\gamma=0{,}999.$ Оскільки $\frac{x\sigma_{\text{Г}}}{\sqrt{n}}=\varepsilon,$ то дістанемо: $n=\frac{x^2\sigma_{\text{Г}}^2}{\varepsilon^2}$. Величину x знаходимо з рівності

$$\Phi(x)=0{,}5\gamma=0{,}5\cdot 0{,}999=0{,}4995\,,\quad x\approx 3{,}4.$$

Тоді $n=\frac{(3{,}4)^2\cdot 5^2}{(0{,}01)^2}=2\,890\,000.$

2.6 Побудова довірчого інтервалу для $\bar{x}_{\text{Г}}$ при невідомому значенні $\sigma_{\text{Г}}$ із заданою надійністю γ

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

$$t=\frac{\bar{x}_{\text{В}}-a}{\frac{S}{\sqrt{n}}},\tag{2.21}$$

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

Тоді (2.21) набирає такого вигляду:

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

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

Обчисливши за даним статистичним розподілом $\bar{x}_{\text{В}}$, S і визначивши за таблицею розподілу Стьюдента значення t_γ, будуємо довірчий інтервал

$$\bar{x}_{\text{В}}-\frac{t_\gamma\cdot S}{\sqrt{n}}<a<\bar{x}_{\text{В}}+\frac{t_\gamma S}{\sqrt{n}}.\tag{2.22}$$

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

Приклад. 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 «*a*» (середнього часу безвідказної роботи приладу).

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

Обчислимо $\bar{x}_{\text{В}}$:

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

Отже, дістали $\bar{x}_{\text{В}} = 222{,}5$ год.

Визначимо $D_{\text{В}}$:

$$\frac{\sum t_i^2 n_i}{n} = \frac{100^2\cdot 2 + 170^2\cdot 5 + 240^2\cdot 10 + 310^2\cdot 2 + 380^2\cdot 1}{20} = \frac{1\,077\,100}{20} = 53855.$$

$$D_{\text{В}} = \frac{\sum t_i^2 n_i}{n} - \left(\bar{x}_{\text{В}}\right)^2 = 53\,855 - \left(222{,}5\right)^2 = 53\,855 - 49\,506{,}25 = 4348{,}75.$$

Отже, $D_{\text{В}}$ = 4348,75.

Виправлене середнє квадратичне відхилення дорівнюватиме:

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

За таблицею значень $\int_0^t f(x)\,dt = \gamma = 0{,}99$ (додаток В) розподілу Стьюдента за заданою надійністю $\gamma = 0{,}99$ і числом ступенів свободи $k = n - 1$ = 20 – 1 = 19 знаходимо значення $t(\gamma = 0{,}99, k = 19) =$ 2,861.

Обчислимо кінці довірчого інтервалу:

$$\bar{x}_{\text{В}} - \frac{t_\gamma S}{\sqrt{n}} = 222{,}5 - \frac{2{,}861\cdot 67{,}66}{\sqrt{20}} = 222{,}5 - \frac{2{,}861\cdot 67{,}66}{4{,}472} = 179{,}2 \text{ год.}$$

$$\bar{x}_{\text{В}} + \frac{t_\gamma S}{\sqrt{n}} = 222{,}5 + \frac{2{,}861\cdot 67{,}89}{\sqrt{20}} = 222{,}5 + \frac{2{,}861\cdot 67{,}66}{4{,}472} = 265{,}8 \text{ год.}$$

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

$$179{,}2 < a < 265{,}8.$$

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

Приклад. У таблиці наведено відхилення діаметрів валиків, оброблених на верстаті, від номінального розміру:

h = 5 мк	0 — 5	5 — 10	10 — 15	15—20	20 — 25

n_i	15	75	100	50	10

Із надійністю $\gamma = 0{,}99$ побудувати довірчий інтервал для $\overline{X}_{\text{Г}} = a$.

Розв'язання. Для побудови довірчого інтервалу необхідно знайти $\overline{x}_{\text{В}}$, S.

Для цього від інтервального статистичного розподілу, наведеного в умові задачі, необхідно перейти до дискретного, а саме:

x_i^*	2,5	7,5	12,5	17,5	22,5
n_i	15	75	100	50	10

Обчислимо $\overline{x}_{\text{В}}$:

$$\overline{x}_{\text{В}} = \frac{\sum x_i^* n_i}{n} = |\text{Оскільки } n = \sum n_i = 250| =$$

$$= \frac{2{,}5\cdot 15 + 7{,}5\cdot 75 + 12{,}5\cdot 100 + 17{,}5\cdot 50 + 22{,}5\cdot 10}{250} =$$

$$= \frac{37{,}5 + 562{,}5 + 1250 + 875 + 225}{250} = \frac{2950}{250} = 11{,}8.$$

Отже, $\overline{x}_{\text{В}} = 11{,}8$ мк.

Визначимо $D_{\text{В}}$:

$$\frac{\sum (x_i^*)^2 n_i}{n} = \frac{(2{,}5)^2\cdot 15 + (7{,}5)^2\cdot 75 + (12{,}5)^2\cdot 100 + (17{,}5)^2\cdot 50 + (22{,}5)^2\cdot 10}{250} =$$

$$= \frac{93{,}75 + 4218{,}75 + 15625 + 15312{,}5 + 5062{,}5}{250} = \frac{40312{,}5}{250} = 161{,}25.$$

$$D_{\text{В}} = \frac{\sum (x_i^*)^2 n_i}{n} - (\overline{x}_{\text{В}})^2 = 161{,}25 - (11{,}8)^2 = 161{,}25 - 139{,}24 = 22{,}01.$$

Обчислимо виправлене середнє квадратичне відхилення S:

$$S = \sqrt{\frac{n}{n-1} D_{\text{В}}} = \sqrt{\frac{250}{250-1}\cdot 22{,}01} \approx 4{,}7 \text{ мк.}$$

З огляду на великий (n = 250) обсяг вибірки можна вважати, що розподіл Стьюдента близький до нормального закону. Тоді за таблицею значення функції Лапласа

$$\Phi(t_\gamma) = 0{,}495 \to t_\gamma = 2{,}58.$$

Обчислимо кінці інтервалів:

$$\overline{x}_{\text{В}} - \frac{t_\gamma S}{\sqrt{n}} = 11{,}8 - \frac{2{,}58\cdot 4{,}7}{\sqrt{250}} = 11{,}8 - \frac{2{,}58\cdot 4{,}7}{15{,}8} = 11{,}8 - 0{,}77 = 11{,}03 \text{ мк.}$$

$$\overline{x}_{\text{В}} + \frac{t_\gamma S}{\sqrt{n}} = 11{,}8 + \frac{2{,}58\cdot 4{,}7}{\sqrt{250}} = 11{,}8 + \frac{2{,}58\cdot 4{,}7}{15{,}8} = 11{,}8 + 0{,}77 = 12{,}57 \text{ мк.}$$

Отже, довірчий інтервал для середнього значення відхилень буде таким:

$$11{,}03 < a < 12{,}57.$$

Звідси з надійністю $\gamma = 0{,}99$ (99%) можна стверджувати, що

$a \in$ [11,03 мк; 12,57 мк].

2.7 Побудова довірчих інтервалів із заданою надійністю γ для $D_г$, $\sigma_г$

У разі, коли ознака X має нормальний закон розподілу, для побудови довірчого інтервалу із заданою надійністю γ для $D_г$, $\sigma_г$ застосовуємо випадкову величину

$$\chi^2 = \frac{n-1}{\sigma_Г^2} S^2, \tag{2.23}$$

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

Оскільки випадкові події

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

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

$$P\left(\chi_1^2 < \chi^2 < \chi_2^2\right) = P\left(\frac{1}{\chi_2^2} < \frac{1}{\chi^2} < \frac{1}{\chi_1^2}\right). \tag{2.24}$$

Підставляючи в (2.24) $\chi^2 = \frac{n-1}{\sigma^2} S^2$, дістанемо

$$P\left(\frac{1}{\chi_2^2} < \frac{1}{\chi^2} < \frac{1}{\chi_1^2}\right) = P\left(\frac{1}{\chi_2^2} < \frac{1}{\frac{n-1}{\sigma_Г^2} S^2} < \frac{1}{\chi_1^2}\right) =$$

$$= P\left(\frac{1}{\chi_2^2} < \frac{\sigma_Г^2}{(n-1)S^2} < \frac{1}{\chi_1^2}\right) = P\left(\frac{(n-1)S^2}{\chi_2^2} < \sigma_Г^2 < \frac{(n-1)S^2}{\chi_1^2}\right) = \gamma.$$

Отже, довірчий інтервал для $\sigma_Г^2 = D_Г$ матиме вигляд

$$\frac{(n-1)S^2}{\chi_2^2} < D_Г < \frac{(n-1)S^2}{\chi_1^2}. \tag{2.25}$$

Тоді довірчий інтервал для $\sigma_Г$ випливає із (425) і буде таким:

$$\frac{S\sqrt{n-1}}{\chi_2} < \sigma_Г < \frac{S\sqrt{n-1}}{\chi_1}. \tag{2.26}$$

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

$$P\left(\chi^2 > \chi_1^2\right) = 1 - \frac{\alpha}{2}; \tag{2.27}$$

$$P\left(\chi^2 > \chi_2^2\right) = \frac{\alpha}{2}, \tag{2.28}$$

де $\alpha = 1 - \gamma$.

Питання для самоперевірки

1 Що називається точковою статистичною оцінкою?

2 Що таке незміщена точкова статистична оцінка?

3 Що таке зміщена точкова статистична оцінка?

4 Що називають ефективною точковою статистичною оцінкою?

5 Що називають ґрунтовною точковою статистичною оцінкою?

6 У чому сутність методу аналогій?

7 У чому сутність методу найменших квадратів?

8 У чому сутність методу максимальної правдоподібності?

9 Що є точковою незміщеною статистичною оцінкою для $\overline{X}_{\Gamma}$?

10 Що означає точкова незміщена статистична оцінка для D_{Γ}?

11 Що називається виправленою дисперсією, виправленим середнім квадратичним відхиленням?

12 Чому дорівнює $M\left(\overline{x}_{\text{В}}\right)$?

13 Чому дорівнює $M\left(D_{\text{В}}\right)$?

14 Чому дорівнює $M\left(S_{\text{В}}^2\right)$?

15 Який закон розподілу ймовірностей має випадкова величина $\overline{x}_{\text{В}}$?

16 Який закон розподілу має випадкова величина $\frac{n-1}{\sigma_{\Gamma}^2}S^2$?

17 Який закон розподілу ймовірностей має випадкова величина $\frac{\sqrt{n-1}}{\sigma_{\Gamma}}S$?

18 Визначення інтервальної статистичної оцінки для параметрів генеральної сукупності.

19 Що називають точністю і надійністю оцінки?

20 Що називають довірчим інтервалом?

21 Як побудувати довірчий інтервал із заданою надійністю γ при відомому значенні σ_{Γ}?

22 Як побудувати довірчий інтервал для $\overline{X}_{\Gamma}$ із заданою надійністю γ при невідомому значенні σ_{Γ}?

Задачі для самостійного розв’язання

1 Серед студентів 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 γ = 0,99 побудувати довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 2$.

Відповідь: $173{,}81 < \overline{X}_{\Gamma} < 176{,}19$.

2 У 25-ти осіб було виміряно кров’яний тиск x_i (в умовних одиницях). Результати вимірювання наведено у вигляді дискретного статистичного розподілу:

x_i	1,5	1,8	2,3	2,5	2,9	3,3
n_i	2	3	5	8	4	3

Із надійністю γ = 0,999 побудувати довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 1$.

Відповідь: $1{,}78 < \overline{X}_{\Gamma} < 3{,}14$.

3 Виміряна максимальна місткість конденсаторів x_i (у пікофарадах). Результати вимірювання наведено у вигляді інтервального статистичного розподілу:

x_i пф	4,0—4,2	4,2—4,4	4,4—4,6	4,6—4,8	4,8—5,0
n_i	2	4	5	9	5

З надійністю $\gamma = 0{,}999$ побудувати довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 0{,}5$.

Відповідь: $4{,}24 < \overline{X}_{\Gamma} < 4{,}92$.

4 У 30 телевізорів була виміряна чутливість x_i. Результати вимірювання подано як дискретний статистичний розподіл:

x_i, мкВ	200	250	300	350	400	450	500
n_i	2	7	6	8	4	2	1

Із надійністю γ = 0,99 побудувати довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 4$.

Відповідь: $376 < \overline{X}_{\Gamma} < 380{,}27$.

5 У 25-ти випадково відібраних деталей була виміряна відстань у мікронах від центру маси, що міститься на її осі, до зовнішньої поверхні. Результати вимірювання наведено у вигляді інтервального статистичного розподілу:

Межі інтервалів x_i, мк	80—96	96—112	112—128	128—144	144—160
n_i	2	5	8	6	4

З надійністю γ = 0,999 побудувати довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 3$.

Відповідь: $130{,}58 < \overline{X}_{\Gamma} < 134{,}42$.

6 З партії однотипних запобіжників відібрано 24 шт. Вимірювання відхилення від номіналу в кілоомах x_i наведено як дискретний статистичний розподіл:

x_i, кОм	– 1	– 2	1	2	3	4	5
n_i	3	4	4	5	4	3	1

Із надійністю γ = 0,99 побудувати довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 3$.

Відповідь: $-0{,}3 < \overline{X}_{\Gamma} < 2{,}96$.

7 З партії однотипних плашок навмання було вибрано 28 шт., і в кожній із них була виміряна глибина пазу (канавки) x_i. Результати вимірювання наведено як інтервальний статистичний розподіл:

x_i, мм	2,4—2,6	2,6—2,8	2,8—3,0	3,0—3,2	3,2—3,4
n_i	5	8	9	5	1

З надійністю γ = 0,999 побудувати довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 0{,}8$.

