МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Харківський державний університет харчування та торгівлі

МАТЕМАТИЧНІ МЕТОДИ В ТЕХНОЛОГІЇ

Методичні вказівки та індивідуальні контрольні завдання

Харків
ХДУХТ
2016

Методичні вказівки та індивідуальні контрольні завдання з курсу «Математичні методи в технології» [Електронний ресурс] / укладачі Л. О. Пархоменко, Д. О. Торяник. – Електрон. дані. – Х. : ХДУХТ, 2016. – 1 електрон. опт. диск (CD-ROM); 12 см. – Назва з тит. екрана.

Укладачі: Л. О. Пархоменко, ст. викл, Д. О. Торяник, доц.

Рецензент: д-р техн. наук, проф. М. С. Синєкоп

Кафедра фізико-математичних та інженерно-технічних дисциплін,
кафедра технологій харчування

Схвалено науково-методичною комісією ННІХТБ

Протокол від 15 грудня 2015 року № 3

Схвалено вченою радою ХДУХТ

Протокол від 25 лютого 2016 року № 7

Схвалено редакційно-видавничою радою ХДУХТ

Протокол від 24 лютого 2016 року № 3

Зміст

Передмова

Розв’язання багатьох технологічних задач потребує застосування методів математичної статистики. Статистичні розрахунки без допомоги ПЕОМ є складними і потребують застосування багатьох таблиць функцій стандартних розподілів та великої кількості обчислень. Спеціалізовані математичні пакети не можуть застосовуватись для навчання, оскільки потребують достатньо високого рівня підготовки з математичної статистики та знання основ програмування. Тому для засвоєння студентами дисципліни «Математичні методи в технологіях» пропонується використання універсального математичного пакету MathCAD Professional.

На кожному лабораторному занятті з дисципліни студент отримує індивідуальне завдання, яке виконує самостійно під керівництвом викладача. З метою кращого засвоювання матеріалу до кожної лабораторної роботи пропонується індивідуальне завдання, яке потрібно виконати до початку заняття без застосування комп’ютера. В «Методичних вказівках...» розглядаються необхідні теоретичні відомості з розділу «Елементи математичної статистики» та наводяться приклади розв’язання задач з дисципліни і варіанти індивідуальних завдань.

Отже, методичні вказівки сприяють отриманню навичок застосування елементів математичної статистики до розв’язання технологічних задач.

1. Варіаційні ряди та їх характеристики

1.1. Генеральна та вибіркова сукупності

Математична статистика – це розділ математики, присвячений вивченню закономірностей, які мають місце в масових явищах. Задачі математичної статистики полягають у розробці методів збору та обробки статистичних даних для отримання наукових та практичних висновків.

Нехай треба дослідити сукупність однорідних об’єктів відносно деякої якісної або кількісної ознаки. Сукупність предметів або явищ, що об’єднані спільною властивістю, називається *об’єктом спостереження*.

Сукупність називається *генеральною*, якщо дослідженню підлягають усі об’єкти сукупності.

Відібрана випадковим чином частина елементів генеральної сукупності, що підлягає дослідженню, називається *вибірковою сукупністю* або *вибіркою*.

Вибірки бувають *повторними* (відібраний об’єкт повертається до генеральної сукупності) та *безповторними* (об’єкт до сукупності не повертається).

Об’ємом сукупності (вибіркової або генеральної) називається кількість об’єктів сукупності.

Значення ознаки, що змінюється при переході від одного до іншого елемента вибірки, називається *варіантою* та позначається зазвичай малими латинськими літерами x, y, z тощо.

Ряд значень ознаки (ряд варіант), розташований у порядку зростання (спадання), називається *варіаційним рядом*. Варіаційні ряди бувають *дискретними* та *інтервальними*.

1.2. Дискретний варіаційний ряд

Дискретний варіаційний ряд зазвичай задається у вигляді таблиці – *статистичного розподілу частот*, що містить перелік точкових значень ознаки (варіант) та відповідних їм *частот*:

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

Частота n_i вказує на те, скільки разів значення ознаки (варіанта) x_i $(i=1,2,...,k)$ зустрілося у вибірковій сукупності.

Об'єм вибірки можна знайти, обчисливши суму всіх частот:

$$N = n_1 + n_2 + ... + n_k$$

або

$$N = \sum_{i=1}^{k} n_i . \qquad (1.1)$$

Також замість частот використовують відносні частоти. *Відносна частота* ω_i визначається відношенням частоти до суми всіх частот – об'єму вибірки:

$$\omega_i = \frac{n_i}{N} \quad (i=1,2,...,k) . \qquad (1.2)$$

При цьому

$$\omega_1 + \omega_2 + ... + \omega_k = 1$$

або

$$\sum_{i=1}^{k} \omega_i = 1. \qquad (1.3)$$

Статистичний розподіл відносних частот має наступний вигляд:

x_i	x_1	x_2	...	x_k
ω_i	ω_1	ω_2	...	ω_k

Різницю r між максимальним і мінімальним елементами вибірки називають *розмахом вибірки*:

$$r = x_{\max} - x_{\min}\,. \tag{1.4}$$

Приклад. В результаті експерименту отримано дані про вміст крохмалю в зразках рецептурної суміші (%):

0.1, 0.1, 0.2, 0.3, 0.2, 0.1, 0.2, 0.4, 0.1.

Побудувати дискретний варіаційний ряд, представляючи дані у вигляді статистичного розподілу частот та відносних частот; обчислити розмах вибірки.

Розв'язання. Для побудови варіаційного ряду розташуємо елементи вибірки в порядку зростання:

0.1, 0.1, 0.1, 0.1, 0.2, 0.2, 0.2, 0.3, 0.4.

Представимо дані у вигляді статистичного розподілу частот, визначивши варіанти та їх частоти:

x_i	0.1	0.2	0.3	0.4
n_i	4	3	1	1

Обчислимо об'єм вибірки за формулою (1.1):

$$N = \sum_{i=1}^{4} n_i = n_1 + n_2 + n_3 + n_4 = 4+3+1+1=9\,.$$

Визначимо відносні частоти за формулою (1.2):

$$\omega_1 = \frac{n_1}{N} = \frac{4}{9},\quad \omega_2 = \frac{n_2}{N} = \frac{3}{9} = \frac{1}{3},\quad \omega_3 = \frac{n_3}{N} = \frac{1}{9},\quad \omega_4 = \frac{n_4}{N} = \frac{1}{9}\,.$$

Контроль обчислень здійснюємо, використовуючи формулу (1.3):

$$\sum_{i=1}^{4} \omega_i = \omega_1 + \omega_2 + \omega_3 + \omega_4 = \frac{4}{9} + \frac{1}{3} + \frac{1}{9} + \frac{1}{9} = 1\,.$$

Представимо дані у вигляді статистичного розподілу відносних частот:

x_i	0.1	0.2	0.3	0.4
ω_i	4/9	1/3	1/9	1/9

Розмах вибірки обчислимо за формулою (1.4):

$$r = x_{\max} - x_{\min} = 0.4 - 0.1 = 0.3.$$

1.3. Інтервальний варіаційний ряд

Якщо об’єм вибірки великий, для спрощення обчислень елементи вибірки об’єднують в групи. Для цього інтервал, який містить всю множину елементів вибірки, розбивають на кілька інтервалів довжини h_i, які не перетинаються. Таке завдання вибірки називають *інтервальним варіаційним рядом*:

$[x_i; x_{i+1})$	$[x_1; x_2)$	$[x_2; x_3)$	...	$[x_k; x_{k+1}]$
n_i	n_1	n_2	...	n_k

(1.4)

Для визначеності покладаємо лівий кінець інтервалів закритим, а правий – відкритим (останній інтервал містить два закритих кінці). Частота n_i дорівнює кількості варіант, що потрапили у відповідний інтервал.

Якщо інтервали у варіаційних рядах мають однакову довжину, їх називають *рівновеликими*, у іншому випадку – *нерівновеликими*.

При побудові інтервального ряду з рівновеликими інтервалами оптимальну кількість інтервалів можна визначити, наприклад, за *формулою Стерджеса*

$$k = 1 + [3.322 \lg N],$$

де квадратними дужками позначено цілу частину отриманого числа, N – об’єм вибірки.

Довжину інтервалів рівновеликого інтервального ряду можна знайти за формулою

$$h = \frac{z}{k}, \tag{1.5}$$

де $k -$ заздалегідь задана або отримана за формулою Стерджеса кількість інтервалів, $z -$ розмах вибірки.

Зауважимо, що для забезпечення покриття інтервалами усіх варіант треба округлювати величину h в бік більших значень.

За визначеною довжиною інтервалу кінці рівновеликих інтервалів знаходяться наступним чином:

$$x_1 = x_{\min};$$
$$x_2 = x_1 + h;$$
$$x_3 = x_2 + h;$$
$$...$$
$$x_{k+1} = x_k + h.$$

Приклад. В результаті складання іспиту студентами групи в кількості 25 осіб отримано такі бали: 80, 61, 45, 50, 66, 76, 40, 60, 89, 96, 70, 57, 64, 70, 81, 60, 77, 80, 58, 88, 61, 64, 69, 60, 70. Побудувати інтервальний варіаційний ряд розподілу оцінок в групі.

Розв'язання. Розіб'ємо дані на 4 нерівновеликі інтервали, що відповідають оцінкам «незадовільно», «задовільно», «добре» та «відмінно» і визначимо частоти – кількість студентів, оцінки яких потрапили у відповідний інтервал:

$[x_i; x_{i+1})$	[40;60)	[60;74)	[74;90)	[90;100]
n_i	5	12	7	1

Контроль обчислень: $\sum_{i=1}^{4} n_i = 5+12+7+1=25$.

Таким чином, оцінку «незадовільно» мають 5 студентів, «задовільно» – 12 студентів, «добре» – 7 студентів та «відмінно» – 1 студент.

1.4. Графічні характеристики вибіркових даних

Дискретний варіаційний ряд

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

можна представити графічно у вигляді *полігону частот*. *Полігоном частот* називають ламану лінію, вершинами якої є точки з координатами $(x_1;n_1),(x_2;n_2),...,(x_k;n_k)$.

Інтервальний варіаційний ряд

$[x_i; x_{i+1})$	$[x_1; x_2)$	$[x_2; x_3)$	...	$[x_k; x_{k+1}]$
n_i	n_1	n_2	...	n_k

можна представити графічно у вигляді *гістограми частот* – ступінчатої фігури, яка складається з прямокутників, основами яких є інтервали варіант довжини $h_i = x_{i+1} - x_i$, а висоти дорівнюють відношенню $\frac{n_i}{h_i}$ (щільність частоти). Площа i-го прямокутника гістограми частот дорівнює сумі частот варіант i-го інтервалу, отже, площа гістограми частот дорівнює сумі усіх частот, тобто об'єму вибірки. Інтервальний варіаційний ряд можна зобразити й у вигляді полігону. Для цього середини верхніх основ прямокутників з яких складається гістограма та дві точки на осі абсцис, що відповідають серединам попереднього перед першим та наступного за останнім інтервалів, з'єднують відрізками прямих ліній. Площа під отриманою ламаною дорівнює площі гістограми.

Також замість полігона і гістограми частот будують *полігон і гістограму відносних частот*, побудова яких відрізняється лише тим, що на осі ординат полігона частот відкладається відносна частота (1.2), а висоти прямокутників гістограми будуть дорівнювати відповідно $\frac{\omega_i}{h_i}$ (щільність відносної частоти). Гістограма відносних частот є статистичним наближенням щільності розподілу $f(x)$ ознаки генеральної сукупності. Площа гістограми відносних частот буде дорівнювати одиниці.

Приклад. В результаті опиту 50 працівників було встановлено кількісний склад їх родин: 5; 3; 2; 1; 4; 6; 3; 7; 9; 1; 3; 2; 5; 6; 8; 2; 5; 2; 3; 6; 8; 3; 4; 4; 5; 6; 5; 4; 7; 5; 6; 4; 8; 7; 4; 5; 7; 8; 6; 5; 7; 5; 6; 6; 7; 3; 4; 6; 5; 4.

1) Представити дані у вигляді статистичного розподілу частот та у вигляді інтервального ряду в кількості 4 інтервалів.
2) Побудувати полігон частот для дискретного ряду та гістограму частот для інтервального ряду.

Розв'язання.

