МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ, МОЛОДІ ТА СПОРТУ
ХАРКІВСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ХАРЧУВАННЯ ТА ТОРГІВЛІ

Математичні методи в технології

МЕТОДИЧНІ ВКАЗІВКИ
з розв'язання задач
з використанням ПЕОМ

для студентів технологічних спеціальностей

Харків – 2011

Рекомендовано до видання
кафедрою вищої математики,
протокол № 12 від 23.03.11р.

Обговорено і схвалено на засіданні
науково-методичної комісії НН ІХТБ,
протокол № 4 від 13. 04. 11 р

Рецензент: д.т.н., проф. Синєкоп М.С.

Зміст

ПЕРЕДМОВА

Статистичні розрахунки без допомоги ЕОМ є складними і потребують застосування багатьох таблиць функцій і квантилів стандартних розподілів. Тому вони не дають можливості відчути елемент нового в матеріалі, що вивчається, змінювати довільно умови задач тощо. Спеціалізовані математичні пакети не можуть застосовуватись для навчання, оскільки їх застосування потребує достатньо високого рівня підготовки з математичної статистики. Тому ми пропонуємо використовувати універсальний математичний пакет MathCAD Professional.

На кожному лабораторному занятті студент отримує індивідуальне завдання, яке виконує самостійно під керуванням викладача. З метою кращого засвоювання матеріалу до кожної лабораторної пропонується індивідуальне завдання, яке потрібно виконати вдома без застосування комп’ютера. В першому розділі методичних вказівок надані необхідні теоретичні питання з курсу «Математичні методи в технології», в другому основні відомості щодо застосування пакету MathCAD. Методичні вказівки містять варіанти індивідуальних завдань і приклади, які показують способи розв’язання задач за допомогою пакету MathCAD.

Отже, методичні вказівки дозволяють, по-перше, інтенсифікувати практичну складову навчання математичній статистиці, по-друге, навчити студентів навичкам використання математичного пакету MathCAD.

РОЗДІЛ 1. ОСНОВНІ ПОНЯТТЯ МАТЕМАТИЧНОЇ СТАТИСТИКИ

1.1. Вибірковий метод

Математична статистика – це розділ математики, присвячений вивченню закономірностей, які мають місце в масових явищах та статистичних сукупностях. Зміст математичної статистики складають математичні методи систематизації, обробки та аналізу масових статистичних даних незалежно від їх якісного вмісту. Методи математичної статистики можуть бути застосовані для обробки та аналізу будь-яких статистичних даних. Математичні методи безпосередньо пов’язані з ймовірною оцінкою результатів спостережень та визначенням математичної ймовірності. У зв’язку з цим висновки математичної статистики відносно масових явищ та процесів носять ймовірний характер і спираються на апарат теорії ймовірності.

Вивчення множини схожих об’єктів може проводитися як за якісними, так і за кількісними ознаками. Наприклад, якщо досліджується партія деталей, то якісною ознакою може бути стандартність (або нестандартність) деталі, а кількісною ознакою – розмір деталі.

Перевірку можна проводити двома різними способами: можна провести суцільний контроль усіх об’єктів; можна провести контроль лише деякої частини об’єктів. Вся множина об’єктів, яким належить пройти контроль та обстеження, називається **генеральною сукупністю**. При другому способі множина випадковим чином відібраних об’єктів називається **вибірковою сукупністю** або просто **вибіркою**. Число елементів вибірки називається **об’ємом вибірки** і позначається n.

Значення ознаки, яка при переході від одного елемента сукупності до іншого змінюється (варіюється), називається **варіантою** і зазвичай позначається малими латинськими літерами x, y, z. Порядковий номер варіанти (значення ознаки) називається **рангом**: x_1 – 1-ша варіанта (1-е значення ознаки), x_2 – 2-га варіанта (2-е значення ознаки), x_i – i-та варіанта (i-те значення ознаки). **Частота варіанти** (n_i) показує, скільки разів зустрічається та чи інша варіанта (значення ознаки) у статистичній сукупності.

Сума всіх частот дорівнює **об’єму вибірки** $\sum_{i=1}^{k} n_i = n$. Величина $\omega_i = \frac{n_i}{n}$ називається **відносною частотою**, $\sum_{i=1}^{k} \omega_i = 1$.

Сама вибірка може здійснюватися двома способами. При формуванні вибірки відібрані об’єкти можна повертати в генеральну сукупність, а можна не повертати. **Повторною** називають вибірку, коли відібраний об’єкт повертається в генеральну сукупність, **безповторною** – коли відібраний об’єкт не повертається.

Вибірка повинна відбивати досить повно особливості усіх об’єктів генеральної сукупності, тобто вибірка повинна бути **репрезентативною**.

Результати спостережень, у загальному випадку безладно розташований ряд чисел, для вивчення необхідно впорядкувати (ранжирувати). Розташування значень ознаки за не спаданням називається **ранжируванням** даних.