Відповідь: $2{,}31 < \overline{X}_{\Gamma} < 3{,}33$.

8 28-ім однотипних приладів були випробувані щодо часу їх безвідмовної роботи x_i. Результати вимірювання наведено як дискретний статистичний розподіл:

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

З надійністю γ = 0,99 побудувати довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 4$.

Відповідь: $112{,}63 < \overline{X}_{\Gamma} < 116{,}65$.

9 У 30-ти випадково вибраних валиках виміряні відхилення їх діаметрів від номіналу x_i. Результати вимірювань наведено як інтервальний статистичний розподіл:

x_i, мм	5—10	10—15	15—20	20—25	25—30
n_i	2	6	10	8	4

З надійністю γ = 0,999 побудувати довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 0{,}8$ мм.

Відповідь: $18{,}004 < \overline{X}_{\Gamma} < 18{,}996$.

10 У будинку відпочинку випадковим способом було відібрано 25 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	5	7	8	5

Із надійністю γ = 0,99 побудувати довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 2$.

3 СТАТИСТИЧНА ПЕРЕВІРКА ГІПОТЕЗ

3.1 Основні поняття

Більшість економіко-математичних моделей потребує багатократного покращення та уточнення. Для цього необхідно проведення відповідних розрахунків, пов'язаних перевіркою достовірності тих чи інших передумов, аналізом якості знайдених оцінок, достовірності одержаних висновків. Зазвичай ці розрахунки поводяться за схемою статистичної перевірки гіпотез.

У багатьох випадках необхідно знайти закон розподілу генеральної сукупності. Якщо закон розподілу невідомий, але є підстави припустити, що він має визначений вид (назвемо його A),висувають гіпотезу: генеральна сукупність (випадкова величина X)розподілена за законом A. Наприклад, можна припустити, що прибуток населення, денна кількість покупців у магазині, розмір деталей, що виробляються, мають нормальний закон розподілу.

Можливий випадок, коли закон розподілу відомий, а його параметри невідомі. Якщо є підстави припустити, що невідомий параметр Θ дорівнює числу Θ_0, яке очікують, то висувають гіпотезу: $\Theta = \Theta_0$. Наприклад, можна висунути припущення про величину середнього прибутку населення, середнього очікуваного прибутку за акціями, про розсіяння в прибутках і т. ін.

Статистичною називають гіпотезу про закон розподілу або про параметри відомого розподілу. У першому випадку гіпотеза називається непараметричною, а у другому – параметричною.

Гіпотеза H_0, яка підлягає перевірці, називається нульовою (основною). Поряд з нульовою розглядають гіпотезу H_1, яка буде сприйматися у разі відхилення H_0. Така гіпотеза називається альтернативною (конкуруючою). Наприклад, якщо перевіряється гіпотеза $H_0 : \Theta = \Theta_0$, то у якості альтернативної можуть розглядатися такі гіпотези:

$$H_1^{(1)} : \Theta \neq \Theta_0;\ H_1^{(2)} : \Theta > \Theta_0;\ H_1^{(3)} : \Theta < \Theta_0;\ H_1^{(4)} : \Theta = \Theta_1 .$$

Вибір альтернативної гіпотези визначається конкретною формульовкою задачі, а нульова гіпотеза часто спеціально підбирається так, щоб відхилити її і прийняти тим самим альтернативну гіпотезу. Наприклад, щоб прийняти

гіпотезу про існування кореляції між двома економічними показниками (скажімо, між інфляцією і безробіттям), можна відхилити гіпотезу про відсутність такої кореляції, взявши її у якості нульової гіпотези.

Сутність перевірки статистичної гіпотези полягає у тому, щоб з'ясувати узгоджуються чи ні дані спостережень і висунутої гіпотези. Чи можна розбіжність між гіпотезою і результатом вибіркових спостережень розглядати як випадкову похибку, обумовлену механізмом випадкового відбору? Ця задача розв'язується за допомогою спеціальних методів математичної статистики – методів статистично перевірки гіпотез.

При перевірці гіпотези вибіркові дані можуть протирічати гіпотезі H_0. Тоді вона відхиляється. Якщо ж статистичні дані узгоджуються з висунутою гіпотезою, то вона сприймається. Статистична перевірка гіпотез на основі вибіркових даних необхідно пов'язана з ризиком прийняття хибного рішення. При цьому можливі помилки двох типів.

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

Помилка другого роду полягає у тому, що буде прийнята нульова гіпотеза, у той час, як в дійсності справедлива альтернативна гіпотеза.

Наслідки вказаних помилок нерівнозначні. Перша приводить до більш обережного консервативного рішення, а друга – до неоправданого ризику. Що краще чи гірше – залежить від конкретної задачі і змісту нульової гіпотези. Наприклад, якщо H_0 полягає в погодженні того, що продукція підприємства якісна і допущена помилка першого роду, то буде забракована придатна продукція. Допустивши помилку другого роду, ми відправимо споживачу брак. Очевидно, наслідки другої помилки більш серйозні з точки зору іміджу фірми і її довгострокових перспектив.

Позбавитися помилок першого і другого роду неможливо, оскільки вибірка обмежена. Тому прагнуть мінімізувати наслідки від цих помилок. Зауважимо, що одночасне зменшення ймовірностей даних помилок неможливе, оскільки задачі їх зменшення є конкуруючими, і зменшення ймовірності

допустити одну з них призведе до збільшення ймовірності допустити другу. У більшості випадків єдиний спосіб зменшити ймовірності помилок полягає у збільшенні об'єму вибірки.

Ймовірність допустити помилку першого роду зазвичай позначають буквою α, і її називають рівнем значущості. Ймовірність допустити помилку другого роду позначають через β. Тоді ймовірність не допустити помилку другого роду (1 – β) називають потужністю критерія.

Зазвичай значення α задають "круглими" числами (наприклад, 0,1; 0, 05; 0,01 і т. ін.), а потім прагнуть побудувати критерій найбільшої потужності. Отже, якщо α = 0,05, то це означає, що дослідник не хоче допустити помилку першого роду більш ніж у 5 випадках з 100.

3.2 Критерії перевірки. Критична область

Перевірку статистичної гіпотези виконують на основі вибірки. Для цього використовують спеціально вибрану випадкову величину (статистику, критерій), точне або наближене значення якої відомо. Цю величину позначають так:

U (або Z) – якщо вона має стандартизований нормальний розподіл;

T – якщо вона розподілена за законом Ст'юдента;

χ^2 – якщо вона розподілена за законом χ^2;

F – якщо вона має розподіл Фішера.

З ціллю загальності, далі будемо позначати таку випадкову величину через K.

Отже, статистичним критерієм називають випадкову величину K, яка слугує для перевірки нульової гіпотези. Після вибору критерію множину всіх його можливих значень розбивають на дві множини, які не перетинаються: одна з них містить значення критерію, при яких нульова гіпотеза відхиляється, а друга – при яких вона не відхиляється. Сукупність значень критерію, при яких нульову гіпотезу відхиляють, називають критичною областю. Сукупність

значень критерію, при яких нульову гіпотезу сприймають, називають областю прийняття гіпотези.

Основний принцип перевірки статистичних гіпотез можна сформулювати так: якщо спостережуване значення критерію *K* (обчислене за вибіркою) належить критичній області, то нульову гіпотезу відхиляють. Якщо ж спостережуване значення критерію *K* належить області прийняття гіпотези, то нульову гіпотезу сприймають.

Точки, які розділяють критичну область і область прийняття гіпотези, називають критичними.

В основу визначення критичних точок покладено принцип неможливості малоймовірних подій.

Нехай для перевірки нульової гіпотези H_0 слугує критерій *K*. Припустимо, що густина розподілу ймовірності випадкової величини у випадку справедливості H_0 має вигляд $f(k|H_0)$, а математичне сподівання *K* дорівнює k_0. Тоді ймовірність того, що випадкова величина *K* влучить в інтервал $(k_{1-\alpha/2}, k_{\alpha/2})$ можна знайти так

$$P(k_{1-\alpha/2} < K < k_{\alpha/2}) = \int_{k_{1-\alpha/1}}^{k_{\alpha/2}} f(k|H_0)dk \qquad (3.1)$$

Задамо цю ймовірність як 1 – α і знайдемо критичні точки *K* – розподілу $k_{1-\alpha/2}, k_{\alpha/2}$ з умов:

$$P(K \le k_{1-\alpha/2}) = \int_{-\infty}^{k_{1-\alpha/2}} f(k|H_0)dk = \frac{\alpha}{2} \qquad (3.2)$$

$$P(K \ge k_{\alpha/2}) = \int_{k_{\alpha/2}}^{+\infty} f(k|H_0)dk = \frac{\alpha}{2} \qquad (3.3)$$

Отже,

$$P(k_{1-\alpha/2} < K < k_{\alpha/2}) = 1-\alpha,$$

$$P((K \le k_{1-\alpha/2}) \cup (K \ge k_{\alpha/2})) = \alpha.$$

Задамо ймовірність α настільки малою (0,95; 0,01), щоб влучення випадкової величини *K* за межі інтервалу $(k_{1-\alpha/2}, k_{\alpha/2})$ можна б було розглядати

як малоймовірну подію. Тоді, виходячи з принципу практичної неможливості малоймовірних подій, можна вважати, що у разі справедливості H_0 при її перевірці за допомогою критерію K за даними однієї вибірки спостережуване значення K повинно влучити в інтервал $(k_{1-\alpha/2}, k_{\alpha/2})$. Якщо ж спостережуване значення влучить за межі вказаного інтервалу, то настане малоймовірна, практично неможлива подія. Це дає підставу вважати, що з ймовірністю 1– α нульова гіпотеза H_0 несправедлива.

Точки $k_{1-\alpha/2}, k_{\alpha/2}$ є критичними.

Критична область $(-\infty; k_{1-\alpha/2}) \cup (k_{\alpha/2}; +\infty)$ називається двобічною критичною областю (рис.1).

Вона визначається у випадку, коли альтернативна гіпотеза має вигляд: $H_1 : \Theta \neq \Theta_0$.

Окрім двобічної, розглядають однобічні критичні області – правобічні та лівобічні.

Лівобічною називають критичну область $(-\infty; k_{1-\alpha})$, яка визначається співвідношенням $P(K < k_{1-\alpha}) = \alpha$. Вона використовується у випадку, коли альтернативна гіпотеза має вигляд: $H_1 : \Theta < \Theta_0$ (рис.2).

Правобічною називають критичну область $(k_\alpha; +\infty)$, яка визначається співвідношенням $P(K > k_\alpha) = \alpha$. Вона використовується у випадку, коли альтернативна гіпотеза має вигляд: $H_1 : \Theta > \Theta_0$ (рис.3).

3.3 Загальна схема перевірки гіпотез

1 Формулювання нульової (H_0) і альтернативної (H_1) гіпотез.

2 Вибір відповідного рівня значущості α.

3 Вибір критерію K для перевірки H_0.

4 Визначення критичної області і області прийняття гіпотези.

5 Знаходження спостережуваного значення критерію $K_{спос.}$

6 Прийняття статистичного рішення.

Перевірка гіпотези у випадку двобічної критичної області тісно пов'язана з інтервальним оцінюванням. При одному і тому ж рівні значущості α і об'ємі вибірки n влучення гіпотетичного значення досліджуваного параметра в довірчий інтервал рівнозначно влученню відповідного критерію в область прийняття рішення. Отже, для перевірки гіпотези в цьому випадку можна скористатися довірчим інтервалом. Якщо гіпотетичне значення досліджуваного параметра влучить в цей інтервал, то приходять до висновку, що немає підстав для відхилення висунутої гіпотези.

3.4 Перевірка гіпотези про математичне сподівання нормально розподіленої випадкової величини при відомій дисперсії

Нехай генеральна сукупність X розподілена нормально, причому її математичне сподівання a невідоме, а дисперсія σ^2 відома. Також є підстава того, що $a = a_0$. Тоді

$$H_0 : a = a_0$$

$$H_1^{(1)} : a \neq a_0 \quad (H_1^{(2)} : a > a_0;\ H_1^{(3)} : a < a_0)$$

Для перевірки H_0 беруть вибірку об'єму n: $x_1, x_2, \ldots, x_n$ і у якості критерію будують статистику

$$U = \frac{\bar{x} - a_0}{\sigma / \sqrt{n}}, \tag{3.4}$$

де

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \sigma = \sqrt{\sigma^2}.$$

Доведемо, що у разі справедливості H_0, статистика U розподілена за стандартизованим нормальним законом $U \sim N(0; 1)$.

1) Нехай у якості альтернативної гіпотези розглядається гіпотеза $H_1^{(1)} : a \neq a_0$. Тоді критичні точки $u_{\alpha/2}$ і $u_{1-\alpha/2} = -u_{\alpha/2}$ знаходяться за таблицею значень функції Лапласа за умови $\Phi(u_{\alpha/2}) = \frac{1-\alpha}{2}$.

Якщо $|u_{спос}| = \left|\frac{\bar{x} - a_0}{\sigma / \sqrt{n}}\right| < u_{\alpha/2}$, то немає підстав для відхилення H_0.

Якщо $|u_{спос}| > u_{\alpha/2}$ – гіпотеза H_0 відхиляється на користь альтернативної гіпотези $H_1^{(1)}$.

2) Нехай $H_1^{(2)} : a > a_0$. Тоді критичну точку u_α правобічної критичної області знаходять за умови $\Phi(u_\alpha) = \frac{1-2\alpha}{2}$.

Якщо $U_{спос} < u_\alpha$ – немає підстав для відхилення H_0.

Якщо $U_{спос} \geq u_\alpha$ – H_0 відхиляють на користь $H_1^{(2)}$.

3) Нехай $H_1^{(3)} : a < a_0$. Тоді критична точка $u_{1-\alpha} = -u_\alpha$.

Якщо $U_{спос} > u_{1-\alpha}$ – немає підстав для відхилення H_0.