1) Значення ознаки, що досліджується – кількісний склад родини – позначимо через x_i, відповідні частоти – через n_i. Занесемо дані в таблицю та отримаємо дискретний варіаційний ряд:

x_i	1	2	3	4	5	6	7	8	9
n_i	2	4	6	8	10	9	6	4	1

Розіб'ємо дані на $k = 4$ рівновеликі інтервали. Для знаходження довжини інтервалів використаємо формулу (1.5).

$$z = x_{\max} - x_{\min} = 9 - 1 = 8 \ ;$$

$$h = \frac{z}{k} = \frac{8}{4} = 2 .$$

Знайдемо кінці інтервалів:

$$x_1 = x_{\min} = 1;$$

$$x_2 = x_1 + h = 1 + 2 = 3;$$

$$x_3 = x_2 + h = 3 + 2 = 5;$$
$$x_4 = x_3 + h = 5 + 2 = 7;$$
$$x_5 = x_4 + h = 7 + 2 = 9.$$

Інтервальний ряд має наступний вигляд:

$[x_i; x_{i+1})$	[1;3)	[3;5)	[5;7)	[7;9]
n_i	6	14	19	11

Контроль обчислень: $\sum_{i=1}^{4} n_i = 6 + 14 + 19 + 11 = 50$.

2) Побудуємо полігон частот для дискретного ряду та гістограму частот для інтервального ряду. Полігон частот має вигляд:

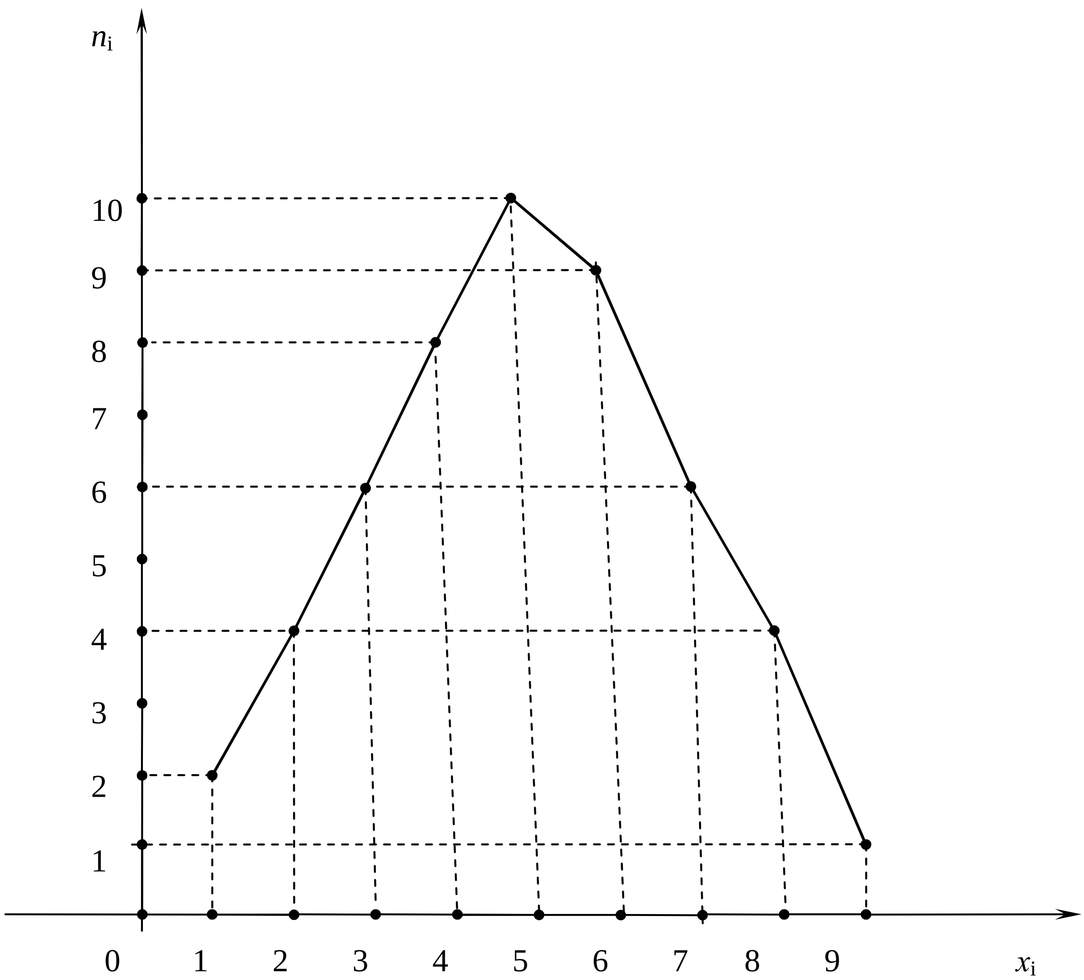

Рис.1. Полігон частот

Для побудови гістограми знайдемо щільності частот $\frac{n_i}{h}$:

$[x_i; x_{i+1})$	[1;3)	[3;5)	[5;7)	[7;9]
n_i	6	14	19	11
$\frac{n_i}{h}$	3	7	9,5	5,5

Тоді гістограма має вигляд:

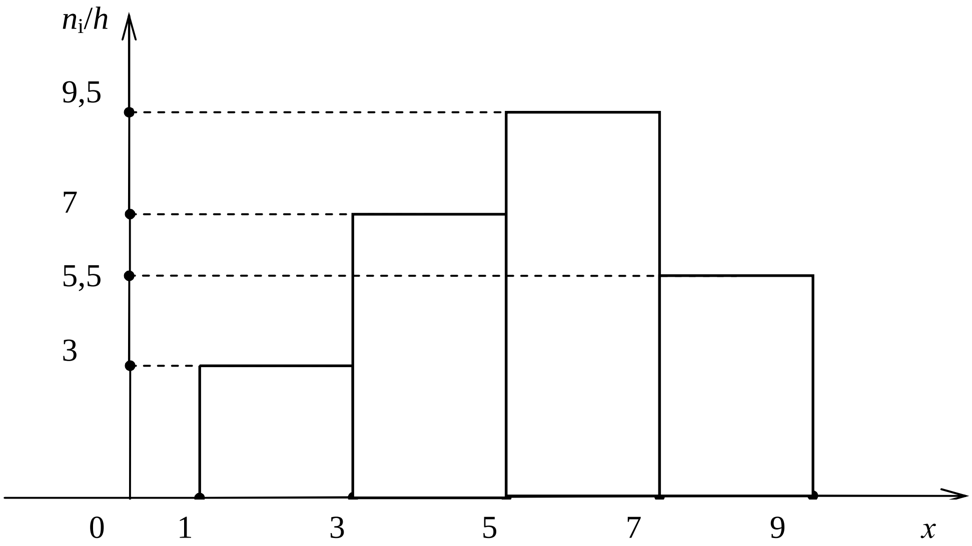

Рис.2. Гістограма частот

Розглянемо статистичний аналог функції розподілу дискретної випадкової величини – *емпіричну функцію розподілу*.

Нехай для дослідження отримано вибіркові дані x_1, x_2, ..., x_k. Визначимо емпіричну функцію розподілу наступним чином:

$$F^*(x) = \frac{n_x}{N}.$$

Тут n_x − кількість елементів x_i, що задовольняють нерівності $x_i < x$, N – об'єм вибірки.

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

1) $0 \le F^*(x) \le 1$;
2) $F^*(x)$ є неспадною кусково-сталою функцією;
3) $F^*(x) = \begin{cases} 0, & x \le x_{\min}; \\ 1, & x > x_{\max}. \end{cases}$ де $x_{\min}$ – найменший, $x_{\max}$ – найбільший елемент.

Приклад. Отримано дані про вміст хлориду натрію (%) у зразках рецептурної суміші: 0.1, 0.2, 0.3, 0.2, 0.5, 0.5, 0.3, 0.3, 0.2, 0.2. Визначити емпіричну функцію розподілу та побудувати її графік.

Розв'язання. Представимо вибіркові дані у вигляді статистичного розподілу частот:

x_i	0.1	0.2	0.3	0.5
n_i	1	4	3	2

Знайдемо об'єм вибірки: $N = \sum_{i=1}^{4} n_i = 1 + 4 + 3 + 2 = 10$.

Визначимо емпіричну функцію розподілу $F^*(x)$:

коли $x \le 0.1$, $F^*(x) = 0$;

коли $0.1 < x \le 0.2$, поява значень, менших за x, спостерігалась при появі значення $x_1 = 0.1$ $n_1 = 1$ раз, значить, в цьому інтервалі $F^*(x) = \frac{n_1}{N} = \frac{1}{10}$;

коли $0.2 < x \le 0.3$, поява значень, менших за x, спостерігалась при появі $x_1 = 0.1$ та $x_2 = 0.2$ в кількості $n_1 + n_2 = 1 + 4 = 5$ разів, значить, в цьому інтервалі $F^*(x) = \frac{n_1 + n_2}{n} = \frac{5}{10} = \frac{1}{2}$;

коли $0.3 < x \le 0.5$, поява значень, менших за x, спостерігалась при появі $x_1 = 0.1$, $x_2 = 0.2$ та $x_3 = 0.3$ в кількості $n_1 + n_2 + n_3 = 1 + 4 + 3 = 8$ разів, значить, в цьому інтервалі $F^*(x) = \frac{n_1 + n_2 + n_3}{n} = \frac{8}{10} = \frac{4}{5}$;

коли $x > 0.5$, $F^*(x) = 1$.

Отримали таку емпіричну функцію розподілу вибіркових даних:

$$F^*(x) = \begin{cases} 0, & x \le 0.1 \\ \frac{1}{10}, & 0.1 < x \le 0.2 \\ \frac{1}{2}, & 0.2 < x \le 0.3 \\ \frac{4}{5}, & 0.3 < x \le 0.5 \\ 1, & x > 0.5 \end{cases}$$

графік якої має вигляд.

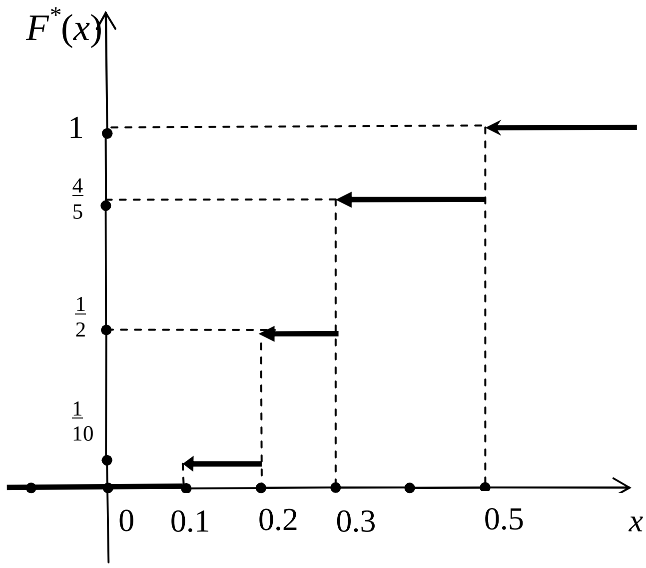

Рис. 3. Емпірична функція розподілу

2. Статистичні оцінки параметрів розподілу

2.1. Точкові оцінки параметрів розподілу

При розв'язанні багатьох технологічних задач виникає потреба дослідити кількісну ознаку X генеральної сукупності, використовуючи вибіркові дані. Іноді, з деяких міркувань, вдається встановити закон розподілу випадкової величини X. Тоді виникає задача оцінювання параметрів цього закону розподілу − математичного сподівання $M(X)$, дисперсії $D(X)$, середнього квадратичного відхилення $\sigma(X)$.

Для того, щоб шукані статистичні оцінки давали достатньо гарні наближення параметрів, що оцінюються, вони повинні задовольняти таким вимогам:

- *незсунутості* (тобто математичне сподівання такої оцінки повинно бути рівним оцінюваному параметру за будь-якого об'єму вибірки);
- *ефективності* (тобто при заданому об'єму вибірки N така оцінка повинна мати найменшу з можливих дисперсій);
- *спроможності* (тобто якщо $n \to \infty$, статистична оцінка за ймовірністю повинна прямувати до оцінюваного параметра).

Якщо досліджувану ознаку X генеральної сукупності розглядати як випадкову величину, то незсунутою оцінкою її математичного сподівання $M(X)$ є *вибіркове середнє значення* $\bar{x}$, зсунутою оцінкою дисперсії $D(X)$ – *вибіркова дисперсія* D, незсунутою оцінкою дисперсії $D(X)$ – *«виправлена»* *вибіркова дисперсія* $\widehat{D}$. Для оцінки середнього квадратичного відхилення $\sigma(X)$ ознаки генеральної сукупності використовують *вибіркове середнє квадратичне відхилення* σ та *«виправлене» вибіркове середнє квадратичне відхилення* S.

В залежності від того, в якому вигляді представлено вибіркові дані, використовуються різні формули обчислення оцінок параметрів розподілу.

1. Вибіркові дані представлено у вигляді негрупованого дискретного варіаційного ряду $x_1, x_2, ..., x_N$.