Розташовані у порядку зростання числа називаються **варіаційним рядом.**

Розміщена у порядку зростання варіант послідовність пар чисел, складених з варіант та їх частот (або відносних частот), називається **статистичним рядом.** Статистичні ряди бувають дискретними та інтервальними. *Дискретні* статистичні ряди будують звичайно в тому випадку, якщо значення досліджуваного ознаки відрізняються один від одного не менше ніж на деяку кінцеву величину. У дискретних статистичних рядах задаються точкові значення ознаки. Загальний вигляд дискретного статистичного ряду:

x_i	x_1	x_2	...	x_k
n_i	n_1	n_2	...	n_k
ω_i	ω_1	ω_2	...	ω_k

Інтервальні статистичні ряди будують звичайно в тому випадку, якщо значення досліджуваного ознаки можуть відрізнятися один від одного на як завгодно малу величину або якщо об'єм вибірки досить великий. Значення ознаки в них задаються у вигляді інтервалів $[a_i a_{i+1})$. Зазвичай при цьому *правий кінець* кожного інтервалу *виключають* з відповідної множини, а *лівий включають*. Для побудови статистичного ряду інтервал, який містить всю множину елементів вибірки, розбивають на кілька інтервалів довжини $h_i = a_{i+1} - a_i$, які не перетинаються. Якщо інтервали у варіаційних рядах мають однакову довжину, їх називають **рівновеликими**, у іншому випадку – **нерівновеликими**.

Різницю R між максимальним і мінімальним елементами вибірки називають **розмахом вибірки** $R = x_{max} - x_{min}$. Якщо будується ряд з рівними інтервалами, кількість інтервалів часто визначають за формулою **Стерджесса**:

$$k = 1 + 3{,}322 \cdot lg\, n, \tag{1}$$

де n – об'єм вибірки, значення k округлюють до цілого. Тоді довжина інтервалу знаходиться за формулою $h = \frac{R}{k}$.

x_i	$[a_1 a_2)$	$[a_2 a_3)$	...	$[a_k a_{k+1})$
n_i	n_1	n_2	...	n_k

Часто у таблиці вказують не межі інтервалів, за якими групуються результати вимірювань, а їх середини xs_i.

Для будь-якого дійсного числа x позначимо через n_x кількість чисел x_i з вибірки $x_1, x_2, ..., x_k$, які задовольняють нерівності $x_i < x$. Отже на усій числовій прямій визначена функція n_x. Покладемо $F^*(x) = \frac{n_x}{n}$.

Ця функція є **емпіричною функцією розподілу,** що визначає для кожного значення x відносну частоту події $X < x$.

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

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

2) $F*(x)$ є не спадною функцією;

3) $F*(x)=\begin{cases}0, & x \le x_{\min}; \\ 1, & x > x_{\max}.\end{cases}$ де $x_{\min}$ – найменша, $x_{\max}$ – найбільша варіанта.

Емпірична функція розподілу завжди сходинкова. При цьому величина сходинки, якщо таке значення зустрічається 1 раз, дорівнює $1/n$, і дорівнює k/n у точках, де k значень збігаються.

Дискретний статистичний ряд графічно можна представити за допомогою полігону розподілу частот або відносних частот. **Полігоном частот** називають ламану лінію, вершинами якої є точки з координатами (x_1,n_1), (x_2,n_2), ... , (x_k,n_k). Для його побудови на осі абсцис відкладають варіанти x_i, а на осі ординат – відповідні їм частоти n_i. Побудовані точки (x_i, n_i) з'єднують відрізками прямих і отримують полігон частот.

Інтервальний ряд графічно зображують за допомогою гістограми частот, яка є статистичним наближенням щільності розподілу $f(x)$.

Гістограмою частот називають ступінчату фігуру, яка складається з прямокутників, основами яких є інтервали варіант довжини $h_i = a_{i+1} - a_i$, а висоти дорівнюють відношенню $\frac{n_i}{h_i}$ (щільність частоти).

Площа i–го прямокутника дорівнює частоті n_i, отже площа гістограми частот дорівнює сумі усіх частот, тобто об'єму вибірки.

Іноді замість полігона і гістограми частот будують полігон і гістограму відносних частот, побудова яких відрізняється лише тим, що на осі ординат відкладається не n_i, а відносна частота ω_i, висоти прямокутників гістограми будуть дорівнювати відповідно $\frac{\omega_i}{h_i}$ (щільність відносної частоти), а площа всієї гістограми відносних частот буде дорівнювати одиниці.

Приклад 1. При обстеженні 50 членів сімей робітників і службовців встановлена наступна кількість членів сім'ї: 5;3;2;1;4;6;3;7;8;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) Складіть статистичний ряд розподілу частот.

2) Побудуйте полігон розподілу частот, емпіричну функцію розподілу.

Розв'язання. 1) У цьому завданні ознака, що вивчається, є дискретною, оскільки розмір сімей не може відрізнятися один від одного менше ніж на одну людину. Отже, необхідно побудувати дискретний статистичний ряд. Щоб зробити це, необхідно знайти, скільки разів зустрічаються ті або інші значення ознаки, і розташувати їх в порядку зростання. Значення ознаки, що вивчається, розмір сім'ї – позначимо x_i, частоти (кількість таких сімей) – n_i.