Якщо $U_{спос} \leq u_{1-\alpha}$ – H_0 відхиляють на користь $H_1^{(3)}$.

Приклад 1. На основі довготривалих спостережень за вагою X пакетів горішків, які заповнюються автоматично, з'ясовано, що стандартне відхилення ваги пакетів $\sigma = 10$ г. Зважено 25 пакетів, при цьому їх середня вага склала $\bar{x} = 244$ г. Перевірити гіпотезу $M(X) = 250$ г при рівні значущості $\alpha = 0{,}05$, якщо відомо, що X розподілена за нормальним законом.

Розв'язання.

Розв'язання.

$$H_o : a=250$$

$$H_1 : a \neq 250$$

За формулою (4)

$$U_{спос.} = \frac{244-250}{10/\sqrt{25}} = -3.$$

У данному випадку будуємо двобічну критичну область. За таблицею функції Лапласа знайдемо критичну точну $U_{\alpha/2} = U_{0,025} = 1{,}96$. Оскільки $U_{спос} = 3 > 1{,}96 = U_{kp}$, то H_o відхиляється на користь H_1. Це свідчить про те, що станок потребує під наладки.

3.5 Перевірка гіпотези про математичне сподівання нормально розподіленої випадкової величини при невідомій дисперсії.

Нехай генеральна сукупність X має нормальний розподіл, причому її математичне сподівання a і дисперсія σ^2 невідомі. Дана ситуація більш реалістична у порівнянні з попередньою. Нехай є підстави стверджувати, що $a = a_0$. Тоді

$$H_o : a = a_0$$

$$H^{(1)}_1 : a \neq a_0 \ (H^{(2)}_1 : a > a_0;\ H^{(3)}_1 : a < a_0\ ).$$

Для перевірки беруть вибірку об'єму n: $x_1, x_2, ... x_n$; обчислюють вибіркове середнє $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ і виправлену вибіркову дисперсію $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$, якій відповідає стандартне відхилення $S = \sqrt{S^2}$. Далі будуємо t- статичну

$$T = \frac{\bar{x}}{S/\sqrt{n}}, \qquad (5)$$

Яка, у разі справедливості H_o, має розподіл Ст'юдента з $\nu = n - 1$ степенями свободи. Критичну будують в залежності від альтернативної гіпотези.

1) Нехай $H^{(1)}_1$: $a = a_0$. За таблицею критичних точок розподілу Ст'юдента за даним рівнем значущості і числу степенів свободи $\nu = n - 1$ знаходять критичні точки $t_{1-\alpha/2,n-1}$ і $t_{1-\alpha/2,n-1} = -t_{1-\alpha/2,n-1}$.

Якщо $|T_{спос.}| = \left|\frac{\bar{x} - a_0}{s/\sqrt{n}}\right| < t_{\alpha/2,n-1}$

То немає підстав для відхилення H_o.

Якщо $|T_{спос.}| > t_{\alpha/2,n-1}$ – H_o відхиляють на користь $H^{(1)}_1$.

2) Нехай $H^{(2)}_1$: $a > a_0$. Знаходять критичну точку $t_{\alpha,n-1}$ правобічної критичної області.

Якщо $T_{спос.} < t_{\alpha,n-1}$ – H_o сприймається

Якщо $T_{спос.} \geq t_{\alpha,n-1}$-- H_o відхиляється на користь $H^{(2)}_1$.

3) Нехай $H^{(3)}_1: a < a_0$. Знаходять критичну точку $t_{1-\alpha,n-1}$=-$t_{\alpha,n-1}$. Лівобічної критичної області.

Якщо $T_{спос.} < -t_{\alpha,n-1}$ - H_o сприймається.

Якщо $T_{спос.} \geq -t_{\alpha,n-1}$ - H_o відхиляється на користь $H^{(3)}_1$.

Приклад 2. Аналізується дохід *X* фірм отраслі, який має нормальний закон розподілу. Припускають, що середній дохід у даній отраслі складає не менше 1 млн $. За вибіркою з 49 фірм одержали такі дані: $\bar{x}$=0,9 млн $ і S=1,15 млн $. Чи не протиречать ці результати висунутій гіпотезі при рівні значущості $\alpha = 0{,}01$?

Розв'язання.

$$H_o: a=1$$

$$H_1: \ a<1.$$

Для перевірки гіпотези знаходимо $T_{спос.} = \frac{0{,}9-1}{0{,}15/7}$=-4,67.

Критичну точку лівобічної критичної області визначаємо за таблицею критичних точок розподілу Ст'юдента $t_{kp.} = -t_{0{,}01;48} = -24{,}04$. Оскільки

$T_{спос.}$=-4,67<-2,404=$t_{kp.}$,

то H_o повинна бути відхилена на користь H_1. Це вказує на те, що середній дохід в отраслі менше, ніж 1 млн.$.

3.6 Перевірка гіпотези про величину дисперсії нормально розподіленої сукупності

Прийняття того чи іншого рішення в економіці часто нав'язують з аналізом можливих результатів, точніше розсіяння можливих результатів. Наприклад, при придбанні акцій будь-якої компанії важливо оцінити ризик від такої покупки, який визначається розсіюванням річних девідендів за даними акціями в довгостроковому періоді часу. Таку оцінку можна здійснити на базі аналізу дисперсії випадкової величини-розміру дивідендів. Отже, при вивченні

багатьох економічних проблем є необхідність чісти справу з висінням і перевіркою гіпотези про величину дисперсії.

Нехай випадкова величина $X \square N(a,\delta^2)$; а і δ^2 - невідомі. Перевіряється гіпотеза про рівність дисперсії δ^2 нормально розподіленої генеральної сукупності Х гіпотетичному значенню $\delta_0{}^2$. Тоді:

$$\text{Н}_{\text{о}}: \ \delta^2 = \delta_0{}^2$$

$$\text{Н}_1: \quad \delta^2 \neq \delta_0{}^2 \ (\text{Н}_1{}^{(2)}: \quad \delta^2 > \delta_0{}^2; \ \text{Н}_1{}^{(3)}: \quad \delta^2 < \delta_0{}^2).$$

Для перевірки Н $_{\text{о}}$ беруть вибірку об'єма n: $x_1, x_2, ...x_n$; знаходять вибіркове середнє $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, виправлену вибіркову дисперсію $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$. Тоді критерій перевірки Н $_{\text{о}}$ має такий вигляд: $X^2 = \frac{(n-1)S^2}{\sigma_0^2}$ (6)

У разі справедливості Н $_{\text{о}}$ статистика X^2 має X^2-розподіл з $\nu = n-1$ степенями свободи.

1) При Н$_1$: $\delta^2 \neq \delta_0{}^2$ за таблицею критичних точок X^2-розподілу за даним рівнем значущості α і числу степенів свободи $\nu = n-1$ знаходять критичні точки $X^2_{1-\alpha/2,n-1}$ і $X^2_{\alpha/2,n-1}$ двобічної критичної області.

Якщо $X^2_{1-\alpha/2,n-1} < X^2_{спос.} < X^2_{\alpha/2,n-1}$ - немає підстав для відхилення Н $_{\text{о}}$.

Якщо $X^2_{спос.} \leq X^2_{1-\alpha/2,n-1}$ або $X^2_{спос.} \geq X^2_{\alpha/2,n-1}$ - Н $_{\text{о}}$ відхиляється на користь Н$^{(1)}{}_1$.

3) При Н$^{(2)}{}_1$: $\delta^2 > \delta_0{}^2$ визначають критичну точку $X^2_{\alpha,n-1}$ правобічної критичної області.

Якщо $X^2_{спос.} < X^2_{\alpha,n-1}$ - немає підстав для відхилення Н $_{\text{о}}$.

Якщо $X^2_{спос.} \leq X^2_{1-\alpha,n-1}$ - Н $_{\text{о}}$ відхиляється на користь $H^{(3)}{}_1$.

Приклад 3.Точність роботи верстата – автомата, який наповнює пакети порошком, визначається співна дінням ваги пакетів. Дисперсія ваги не повинна перебільшувати $25(\text{г})^2$. За вибіркою з 20 пакетів знайдена виправлена дисперсія $S^2 = \frac{1}{20-1}\sum_{i=1}^{20}(x_i - \bar{x})^2 = 30\,(\text{г})^2$.

Необхідно з'ясувати, чи потребує верстат під наладки.

Розв'язання. Сформулюємо нульову і альтернативну гіпотези:

$$H_o: \delta^2 = 25;$$

$$H_1: \delta^2 > 25.$$

Знайдемо спостережуване значення критерія X^2 за формулою (6):

$X^2_{спос.} = \frac{19 \cdot 30}{25} = 22,8$, $X^2_{кр} = X^2_{0,05;19} = 30,14$.

Оскільки $X^2_{спос.}$=22,8<30,14= $X^2_{кр}$, то немає підстав для відхилення H_o. Це означає, що верстат не потребує під наладки.

3.7 Перевірка гіпотези про рівність математичних сподівань двох нормально розподілених сукупностей при невідомих дисперсіях

При аналізі багатьох економічних показників є потреба порівнювати дві генеральні сукупності. Наприклад, можна порівняти рівні життя в двох країнах за розміром доходу на душу населення, можна порівняти два варіанта інвестицій за розміром середніх дивідендів; якість знань студентів двох університетів – за середнім балом на комплексному тестовому екзамені. У цих випадках логічно провести порівняння за схемою аналіза рівності математичних. Сподівань двох генеральних сукупностей X і Y.

Нехай X $N(a_x, \sigma_x^2)$ і Y $N(a_y, \sigma_y^2)$, причому їх дисперсії σ_x^2, σ_y^2 відомі. За двома вибірками $x_1, x_2, ...x_n$ і $y_1, y_2, ...y_n$ об'ємів n і k відповідно необхідно перевірити гіпотезу $M(X) = M(Y)$, тобто

H_o: $M(X) = M(Y)$

H_1^1: $M(X) \neq M(Y)$ ($H_1^{(2)}$: $M(X) > M(Y)$; $H_1^{(3)}$: $M(X) < M(Y)$).

У якості критерія перевірки H_o беруть випадкову величину U:

$$U = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{k}}} \qquad (7)$$

У разі справедливості H_o, випадкова величина U $N(0;1)$.

1) При H_1^1: $M(X) \neq M(Y)$ за таблицею функції Лапласа визначають дві критичні точки $u_{1-\alpha/2}$ і $u_{\alpha/2}$ за умови:

$\Phi(u_{\alpha/2}) = \frac{1-\alpha}{2}$, $u_{1-\alpha/2} = u_{\alpha/2}$.

Якщо $|U_{спос.}| < u_{\alpha/2}$ - H_o сприймається.

Якщо $|U_{спос.}| \geq u_{\alpha/2}$ - H_o відхиляється на користь H_1^1

2) При $H^{(2)}{}_1$: $M(X) > M(Y)$ критичну точку u_α правобічної критичної області знаходять з рівності $\Phi(u_\alpha) = \frac{1-2\alpha}{2}$.

Якщо $U_{спос.} < u_\alpha$ - H_o сприймається.

Якщо $U_{спос.} \geq u_\alpha$ - H_o відхиляється на користь $H_1^{(2)}$.

4) При $H_1^{(3)}$: $M(X) < M(Y)$ критична точка $u_{1-\alpha}$ лівобічної критичної області визначається за умови $u_{1-\alpha} = -u_\alpha$.

Якщо $U_{спос.} > u_\alpha$ - H_o сприймається.

Якщо $U_{спос.} \leq u_\alpha$ - H_o відхиляється на користь $H_1^{(3)}$.

3.8 Перевірка гіпотези про рівність математичних сподівань двох нормально розподілених сукупностей при невідомих дисперсіях

Більш реальним у порівнянні з попереднім є випадок, коли дисперсії випадкових величин невідомі.

Нехай $X \square N(a_x, \sigma_x^2)$ і $Y \square N(a_y, \sigma_y^2)$, причому їх дисперсії σ_x^2, σ_y^2 невідомі. Висувається гіпотеза про рівність математичних сподівань.

H_o : $M(X) = M(Y)$

$H_1^{(1)}$: $M(X) \neq M(Y)$ ($H_1^{(2)}$: $M(X) > M(Y)$; $H_1^{(3)}$: $M(X) < M(Y)$).

У якості критерія перевірки H_o беруть випадкову величину T:

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{(n-1) \cdot S_x^2 + (k-1) S_y^2}} \cdot \sqrt{\frac{nk(n+k-2)}{n+k}}, \qquad (8)$$

де n, k – об'єми виборок $x_1, x_2, ... x_n$ і $y_1, y_2, ... y_n$ відповідно;

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $S_x{}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$,

$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $S_y{}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$

Якщо H_o справедлива, то статистика T має t – розподіл Ст'юдента з $\nu = n + k - 2$ степенями свободи.

1) При $H_1^{(1)}: M(X) \neq M(Y)$ за допомогою таблиці критичних точок розподілу Ст'юдента за даним рівнем значущості α і числом студентів свободи $\nu = n + k - 2$ визначається критичні точки $t_{1-\alpha/2, n+k-2}$, і $t_{\alpha/2, n+k-2}$ ($t_{1-\alpha/2, n+k-2}$=-$t_{\alpha/2, n+k-2}$) двобічної критичної області.

Якщо $|T_{спос.}| < t_{\alpha/2, n+k-2}$ - H_o сприймається

Якщо $|T_{спос.}| \geq t_{\alpha/2, n+k-2}$ - H_o відхиляється на користь $H_1^{(1)}$.

2) При $H_1^{(2)}$: $M(X) = M(Y)$ знаходять критичну точку $t_{\alpha, n+k-2}$ правобічної критичної області.

Якщо $T_{спос.} < t_{\alpha, n+k-2}$ - H_o сприймається

Якщо $T_{спос.} \geq t_{\alpha, n+k-2}$ - H_o відхиляється на користь $H_1^{(2)}$.