В цьому випадку

$$\bar{x}_в = \frac{1}{N}\sum_{i=1}^{N} x_i \quad \text{− вибіркове середнє значення} \qquad (2.1)$$

$$D = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

або − вибіркова дисперсія (2.2)

$$D = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \bar{x}^2$$

$$S^2 = \frac{N}{N-1}D$$

або − *«виправлена»* вибіркова дисперсія; (2.3)

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

$\sigma = \sqrt{D}$ − вибіркове середнє квадратичне відхилення;

$S = \sqrt{S^2}$ − «виправлене» вибіркове середнє квадратичне відхилення.

2. Вибіркові дані представлено у вигляді статистичного розподілу частот

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

В цьому випадку

$$\bar{x}_в = \frac{1}{N}\sum_{i=1}^{k} x_i n_i \quad \text{− вибіркове середнє значення;} \qquad (2.4)$$

$$D = \frac{1}{N}\sum_{i=1}^{k}(x_i - \overline{x})^2 n_i$$

або − вибіркова дисперсія; (2.5)

$$D = \frac{1}{N}\sum_{i=1}^{k} x_i^2 n_i - \overline{x}^2$$

$$S^2 = \frac{N}{N-1} D$$

або − «*виправлена*» вибіркова дисперсія; (2.6)

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{k}(x_i - \overline{x})^2 n_i$$

$\sigma = \sqrt{D}$ − вибіркове середнє квадратичне відхилення;

$S = \sqrt{S^2}$ − «виправлене» вибіркове середнє квадратичне відхилення.

3. Вибіркові дані представлено у вигляді інтервального ряду

$[x_i; x_{i+1})$	$[x_1; x_2)$	$[x_2; x_3)$	...	$[x_k; x_{k+1}]$
n_i	n_1	n_2	...	n_k

Для знаходження точкових оцінок параметрів розподілу обчислюються середини інтервалів x_i^* $(i = 1, 2, ..., k)$ за формулою

$$x_i^* = \frac{x_i + x_{i+1}}{2}.$$

Статистичні оцінки параметрів розподілу набувають вигляду

$$\overline{x}_в = \frac{1}{N}\sum_{i=1}^{k} x_i^* n_i$$ − вибіркове середнє значення; (2.7)

$$D = \frac{1}{N}\sum_{i=1}^{k}(x_i^* - \overline{x})^2 n_i$$

або − вибіркова дисперсія; (2.8)

$$D = \frac{1}{N}\sum_{i=1}^{k}(x_i^*)^2 n_i - \overline{x}^2$$

$$S^2 = \frac{N}{N-1}D$$

або − *«виправлена»* вибіркова дисперсія;

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{k}(x_i^* - \overline{x})^2 n_i \ .$$

Середнє квадратичне відхилення та «виправлене» середнє квадратичне відхилення обчислюються аналогічно до попередніх випадків.

В статистичному аналізі широко застосовують такі точкові оцінки, як медіана і мода. Величина моди та медіани залежить тільки від характеру частот, тобто від структури розподілу.

Модою M_0 називається значення ознаки, яке має найбільшу частоту в статистичному розподілі. У дискретних варіаційних рядах мода визначається без додаткових розрахунків за значенням варіанти, що має найбільшу частоту. В інтервальних варіаційних рядах мода визначається за формулою:

$$M_0 = x_0 + h\frac{n_2 - n_1}{(n_2 - n_1) + (n_2 - n_3)},$$

де x_0 – нижня (мінімальна) межа модального інтервалу (інтервалу з найбільшою частотою); h – величина інтервалу; n_1 – частота інтервалу, попереднього до модального; n_2 – частота модального інтервалу; n_3 – частота інтервалу, наступного після модального.

Медіаною називається таке значення ознаки, що поділяє варіаційний ряд на дві рівні частини, тобто це значення, яке знаходиться у середині ряду розподілу. Якщо в дискретному варіаційному ряді $2m+1$ елементів, то значення ознаки для елементу $m+1$ буде медіанним:

$$M_e = x_{m+1}.$$

Якщо в ряді парна кількість $2m$ елементів, медіану визначають як середню арифметичну величину з двох серединних значень:

$$M_e = \frac{x_m + x_{m+1}}{2}.$$

В інтервальних варіаційних рядах розподілу медіана визначається за формулою:

$$M_e = x_{Me} + h\frac{0.5N - S_{Me-1}}{n_{Me}},$$

де x_{Me} – нижня (мінімальна) межа медіанного інтервалу; h – величина інтервалу; n – об'єм сукупності; S_{Me-1}– накопичена сума частот варіант інтервалу, попереднього до медіанного; n_{Me} – сума частот варіант медіанного інтервалу.

Приклад. При дослідженні партії кулінарних виробів на вміст жиру відібрано 10 зразків. Отримано наступні дані:

x_i	10	12	13	14	15
n_i	1	3	2	1	3

Тут x_i(%)− вміст жиру в i-му зразку, n_i− кількість відповідних зразків.

Обчислити:

1) середнє значення вмісту жиру у відібраних зразках;
2) вибіркову дисперсію та «виправлену» вибіркову дисперсію;
3) вибіркове середнє квадратичне відхилення та «виправлене» вибіркове середнє квадратичне відхилення;
4) моду та медіану.

Розв'язання.

1) Вибіркове середнє значення обчислимо за формулою (2.4):

$$\bar{x}_в = \frac{1}{N}\sum_{i=1}^{k} x_i n_i = \frac{1}{10}(10\cdot 1 + 12\cdot 3 + 13\cdot 2 + 14\cdot 1 + 15\cdot 3) = 13.1.$$

2) Для обчислення вибіркової дисперсії використаємо другу з формул (2.5):

$$D = \frac{1}{N}\sum_{i=1}^{k} x_i^2 n_i - \bar{x}^2 = \frac{1}{10}\left(10^2\cdot 1 + 12^2\cdot 3 + 13^2\cdot 2 + 14^2\cdot 1 + 15^2\cdot 3\right) - (13{,}1)^2 = 2.5;$$

3) «Виправлену» вибіркову дисперсію знайдемо за першою з формул (2.6):

$$\mathrm{S}^2 = \frac{N}{N-1}D = \frac{10}{9}\cdot 2.5 = 2.8.$$

4) Вибіркове середнє квадратичне відхилення:

$$\sigma = \sqrt{D} = \sqrt{2.5} = 1.6.$$

5)«Виправлене» вибіркове середнє квадратичне відхилення:

$$S = \sqrt{S^2} = \sqrt{2.8} = 1.7.$$

6)Маємо дві варіанти з найбільшою частотою, тому

$$M_{01} = 12,\; M_{02} = 15.$$

Оскільки об'єм вибірки – число парне ($N = 10 = 2 \cdot 5$), маємо:

$$M_e = \frac{x_5 + x_6}{2} = \frac{13+13}{2} = 13.$$

Приклад. Наведено результати вимірювань зросту (в см) випадково відібраних 100 студентів.

Зріст	154–158	158–162	162–166	166–170	170–174	174–178	178–182
Число студентів	10	14	26	28	12	8	2

Знайти середній зріст та вибіркову дисперсію зросту обстежуваних студентів.

Розв'язання. Знайдемо вибіркове середнє значення за формулою (2.7) та вибіркову дисперсію за першою з формул (2.8). Для цього визначимо середини інтервалів x_i^*:

x_i^*	156	160	164	168	172	176	180

Вибіркове середнє значення:

$$\bar{x}_в = \frac{1}{N}\sum_{i=1}^{k} x_i^* n_i =$$

$$= \frac{1}{100}\left(156 \cdot 10 + 160 \cdot 14 + 164 \cdot 26 + 168 \cdot 28 + 172 \cdot 12 + 176 \cdot 8 + 180 \cdot 2\right) = 166 \text{ (см)};$$

Вибіркова дисперсія:

$$D = \frac{1}{N}\sum_{i=1}^{k}(x_i^* - \bar{x})^2 n_i = \frac{1}{100}((156-166)^2 \cdot 10 + (160-166)^2 \cdot 14 +$$

$$+ (164-166)^2 \cdot 26 + (168-166)^2 \cdot 28 + (172-166)^2 \cdot 12 +$$

$$+ (176-166)^2 \cdot 8 + (180-166)^2 \cdot 2) = 33.44 \text{ (см}^2\text{)}.$$

Відповідь: середній зріст обстежуваних студентів складає 166 см, дисперсія – 33.44 см^2.

2.2. Інтервальні статистичні оцінки параметрів розподілу

Середнє вибіркове значення, вибіркова дисперсія і вибіркове середнє квадратичне відхилення є точковими оцінками параметрів розподілу ознаки генеральної сукупності. Точкова оцінка залежить від об'єму вибірки і може сильно відрізнятися від дійсного значення параметру, що оцінюється, тобто може привести до грубих помилок. Тому при дослідженні вибірок невеликого об'єму виникає необхідність використовувати так звані *інтервальні оцінки*.

Розглянемо точкову оцінку математичного сподівання a (або, що теж саме, генерального середнього значення) деякої кількісної ознаки X генеральної сукупності – вибіркове середнє значення $\bar{x}_в$. Характеристикою *точності оцінки* будемо вважати величину δ, потрібну для виконання нерівності $|\bar{x}_в - a| < \delta$. Внаслідок випадковості того, що той чи інший об'єкт потрапляє до вибіркової сукупності, говорити про виконання останньої нерівності можна лише з деякою ймовірністю γ, яка називається *надійністю* (*надійною або довірчою ймовірністю*) оцінки, тобто

$$P(|\bar{x}_в - a| < \delta) = \gamma \text{ або } P(\bar{x}_в - \delta < a < \bar{x}_в + \delta) = \gamma .$$

Це означає, що невідоме математичне сподівання a розподілу ознаки генеральної сукупності з імовірністю γ покривається інтервалом $(\bar{x}_в - \delta;\ \bar{x}_в + \delta)$, який називається *надійним інтервалом* або *довірчим інтервалом* для оцінки невідомого математичного сподівання**.**

Особливу цікавість викликає надійний інтервал для оцінки математичного сподівання кількісної ознаки X генеральної сукупності, розподіленої за нормальним законом. Можливі наступні випадки:

1. Якщо заздалегідь відома величина середнього квадратичного відхилення σ, то межі надійного інтервалу для оцінки математичного сподівання a мають вигляд:

$$\left(\bar{x}_в - t\frac{\sigma}{\sqrt{N}};\ \bar{x}_в + t\frac{\sigma}{\sqrt{N}} \right), \tag{2.9}$$

де $\bar{x}_в$ – середнє вибіркове значення; N – об'єм вибірки; σ – відоме середнє квадратичне відхилення розподілу ознаки генеральної сукупності; t – величина, що визначається за таблицею значень функції Лапласа $\Phi(t) = \frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{t^2}{2}}dt$ (*таблиця 1*) із співвідношення

$$\Phi(t) = \frac{\gamma}{2}, \tag{2.10}$$

де γ – заздалегідь обрана надійна ймовірність.

2. Якщо середнє квадратичне відхилення σ досліджуваної ознаки заздалегідь невідомо, то використовується його вибіркова оцінка S – «виправлене» середнє квадратичне відхилення, яке знаходять за даними вибірки. У цьому випадку надійний інтервал для оцінки математичного сподівання a має вигляд

$$\left(\bar{x}_в - t_\gamma \frac{S}{\sqrt{N}};\ \bar{x}_в + t_\gamma \frac{S}{\sqrt{N}} \right), \tag{2.11}$$

де $\bar{x}_в$ – вибіркове середнє значення; S – «виправлене» вибіркове середнє квадратичне відхилення; N – об'єм вибірки; t_γ – величина, що визначається за таблицею розподілу Ст'юдента (*таблиця 2*) для $N-1$ ступенів свободи та надійної ймовірності γ.

Зауважимо, що у деяких таблицях замість надійної ймовірності γ розглядається рівень значущості $\alpha = 1-\gamma$.

Приклад. Побудувати з надійністю $\gamma = 0{,}99$ довірчий інтервал для оцінки середнього вмісту білка (%) у партії хлібобулочних виробів, якщо:

1) середнє квадратичне відхилення вмісту білка усієї партії складає 4 %, середній вміст білка у досліджуваних 16 зразках складає 10.2 %.

2) для досліджування відібрано зразки та визначено вміст білка (%) в них: 7, 13, 8, 10, 12, 9, 11, 12, 8.

Вміст білка у виробах вважається нормально розподіленою випадковою величиною.

Розв'язання.