3)При $H_1^{(3)}$: $M(X) < M(Y)$ знаходять критичну точку лівобічної критичної області $t_{1-\alpha, n+k-2}$=-$t_{\alpha, n+k-2}$.

Якщо $T_{спос.} > -t_{\alpha, n+k-2}$ - H_o сприймається

Якщо $T_{спос.} \leq -t_{\alpha, n+k-2}$ - H_o відхиляється на користь $H_1^{(3)}$.

Приклад 4. В університеті проведено аналіз успішності серед студентів і студенток за останні 25 років. Випадкові величини X і Y – відповідно їх сумарний бал за час навчання. Одержано такі результати: $\bar{x}$=400б $S_x^2 = 300; \bar{y} = 420, S_y^2 = 150.$

Чи можна стверджувати, що дівчата в середньому вчаться краще хлопців? Прийняти α =0,05.

Розв’язання. Для відповіді на питання фактично необхідно перевірити таку гіпотезу:

H_o : $M(X) = M(Y)$

H_1 : $M(X) < M(Y)$.

За формулою (8) знаходимо $T_{спос.}$ з урахування того, що n=k=25:

$$T_{спос.} = \frac{400-420}{\sqrt{24\cdot 300+24\cdot 150}}\cdot\sqrt{\frac{25\cdot 25(25+25-2)}{25+25}} = -4,71;$$

$$t_{kp} = -t_{0,05;25+25-2} = -1,68.$$

Оскільки $T_{спос.}$=-4,71<-1,68=t_{kp}, то H_o повинна бути відхилена на користь H_1. Це дає підставу стверджувати, що у даному університеті дівчата в середньому вчаться краще хлопців.

3.9 Перевірка гіпотези про рівність дисперсій двох нормально розподілених сукупностей

Часто при порівнянні двох економічних показників на перший план виходить аналіз розсіювання значень випадкових величин. Наприклад, при вирішенні питання про інвестиції гостро стоїть проблема ризику в двох країнах середньодушові доходи можуть бути приблизно однаковими. Зіставив розсіювання доходів, ми одержуємо більш точну уяву про них. Аналіз, аналогічний описаному вище, доцільно проводити шляхом порівняння дисперсій випадкових величин.

Нехай X□ $N(a^2{}_x, \sigma_x^2)$ і Y□ $N(a_y, \sigma_y^2)$, причому їх дисперсії σ_x^2, σ_y^2 невідомі.

Висувається гіпотеза про рівність дисперсій σ_x^2, σ_y^2:

H_o : $\delta_x^2 = \delta_y^2$

$H_1^{(1)}$: $\delta_x^2 \neq \delta_y^2$ ($H_1^{(2)}$: $\delta_y^2 > \delta_y^2$).

За незалежними вибірками $x_1, x_2, ...x_n$ і $y_1, y_2, ...y_n$ об’ємів n і k відповідно визначається $\bar{x}, \bar{y}, S_x^2, S_y^2$ (для визначеності нехай $S_x^2 \geq S_y^2$, у протилежному випадку ці величини можна перепозначити).

У якості критерію перевірки H_o беруть випадкову величину F:

$$F = \frac{S_x^2}{S_y^2}, \qquad (9)$$

Тобто відношення більшої виправленої дисперсії до меншої, якщо H_o справедлива, то ця статистика F має F – розподіл Фішера з $\nu_1 = n-1$ і $\nu_2 = k-1$ ступенями свободи.

1) При $H_1^{(1)}$: $\delta_x^2 \neq \delta_y^2$ за таблицею критичних значень розподілу Фішера за рівнем значущості α і числом ступенів свободи ν_1 і ν_2 знаходять критичну точку $F_{\alpha/2,\nu_1,\nu_2}$.

Якщо $F_{спос.} < F_{\alpha/2,\nu_1,\nu_2}$ – H_o сприймається

Якщо $F_{спос.} \geq F_{\alpha/2,\nu_1,\nu_2}$ -- H_o відхиляється на користь $H_1^{(1)}$.

2) При $H^{(2)}{}_1$: $\delta^2 > \delta_0{}^2$ визначають критичну точку F_{α,ν_1,ν_2}.

Якщо $F_{спос.} < F_{\alpha,\nu_1,\nu_2}$ – H_o сприймається

Якщо $F_{спос.} \geq F_{\alpha,\nu_1,\nu_2}$ -- H_o відхиляється на користь $H_1^{(2)}$.

Приклад 5. За умовами прикладу 4 визначити, чи є підстави припускати, що дисперсії X і Y істотно розрізняються – тобто розсіювання оцінок у хлопців більше, ніж у дівчат.

Розв'язання. За умови задачі висувається слідуюча гіпотеза:

H_o: $\delta_x^2 = \delta_y^2$

$H_1^{(1)}$: $\delta_y^2 > \delta_y^2$

Для перевірки гіпотези за формулою (9) знаходимо $F_{спос.}$=300/150=2. Критична точка розподілу Фішера $F_{kp} = F_{0,05;24;24} = 1{,}98$.

Оскільки $F_{спос.}$=2>1,98=F_{kp}, то H_o повинна бути відхилена на користь H_1, тобто є підстави припустити, що розсіювання оцінок у хлопців даного університету істотно більше розсіювання оцінок у дівчат.

Перевірка гіпотези про значущість коефіцієнту кореляції

Одним з важливих елементів економетричного аналізу є установлення зв'язку між різними показниками (між ціною і попитом, доходом і споживанням, інфляцією і безробіттям). Зазвичай аналіз починають з найпростішої – лінійної залежності. Для того, щоб установити наявність значущої лінійної залежності між двома випадковими величинами X і Y необхідно перевірити гіпотезу про статистичну значущість кореляції. У цьому випадку використовують таку гіпотезу:

$$H_o: \delta_{xy} = 0$$

$$H_1^{(1)}: \quad \delta_{xy} \neq 0$$

Для перевірки H_o за вибіркою $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ об'єму і будують статистику

$$T = \frac{r_{xy} \cdot \sqrt{n-2}}{\sqrt{1-r_{xy}^2}}, \tag{10}$$

де r_{xy} - вибірковий коефіцієнт кореляції.

Якщо гіпотеза H_o справедлива, то T має розподіл Ст'юдента з $\nu = n-2$ ступенями свободи. За таблицею критичних точок розподілу Ст'юдента за даним рівнем значущості α і числу степенів свободи n-2 визначають критичну точку $t_{\alpha/2, n-2}$.

Якщо $|T| = \left| \frac{r_{xy} \cdot \sqrt{n-2}}{\sqrt{1-r_{xy}^2}} \right| < t_{\alpha/2, n-2}$, то немає підстав для відхилення H_o. Якщо $|T_{спос.}| \geq t_{\alpha/2, n-2}$ то H_o відхиляється на користь альтернативної гіпотези $H_1^{(1)}$.

Якщо H_o відхиляється, то це означає, що коефіцієнт кореляції статистично значущий (істотно відрізняється від нуля). Отже, X і Y – корильовані, тобто між ними існує лінійна залежність.

Приклад 6. Вивчається наявність лінійної залежності між рівнем інфляції X і безробіттям У у деякій країні за 11 років. За статистичними даними розраховано вибірковий коефіцієнт кореляції r_{xy}=-0,34. Чи існує значуща

лінійна залежність між вказаними показниками у даній країні? Прийняти $\alpha = 0,02$.

Розв'язання. Проаналізуємо таку гіпотезу

H_o : δ_{xy} =0

$H_1^{(1)}$: $\delta_{xy} \neq 0$.

За формулою (10) знаходимо $T_{спос} = \frac{= 0,34 \cdot \sqrt{11-2}}{\sqrt{1-(-0,34)^2}} = -3,254$.

За допомогою таблиці критичних точок розподілу Ст'юдента знаходимо $t_{\alpha/2, n-2} = t_{0,01;9}$=2,821. Оскільки $|T_{спос}| = 3,254 > 2,821 = t_{kp}$ то коефіцієнт кореляції r_{xy} статистично значущий. Отже δ_{xy} істотно відрізняється від нуля, а між рівнями інфляції X і безробіттям У існує обернена лінійна залежність.

Запитання для самоперевірки

1. Дайте означення статистичної гіпотези.
2. Чим відрізняється параметрична гіпотеза від непараметричної?
3. Дайте визначення нульової та альтернативної гіпотез.
4. Що називається статистичним критерієм?
5. Що таке помилка першого і другого роду? Як можна зменшити ймовірності цих помилок?
6. Що таке рівень значущості α?
7. Що таке потужність критерію?
8. Приведіть загальну схему перевірки гіпотез.

Задачі для самостійного розв'язання

1. Верстат – автомат наповнює пакет чіпсами по 250 гр.. Вважають, що верстат потребує під наладки, якщо стандартне відхилення від номінальної ваги більш ніж 5 г.. Контрольні зважування 10 пакетів дали такі результати: 245, 248, 250, 252, 256, 243, 251, 244, 253. Чи потребує верстат під наладки?

2. Витрати бензину на автомобілі деякої фірми мають нормальний закон розподілу з a_x=7,5 л і σ_x=0,5 л. Фірма виробила нову модифікацію автомобіля і

стверджує, що у нього середні витрати a_x бензину знижено до 7 л при тому ж значенні σ. Вибірка з 15 автомобілів кожної моделі дала слідуючі середні витрати: $\bar{x}$=7,45; $\bar{y}$=7,15.

3. Два університету А і В готовлять спеціалістів аналогічних спеціальностей. Міністерство освіти вирішило перевірити якість підготовки в обох університетах. Було проведено тестовий екзамен для студентів п'ятого курсу . Відібрані випадково студенти показали слідуючи суми балів:

А: 41,50,35,45,53,30,57,20,50,44,36,48,55,28,40,50.

В: 40,57,52,38,25,47,52,48,55,48,53,39,46,51,45,55,43,51.

Чи мажна стверджувати при рівні значущості α=0,05, що один з університетів забезпечує краще підготовку?

4. На основі спостережень за роботою 25 кандидатів на посаду секретаря-референта з'ясовано, що в середньому вони витрачали 7 хвилин на набір однієї сторінки складного тексту на комп'ютері при вибірному стандартному віди ленні S=2 хвилин.

Проте, припускали, що час X набору однієї сторінки 5,5 хвилин. Чи не протирічать одержані діні цій гіпотезі, якщо час X набору текста має нормальний закон розподілу? Прийняти рівень значущості $\alpha = 0,01$.

5. Аналіз залежності між двома показниками X і У за 25 спостереженнями дав такі результати: $\bar{x}$=100; $\bar{y}$=75; $\sum_{i=1}^{25}(y_i - \bar{y})^2 = 625$, $\sum_{i=1}^{25} x_i y_i = 187000$; $\sum_{i=1}^{25}(y_i - \bar{y})^2 = 484$. Оцінити наявність лінійної залежності між X і У. Чи буде коефіцієнт кореляції σ_{xy} статистично значущим?

6. Припускають, що місячна зарплатня співробітників фірми складає 1000$ при стандартному відхиленні σ=100$. Вибірка з 36 осіб дала такі результати: $\bar{x} = 900\$$, S=150$. Чи можна за результатами спостережень стверджувати, що середня зарплатня співробітників фірми менше рекламованої, а розсіюванні зарплатні більше.

4 Елементи кореляційного аналізу. Парна лінійна регресія

Можливі два варіанта взаємозв'язків між двома змінними X і Y. У першому випадку обидві змінні розглядаються як рівноцінні в тому сенсі, що вони не поділяються на залежну і незалежну змінні. Головним у цьому разі є питання про присутність і силу взаємозв'зку між цими змінними. Наприклад, між ціною товару й об'ємом попиту на нього, між урожаєм картоплі й урожаєм зерна, між інтенсивністю руху транспорта і числом аварій. При дослідженні сили лінійної залежності між такими змінними звертаються до кореляційного аналізу, основною мірою якого є коефіцієнт кореляції.

У другому випадку взаємозв'язків виділяють одну з величин як незалежну, а іншу як залежну. У цьому разі зміна першої з них може бути причиною для зміни другої. Наприклад, зростання доходу веде до збільшення споживання; зростання ціни – до зниження попиту; зниження відсоткової ставки збільшує інвестиції. Але така залежність не є однозначною в тому сенсі, що кожному конкретному значенню незалежній змінній може відповідати не одна, а множина значень з деякої області. Іншими словами, кожному конкретному значенню незалежної змінної відповідає деякий ймовірносний розподіл залежної змінної (яку розглядають як випадкову величину). Залежність такого типу виражається співвідношенням $M(Y|x) = f(x)$ і називається функцією регресії Y на X. $M(Y|x)$ – умовне математичне сподівання. Інакше кажучи, регресія – це функціональна залежність між незалежною змінною й умовним математичним сподіванням залежної змінної, яка будується з метою передбачення цього середнього значення при фіксованому значенні першої.

Для відображення того факта, що реальне значення залежної змінної не завжди співпадає з її умовним математичним сподіванням і можуть бути різними при одному й тому ж значенні незалежної змінної, фактична залежність повинна бути доповнена деяким доданком ε, який по суті є випадковою величиною й вказує на стохастичну сутність залежності. Отже,

зв'язок між залежною і незалежною змінними виражається співвідношенням $Y = M(Y|x) + \varepsilon,$ яке називається регресійною моделлю.

Найбільш істотними причинами включення випадковою фактора в регресійну модель є:

1. Невключення в модель усіх пояснювальних змінних, оскільки будь-яка регресійна модель є спрощенням реальної ситуації.
2. Неправильний вибір функціональної форми моделі.
3. Помилки вимірювань.
4. Обмеженність статистичних даних.
5. Непередбаченість людського фактора.

Розв'язання задачі побудови якісного рівняння регресії, яке відповідає емпірічним даним можно розбити на три кроки:

1. вибір формули рівняння регресії;
2. оцінка параметрів вибраного рівняння;
3. аналіз якості рівняння і перевірка адекватності рівняння емпіричним даним.

Вибір формули зв'язку змінних називається специфікацією рівняння регресії. У випадку парної регресії вибір формули зазвичай здійснюється за графічним зображенням реальних статистичних даних у вигляді точок в декартовій системі координат, яке називається кореляційним полем (рисунок 1).

На рисунку 1а взаємозв'язок між X і Y близький до лінійного. У даному випадку в якості залежності між X і Y можна взяти лінійну функцію $Y = b_0 + b_1 X$.

На графіку 1б реальний зв'язок між X і Y скоріше всього описується квадратичною функцією $Y = aX^2 + bX + c.$

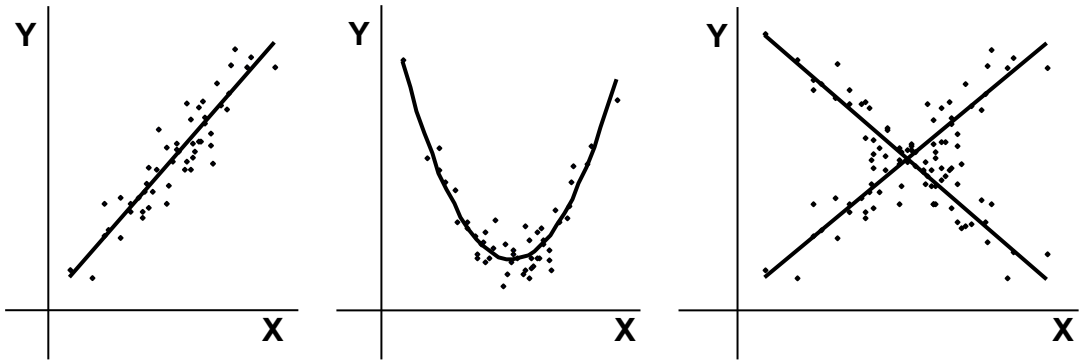
На графіку 1в явний зв'язок між X і Y відсутній.

Якщо функція регресії лінійна, то кажуть про лінійну регресію. Отже, лінійна регресія являє собою лінійну функцію між умовним математичним

сподіванням $M(Y|X = x_i)$ залежної змінної Y і однієї пояснювальної змінної X (x_i – значення незалежної змінної в i – тому спостереженні, $i = 1,2,...,n$)

$$M(Y|X = x_i) = b_0 + b_1 x_i.$$

Зауважимо, що принциповою в даному випадку є лінійність за параметрами b_0 і b_1.



а *б* *в*

Рис. 4.1

Для відображення того факту, що кожне індивідуальне значення y_i відхиляється від відповідного умовного математичного сподівання, необхідно ввести у співвідношення $M(Y|X = x_i) = b_0 + b_1 x_i$ випадковий доданок ε_i

$$y_i = M(Y|X = x_i) + \varepsilon_i = b_0 + b_1 x_i + \varepsilon_i.$$

Таке співвідношення називається теоретичною лінійною регресійною моделлю; b_0 і b_1 - теоретичними параметрами (коефіцієнтами); ε_i - випадковим відхиленням.

У загальному вигляді теоретичну лінійну регремійну модель подають у вигляді

$$Y = b_0 + b_1 X + \varepsilon.$$

Для визначення значень теоретичних коефіцієнтів регресії необхідно знати всі значення змінних X і Y генеральної сукупності, що практично неможливо.

Отже, задача лінійного регресійного аналізу полягає в тому, щоб за статистичними даними $(x_i, y_i), i = 1,2,...,n$ для змінниз X і Y:

1. одержати найкращі оцінки невідомих параметрів b_0 і b_1;
2. перевірити статистичні гіпотези про параметри моделі;
3. перевірити, чи достатньо гарно модель погоджується зі статистичними даними.

Таким чином, за вибіркою обмеженого об'єма можна побудувати так зване емпірічне рівняння регресії

$$\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_i,$$

де $\hat{y}_i$ – оцінка (наближене значення) умовного математичного сподівання $M(Y|X = x_i)$; $\hat{b}_0$ і $\hat{b}_1$ – оцінки невідомих параметрів b_0 і b_1, які називають емпірічними коефіцієнтами регресії. Отже, в конкретному випадку

$$y_i = \hat{b}_0 + \hat{b}_1 x_i + e_i,$$

де відхилення e_i – оцінка теоретичного випадкового відхилення ε_i.

Найбільш поширеним і теоретично обгрунтованим методом знаходження коефіцієнтів є метод найменших квадратів (МНК), згідно з яким коефіцієнти $\hat{b}_0$ і $\hat{b}_1$ знаходяться за умови

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{b}_0 - \hat{b}_1 x_i)^2 \to \min.$$

Згідно МНК оцінки параметрів $\hat{b}_0$ і $\hat{b}_1$ визначаються так:

$$\hat{b}_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2}, \ \hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x}.$$

Тут $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\overline{xy} = \frac{1}{n}$.

Коефіцієнт $\hat{b}_1$ можно також обчислити й за формулою $\hat{b}_1 = r_{xy} \cdot \frac{S_y}{S_x}$,

де $r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\overline{x^2} - \bar{x}^2} \cdot \sqrt{\overline{y^2} - \bar{y}^2}}$ – вибірковий коефіцієнт кореляції;

$S_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$, $S_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}$ – стандартні відхилення.

Властивості оцінок коефіцієнтів регресії, а, отже, і якості побудованої регресії істотно залежать від властивостей випадкової складової ε. Доведено, що для одержання за МНК найкращих результатів (оцінки обгрунтовані, незміщені й ефективні) необхідно, щоб виконувався ряд передбачень відносно випадкового відхилення:

1. математичне сподівання випадкового відхилення ε_i дорівнює нулю, а саме $M(\varepsilon_i) = 0, i = 1,2,..n;$
2. дисперсія випадкових відхилень ε_i стала $D(\varepsilon_i) = \sigma^2, i = 1,2,..n;$
3. випадкові відхилення ε_i і ε_j є незалежними одне від одного, $i \neq j, \mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0;$
4. випадкове відхилення повинно бути незалежно від пояснюючих змінних $\mathrm{cov}(\varepsilon_i, x_i) = 0.$

Припускають також, що випадкова величина ε розподілена за нормальним законом.

Емпірічне рівняння регресії визначається на основі скінченного числа статистичних даних. Отже, коефіцієнти емпірічного рівняння регресії є випадковими величинами. При проведенні статистичного аналізу виникає необхідність порівняння емпірічних коефіцієнтів регресії $\hat{b}_0$ і $\hat{b}_1$ з деякими теоретично очікуваними значеннями b_0 і b_1 цих коефіцієнтів. Даний аналіз проводиться за схемою статистичної перевірки гіпотез.

Для перевірки гіпотези

$$H_0 : \hat{b}_1 = b_1$$

$$H_1 : \hat{b}_1 \neq b_1$$

використовують статистику $t = \dfrac{\hat{b}_1 - b_1}{S_{\hat{b}_1}}$, яка при справедливості H_0 має розподіл Стьюдента з числом степеней свободи $k = n - 2$, $n -$ об'єм вибірки. Отже, $H_0 : \hat{b}_1 = b_1$ відхиляється на основі даного критерія, якщо

$$\left|t_{факт}\right| = \left|\frac{\hat{b}_1 - b_1}{S_{\hat{b}_1}}\right| \geq t_{\alpha/2, n-2},$$

де $\alpha -$рівень значущості. При невиконанні цієї нерівності немає підстав для відхилення H_0.

$S_{\hat{b}_1}$ $-$стандартна помилка оцінки $\hat{b}_1$, яка обчислюється так

$$S_{\hat{b}_1} = \sqrt{\frac{S^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}},$$

де

$$S^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2}.$$

$t_{\alpha/2, n-2}$ $-$критичне (табличне) значення статистики Стьюдента, яке відповідає рівню значущості α і числу степеней свободи $k = n - 2$.

Найбільш важливою на початковому етапі статистичного аналізу побудованої моделі є задача виявлення присутності лінійного зв'язку між Y і X. Ця проблема може бути вирішена за такою схемою:

$$H_0 : \hat{b}_1 = 0$$

$$H_1 : \hat{b}_1 \neq 0.$$

Таку гіпотезу називають гіпотезою про статистичну значущість коефіцієнтів регресії. При цьому, якщо H_0 сприймається, то є підстава стверджувати, що величина Y не залежить від X. У цьому випадку кажуть, що коефіцієнт $\hat{b}_1$ статистично незначущий (він близький до нуля). При відхиленні H_0 припускають, що коефіцієнт $\hat{b}_1$ значущий, тобто є лінійний зв'язок між Y і X. За аналогічною схемою на основі $t-$статистики превіряють гіпотезу про статистичну значущість коефіцієнта $\hat{b}_0$, при цьому стандартна помилка $S_{\hat{b}_0}$ обчислюється так

$$S_{\hat{b}_0} = \sqrt{\frac{S^2 \sum_{i=1}^{n} x_i^2}{n \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}} = \sqrt{\overline{x^2} \cdot S_{\hat{b}_1}^2}.$$

Слід зазначити, що для парної регресії більш важливим є аналіз статистичної значущості коефіцієнта $\hat{b}_1$, оскільки саме в ньому завуальований вплив змінної X на результат Y.

Довірчі інтервали, які з надійністю $\gamma = 1 - \alpha$ покривають параметри b_0 і b_1 мають такий вигляд

$$(\hat{b}_0 - t_{\alpha/2, n-2} \cdot S_{\hat{b}_0}; \hat{b}_0 + t_{\alpha/2, n-2} \cdot S_{\hat{b}_0}),$$

$$(\hat{b}_1 - t_{\alpha/2, n-2} \cdot S_{\hat{b}_1}; \hat{b}_1 + t_{\alpha/2, n-2} \cdot S_{\hat{b}_1}).$$

Однією з центральних задач регресійного моделювання є передбачення (прогнозування) значень залежної змінної при конкретних значеннях пояснювальної змінної. Можливі два підходу: або передбачити умовне математичне сподівання залежної змінної (прогнозування середнього значення), або прогнозувати деяке конкретне значення залежної змінної (передбачення індивідуального значення).

Нехай побудовано рівняння регресії $\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_i$, на основі якого необхідно передбачити умовне математичне сподівання $M(Y|X = x_p)$ змінної Y при $X = x_p$. у даному випадку значення $\hat{y}_p = \hat{b}_0 + \hat{b}_1 x_p$ є оцінкою $M(Y|X = x_p)$. Тоді природньо виникає питання, як сильно може відхилитися модельне середнє значення $\hat{y}_p$, розраховане за емпірічним рівнянням, від відповідного умовного математичного сподівання? Відповідь на це питання дається на основі інтервальних оцінок, побудованих із заданою надійністю $\gamma = 1 - \alpha$ при будь-якому конкретному значення x_p пояснювальної змінної.

Довірчий інтервал для математичного сподівання $M(Y|X = x_p)$ має такий вигляд

$$(\hat{b}_0 + \hat{b}_1 x_p - t_{\alpha/2,n-2} \cdot S \cdot \sqrt{\frac{1}{n} + \frac{(\bar{x} - x_p)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}; \hat{b}_0 + \hat{b}_1 x_p + t_{\alpha/2,n-2} \cdot S \cdot \sqrt{\frac{1}{n} + \frac{(\bar{x} - x_p)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}).$$

На практиці іноді більш важливо знайти дисперсію Y, ніж її середнє значення або довірчі інтервали для умовних математичних сподівань. Це дозволяє визначити допустимі межі для конкретного значення Y.

Довірчий інтервал для конкретного значення змінної Y, яке відповідає значенню $X = x_p$ визначається так

$$(\hat{b}_0 + \hat{b}_1 x_p - t_{\alpha/2,n-2} \cdot S \cdot \sqrt{1 + \frac{1}{n} + \frac{(\bar{x} - x_p)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}; \hat{b}_0 + \hat{b}_1 x_p + t_{\alpha/2,n-2} \cdot S \cdot \sqrt{1 + \frac{1}{n} + \frac{(\bar{x} - x_p)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}).$$

Цей інтервал визначає межі, поза якими може знаходитися не більше $100 \cdot \alpha\%$ точок спостережень при $X = x_p$. Зауважимо, що даний інтервал ширший довірчого інтервалу для умовного математичного сподівання (на рисунку 2 межі цього інтервалу показано пунктирною лінією).

Аналізуючи побудовані інтервали, слід зазначити, що найбільш вузькими вони будуть при $x_p = \bar{x}$. По мірі віддалення x_p від серенього значення довірчі

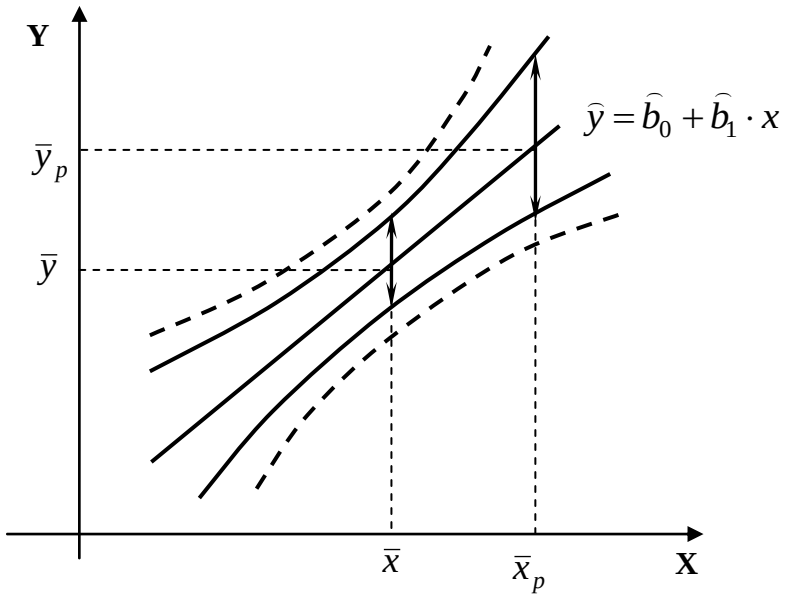


Рис. 4.2

інтервали розширяються (рисунок 4.2). Отже, необхідно достатньо обережно

екстраполювати одержані результати на прогнозові області. З іншого боку, по мірі зростання числа спостережень n ці інтервали звужуються до лінії регресії.