1) За умовою $\sigma = 4$, $\bar{x}_в = 10.2$, $N = 16$. Надійний інтервал для оцінки середнього вмісту білка знаходиться за формулою (2.9). Знайдемо невідому величну t з рівняння (2.10):

$$\Phi(t) = \frac{0.99}{2} = 0.495,$$

з *таблиці 1* знаходимо

$$t = 2.57.$$

Будуємо надійний інтервал:

$$10.2 - 2.57\frac{4}{\sqrt{16}} < a < 10.2 + 2.57\frac{4}{\sqrt{16}}$$

або

$$7.63 < a < 12.77.$$

2) Для побудови надійного інтервалу скористаємося формулою (2.11). Для цього визначимо:

- об'єм вибірки $N = 9$;
- вибіркове середнє значення за формулою (2.1):

$$\bar{x}_в = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{9}(7+13+8+10+12+9+11+12+8) = 10;$$

- «виправлене» вибіркове середнє квадратичне відхилення:

$$S = \sqrt{S^2};$$

$$\mathrm{S}^2 = \frac{1}{9-1}\Big((7-10)^2 + (13-10)^2 + (8-10)^2 \cdot 2 + (10-10)^2 + (12-10)^2 \cdot 2 +$$
$$+ (9-10)^2 + (11-10)^2\Big) = 4.5;$$

$$S = \sqrt{S^2} = \sqrt{4.5} = 2.12;$$

- значення t_γ з таблиці розподілу Ст'юдента з 9–1= 8 ступенями свободи та довірчою ймовірністю $\gamma = 0.99$:

$$t_\gamma = 3.36.$$

Маємо:

$$10 - 3.36 \cdot \frac{2.12}{\sqrt{9}} < a < 10 + 3.36 \cdot \frac{2.12}{\sqrt{9}}$$

або

$$7.63 < a < 12.37.$$

Відповідь: середній вміст білка (%) партії хлібобулочних виробів з імовірністю 0.99 покривається:

1) інтервалом (7.63; 12.77);
2) інтервалом (7.63; 12.37).

3. Елементи теорії кореляції

Функціональною залежністю називається такий зв'язок між змінними величинами, коли залежна величина (функція) повністю визначається значеннями незалежних величин (аргументів). Вид залежності між аргументами і функцією, як правило, задається у вигляді формули, яка дає можливість однозначно обчислити значення функції, якщо підставити у формулу значення аргументів.

Статистичною називають залежність, коли зміна однієї з величин викликає зміну розподілу іншої. Якщо при змінюванні однієї з величин змінюється середнє значення іншої, має місце *кореляційна залежність*. Якщо за умовну середню прийняти середнє арифметичне значень ознаки Y, що відповідають значенню x ознаки X $\bar{y}_x = \frac{y_1 + y_2 + ... + y_k}{k}$, то *кореляційною залежністю* Y від X називають функціональну залежність умовної середньої $\bar{y}_x$ від x: $\bar{y}_x = f(x)$. Таке рівняння називається *рівнянням регресії* Y на X; функція $f(x)$ називається *функцією регресії* Y на X, а її графік – *лінією регресії* Y на X.

Аналогічно регресії Y на X може бути визначена регресія X на Y з рівнянням $\bar{x}_y = \Phi(y)$ і графіком, що відрізняється від графіка $\bar{y}_x = f(x)$.

Існують дві основні задачі теорії кореляції: встановити вид кореляційного зв'язку (лінійний, квадратичний тощо) та оцінити силу кореляційного зв'язку.

На практиці часто кореляційний зв'язок між ознаками генеральної сукупності виявляється лінійним. У цьому випадку функція регресії є лінійною функцією $\bar{y}_x = kx + b$, а її графіком є пряма лінія (пряма регресії).

Кутовий коефіцієнт прямої лінії регресії Y на X називають *вибірковим коефіцієнтом регресії* і позначають через ρ_{yx}.

Класичним методом знаходження параметрів ρ_{yx} та b є метод найменших квадратів (МНК). В результаті отримаємо формули:

$$\rho_{yx} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x^2} \tag{3.1}$$

або

$$\rho_{yx} = \frac{n\sum xy - \sum x \cdot \sum y}{n\sum x^2 - \left(\sum x\right)^2},$$

$$b = \bar{y} - \rho_{yx}\bar{x} \tag{3.2}$$

або

$$b = \frac{\sum x^2 \sum y - \sum x \cdot \sum xy}{n\sum x^2 - \left(\sum x\right)^2}.$$

Сила лінійної кореляційної залежності між ознаками генеральної сукупності X та Y характеризується *вибірковим коефіцієнтом кореляції* r, який вказує на ступінь лінійної залежності між ознаками. Його величина визначається за формулою

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} \tag{3.3}$$

або

$$r = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sigma_x \sigma_y}.$$

Якщо відомий вибірковий коефіцієнт кореляції r, вибірковий коефіцієнт регресії ρ_{yx} можна знайти за формулою

$$\rho_{yx} = r\frac{\sigma_y}{\sigma_x}. \tag{3.4}$$

Слід пам'ятати *властивості коефіцієнта кореляції*:

1. $-1 \le r \le 1$.
2. Якщо $r = 0$, то лінійний кореляційний зв'язок відсутній, при цьому нелінійний кореляційний зв'язок є можливим.
3. Якщо абсолютна величина коефіцієнта кореляції дорівнює одиниці, тобто $|r| = 1$, то між X і Y існує функціональний лінійний зв'язок.

Для оцінки тісноти кореляційного зв'язку між X і Y можна користуватись таблицею Чеддока.

Діапазон змін $\lvert r\rvert$	0.1−0.3	0.3−0.5	0.5−0.7	0.7−0.9	0.9−0.99
Характер тісноти зв'язку	дуже слабкий	слабкий	середній	високий	дуже високий

Приклад. Досліджується залежність піноутворюючої здатності Y (%) молочної суміші від вмісту крохмалю X (%).

X	0.8	1	1.5	1	0.5	0.7	1
Y	100	100	120	110	80	90	90

Оцінити кореляційний зв'язок між ознаками X і Y, побудувати рівняння лінійної регресії Y на X.

Розв'язання. Об'єм вибірок $N_X = N_Y = N = 7$. Обчислимо коефіцієнт кореляції за формулою (3.3).

Вибіркове середнє значення для вибірки X:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{7} x_i = \frac{1}{7}(0.8+1+1.5+1+0.5+0.7+1) = 0.93.$$

Вибіркове середнє значення для вибірки Y:

$$\overline{y} = \frac{1}{N}\sum_{i=1}^{7} y_i = \frac{1}{7}\left(100+100+120+110+80+90+90\right) = 98.6.$$

Вибіркове середнє значення для вибірки XY:

$$\overline{xy} = \frac{1}{N}\sum_{i=1}^{7} x_i y_i =$$

$$\frac{1}{7}\left(0.8\cdot 100+1\cdot 100+1.5\cdot 120+1\cdot 110+0.5\cdot 80+0.7\cdot 90+1\cdot 90\right) = 94.7.$$

Середнє квадратичне відхилення для вибірки X:

$$\sigma_X = \sqrt{\frac{1}{N}\sum_{i=1}^{7}(x_i-\overline{x})^2} =$$

$$=\sqrt{\frac{1}{7}\left((0.8-0.93)^2+(1-0.93)^2+(1.5-0.93)^2+(1-0.93)^2+(0.5-0.93)^2+(0.7-0.93)^2+(1-0.93)^2\right)} =$$

$$=0.29.$$

Середнє квадратичне відхилення для вибірки Y:

$$\sigma_Y = \sqrt{\frac{1}{N}\sum_{i=1}^{7}(y_i-\overline{y})^2} =$$

$$=\sqrt{\frac{1}{7}\left((100-98.6)^2+(100-98.6)^2+(120-98.6)^2+(110-98.6)^2+(80-98.6)^2+(90-98.6)^2+(90-98.6)^2\right)} =$$

$$=12.45.$$

Коефіцієнт кореляції:

$$r = \frac{\overline{xy}-\overline{x}\cdot\overline{y}}{\sigma_X\cdot\sigma_Y} = \frac{94.7-0.93\cdot 98.6}{0.29\cdot 12.45} = 0.83.$$

Робимо висновок, що зв’язок між ознаками X і Y є високим**.**

Знайдемо рівняння лінійної регресії Y на X у вигляді $y_x = \rho_{yx}x + b$. Вибірковий коефіцієнт регресії обчислимо за формулою (3.1):

$$\rho_{yx} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x^2} = \frac{94.7 - 0.93 \cdot 98.6}{0.29^2} = 35.7 .$$

Знаходимо коефіцієнт b за формулою (3.2):

$$b = \bar{y} - \rho_{yx}\bar{x} = 98.6 - 35.7 \cdot 0.93 = 65.4$$

Рівняння лінійної регресії Y на X: $y_x = 35.7x + 65.4$.
Нижче зображений графік лінії регресії та кореляційного поля.

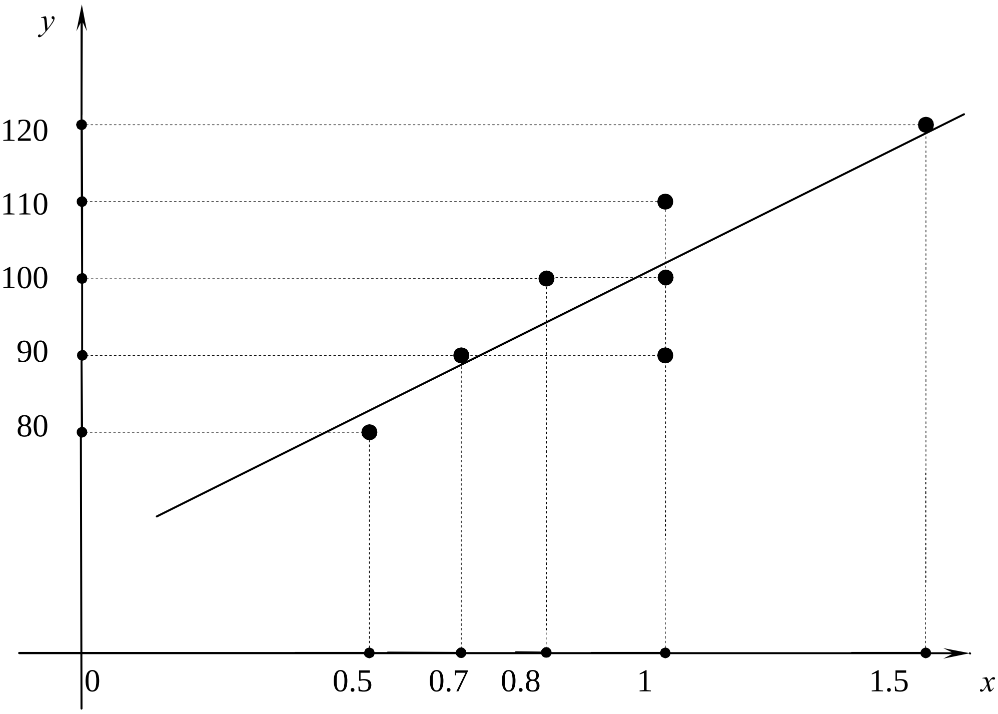

Рис. 4. Наближення вибіркових даних лінією регресії

4. Статистична перевірка статистичних гіпотез

Статистичною гіпотезою називають будь-яке висловлювання про розподіл генеральної сукупності або про параметри відомих розподілів, що перевіряються за вибіркою. Процес використання вибіркових даних для перевірки гіпотези називають *статистичним доведенням*.

Основною (нульовою) називається припущена гіпотеза, яка позначається H_0. Разом з припущеною гіпотезою завжди можна розглядати *альтернативну (конкуруючу)* гіпотезу H_1. Наприклад, якщо H_0: $M(X) = 6$, то H_1: $M(X) \neq 6$, або $M(X) < 6$, або $M(X) > 6$.

Вибір між гіпотезами H_0 та H_1 може супроводжуватися двома видами похибок. Якщо за висновком буде відкинута правильна гіпотеза H_0, то кажуть, що відбулася *похибка першого роду*. Якщо за висновком буде прийнята неправильна гіпотеза H_0, то кажуть, що відбулася *похибка другого роду*.

Правило, за яким приймають рішення про правильність гіпотези H_0, називається *критерієм*.

Ймовірність здійснити похибку першого роду позначається α і називається *рівнем значущості критерію*.

Ймовірність здійснити похибку другого роду позначають β. Ймовірність $1-\beta$ називають *потужністю критерію*. Найчастіше рівень значущості приймають рівним 0.05 або 0.01. Якщо прийнято рівень значущості рівним 0.05, то це означає, що в п’яти випадках із 100 ми ризикуємо одержати похибку першого роду (відкинути правильну гіпотезу).

Перевірку статистичної гіпотези можна здійснити лише з використанням даних вибірки. Для цього слід вибрати деяку випадкову статистичну характеристику (вибіркову функцію), точний або наближений розподіл якої відомий, і за допомогою цієї характеристики перевірити основну гіпотезу.