Приклад 24 Для аналізу залежності об'єма споживання Y (у.о.) домогосподарств від доходу X (у.о.) відібрана вибірка об'єма $n = 12$ (помісячно протягом року), результати якої наведені в таблиці 1. Необхідно:

1 визначити тип залежності;

2 оцінити параметри рівняння регресії Y на X;

3 оцінити силу лінійної залежності між X і Y;

4 оцінити статистичну значущість коефіцієнтів регресії при рівні значущості $\alpha = 0{,}05$;

5 розрахувати 95%-і довірчі інтервали для теоретичних коефіцієнтів регресії;

6 спрогнозувати споживання при доході $X = 160$; визначити 95%-і довірчі інтервали для цього прогнозу.

Таблиця 4.1

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	107	109	110	113	120	122	123	128	136	140	145	150
y_i	102	105	108	110	115	117	119	125	132	130	141	144

Розв`язання.

1 Для визначення типу залежності побудуємо кореляційне поле (рисунок 4. 3).

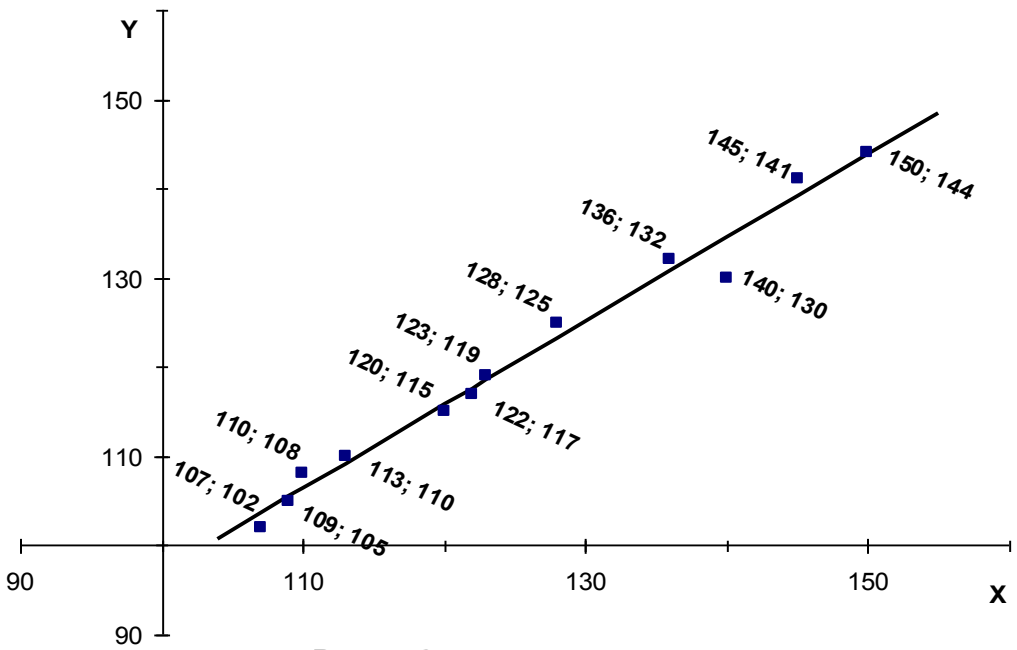


Рис. 4.3

За розташуванням точок на кореляційн6ому полі припускаємо, що залежність між X і Y лінійна: $\hat{Y} = \hat{b}_0 + \hat{b}_1 X$.

2 Для наочності обчислень за МНК побудуємо таблицю 4.2.

Таблиця 4.2

i	x_i	y_i	x_i^2	$x_i y_i$	y_i^2	$\hat{y}_i$	e_i	e_i^2
1	107	102	11449	10914	10404	103,63	-1,63	2,66
2	109	105	11881	11445	11025	105,49	-0,49	0,24
3	110	108	12100	11880	11664	106,43	1,57	2,46
4	113	110	12769	12430	12100	109,23	0,77	0,59
5	120	115	14400	13800	13225	115,77	-0,77	0,59
6	122	117	14884	14274	13689	117,63	-0,63	0,40
7	123	119	15129	14637	14161	118,57	0,43	0,18
8	128	125	16384	16000	15625	123,24	1,76	3,10
9	136	132	18496	17952	17424	130,71	1,29	1,66
10	140	130	19600	18200	16900	134,45	-4,45	19,8
11	145	141	21025	20445	19881	139,11	1,89	3,57
12	150	144	22500	21600	20736	143,78	0,22	0,05
Сума	1503	1448	190617	183577	176834	-	0	35,3
Середнє	125,25	120,67	15884,75	15298,08	14736,17	-	-	-

У таблиці значення округляються до сотих; ураховуються похибки округлень.

Згідно МНК, маємо

$$\hat{b}_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{15298{,}08 - 125{,}25 \cdot 120{,}67}{15884{,}75 - (125{,}25)^2} = \frac{184{,}1625}{197{,}1875} = 0{,}9339;$$

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x} = 120{,}67 - 0{,}9339 \cdot 125{,}25 = 3{,}699.$$

Отже, рівняння парної лінійної регресії має такий вигляд: $\hat{Y} = 3{,}699 + 0{,}9339X$. Зобразимо дану пряму регресії на кореляційному полі. За цим рівнянням розрахуємо $\hat{y}_i$, а також $e_i = y_i - \hat{y}_i$.

3 Для аналізу сили лінійної залежності обчислимо коефіцієнт кореляції:

$$r_{xy} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sqrt{\overline{x^2} - \overline{x}^2} \cdot \sqrt{\overline{y^2} - \overline{y}^2}} = \frac{184{,}1625}{14{,}04 \cdot 13{,}23} = 0{,}9914.$$

Знайдене значення коефіцієнта кореляції дозволяє зробити висновок про сильну лінійну залежність між змінними X і Y.

4 Оцінимо статистичну значущість коефіцієнтів $\hat{b}_0$ і $\hat{b}_1$.

Маємо:

$$S^2_{\hat{b}_1} = \frac{S^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{S^2}{n(\overline{x^2} - \overline{x}^2)} = \frac{\sum_{i=1}^{n} e_i^2}{n(n-2)(\overline{x^2} - \overline{x}^2)} = \frac{\sum_{i=1}^{n}(y_i - \hat{b}_0 - \hat{b}_1 x_i)^2}{n(n-2)(\overline{x^2} - \overline{x}^2)} =$$

$$= 0{,}0023;$$

$$S_{\hat{b}_1} = \sqrt{0{,}0023} = 0{,}0485; \; t_{\hat{b}_1} = \frac{\hat{b}_1}{S_{\hat{b}_1}} = \frac{0{,}9339}{0{,}0485} = 19{,}2557.$$

Критичне значення для рівня значущості $\alpha = 0{,}05$ дорівнює $t_{крит.} = t_{\alpha/2, n-2} = t_{0{,}025;10} = 2{,}23.$

Порівняємо модуль спостережуваного значення $\left|t_{\hat{b}_1}\right| = 19{,}2557$ з критичним значенням $t_{0{,}025;10}$. Оскільки $t_{\hat{b}_1} = 19{,}2557 > 2{,}23 = t_{крит}$, то нульова гіпотеза $H_0 : \hat{b}_1 = 0$ повинна бути відхилена на користь альтернативної гіпотези $H_1 : \hat{b}_1 \neq 0$ при вибраному рівні значущості. Це підтверджує статистичну значущість коефіцієнта регресії $\hat{b}_1$.

Аналогічно перевіряється статистична значущість коефіцієнта $\hat{b}_0$:

$$S^2_{\hat{b}_0} = \overline{x^2} \cdot S^2_{\hat{b}_1} = 0{,}0023 \cdot 15884{,}75 = 36{,}5349;$$

$$S_{\hat{b}_0} = \sqrt{36{,}5349} = 6{,}044; \; t_{\hat{b}_0} = \frac{\hat{b}_0}{S_{\hat{b}_0}} = \frac{3{,}699}{6{,}044} = 0{,}612.$$

Оскільки $\left|t_{\hat{b}_0}\right| = 0{,}612 < 2{,}23 = t_{крит.}$, то гіпотеза $H_0 : \hat{b}_0 = 0$ про статистичну незначущість коефіцієнта $\hat{b}_0$ не відхиляється. Це означає, що у

даному випадку вільним членом рівняння регресії можна знехтувати, тобто можна розглядати регресію $Y = \hat{b}_1 \cdot X$.

5 Згідно формулам

$$(\hat{b}_0 - t_{\alpha/2,n-2} \cdot S_{\hat{b}_0}; \hat{b}_0 + t_{\alpha/2,n-2} \cdot S_{\hat{b}_0}),\ (\hat{b}_1 - t_{\alpha/2,n-2} \cdot S_{\hat{b}_1}; \hat{b}_1 + t_{\alpha/2,n-2} \cdot S_{\hat{b}_1})$$

довірчі інтервали будуть такими:

$(3{,}699 - 2{,}23 \cdot 6{,}044; 3{,}699 + 2{,}23 \cdot 6{,}044) = (-9{,}779; 17{,}177) -$ для коефіцієнта $\hat{b}_0$;

$(0{,}9339 - 2{,}23 \cdot 0{,}0485; 0{,}9339 + 2{,}23 \cdot 0{,}0485) = (0{,}826; 1{,}042) -$ для коефіцієнта $\hat{b}_1$.

Фактично довірчі інтервали визначають значення теоретичних коефіцієнтів регресії b_0 і b_1, які будуть прийнятні з надійністю 0,95 при знайдених оцінках $\hat{b}_0 = 3{,}699$ і $\hat{b}_1 = 0{,}9339$.

6 Прогнозове споживання при доході $x = 160$ складе $\hat{y}(160) = 3{,}699 + 0{,}9339 \cdot 160 =$ $= 153{,}12$.

Розрахуємо 95%-і довірчий інтервал для умовного математичного сподівання $M(Y|X = x_p)$ при $X = x_p = 160$. Скориставшись формулою

$$(\hat{b}_0 + \hat{b}_1 x_p - t_{\alpha/2,n-2} \cdot S \cdot \sqrt{\frac{1}{n} + \frac{(\bar{x} - x_p)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}; \hat{b}_0 + \hat{b}_1 x_p + t_{\alpha/2,n-2} \cdot S \cdot \sqrt{\frac{1}{n} + \frac{(\bar{x} - x_p)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}).$$

знайдемо межі інтервалу:

$$3{,}699 + 0{,}9339 \cdot 160 \pm 2{,}23 \cdot 1{,}8788 \cdot \sqrt{\frac{1}{12} + \frac{(125{,}25 - 160)^2}{2102{,}1875}}.$$

Таким чином, довірчий інтервал для середнього значення Y при $X = 160$ має вигляд: (149,728;156,5193). Іншими словами, середнє споживання при доході 160 з імовірністю 0,95 буде знаходитися в інтервалі (149,728;156,5193).

Розрахуємо межі інтервала, в якому буде зосереджено не менше 95% можливих об’ємів споживання при необмежено великому числі спостережень і рівні доходу $X = 160$. Для цього скористаємося формулою

$$(\hat{b}_0+\hat{b}_1x_p-t_{\alpha/2,n-2}\cdot S\cdot\sqrt{1+\frac{1}{n}+\frac{(\bar{x}-x_p)^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}};\hat{b}_0+\hat{b}_1x_p+t_{\alpha/2,n-2}\cdot S\cdot\sqrt{1+\frac{1}{n}+\frac{(\bar{x}-x_p)^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}).$$

$$3{,}699+0{,}9339\cdot 160\pm 2{,}23\cdot 1{,}8788\cdot\sqrt{1+\frac{1}{12}+\frac{(125{,}25-160)^2}{2102{,}1875}}.$$

Тоді інтервал, у якому буде знаходитися, в крайньому разі, 95% індивідуальних об’ємів споживання при доході $X=160$, має вигляд: (147,4898;158,7082).

Слід звернути увагу на те, шо він включає в себе довірчий інтервал для умовного середнього споживання.

Індивідуальні завдання

22.01-22.30

1 Визначити тип залежності;

2 оцінити параметри рівняння регресії Y на X;

3 оцінити силу лінійної залежності між X і Y;

4 оцінити статистичну значущість коефіцієнтів регресії при рівні значущості $\alpha=0{,}05$;

5 розрахувати 95%-і довірчі інтервали для теоретичних коефіцієнтів регресії;

6 спрогнозувати результативну ознаку $X=x_p$; визначити 95%-і довірчі інтервали для цього прогнозу.

22.01

Залежність між ростом Y та масою дітей X наведена в таблиці:

$(X=x_i)$, кг	0,531	0,524	0,541	0,550	0,559	0,620	0,632	0,672	0,682
$(Y=y_i)$, м	0,620	0,580	0,640	0,650	0,670	0,680	0,695	0,699	0,710

Продовження таблиці

$(X=x_i)$, кг	0,689	0,692	0,694	0,698	0,690	0,710	0,720	0,725	0,730
$(Y=y_i)$, м	0,715	0,725	0,781	0,790	0,795	0,800	0,810	0,850	0,860

$X=x_p=0{,}75.$

22.02

Залежність кількості проданих пар чоловічого взуття Y від його розміру X наведена в таблиці:

$(X = x_i)$	44	43	42	41	40	39	38	37
$(Y = y_i)$, шт	10	25	68	136	152	162	170	180

$X = x_p = 45.$

22.03

Вимірювання температури в грудні, здійснені в двох містах, що умовно позначені A і B, наведено в таблиці:

Місто B $(X = x_i)$, °C	-20,2	-20,5	-21,4	-21,8	-22,0	-22,5	-22,8	-22,8
Місто A $(Y = y_i)$, °C	-10,2	-11,5	-12,4	-12,8	-13,0	-13,5	-14,2	-14,6

Продовження таблиці

Місто B $(X = x_i)$, °C	-23,2	-24,1	-24,5	-25,1	-25,8	-26,0	-26,5	-27,0
Місто A $(Y = y_i)$, °C	-14,6	-15,7	-16,4	-17,2	-17,5	-18,2	-18,6	-18,9

$X = x_p = -27{,}5.$

22.04

Зі старшого класу навмання вибраної середньої школи було відібрано групу учнів. Дані про їх середньорічні оцінки з математики та решти дисциплін у балах наведено в таблиці:

$(X = x_i)$	30	35	31	38	41	48	50	55	51	58
$(Y = y_i)$	45	25	48	52	54	51	59	60	62	69

Продовження таблиці

$(X = x_i)$	60	59	65	73	78	71	79	80	81
$(Y = y_i)$	72	78	76	80	82	85	81	90	93

$X = x_p = 82.$

22.05

Конденсатор було заряжено до повної напруги в певний момент часу t, після цього він починає розряджатися. Залежність напруги Y від часу розряджання X наведено в таблиці:

$(X = x_i)$	0	1	2	3	4	5	6
$(Y = y_i)$	100	85	70	65	60	55	50

Продовження таблиці

$(X = x_i)$	7	8	9	10	11	12	13
$(Y = y_i)$	45	40	35	30	25	22	20

$X = x_p = 14.$

22.06

Залежність урожайності пшениці Y від глибини зволоження X наведено в таблиці:

$(X = x_i)$, см	0	5	8	10	12	14	16	18	20	22
$(Y = y_i)$, ц/га	10	12	14	16	18	20	22	24	26	28

Продовження таблиці

$(X = x_i)$, см	24	26	28	30	32	34	36	38	40	42
$(Y = y_i)$, ц/га	30	32	34	36	38	40	42	44	46	48

$X = x_p = 43.$

22.07

Показники товарообігу Y та суми витрат X, які досліджувалися в 19-ти магазинах, наведено в таблиці:

$(X = x_i)$, грн.	30	25	31	32	38	41	40	46	49	54
$(Y = y_i)$, грн.	480	510	530	540	555	564	570	575	580	585

Продовження таблиці

$(X = x_i)$,грн.	58	60	64	75	78	82	83	85	90
$(Y = y_i)$, грн.	590	596	605	618	625	635	640	650	660

$X = x_p = 91.$

22.08

Результати вимірювання чутливості Y відеоканалу та звукового каналу X наведено в таблиці:

$(X = x_i)$	170	180	200	230	240	250	280	300	310	320
$(Y = y_i)$	240	200	190	180	170	160	150	140	130	120

Продовження таблиці

$(X = x_i)$	330	350	380	400	410	420	430	440	450	460
$(Y = y_i)$	110	100	90	80	70	65	60	55	50	45

$X = x_p = 470.$

22.09

Залежність величини зносу різця Y від тривалості роботи X показано в таблиці:

$(X = x_i)$, год	6	7	8	9	10	11	12	13	14
$(Y = y_i)$, мм	30,0	29,1	28,4	28,1	28,0	27,7	27,5	27,2	27,0

Продовження таблиці

$(X = x_i)$, год	15	16	17	18	19	20	21	22	23
$(Y = y_i)$, мм	26,8	26,5	26,3	26,1	25,7	25,3	24,3	24,1	24,0

$X = x_p = 24.$

22.10

Залежність кров'яного тиску Y людини (в умовних одиницях) від довжини руки X наведена в таблиці:

$(X = x_i)$, см	62,1	61,1	61,0	60,5	60,0	59,0	58,5	58,0	57,5
$(Y = y_i)$, умов. од.	115	116	117	118	119	120	121	122	123

Продовження таблиці

$(X = x_i)$, см	56,5	56,0	55,5	55,0	54,5	54,0	53,5	53,0	52,5
$(Y = y_i)$, умов. од.	124	125	126	127	128	129	130	135	150

$X = x_p = 62{,}5.$

22.11

Залежність пружності Y сталевих болтів від змісту в них нікелю X наведена в таблиці:

$(X = x_i)$, %	2,20	2,35	2,42	2,58	2,65	2,69	2,74	2,88	2,91
$(Y = y_i)$, %	35,4	35,0	35,8	36,2	36,7	36,9	37,3	37,8	38,2

Продовження таблиці

$(X = x_i)$, %	2,95	2,99	3,00	3,11	3,21	3,29	3,34	3,44	3,50
$(Y = y_i)$, %	39,1	40,5	42,4	43,8	45,6	46,9	48,5	49,4	50,0

$X = x_p = 3{,}52.$

22.12

Результати порівняння нового методу газового аналізу зі старим X наведено в таблиці:

$(X = x_i)$, умов. од.	2,07	2,12	2,11	2,58	2,89	2,92	3,01	3,12
$(Y = y_i)$, умов. од.	2,88	2,91	2,92	2,96	3,01	3,11	3,21	3,25

Продовження таблиці

$(X = x_i)$, умов. од.	3,21	3,29	3,31	3,35	3,41	3,48	3,81
$(Y = y_i)$, умов.од.	3,32	3,36	3,42	3,46	3,58	3,88	4,12

$X = x_p = 3{,}9.$

22.13

Показники річної продуктивності праці в розрахунку на одного робітника Y і енергомісткості праці X на підприємствах однієї галузі наведено в таблиці:

$(X = x_i)$, кВт/робітн.	1,8	2,1	2,8	3,0	3,2	3,8	3,9	4,2	4,5	4,8
$(Y = y_i)$, тис. грн.	5,4	5,6	6,2	6,8	7,1	7,8	8,5	9,1	10,5	10,9

Продовження таблиці

$(X = x_i)$, кВт/робітн.	5,2	5,8	5,9	6,2	6,9	7,2	7,5	8,5	8,8	9,4
$(Y = y_i)$, тис. грн.	11,0	11,6	12,1	12,7	13,2	13,9	14,1	14,6	14,9	15,4

$X = x_p = 9{,}5.$

22.14

Залежність денного споживання масла Y певної особи від розміру її заробітної плати за місяць X наведено в таблиці:

$(X = x_i)$, грн.	70	75	82	89	95	100	105	110	115	120
$(Y = y_i)$, г.	10,5	15,8	17,8	19,5	20,4	21,5	22,2	24,3	25,8	26,5

Продовження таблиці

$(X = x_i)$, грн.	125	130	135	140	145	150	155	160	165	170
$(Y = y_i)$, г.	28,1	30,1	35,2	36,4	37,0	38,5	39,5	40,5	41,0	42,5

$X = x_p = 175.$

22.15

Залежність маси монети Y від часу її обігу в роках X наведено в таблиці:

$(X = x_i)$,років	4,0	5,0	5,5	6,0	6,8	7,5	8,5	10,8	12,0
$(Y = y_i)$, мг	9,35	9,21	9,18	9,50	9,10	9,08	9,05	9,01	9,00

Продовження таблиці

$(X = x_i)$,років	14,5	15,9	25,0	28,5	30,5	36,8	40,0	45,8	50,0
$(Y = y_i)$, мг	8,98	8,94	8,90	8,88	8,82	8,78	8,75	8,70	8,65

$X = x_p = 50{,}5.$

22.16

Залежність вмісту кремнію Y у чавуні від температури шлаку X наведено в таблиці:

$(X = x_i)$, °C	1330	1340	1350	1360	1370	1380	1390	1400	1410
$(Y = y_i)$, %	0,27	0,40	0,36	0,42	0,45	0,51	0,55	0,58	0,61

Продовження таблиці

$(X = x_i)$, °C	1420	1430	1440	1450	1460	1470	1480	1490	1500

$(Y = y_i)$, %	0,64	0,68	0,72	0,76	0,78	0,82	0,88	0,95	1,20

$X = x_p = 1510.$

22.17

Залежність урожайності Y пшениці від кількості внесених добрив X наведено в таблиці:

$(X = x_i)$, кг/га	10	30	40	50	60	70	80	90	100	110	120	130	140
$(Y = y_i)$, ц/га	10	12	14	16	18	20	22	23	26	28	30	32	34

$X = x_p = 150.$

22.18

Залежність міцності волокна бавовни в умовних одиницях Y від його товщини X наведено в таблиці:

$(X = x_i)$, мкм	0,41	0,48	0,56	0,66	0,72	0,79	0,85
$(Y = y_i)$, умов. од.	6.02	6,12	6,22	6,28	6,30	6,35	6,39

Продовження таблиці

$(X = x_i)$, мкм	0,86	0,88	0,92	0,94	0,96	0,98	0,99
$(Y = y_i)$, умов.од.	6,44	6,48	6,52	6,54	6,56	6,60	6,69

$X = x_p = 1{,}0.$

22.19

Залежність граничного навантаження на болт Y від його твердості X наведено в таблиці:

$(X = x_i)$, умов. од.	50,0	50,2	52,8	53,5	54,0	56,8	58,8	59,5	60,5
$(Y = y_i)$, умов. од.	10,10	10,30	10,45	10,90	11,20	11,35	11,90	12,45	12,58

Продовження таблиці

$(X = x_i)$, умов. од.	64,8	65,4	68,4	69,2	70,5	74,5	76,8	78,5	80,0
$(Y = y_i)$, умов. од.	12,96	13,44	13,60	13,95	14,50	14,98	15,48	15,96	16,50

$X = x_p = 80{,}5.$

22.20

Вплив температури Y в °С середовища на добовий хід хронометра X наведено в таблиці:

$(X = x_i)$	5,0	5,5	6,0	6,5	7,0	7,5	8,0	8,5	9,0	9,5	10,0
$(Y = y_i)$, °С	2,60	2,30	2,11	2,01	1,92	1,82	1,55	1,34	1,30	1,28	1,22

Продовження таблиці

$(X = x_i)$	10,5	11,0	11,5	12,0	12,5	13,0	14,0	18,0	24,0	30,0
$(Y = y_i)$, °С	1,18	1,12	1,10	0,98	0,92	0,90	0,89	0,88	0,80	0,79

$X = x_p = 30{,}5.$

22.21

Залежність вмісту свинцю Y в руді від вмісту X срібла наведено в таблиці:

$(X = x_i)$, %	2,0	7,5	12,5	14,5	16,0	18,5	20,0	20,5	22,0	24,5
$(Y = y_i)$, %	2,0	2,5	3,0	3,5	4,0	4,5	5,0	5,5	6,0	6,5

Продовження таблиці

$(X = x_i)$, %	26,0	28,5	30,0	32,5	34,0	36,5	38,0	40,5	42,0	45,0
$(Y = y_i)$, %	7,0	7,5	8,0	8,5	9,0	10,5	12,5	14,5	15,0	16,5

$X = x_p = 45{,}5.$

22.22

Залежність числа гризунів Y, які загинули від наявності отрути в їжі при концентрації X, наведено в таблиці:

$(X = x_i)$, %	3,0	3,5	4,0	4,5	5,0	6,0	6,5	7,0	7,5
$(Y = y_i)$,	32	36	38	42	46	49	55	59	62

Продовження таблиці

$(X = x_i)$, %	8,0	8,5	9,0	9,5	10,0	10,5	11,0	11,5	12,0
$(Y = y_i)$,	68	70	73	75	81	88	92	94	98

$X = x_p = 12{,}5.$

22.23

Залежність між собівартістю Y та кількістю виготовлених виробів X наведено в таблиці:

$(X = x_i)$, тис. шт.	1,5	1,4	1,2	1,1	0,9	0,8
$(Y = y_i)$, тис. грн.	2,2	3,5	3,7	3,8	4,5	5,7

$X = x_p = 1{,}7.$

22.24

Залежність урожайності пшениці Y від глибини оранки X наведено в таблиці:

$(X = x_i)$, см	8,1	8,3	8,2	9,1	10,3	10,8
$(Y = y_i)$, ц/га	7	8	9	10	11	12

$X = x_p = 11{,}0.$

22.25

Залежність кількості споживання масла на добу певною категорією пенсіонерів Y від розміру X отриманої місячної пенсії наведено в таблиці:

$(X = x_i)$, грн.	29	38	49	54	62	70	79	98
$(Y = y_i)$, г	15,99	19,75	23,10	26,44	29,79	33,13	36,89	44,54

$X = x_p = 99.$

22.26

Залежність вмісту срібла в руді Y від вмісту свинцю X наведено в таблиці:

$(X = x_i)$, %	2,5	7,5	12,5	17,5	22,5	27,5	32,5	37,5
$(Y = y_i)$, %	2	6	10	14	18	22	26	30

$X = x_p = 38.$

22.27

Залежність вмісту кремнію Y від температури X наведено в таблиці:

$(X = x_i)$, °C	1330	1340	1350	1360	1370	1380	1390	1400	1410
$(Y = y_i)$, %	0,27	0,26	0,27	0,28	0,29	0,3	0,31	0,32	0,33

$X = x_p = 1420.$

22.28

Залежність міцності волокна бавовни Y від граничного навантаження X наведено в таблиці:

$(X = x_i)$, г	3,75	4,25	4,75	5,25	5,75	6,25	6,75	7,00	7,25
$(Y = y_i)$, умов. од.	4100	4300	4500	4700	4900	5100	5200	5300	5500

$X = x_p = 7{,}3.$

22.29

Залежність собівартості Y від вирощеного врожаю соняшнику X наведено в таблиці:

$(X = x_i)$, ц/га	1,23	1,33	1,43	1,53	1,63	1,73	1,83	1,93
$(Y = y_i)$, грн/ц	10,36	11,56	13,29	14,51	15,6	14,25	17,36	16,23