Статистичним критерієм узгодження перевірки гіпотези (або просто *критерієм*) називається випадкова величина *K*, розподіл якої (точний або наближений) відомий і яка застосовується для перевірки основної гіпотези.

Спостереженим значенням критерію узгодження називається значення відповідного критерію, обчислене за даними вибірки.

Після вибору певного критерію узгодження множину всіх його можливих значень поділяють на дві підмножини, що не перетинаються: одна з них містить значення критерію, за яких основна гіпотеза відхиляється, а друга – за яких приймається.

Критичною областю називається сукупність значень критерію, за яких основна гіпотеза відхиляється.

Областю прийняття гіпотези (областю припустимих значень) називається сукупність значень критерію, за яких гіпотезу приймають.

Критерій узгодження *K* – одновимірна випадкова величина, усі її можливі значення належать деякому інтервалу. Тому критична область та область прийняття гіпотези також будуть інтервалами, а це означає, що існують точки, які ці інтервали відокремлюють.

Критичними точками (межами) критерію K називаються точки $k_{кр}$ (k_α), які відокремлюють критичну область від області прийняття гіпотези. Розрізняють однобічну (правобічну, лівобічну) та двобічну критичні області.

Правобічною називається критична область, що визначається інтервалом $(k_{кр};+\infty)$, де $k_{кр}$ знаходиться з умови $P\left(K > k_{кр}\right)=\alpha$.

Лівобічною називається критична область, що визначається інтервалом $(-\infty;k_{кр})$, де $k_{кр}$ знаходиться з умови $P\left(K < k_{кр}\right)=\alpha$.

Двобічною називається критична область, що визначається інтервалом $(-\infty; k_{1\,кр}) \cup (k_{2\,кр}; +\infty)$, де $k_{1\,кр}, k_{2\,кр}$ знаходять з умови $P\left(K < k_{1\,кр}\right) + P\left(K > k_{2\,кр}\right) = \alpha$.

Для кожного критерію узгодження є відповідні таблиці, які дозволяють знайти точку $k_{кр}$.

Якщо рівень значущості α вже обраний, то критичну область доцільно будувати так, щоб потужність критерію була максимальною. Виконання цієї умови забезпечує мінімальну ймовірність похибки другого роду. Слід зауважити, що єдиним способом одночасного зменшення ймовірностей похибок першого та другого роду є збільшення об'єму вибірки.

Для перевірки правильності основної статистичної гіпотези H_0 необхідно:

1) визначити гіпотезу H_1, альтернативну до гіпотези H_0;

2) вибрати статистичну характеристику перевірки;

3) визначити припустиму ймовірність похибки першого роду, тобто рівень значущості α;

4) знайти за відповідною таблицею критичну область (критичну точку) для вибраної статистичної характеристики.

4.1. Деякі критерії перевірки статистичних гіпотез

Порівняння двох дисперсій нормальних генеральних сукупностей X та Y

Гіпотези	Значення критерію	приймають H_0
H_0: D(X) = D(Y)	$F_{сп} = \dfrac{S_б{}^2}{S_м{}^2}$	
H_1: D(X) > D(Y)	$F_{кр}(\alpha, k_1, k_2)$	$F_{сп} < F_{кр}$
H_1: D(X) ≠ D(Y)	$F_{кр}(\alpha/2, k_1, k_2)$	$F_{сп} < F_{кр}$
$S_б^2$ – більша «виправлена» дисперсія; $S_м^2$ – менша «виправлена» дисперсія; $k_1 = n_1 - 1$ – число ступенів свободи для вибірки з більшою дисперсією; $k_2 = n_2 - 1$ – число ступенів свободи для вибірки з меншою дисперсією; $F_{кр}$ знаходять в таблиці критичних точок розподілу Фішера-Снедекора (*таблиця 3*)		

Приклад. Розглядаються дві генеральних сукупності X і Y, що підкоряються нормальному розподілу. З кожної сукупності добуто вибірку відповідних об'ємів $n_X = 11$ і $n_Y = 14$. За вибірковими даними знайдені «виправлені» середні квадратичні відхилення $S_X{}^2 = 0.76$ та $S_Y{}^2 = 0.38$. На рівні значущості $\alpha = 0.05$ перевірити гіпотезу H_0: $D(X) = D(Y)$ про рівність генеральних дисперсій при конкуруючій гіпотезі H_1: $D(X) > D(Y)$.

Розв'язання. Маємо правобічну критичну область. Обчислимо спостережене значення критерію як відношення більшої «виправленої» дисперсії до меншої:

$$F_{сп} = \frac{S_б{}^2}{S_м{}^2} = \frac{0.76}{0.38} = 2.$$

На рівні значущості $\alpha = 0.05$ із ступенями свободи $k_1 = n_X - 1 = 11 - 1 = 10$ (вибірка з більшою дисперсією) і $k_2 = n_Y - 1 = 14 - 1 = 13$ (вибірка з меншою дисперсією) знаходимо критичну точку, користуючись *таблицею 3*: $F_{кр}(0.05;10;13) = 2.67$. Оскільки $F_{сп} < F_{кр}$, з імовірністю $\gamma = 1 - \alpha = 0.95$ приймається гіпотеза H_0. Робимо висновок, що вибіркові дисперсії відрізняються незначуще.

Порівняння двох середніх нормальних генеральних сукупностей X та Y, дисперсії яких невідомі і однакові (невеликі незалежні вибірки, $N < 30$)

Гіпотези	Значення критерію	приймають H_0
H_0: M(X) = M(Y)	$T_{сп} = \dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{(n_1 - 1) \cdot S_1{}^2 + (n_2 - 1) \cdot S_2{}^2}} \cdot \sqrt{\dfrac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$	
H_1: M(X) > M(Y)	$t_{кр\_прав}(\alpha, k)$	$T_{сп} < t_{кр\_прав}$
H_1: M(X) <M(Y)	$t_{кр\_лів}(\alpha, k) = -t_{кр\_прав}(\alpha, k)$	$T_{сп} > -t_{кр\_прав}$
H_1: M(X) ≠M(Y)	$t_{кр\_двоб}(\alpha, k)$	$\lvert T_{сп} \rvert < t_{кр\_двоб}$
$n -$ об'єм вибірки, $\overline{x}$ – вибіркове середнє значення; S^2 – «виправлена» вибіркова дисперсія; $k = n_1 + n_2 - 2$ – число ступенів свободи; $t_{кр}$ знаходять в таблиці критичних точок розподілу Ст'юдента (*таблиця 2*)		

Приклад. На рівні значущості $\alpha = 0.05$ перевірити гіпотезу H_0: $M(X) = M(Y)$ про рівність генеральних середніх нормально розподілених сукупностей X і Y (дисперсії невідомі, але однакові) при конкуруючій гіпотезі H_1: $M(X) > M(Y)$ за невеликими незалежними вибірками

x_i	12.3	12.5	12.8	13.0	13.5		y_i	12.2	12.3	13.0
n_i	1	2	4	2	1		m_i	6	8	2

Розв'язання.
Обчислимо об'єми вибірок:

$$n = \sum_{i=1}^{5} n_i = 1 + 2 + 4 + 2 + 1 = 10; \qquad m = \sum_{i=1}^{3} m_i = 6 + 8 + 2 = 16.$$

Знайдемо спостережене значення критерію за формулою

$$T_{\text{сп}} = \frac{\overline{x} - \overline{y}}{\sqrt{(n-1)\cdot S_X{}^2 + (m-1)\cdot S_Y{}^2}} \cdot \sqrt{\frac{nm(n+m-2)}{n+m}}.$$

Для цього:
- обчислимо вибіркові середні значення:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{5} x_i n_i = \frac{1}{10}\left(12.3\cdot 1 + 12.5\cdot 2 + 12.8\cdot 4 + 13\cdot 2 + 13.5\cdot 1\right) = 12.8;$$

$$\overline{y} = \frac{1}{m}\sum_{i=1}^{3} y_i m_i = \frac{1}{16}\left(12.2\cdot 6 + 12.3\cdot 8 + 13\cdot 2\right) = 12.35;$$

- обчислимо «виправлені» вибіркові дисперсії:

$$S_X{}^2 = \frac{1}{n-1}\sum_{i=1}^{5}(x_i - \overline{x})^2 n_i =$$

$$= \frac{1}{9}\left((12.3-12.8)^2\cdot 1 + (12.5-12.8)^2\cdot 2 + 0\cdot 4 + (13-12.8)^2\cdot 2 + (13.5-12.8)^2\cdot 1\right) = 0{,}111$$

$$S_Y{}^2 = \frac{1}{m-1}\sum_{i=1}^{3}(y_i - \overline{y})^2 m_i =$$

$$= \frac{1}{15}\left((12.2-12.35)^2\cdot 6 + (12.3-12.35)^2\cdot 8 + (13-12.35)^2\cdot 2\right) = 0.067.$$

Спочатку перевіримо гіпотезу про рівність генеральних дисперсій, використовуючи критерій Фішера – Снедекора. Дисперсія ${S_X}^2$ більша, ніж дисперсія ${S_Y}^2$, тому перевіряємо гіпотезу $D(X) = D(Y)$ із конкуруючою $D(X) > D(Y)$. Знайдемо спостережене значення критерію:

$$F_{сп} = \frac{{S_б}^2}{{S_м}^2} = \frac{0.111}{0.067} = 1.657.$$

Визначимо критичну точку із параметрами $\alpha = 0.05$, $k_1 = n - 1 = 9$ (вибірка з більшою дисперсією), $k_2 = m - 1 = 15$ (вибірка з меншою дисперсією), користуючись *таблицею 3* :

$$F_{кр}(0.05;9;15) = 2.59.$$

Оскільки $F_{сп} < F_{кр}$, нема підстав відкидати гіпотезу про рівність дисперсій.

- знайдемо спостережене значення критерію:

$$T_{\text{сп}} = \frac{\overline{x} - \overline{y}}{\sqrt{(n-1)\cdot {S_X}^2 + (m-1)\cdot {S_Y}^2}} \cdot \sqrt{\frac{nm(n+m-2)}{n+m}} =$$

$$= \frac{12.8 - 12.35}{\sqrt{9\cdot 0.111 + 15\cdot 0.067}} \cdot \sqrt{\frac{160\cdot 24}{26}} = 3.86.$$

- оскільки конкуруюча гіпотеза H_1: $M(X) > M(Y)$, маємо правобічну критичну область. За рівнем значущості $\alpha = 0{,}05$ та $k = n + m - 2 = 24$ ступенями свободи знайдемо із *таблиці 2* критичних точок розподілу Ст'юдента $t_{кр\_прав}(0.05;24) = 1.71$. Оскільки $T_{сп} > t_{кр\_прав}$, гіпотеза про рівність математичних сподівань відхиляється та на рівні значущості $\alpha = 0.05$ приймається конкуруюча гіпотеза $M(X) > M(Y)$.

Порівняння «виправленої» вибіркової дисперсії з гіпотетичною генеральною дисперсією σ_0^2 нормальної сукупності X

Гіпотези	Значення критерію	приймають H_0
H_0: $\sigma^2 = \sigma_0^2$	$\chi^2_{сп} = \dfrac{(n-1)S^2}{\sigma_0{}^2}$	
H_1: $\sigma^2 > \sigma_0^2$	$\chi^2_{кр}(\alpha, k)$	$\chi^2_{сп} < \chi^2_{кр}$
H_1: $\sigma^2 \neq \sigma_0^2$	$\chi^2_{лів\ кр}(1-\alpha/2, k)$ $\chi^2_{пр\ кр}(\alpha/2, k)$	$\chi^2_{лів\ кр} < \chi^2_{сп} < \chi^2_{пр\ кр}$
H_1: $\sigma^2 < \sigma_0^2$	$\chi^2_{кр}(1-\alpha, k)$	$\chi^2_{сп} > \chi^2_{кр}$
σ^2 – невідома генеральна дисперсія; σ_0^2 – значення дисперсії, що припускають; S^2 – «виправлена» дисперсія; $k = n-1$ – число ступенів свободи; $\chi^2_{кр}(\alpha, k)$ знаходять в таблиці критичних точок розподілу Хі-квадрат" (*таблиця 4*)		

Приклад. З розподіленої нормально генеральної сукупності добуто вибірку об'єму $n = 21$ і за нею знайдено «виправлену» вибіркову дисперсію $S^2 = 16.2$. Треба на рівні значущості $\alpha = 0.01$ перевірити гіпотезу про рівність генеральної дисперсії H_0: $\sigma^2 = 15$ при конкуруючій гіпотезі H_1: $\sigma^2 > 15$.