$X = x_p = 1{,}95.$

22.30

Залежність урожайності цукрових буряків Y від кількості внесених у грунт поживних речовин X наведено в таблиці:

$(X = x_i)$, кг/га	83	92	112	132	144	154	162	189
$(Y = y_i)$, ц/га	369	380	370	395	420	412	436	420

$X = x_p = 190.$

ДОДАТОК А

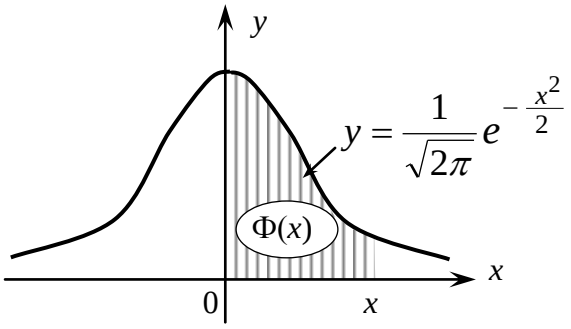
Таблиця значень функції Гауса $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

x	0	1	2	3	4	5	6	7	8	9
0,0	**0,399**	**399**	**399**	**399**	**399**	**398**	**398**	**398**	**398**	**397**
0,1	**397**	**397**	**396**	**396**	**395**	**395**	**394**	**393**	**393**	**392**
0,2	**391**	**390**	**389**	**389**	**388**	**387**	**386**	**385**	**384**	**383**
0,3	**381**	**380**	**379**	**378**	**377**	**375**	**374**	**373**	**371**	**370**
0,4	**368**	**367**	**365**	**364**	**362**	**361**	**359**	**357**	**356**	**354**
0,5	**352**	**350**	**349**	**347**	**345**	**343**	**341**	**339**	**337**	**335**
0,6	**333**	**331**	**329**	**327**	**325**	**323**	**321**	**319**	**317**	**314**
0,7	**312**	**310**	**308**	**306**	**303**	**301**	**299**	**297**	**294**	**292**
0,8	**290**	**287**	**285**	**283**	**280**	**278**	**276**	**273**	**271**	**269**
0,9	**266**	**264**	**261**	**259**	**257**	**254**	**252**	**249**	**247**	**244**
1,0	**242**	**240**	**237**	**235**	**232**	**230**	**228**	**225**	**223**	**220**
1,1	**218**	**216**	**213**	**211**	**208**	**206**	**204**	**201**	**199**	**197**
1,2	**194**	**192**	**190**	**187**	**185**	**183**	**180**	**178**	**176**	**174**
1,4	**150**	**148**	**146**	**144**	**142**	**140**	**137**	**135**	**133**	**132**
1,6	**111**	**109**	**107**	**106**	**104**	**102**	**101**	**099**	**097**	**096**
1,8	**079**	**078**	**076**	**075**	**073**	**072**	**071**	**069**	**068**	**067**
1,9	**066**	**064**	**063**	**062**	**061**	**059**	**058**	**057**	**056**	**055**
2,0	**054**	**053**	**052**	**051**	**050**	**049**	**048**	**047**	**046**	**045**
2,1	**044**	**043**	**042**	**041**	**040**	**040**	**039**	**038**	**037**	**036**
2,2	**036**	**035**	**034**	**033**	**033**	**032**	**031**	**030**	**030**	**029**
2,4	**022**	**022**	**021**	**021**	**020**	**020**	**019**	**019**	**018**	**018**
2,6	**014**	**013**	**013**	**013**	**012**	**012**	**012**	**011**	**011**	**011**
2,8	**008**	**008**	**008**	**007**	**007**	**007**	**007**	**007**	**006**	**006**
2,9	**006**	**006**	**006**	**006**	**005**	**005**	**005**	**005**	**005**	**005**
3,0	**004**	**004**	**004**	**004**	**004**	**004**	**004**	**004**	**004**	**003**
3,1	**003**	**003**	**003**	**003**	**003**	**003**	**003**	**003**	**003**	**003**
3,2	**002**	**002**	**002**	**002**	**002**	**002**	**002**	**002**	**002**	**002**
3,3	**002**	**002**	**002**	**002**	**002**	**002**	**001**	**001**	**001**	**001**
3,4	**001**	**001**	**001**	**001**	**001**	**001**	**001**	**001**	**001**	**001**
3,5	**001**	**001**	**001**	**001**	**001**	**001**	**001**	**001**	**001**	**001**
3,6	**001**	**001**	**001**	**001**	**001**	**001**	**001**	**001**	**001**	**000**

ДОДАТОК Б

Значення функції Лапласа

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt .$$



x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.0	0.00000				
0.05	0.01994	1.05	0.35314	2.05	0.47982
0.10	0.03983	1.10	0.36433	2.10	0.48214
0.15	0.05962	1.15	0.37493	2.15	0.48422
0.20	0.07926	1.20	0.38493	2.20	0.48610
0.25	0.09871	1.25	0.39435	2.25	0.48778
0.30	0.11791	1.30	0.40320	2.30	0.48928
0.35	0.13683	1.35	0.41149	2.35	0.49061
0.40	0.15542	1.40	0.41924	2.40	0.49180
0.45	0.17364	1.45	0.42647	2.45	0.49286
0.50	0.19146	1.50	0.43319	2.50	0.49379
0.55	0.20884	1.55	0.43943	2.55	0.49461
0.60	0.22575	1.60	0.44520	2.60	0.49534
0.65	0.24215	1.65	0.45053	2.65	0.49598
0.70	0.25804	1.70	0.45543	2.70	0.49653
0.75	0.27337	1.75	0.45994	2.75	0.49702
0.80	0.28814	1.80	0.46407	2.80	0.49744
0.85	0.30234	1.85	0.46784	2.85	0.49781
0.90	0.31594	1.90	0.47128	2.90	0.49813
0.95	0.32894	1.95	0.47441	2.95	0.49841
1.00	0.34134	2.00	0.47725	3.00	0.49865
3.1	0.49903	3.2	0.49931	3.3	0.49952
3.4	0.49966	3.5	0.49977	3.6	0.49984
3.7	0.49989	3.8	0.49993	3.9	0.49995
4.0	0.499968	4.5	0.499997	5.0	0.49999997

ДОДАТОК В

Критичні точки розподілення Стьюдента (t – розподіл)

Число ступенів свободи, k	Рівень значущості, α						
	0,20	0,10	0,05	0,02	0,01	0,002	0,001
1	3,08	6,31	12,7	31,82	63,66	127,32	636,62
2	1,89	2,92	4,30	6,97	9,93	14,09	31,60
3	1,64	2,35	3,18	4,54	5,84	7,45	12,94
4	1,53	2,13	2,78	3,75	4,60	5,60	8,61
5	1,48	2,02	2,57	3,37	4,03	4,77	6,86
6	1,44	1,94	2,45	3,14	3,71	4,32	5,96
7	1,42	1,90	2,36	3,00	3,50	4,03	5,41
8	1,40	1,86	2,31	2,90	3,36	3,83	5,04
9	1,38	1,83	2,26	2,82	3,25	3,69	4,78
10	1,37	1,81	2,23	2,76	3,17	3,58	4,59
11	1,36	1,80	2,20	2,72	3,11	3,50	4,44
12	1,36	1,78	2,18	2,68	3,05	3,43	4,32
13	1,35	1,77	2,16	2,65	3,01	3,37	4,22
14	1,34	1,76	2,14	2,62	2,98	3,33	4,14
15	1,34	1,75	2,13	2,60	2,95	3,29	4,07
16	1,34	1,75	2,12	2,58	2,92	3,25	4,02
17	1,33	1,74	2,11	2,57	2,90	3,22	3,97
18	1,33	1,73	2,10	2,55	2,88	3,20	3,92
19	1,33	1,73	2,09	2,54	2,86	3,17	3,88
20	1,33	1,73	2,09	2,53	2,85	3,15	3,85
21	1,32	1,72	2,08	2,52	2,83	3,14	3,82
22	1,32	1,72	2,07	2,51	2,82	3,12	3,79
23	1,32	1,71	2,07	2,50	2,81	3,10	3,77
24	1,32	1,71	2,06	2,49	2,80	3,09	3,75
25	1,32	1,71	2,96	2,48	2,79	3,08	3,73
26	1,32	1,71	2,06	2,48	2,78	3,07	3,71
27	1,31	1,70	2,05	2,47	2,77	3,06	3,69
28	1,31	1,70	2,05	2,47	2,76	3,05	3,67
29	1,31	1,70	2,04	2,46	2,76	3,04	3,66
30	1,31	1,70	2,04	2,46	2,75	3,03	3,65
40	1,30	1,68	2,02	2,42	2,70	2,97	3,55
60	1,30	1,67	2,00	2,39	2,66	2,91	3,46
120	1,29	1,66	1,98	2,36	2,62	2,86	3,37
∞	1,28	1,64	1,96	2,33	2,58	2,81	3,29

ДОДАТОК Г

Таблиця значень $t_\gamma(\gamma, k)$.

	γ												
k	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99	0,999
1	0,158	0,326	0,510	0,727	1,00	1,376	1,963	3,078	6,314	12,706	31,82	63,65	63,66
2	0,142	0,289	0,445	0,617	0,816	1,061	1,336	2,886	2,920	4,303	6,965	9,925	31,60
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	2,638	2,353	3,182	4,541	5,841	12,94
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,694	8,610
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,363	4,032	6,859
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,263	0,401	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,405
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,086	1,363	1,796	2,201	2,718	3,106	4,487
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,728	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,103	2,552	2,872	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,859	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,857	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0.855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646

БІБЛІОГРАФІЧНИЙ СПИСОК

1 Гмурман В.Е. Теория вероятностей и математическая статистика. –М.: Высш. шк., 1977.

2 Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. – М.: Высш. шк., 1977.

3 Маркович Э.С. Курс высшей математики с элементами теории вероятностей и математической статистики. – М.: Высш. шк., 1976.

4 Данко П.Е., Попов А.Г. Высшая математика в упражнениях и задачах. – Т.3. – Высш. шк., 1978.

5 Кремер Н.Ш.Теория вероятностей и математическая статистика. – М.: ЮНИТИ, 2000.

6 Жлуктенко В.І., Наконечний С.І., Савіна С.С. Ж76 Теорія ймовірностей і математична статистика . Навч. метод. Посібник: У 2-х ча-Ч. ІІ Математична статистика. –К.: КНЕУ, 2001

7 Кремер Н.Ш. Теория вероятностей и математическая статистика. Учебник для вузов. – М.: ЮНИТИДАНА, 2001.

8 Харченко А.П., Щелкунова Л.І. Теорія ймовірностей: Навч.-метод. посібник.Х: ХДТУБА,2007–144с.

9 Чорней Р.К. Практикум з теорії ймовірності та математичної статистики. Навч. посіб. для студентів вищ. навч. закл. – К.: МАУП, 2003

ЗМІСТ

Вступ…….. 3
1 Варіаційні ряди……..……..……..... 4
1.1 Статистичний розподіл вибірки……..... 5
1.2 Дискретний статистичний розподіл вибірки……. 9
1.3 Полігон частот і відносних частот……..……..... 10
1.4 Числові характеристики вибірки…….…… 16
1.5 Емпіричні моменти…….. 22
2 Статистичні оцінки параметрів генеральної сукупності……..... 35
2.1 Точкові статистичні оцінки параметрів…..... 35
2.2 Методи визначення точкових оцінок…….. 36
2.3 Закони розподілу ймовірностей для $\bar{x}_в$, S^2, S……. 40
2.4 Інтервальні статистичні оцінки параметрів генеральної сукупності……. 41
2.5 Побудова довірчого інтервалу для $\bar{x}_г$ при відомому значенні $\sigma_г$ із заданою надійністю γ……. 41
2.6 Побудова довірчого інтервалу для $\bar{x}_г$ при невідомому значенні $\sigma_г$ із заданою надійністю γ…….. 45
3 Статистична перевірка гіпотез……..……...…..…….. 52
3.1 Основні поняття…….…….. 52
3.2 Критерій перевірки. Критична область ……...……... 54
3.3 Загальна схема перевірки гіпотез ……..…… 56
3.4. Перевірка гіпотез про математичне сподівання нормально розподіленої випадкової величини при відомій дисперсії…….. 57
3.5 Перевірка гіпотез про математичне сподівання нормально розподіленої випадкової величини при невідомій дисперсії…….. 59
3.6 Перевірка гіпотез про величину дисперсії нормально розподіленої сукупності …… 60
3.7 Перевірка гіпотези про рівність математичних сподівань двох нормально розподілених сукупностей при відомих дисперсіях……. 62
3.8 Перевірка гіпотези про рівність математичних сподівань двох нормально розподілених сукупностей при невідомих дисперсіях……. 63
3.9 Перевірка гіпотези про рівність дисперсій двох нормально розподілених сукупностей…….. 65
4 Елементи кореляційно-регресійного аналізу. Парна регресія……. 70
Індивідуальні завдання…….…….. 82
Додаток А……..…….. 94
Додаток Б...…….. 95
Додаток В……...…….. 96
Додаток Г……... …….. 97
Бібліографічний список……..…...…… 98

Навчальне видання

АРШАВА Олена Олександрівна

ХАРЧЕНКО Анатолій Петрович

ЩЕЛКУНОВА Любов Іванівна

МАТЕМАТИЧНА СТАТИСТИКА

Навчально-методичний посібник

Роботу до видання рекомендував В.В. Шилін

Редактор Л.І. Христенко

План 2017 р., п. 30.18	Формат 60х84 1/16	
Підп. до друку 28.11.17	Обл.-вид. арк.4,8	
Надруковано на різографі	Ум. друк. арк5.0.	Папір друк. № 2
Тираж 50 прим	Зам. № 4861	Безкоштовно

ХНУБА, Україна, 61002, Харків, вул. Сумська,40

Підготовлено та надруковано РВВ Харківського національного університету будівництва та архітектури