Розв'язання. Маємо правобічну критичну область. Визначимо значення критерію, що спостерігається:

$$\chi^2_{сп} = \frac{(n-1)S^2}{\sigma_0{}^2} = \frac{(21-1)16.2}{15} = 21.6.$$

Знайдемо критичну точку за рівнем значущості $\alpha = 0.01$ та числом ступенів свободи $k = n-1 = 20$, скориставшись *таблицею 4* критичних точок розподілу χ^2: $\chi^2_{кр}(0.01;20) = 37.57$. Оскільки $\chi^2_{сп} < \chi^2_{кр}$, з імовірністю $\gamma = 1-\alpha = 0.99$ приймаємо гіпотезу H_0. Робимо висновок, що різниця між гіпотетичною та «виправленою» дисперсіями є незначущою.

Порівняння вибіркової середньої з гіпотетичною генеральною середньою a_0 нормальної сукупності X

<table>
<tr><td rowspan="2">Гіпотези</td><td colspan="2">Значення критерію</td><td rowspan="2">приймають H_0</td></tr>
<tr><td>генеральна дисперсія σ^2 відома</td><td>генеральна дисперсія невідома</td></tr>
<tr><td>H_0:
$a = a_0$</td><td>$U_{cn} = \frac{(\bar{x} - a_0)\sqrt{n}}{\sigma}$</td><td>$T_{cn} = \frac{(\bar{x} - a_0)\sqrt{n}}{S}$</td><td></td></tr>
<tr><td>H_1:
$a > a_0$</td><td>$\Phi(u_{кр}) = (1 - 2\alpha)/2$</td><td>$t_{кр\_прав}(\alpha, k)$</td><td>$U_{cn} < u_{кр}$ або
$T_{cn} < t_{кр}$</td></tr>
<tr><td>H_1:
$a \neq a_0$</td><td>$\Phi(u_{кр}) = (1 - \alpha)/2$</td><td>$t_{кр\_двоб}(\alpha, k)$</td><td>$\lvert U_{cn} \rvert < u_{кр}$ або
$\lvert T_{cn} \rvert < t_{кр}$</td></tr>
<tr><td>H_1:
$a < a_0$</td><td>$\Phi(u_{кр}) = (1 - 2\alpha)/2$</td><td>$t_{кр\_лів}(\alpha, k) = -t_{кр\_прав}(\alpha, k)$</td><td>$U_{cn} > -u_{кр}$ або
$T_{cn} > -t_{кр}$</td></tr>
<tr><td colspan="4">a – невідома генеральна середня;
a_0 – значення середньої, що припускають;
$k = n - 1$ – число ступенів свободи;
$u_{кр}$ знаходять в таблиці значень функції $\Phi(x)$ (*таблиця 1*);
$t_{кр}(\alpha, k)$ знаходять в таблиці критичних точок розподілу Ст’юдента (*таблиця 2*)</td></tr>
</table>

Приклад. З розподіленої нормально генеральної сукупності із відомим середнім квадратичним відхиленням $\sigma = 5.2$ добуто вибірку об’ємом $n = 100$ і за нею знайдено вибіркове середнє значення $\bar{x}_в = 27.56$. На рівні значущості $\alpha = 0.05$ перевірити гіпотезу про рівність генеральної середньої H_0: $a = 26$ при конкуруючій гіпотезі H_1: $a \neq 26$.

Розв’язання. Маємо двобічну критичну область. Визначимо значення критерію, що спостерігається:

$$U_{cn} = \frac{(\bar{x} - a_0)\sqrt{n}}{\sigma} = \frac{(27.56 - 26)\sqrt{100}}{5.2} = 3.$$

Знайдемо критичну точку з рівності

$$\Phi(u_{кр}) = (1 - \alpha)/2 = (1 - 0.05)/2 = 0.475,$$

скориставшись *таблицею 1*: $u_{кр} = 1.96$. Оскільки $U_{сп} > u_{кр}$, гіпотеза H_0 не приймається. На рівні значущості $\alpha = 0.05$ приймаємо гіпотезу H_1. Робимо висновок, що вибіркова середня і гіпотетична середня відрізняються значуще.

4.2. Перевірка гіпотези про закон розподілу генеральної сукупності

Під час дослідження деякої ознаки X генеральної сукупності за основу приймається припущення про те, що вона розподілена за певним законом. Загального підходу стосовно висунення гіпотези про закон розподілу не існує. Проте, виходячи з емпіричного розподілу досліджуваної ознаки, висувається та або інша гіпотеза про закон розподілу.

Законами розподілу випадкової величини, які найчастіше зустрічаються під час дослідженні проблем галузі, є нормальний, показниковий, пуассонівський, логарифмічній та інш.

Розглянемо критерій, який найчастіше зустрічається в практиці розв'язання технологічних задач засобами математичної статистики – *критерій згоди Пірсона* при застосуванні його до перевірки гіпотези про *розподіл генеральної сукупності*.

Суть критерію полягає в порівнянні емпіричних (тих, які одержані дослідним шляхом) частот n_i і теоретичних частот n_i'. Критерій Пірсона відповідає на питання: чи розходження між частотами є випадковим (незначущим) або невипадковим (значущим). При цьому критерій Пірсона не підтверджує однозначно правильність або неправильність гіпотези, а тільки встановлює її згоду або незгоду з даними спостережень за обраним рівнем значущості. Критерій Пірсона можна використовувати для досить великих груп (n ≥ 50), на кожному інтервалі $(x_i; x_{i+1})$ число емпіричних частот n_i $(i = 1..s)$ повинно бути не менше за 5, в протилежному випадку інтервали потрібно об'єднати, а частоти додати.

Правило Пірсона. Для того щоб при заданому рівні значущості α перевірити основну гіпотезу H_0, потрібно:

1) обчислити теоретичні частоти $n_i' = n \cdot P_i$ для варіант вибірки;
2) обчислити спостережене значення критерію χ^2 за формулою
3)

$$\chi^2_{сп} = \sum_{i=1}^{s} \frac{(n_i - n_i')^2}{n_i'}.$$

4) знайти ступені свободи критерію χ^2 за формулою $k = s - 1 - r$, де s – число груп (інтервалів групування) вибірки; r – число параметрів передбаченого закону розподілу.

5) знайти в таблиці критичну точку $\chi^2_{kp}(\alpha,k)$, яка відповідає заданому рівню значущості α та ступеням свободи k;

6) порівняти χ^2_{kp} та χ^2_{cn}: якщо $\chi^2_{cn} < \chi^2_{kp}$, то гіпотеза про розподіл генеральної сукупності приймається з ймовірністю $\gamma = 1-\alpha$, якщо $\chi^2_{cn} > \chi^2_{kp}$, гіпотезу відхиляють на рівні значущості α.

Правила для знаходження теоретичних частот

Нормальний розподіл		
Формула обчислення ймовірності	Ступені свободи	Значення аргументів
$P_i = \Phi(z_{i+1}) - \Phi(z_i)$	$k = s-3$	$z_i = \dfrac{x_i - \bar{x}}{\sigma}$ $z_1 = -\infty;\quad z_{s+1} = \infty$
Показниковий розподіл		
$P_i = e^{-\lambda x_i} - e^{-\lambda x_{i+1}}$	$k = s-2$	$\lambda = \dfrac{1}{\bar{x}}$
Рівномірний розподіл		
$P_1 = \dfrac{1}{b^* - a^*}(x_2 - a^*)$ $P_i = \dfrac{1}{b^* - a^*}(x_{i+1} - x_i)$ $P_s = \dfrac{1}{b^* - a^*}(b^* - x_s)$	$k = s-3$	$a^* = \bar{x} - \sqrt{3}\sigma$ $b^* = \bar{x} + \sqrt{3}\sigma$
Закон Пуассона		
$P_i = \dfrac{\lambda^i \cdot e^{-\lambda}}{i!}$	$k = s-2$	$\lambda = \bar{x}$ $i = 1 \ldots r$ – кількість появ події в n випробуваннях (r – максимальне число події, що спостерігали)

Приклад. Використовуючи критерій Пірсона, на рівні значущості $\alpha = 0.05$ перевірити, чи узгоджується гіпотеза H_0 про нормальний розподіл генеральної сукупності з емпіричним розподілом вибірки

x_i	[3;8)	[8;13)	[13;18)	[18;23)	[23;28)	[28;33)	[33;38)
n_i	6	8	15	40	16	8	7

Розв'язання. Побудуємо статистичний ряд, обчисливши значення варіант як середнє арифметичне кінців інтервалів $x_i^* = \dfrac{x_i + x_{i+1}}{2}$. Маємо розподіл

x_i^*	5.5	10.5	15.5	20.5	25.5	30.5	35.5
n_i	6	8	15	40	16	8	7

Обчислимо об'єм вибірки, вибіркову середню та вибіркове середнє квадратичне відхилення.

$$N = \sum_{i=1}^{7} n_i = 6+8+15+40+16+8+7 = 100;$$

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{7} x_i^* n_i = \frac{1}{100}(5.5\cdot 6 + 10.5\cdot 8 + 15.5\cdot 15 + 20.5\cdot 40 + 25.5\cdot 16 + 30.5\cdot 8 + 35.5\cdot 7) = 20.7;$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{7}(x_i^*)^2 n_i - \bar{x}^2} =$$

$$= \frac{1}{10}\sqrt{5.5^2\cdot 6 + 10.5^2\cdot 8 + 15.5^2\cdot 15 + 20.5^2\cdot 40 + 25.5^2\cdot 16 + 30.5^2\cdot 8 + 35.5^2\cdot 7 - 20.7^2} = 7.28.$$

Обчислимо нормовані значення кінців інтервалів за формулою $z_i = \dfrac{x_i - \bar{x}}{\sigma}$, вважаючи, що $z_1 = -\infty$, $z_7 = \infty$, та побудуємо таблицю для обчислення теоретичних частот (в таблиці Φ(z) – функція Лапласа з *таблиці 1*).

i	x_i	x_{i+1}	z_i	z_{i+1}	$\Phi(z_i)$	$\Phi(z_{i+1})$	$P_i = \Phi(z_{i+1}) - \Phi(z_i)$	$n_i' = nP_i$	n_i
1	3	8	$-\infty$	– 1.74	– 0.5	– 0.4591	0.0409	4.09	6
2	8	13	– 1.74	– 1.06	– 0.4591	– 0.3554	0.1037	10.37	8
3	13	18	– 1.06	– 0.37	– 0.3554	– 0.1443	0.2111	21.11	15
4	18	23	– 0.37	0.32	– 0.1443	0.1255	0.2698	26,98	40
5	23	28	0.32	1	0.1255	0.3413	0.2158	21.58	16
6	28	33	1	1.69	0.3413	0.4545	0.1132	11.32	8
7	33	38	1.69	∞	0.4545	0.5	0.0455	4.55	7
Σ							1	100	100

Обчислимо спостережене значення критерію Пірсона за формулою

$$\chi^2_{cn} = \sum_{i=1}^{7}\frac{(n_i - n_i')^2}{n_i'}.$$

$$\chi^2_{сп} = \frac{(6-4.09)^2}{4.09} + \frac{(8-10.37)^2}{10.37} + \frac{(15-21.11)^2}{21.11} + \frac{(40-26.98)^2}{26.98} + \frac{(16-21.58)^2}{21.58} +$$

$$+ \frac{(8-11.32)^2}{11.32} + \frac{(7-4.55)^2}{4.55} = 13.22 \cdot$$

За *таблицею 4* критичних точок розподілу χ^2 , за рівнем значущості $\alpha = 0.05$ та числом ступенів свободи $k = 7-3 = 4$ знаходимо критичну точку $\chi^2_{kp}(0.05;4) = 0.49$. Оскільки $\chi^2_{сп} > \chi^2_{kp}$, на рівні значущості 0.05 гіпотеза про нормальний розподіл генеральної сукупності відхиляється.

5. Індивідуальні завдання

Задача 1.

При дослідженні зразків м’ясних кулінарних виробів було встановлено концентрацію м’язової тканини у фарші (%), значення якої наведено нижче. Представити дані у вигляді дискретного та інтервального (у кількості k інтервалів) варіаційних рядів.

Варіант 1

30, 40, 30, 35, 30, 45, 40, 50, 30, 35, 30, 40, 40, 45, 50, 55, 30, 40, 30, 45, 30, 45;
$k = 4$.

Варіант 2

45, 40, 35, 35, 40, 45, 45, 50, 55, 55, 30, 40, 30, 40, 40, 45, 50, 55, 30;
$k = 3$.

Варіант 3

25, 30, 25, 40, 35, 20, 25, 30, 35, 20, 45, 40, 35, 45, 20, 30, 25, 20, 25, 30, 35;
$k = 3$.

Варіант 4

30, 25, 30, 20, 25, 25, 30, 20, 35, 40, 20, 25, 20, 30, 40, 25, 30, 35, 20, 45, 40, 20;
$k = 4$.

Варіант 5

35, 40, 30, 40, 45, 45, 30, 40, 30, 45, 50, 30, 40, 45, 45, 40, 50, 30, 35;
$k = 3$.

Варіант 6

40, 45, 20, 25, 30, 40, 30, 45, 20, 35, 30, 45, 30, 20, 25, 30, 25, 30, 35, 20;
$k = 3$.

Варіант 7

35, 50, 40, 50, 45, 35, 35, 40, 45, 50, 50, 45, 35, 35, 40, 45, 35, 30, 35, 30, 30, 50;
$k = 4$.

Варіант 8

45, 50, 45, 55, 45, 55, 60, 50, 45, 55, 50, 60, 55, 50, 60, 55, 45, 40, 50;
$k = 3$.

Варіант 9

50, 45, 50, 50, 45, 40, 55, 40, 60, 45, 50, 50, 40, 50, 45, 60, 50, 45, 55, 60, 45, 50;
$k = 4$.

Варіант 10

55, 30, 50, 35, 40, 55, 35, 40, 45, 35, 45, 50, 50, 40, 35, 40, 55, 55, 30, 40;
$k = 3$.

Варіант 11

30, 45, 35, 35, 30, 45, 40, 55, 30, 35, 35, 40, 40, 45, 50, 55, 30, 50, 40, 45;
$k = 3$.

Варіант 12

30, 25, 20, 40, 25, 25, 35, 25, 35, 40, 20, 25, 30, 30, 45, 25, 30, 35, 25, 45, 40, 25, 30;

$k = 4$.

Варіант 13

35, 45, 30, 45, 45, 40, 30, 40, 35, 45, 50, 30, 45, 45, 40, 40, 30, 30, 55, 40;
$k = 3$.

Варіант 14

35, 55, 45, 55, 45, 55, 60, 55, 45, 55, 50, 65, 55, 50, 60, 50, 45, 40, 60, 40, 45, 55;
$k = 4$.

Варіант 15

35, 40, 30, 45, 40, 45, 45, 55, 55, 55, 30, 45, 30, 40, 40, 45, 50, 55, 35;
$k = 3$.

Задача 2. Для вибірки задачі 1 знайти моду та медіану.

Задача 3. За даними задачі 1 побудувати емпіричну функцію розподілу, полігон частот, полігон відносних частот (для дискретного ряду) та гістограму частот і відносних частот (для інтервального ряду).

Задача 4. За даними задачі 1 обчислити для дискретного та інтервального рядів середні вибіркові значення, вибіркові дисперсії, «виправлені» вибіркові дисперсії, середні квадратичні відхилення, «виправлені» середні квадратичні відхилення.

Задача 5. Взяли n проб молока, що поступило на реалізацію з сільськогосподарського підприємства. Середня жирність молока дорівнює *a%*. З довірчою ймовірністю γ визначити межі, в яких буде знаходитись середня жирність молока всієї партії. Значення S або σ дивись у таблиці даних.

№ з/п	n	*a*	γ		№ з/п	n	*a*	γ	
1	16	3,2	0,95	S=0,2	2	15	3,3	0,99	S=0,36
3	25	2,8	0,99	S=0,3	4	26	3,5	0,99	S=0,24
5	20	2,5	0,94	σ=0,4	6	27	3,6	0,95	S=0,22
7	30	1,6	0,95	S=0,1	8	40	2,8	0,95	σ=0,22
9	36	0,5	0,95	σ=0,25	10	33	2,7	0,99	σ=0,2
11	25	3,8	0,96	σ=0,3	12	54	2,9	0,999	S=0,3
13	30	3,6	0,999	S=0,31	14	12	3,5	0,97	σ=0,31
15	42	2,8	0,9	σ=0,26	16	64	3,4	0,99	S=0,24

Задача 6. Для даних задачі 1 оцінити з надійністю 0,95 математичне сподівання *a* нормально розподіленої ознаки генеральної сукупності за допомогою надійного інтервалу.

Задача 7. За даними, які наведені у таблиці, скласти рівняння регресії y на x, побудувати графік кореляційного поля і лінії регресії та оцінити силу зв'язку між ознаками X і Y.

1	X	10	15	20	25	30	35	40	45	50	55
	Y	12	18	27	24	31	36	39	40	55	54
2	X	20	21	22	23	24	25	26	27	28	29
	Y	10	12	14	17	15	19	21	20	23	22
3	X	10	11	12	13	14	15	16	17	18	19
	Y	30	35	32	40	42	44	43	45	38	46
4	X	5	10	15	20	25	30	35	40	45	50
	Y	6	7	12	14	16	17	18	20	24	10
5	X	0,5	0,6	0,7	0,8	0,9	1	1,1	1,2	1,3	1,4
	Y	1,3	1,21	1,12	1,62	1,73	1,45	1,44	1,52	1,68	1,7
6	X	2	4	6	8	10	12	14	16	18	20
	Y	4,2	7,0	8,0	7,5	9,0	8,5	9,6	9,5	10,0	13,0
7	X	2	4	6	8	10	12	14	16	18	20
	Y	3,5	6,0	7,0	6,2	7,5	8,5	10,0	11,5	13	15
8	X	10	15	20	25	30	35	40	45	50	55
	Y	22	31	39	49	59	63	85	90	99	96
9	X	20	21	22	23	24	25	26	27	28	29
	Y	15	16	14	17	18	25	21	24	23	22
10	X	10	11	12	13	14	15	16	17	18	19
	Y	5	5,2	7	6	7,5	8	9	8,5	10	9,2
11	X	5	10	15	20	25	30	35	40	45	50
	Y	6,8	7,4	12,2	14,2	18,3	17	18,1	20,5	24	10,3
12	X	20	21	22	23	24	25	26	27	28	29
	Y	14	13	14	17	15	19	20	20	23	25
13	X	10	12	15	23	25	34	35	39	45	50
	Y	8	10	15	22	24	30	36	39	40	45
14	X	0,5	0,6	0,7	0,8	0,9	1	1,1	1,2	1,3	1,4
	Y	1,25	1,20	1,22	1,62	1,73	1,45	1,43	1,52	1,68	1,70
15	X	2	4	6	8	10	12	14	16	18	20
	Y	3,4	6,0	7,2	6,5	7,5	8,5	10,0	11,2	12	15

Задача 8. Хронометраж часу розв'язання n_x задач без комп'ютерних технологій дав такі результати: середній час розв'язання m_x хвилин, виправлена вибіркова дисперсія s_x^2 (хв^2). Середній час розв'язання n_y задач із застосуванням комп'ютерних технологій – m_y хвилини, а виправлена вибіркова дисперсія s_y^2 (хв^2). На рівні значущості α перевірити нульову гіпотезу H_0: $M(X)=M(Y)$ при конкуруючий гіпотезі H_1. Чи дозволило застосування комп'ютерних технологій скоротити час розв'язання задач? (Перевірити гіпотезу про рівність дисперсій).

№ з/п	n_x	m_x	s_x^2	n_y	m_y	s_y^2	α	H_1
1	9	57	186,2	15	52	166,4	0,01	$M(X)\neq M(Y)$
2	9	57	186,2	15	52	166,4	0,01	$M(X)>M(Y)$
3	12	60	152,3	16	55	149,7	0,02	$M(X)\neq M(Y)$
4	12	60	152,3	16	55	149,7	0,02	$M(X)>M(Y)$
5	10	45	130,1	9	40	142,5	0,02	$M(X)\neq M(Y)$
6	10	45	130,1	9	40	142,5	0,02	$M(X)>M(Y)$
7	15	48	50,1	11	41	62,3	0,01	$M(X)\neq M(Y)$
8	15	48	50,1	11	41	62,3	0,01	$M(X)>M(Y)$
9	12	45	92,3	10	39	97,2	0,01	$M(X)\neq M(Y)$
10	12	45	92,3	10	39	97,2	0,01	$M(X)>M(Y)$
11	9	38	100,3	8	32	121,2	0,01	$M(X)\neq M(Y)$
12	9	38	100,3	8	32	121,2	0,01	$M(X)>M(Y)$
13	8	62	130,2	11	42	129,5	0,02	$M(X)\neq M(Y)$
14	9	50	185,3	15	51	169,4	0,01	$M(X)>M(Y)$
15	12	58	151,5	16	54	148,3	0,02	$M(X)\neq M(Y)$

Задача 9. Використовуючи критерій Пірсона, на рівні значущості α перевірити, чи узгоджується гіпотеза H_0 про розподіл генеральної сукупності з емпіричним розподілом вибірки.

№ з/п	H_0: нормальний розподіл сукупності							α
1	x_i	[190;195]	[195;200]	[200;205]	[205;210]	[210;215]	[215;220]	
	n_i	5	29	56	65	29	16	0,01
2	x_i	[100;107]	[107;114]	[114;121]	[121;128]	[128;135]	[135;142]	
	n_i	20	23	50	64	30	13	0,01
3	x_i	[50;53]	[53;56]	[56;59]	[59;62]	[62;65]	[65;68]	
	n_i	10	50	48	45	20	12	0,01
4	x_i	[10;12]	[12;14]	[14;16]	[16;18]	[18;20]	[20;22]	
	n_i	8	20	46	46	20	10	0,025
5	x_i	[20;24]	[24;28]	[28;32]	[32;36]	[36;40]	[40;44]	
	n_i	7	37	40	30	21	10	0,05
6	x_i	[0;2]	[2;4]	[4;6]	[6;8]	[8;10]	[10;12]	
	n_i	5	37	40	30	21	10	0,01
7	x_i	[5;7]	[7;9]	[9;11]	[11;13]	[13;15]	[15;17]	
	n_i	20	25	40	30	26	9	0,05
8	x_i	[200;210]	[210;220]	[220;230]	[230;240]	[240;250]	[250;260]	
	n_i	20	37	49	51	31	12	0,01
9	x_i	[400;420]	[420;440]	[440;460]	[460;480]	[480;500]	[500;520]	
	n_i	5	37	43	51	31	18	0,05

№ з/п		H_0: нормальний розподіл сукупності						α
10	x_i	[100;105]	[105;110]	[110;115]	[115;120]	[120;125]	[125;130]	
	n_i	9	30	45	51	31	19	0,05
11	x_i	[90;95]	[95;100]	[100;105]	[105;110]	[110;115]	[115;120]	
	n_i	8	30	52	54	29	46	0,025
12	x_i	[50;52]	[52;54]	[54;56]	[56;58]	[58;60]	[60;62]	
	n_i	10	20	35	45	30	12	0,01
13	x_i	[20;23]	[23;26]	[26;29]	[29;32]	[32;35]	[35;38]	
	n_i	7	20	40	42	21	10	0,05
14	x_i	[10;12]	[12;14]	[14;16]	[16;18]	[18;20]	[20;22]	
	n_i	8	10	40	46	18	12	0,025
15	x_i	[5;8]	[8;11]	[11;14]	[14;17]	[17;20]	[20;23]	
	n_i	20	25	40	30	26	9	0,05

Таблиці

Таблиця 1 – **Значення функції Лапласа** $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$

x	*Φ(x)*	*x*	*Φ(x)*	*x*	*Φ(x)*	*x*	*Φ(x)*
0	0	0,32	0,1255	0,64	0,2389	0,96	0,3315
0,01	0,004	0,33	0,1293	0,65	0,2422	0,97	0,334
0,02	0,008	0,34	0,1331	0,66	0,2454	0,98	0,3365
0,03	0,012	0,35	0,1368	0,67	0,2486	0,99	0,3389
0,04	0,016	0,36	0,1406	0,68	0,2517	1,00	0,3413
0,05	0,0199	0,37	0,1443	0,69	0,2549	1,01	0,3438
0,06	0,0239	0,38	0,148	0,7	0,258	1,02	0,3461
0,07	0,0279	0,39	0,1517	0,71	0,2611	1,03	0,3485
0,08	0,0319	0,40	0,1554	0,72	0,2642	1,04	0,3508
0,09	0,0359	0,41	0,1591	0,73	0,2673	1,05	0,3531
0,10	0,0398	0,42	0,1628	0,74	0,2704	1,06	0,3554
0,11	0,0438	0,43	0,1664	0,75	0,2734	1,07	0,3577
0,12	0,0478	0,44	0,17	0,76	0,2764	1,08	0,3599
0,13	0,0517	0,45	0,1736	0,77	0,2794	1,09	0,3621
0,14	0,0557	0,46	0,1772	0,78	0,2823	1,10	0,3643
0,15	0,0596	0,47	0,1808	0,79	0,2852	1,11	0,3665
0,16	0,0636	0,48	0,1844	0,8	0,2881	1,12	0,3686
0,17	0,0675	0,49	0,1879	0,81	0,291	1,13	0,3708
0,18	0,0714	0,50	0,1915	0,82	0,2939	1,14	0,3729
0,19	0,0753	0,51	0,195	0,83	0,2967	1,15	0,3749
0,20	0,0793	0,52	0,1985	0,84	0,2995	1,16	0,377
0,21	0,0832	0,53	0,2019	0,85	0,3023	1,17	0,379
0,22	0,0871	0,54	0,2054	0,86	0,3051	1,18	0,381
0,23	0,091	0,55	0,2088	0,87	0,3078	1,19	0,383
0,24	0,0948	0,56	0,2123	0,88	0,3106	1,20	0,3849
0,25	0,0987	0,57	0,2157	0,89	0,3133	1,21	0,3869
0,26	0,1026	0,58	0,219	0,9	0,3159	1,22	0,3888
0,27	0,1064	0,59	0,2224	0,91	0,3186	1,23	0,3907
0,28	0,1103	0,60	0,2257	0,92	0,3212	1,24	0,3925
0,29	0,1141	0,61	0,2291	0,93	0,3238	1,25	0,3944
0,30	0,1179	0,62	0,2324	0,94	0,3264	1,26	0,3962
0,31	0,1217	0,63	0,2357	0,95	0,3289	1,27	0,3980

продовження таблиці 1

x	*Φ(x)*	*x*	*Φ(x)*	*x*	*Φ(x)*	*x*	*Φ(x)*
1,28	0,3997	1,61	0,4463	1,94	0,4738	2,54	0,4945
1,29	0,4015	1,62	0,4474	1,95	0,4744	2,56	0,4948
1,30	0,4032	1,63	0,4484	1,96	0,475	2,58	0,4951
1,31	0,4049	1,64	0,4495	1,97	0,4756	2,60	0,4953
1,32	0,4066	1,65	0,4505	1,98	0,4761	2,62	0,4956
1,33	0,4082	1,66	0,4515	1,99	0,4767	2,64	0,4959
1,34	0,4099	1,67	0,4525	2,00	0,4772	2,66	0,4961
1,35	0,4115	1,68	0,4535	2,02	0,4783	2,68	0,4963
1,36	0,4131	1,69	0,4545	2,04	0,4793	2,70	0,4965
1,37	0,4147	1,70	0,4554	2,06	0,4803	2,72	0,4967
1,38	0,4162	1,71	0,4564	2,08	0,4812	2,74	0,4969
1,39	0,4177	1,72	0,4573	2,10	0,4821	2,76	0,4971
1,40	0,4192	1,73	0,4582	2,12	0,483	2,78	0,4973
1,41	0,4207	1,74	0,4591	2,14	0,4838	2,80	0,4974
1,42	0,4222	1,75	0,4599	2,16	0,4846	2,82	0,4976
1,43	0,4236	1,76	0,4608	2,18	0,4854	2,84	0,4977
1,44	0,4251	1,77	0,4616	2,20	0,4861	2,86	0,4979
1,45	0,4265	1,78	0,4625	2,22	0,4868	2,88	0,498
1,46	0,4279	1,79	0,4633	2,24	0,4875	2,90	0,4981
1,47	0,4292	1,80	0,4641	2,26	0,4881	2,92	0,4982
1,48	0,4306	1,81	0,4649	2,28	0,4887	2,94	0,4984
1,49	0,4319	1,82	0,4656	2,30	0,4893	2,96	0,4985
1,50	0,4332	1,83	0,4664	2,32	0,4898	2,98	0,4986
1,51	0,4345	1,84	0,4671	2,34	0,4904	3,00	0,49865
1,52	0,4357	1,85	0,4678	2,36	0,4909	3,20	0,49931
1,53	0,437	1,86	0,4686	2,38	0,4913	3,40	0,49966
1,54	0,4382	1,87	0,4693	2,40	0,4918	3,60	0,499841
1,55	0,4394	1,88	0,4699	2,42	0,4922	3,80	0,499928
1,56	0,4406	1,89	0,4706	2,44	0,4927	4,00	0,499968
1,57	0,4418	1,90	0,4713	2,46	0,4931	4,50	0,499997
1,58	0,4429	1,91	0,4719	2,48	0,4934	5,00	0,5
1,59	0,4441	1,92	0,4726	2,50	0,4938		
1,60	0,4452	1,93	0,4732	2,52	0,4941		

Таблиця 2 – **Критичні точки розподілу Стьюдента**

Рівень значущості α (двобічна критична область)

Число ступенів свободи k	0,1	0,05	0,02	0,01	0,002	0,001
1	6,31	12,71	31,82	63,66	318,31	636,62
2	2,92	4,3	6,96	9,92	22,33	31,6
3	2,35	3,18	4,54	5,84	10,21	12,92
4	2,13	2,78	3,75	4,6	7,17	8,61
5	2,02	2,57	3,36	4,03	5,89	6,87
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3	3,5	4,79	5,41
8	1,86	2,31	2,9	3,36	4,5	5,04
9	1,83	2,26	2,82	3,25	4,3	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,8	2,2	2,72	3,11	4,02	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,6	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,9	3,65	3,97
18	1,73	2,1	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,72	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,5	3,79
23	1,71	2,07	2,5	2,81	3,48	3,77
24	1,71	2,06	2,49	2,8	3,47	3,75
25	1,71	2,06	2,49	2,79	3,45	3,73
26	1,71	2,06	2,48	2,78	3,43	3,71
27	1,7	2,05	2,47	2,77	3,42	3,69
28	1,7	2,05	2,47	2,76	3,41	3,67
29	1,7	2,05	2,46	2,76	3,4	3,66
30	1,7	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,7	3,31	3,55
60	1,67	2	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,16	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,29
	0,05	0,025	0,01	0,005	0,001	0,0005

Рівень значущості α (однобічна критична область)

Таблиця 3 – **Критичні точки розподілу F Фішера – Снедекора**
(k_1 – число ступенів свободи більшої дисперсії,
k_2 – число ступенів свободи меншої дисперсії)

рівень значущості $\alpha = 0{,}01$

k_1 / k_2	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5403	5625	5764	5859	5928	5981	6022	6056	6083	6106
2	98,49	99,01	99,2	99,5	99,3	99,3	99,4	99,4	99,4	99,4	99,4	99,4
3	34,12	30,81	29,5	26,24	28,2	27,9	27,7	27,5	27,4	27,23	27,1	27,1
4	21,2	18	16,7	13,89	15,5	15,2	15	14,8	14,7	14,55	14,5	14,4
5	16,26	13,27	12,1	9,82	11	10,7	10,5	10,3	10,2	10,05	9,96	9,89
6	13,74	10,92	9,78	7,98	8,75	8,47	8,26	8,1	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	6,99	7,46	7,19	6,99	6,84	6,72	6,62	6,54	6,47
8	11,26	8,65	7,59	6,37	6,63	6,37	6,18	6,03	5,91	5,81	5,73	5,67
9	10,56	8,02	6,99	5,8	6,06	5,8	5,61	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,64	5,64	5,39	5,2	5,06	4,94	4,85	4,77	4,71
11	9,86	7,21	6,22	5,32	5,32	5,07	4,89	4,74	4,63	4,54	4,46	4,4
12	9,33	6,93	5,95	5,06	5,06	4,82	4,64	4,5	4,39	4,3	4,22	4,16
13	9,07	6,7	5,74	4,86	4,86	4,62	4,44	4,3	4,19	4,1	4,02	3,96
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,8
15	8,68	6,36	5,42	4,56	4,56	4,32	4,14	4	3,89	3,8	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,2	4,03	3,89	3,78	3,69	3,62	3,55
17	8,4	6,11	5,18	4,67	4,34	4,1	3,93	3,79	3,68	3,59	3,52	3,46

рівень значущості $\alpha = 0{,}05$

k_1 / k_2	1	2	3	4	5	6	7	8	9	10	11	12
1	161,5	199,5	216	224,6	230	234	237	239	241	241,9	243	244
2	18,51	19	19,2	19,25	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6	5,96	5,94	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,7	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,1	4,06	4,03	4
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,6	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,5	3,44	3,39	3,35	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,1	3,07
10	4,96	4,1	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91
11	4,84	3,98	3,59	3,36	3,2	3,09	3,01	2,95	2,9	2,85	2,82	2,79
12	4,75	3,89	3,49	3,26	3,11	3	2,91	2,85	2,8	2,75	2,72	2,69
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,6
14	4,6	3,74	3,34	3,11	2,96	2,85	2,76	2,7	2,65	2,6	2,57	2,53
15	4,54	3,68	3,29	3,06	2,9	2,79	2,71	2,64	2,59	2,54	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42
17	4,45	3,59	3,2	2,96	2,81	2,7	2,61	2,55	2,49	2,45	2,41	2,38

Таблиця 4 – **Критичні точки розподілу** χ^2

Число ступенів свободи k	Рівень значущості α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,63	5,02	3,84	0,0039	0,00098	0,00016
2	9,21	7,38	5,99	0,103	0,051	0,02
3	11,34	9,35	7,81	0,352	0,216	0,115
4	13,28	11,14	9,49	0,711	0,484	0,297
5	15,09	12,83	11,07	1,15	0,831	0,554
6	16,81	14,45	12,59	1,64	1,24	0,872
7	18,48	16,01	14,07	2,17	1,69	1,24
8	20,09	17,53	15,51	2,73	2,18	1,65
9	21,67	19,02	16,92	3,33	2,70	2,09
10	23,21	20,48	18,31	3,94	3,25	2,56
11	24,72	21,92	19,68	4,57	3,82	3,05
12	26,22	23,34	21,03	5,23	4,40	3,57
13	27,69	24,74	22,36	5,89	5,01	4,11
14	29,14	26,12	23,68	6,57	5,63	4,66
15	30,58	27,49	25,00	7,26	6,26	5,23
16	32,00	28,85	26,30	7,96	6,91	5,81
17	33,41	30,19	27,59	8,67	7,56	6,41
18	34,81	31,53	28,87	9,39	8,23	7,01
19	36,19	32,85	30,14	10,1	8,91	7,63
20	37,57	34,17	31,41	10,9	9,59	8,26
21	38,93	35,48	32,67	11,6	10,3	8,9
22	40,29	36,78	33,92	12,3	11,0	9,54
23	41,64	38,08	35,17	13,1	11,7	10,2
24	42,98	39,36	36,42	13,8	12,4	10,9
25	44,31	40,65	37,65	14,6	13,1	11,5
26	45,64	41,92	38,89	15,4	13,8	12,2
27	46,96	43,19	40,11	16,2	14,6	12,9
28	48,28	44,46	41,34	16,9	15,3	13,6
29	49,59	45,72	42,56	17,7	16,0	14,3
30	50,89	46,98	43,77	18,5	16,8	15,0

Список рекомендованої літератури

1. Торяник Д. О. Вища та прикладна математика: навчальний посібник / за ред. проф. Синєкопа / Торяник Д. О., Синєкоп М. С. та ін. – ХДУХТ, 2014. – 330 с.
2. Гмурман В. Е. Теория вероятностей и математическая статистика / В. Е. Гмурман. – М. : Высш. шк., 2003. – 480 с.
3. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике / В. Е. Гмурман. – М. : Высш. шк., 2003. – 405 с.
4. Жилюк Н. О. Методичні вказівки для самостійної роботи студентів з курсу «Теорія ймовірностей та математична статистика» / Н. О. Жилюк, М. С. Синєкоп, С. М. Шинкар. − Х. : ХДУХТ, 2006. – 91 с.
5. Полевич В. В. Статистична перевірка гіпотез: методичні вказівки / В. В. Полевич, Ж. А. Крутовий. − Х. : ХДАТОХ, 1998. – 65 с.

Навчальне електронне видання
комбінованого використання
Можна використовувати в локальному та мережному режимах

МАТЕМАТИЧНІ МЕТОДИ В ТЕХНОЛОГІї

Методичні вказівки та індивідуальні контрольні завдання

Укладачі:
ПАРХОМЕНКО Лариса Олександрівна
ТОРЯНИК Дмитро Олександрович

Відповідальний за випуск зав. кафедри фізико-математичних та інженерно-технічних дисциплін проф. М.І. Погожих

Техн. редактор О.В. Щегельська

План 2016 р., поз. 34

Підп. до друку 14.06.2016 р. Один електронний оптичний диск (CD-ROM);
Супровідна документація. Об'єм даних 1,68 Мб. Тираж 5 прим.

Видавець і виготівник
Харківський державний університет харчування та торгівлі
вул. Клочківська, 333, Харків, 61051.
Свідоцтво суб'єкта видавничої справи ДК № 4417 від 10.10.2012 р.