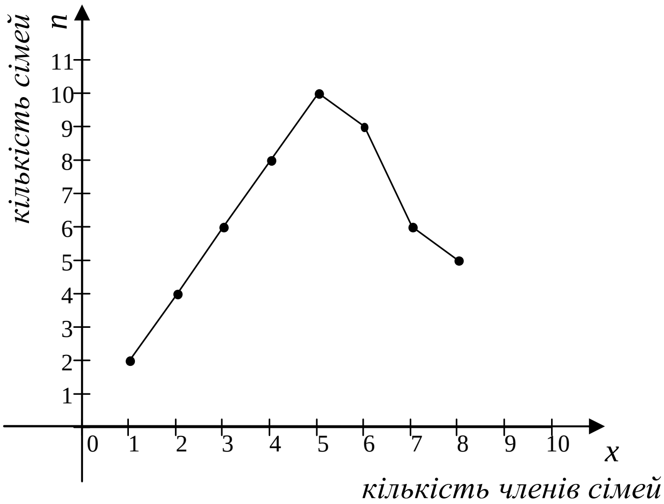
Оформимо розрахунки у вигляді таблиці.

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

Перевіримо правильність підрахунків:

$\sum_{i=1}^{8} n_i = 2 + 4 + 6 + 8 + 10 + 9 + 6 + 5 = 50.$

2) Дискретний варіаційний ряд можна представити графічно, побудувавши полігон розподілу частот. Для цього на графіку потрібно відмітити точки з координатами (1,2); (2,4); (3,6); (4,8); (5,10); (6,9); (7,6); (8,5).



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

коли $x \le x_{min} = 1$, $F^*(x) = 0$;

коли $1 < x \le 2$, поява значень, менших за x, спостерігалась при появі значення $x_1 = 1$ $n_1 = 2$ раз; значить, в цьому інтервалі $F^*(x) = \frac{n_1}{n} = \frac{2}{50}$;

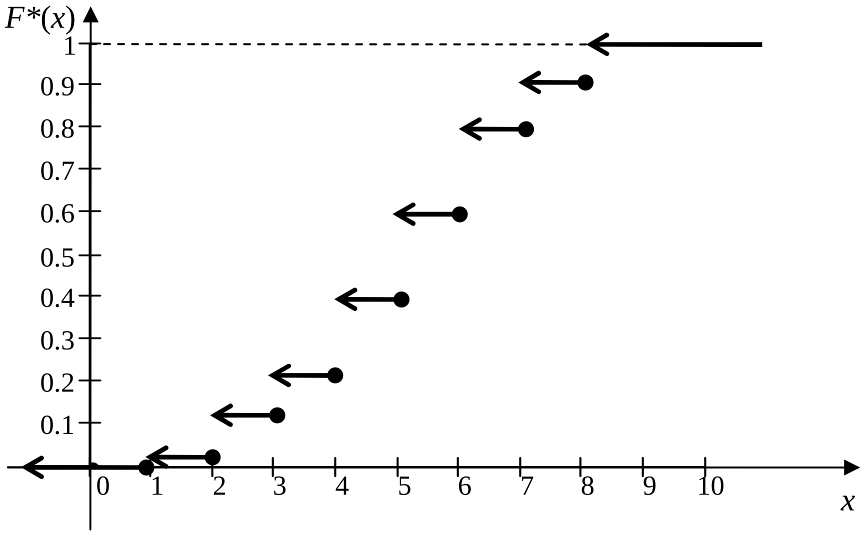
коли $2 < x \le 3$, поява значень, менших за x, спостерігалась при появі $x_1 = 1$ та $x_2 = 2$ в кількості $n_1 + n_2 = 2 + 4 = 6$ раз; значить, в цьому інтервалі $F^*(x) = \frac{n_1+n_2}{n} = \frac{6}{50}$;

коли $3 < x \le 4$, поява значень, менших за x, спостерігалась при появі $x_1 = 1$ та $x_2 = 2$ та $x_3 = 3$ в кількості $n_1 + n_2 + n_3 = 2 + 4 + 6 = 12$ раз; отже, в цьому інтервалі $F^*(x) = \frac{12}{50}$.

Аналогічно знаходимо значення на інших інтервалах. Маємо:

$$F^*(x) = \begin{cases} 0 & \text{при} \quad x \le 1 \\ 2/50 & \text{при} \quad 1 < x \le 2 \\ 6/50 & \text{при} \quad 2 < x \le 3 \\ 12/50 & \text{при} \quad 3 < x \le 4 \\ 20/50 & \text{при} \quad 4 < x \le 5 \\ 30/50 & \text{при} \quad 5 < x \le 6 \\ 39/50 & \text{при} \quad 6 < x \le 7 \\ 45/50 & \text{при} \quad 7 < x \le 8 \\ 1 & \text{при} \quad x \ge 8 \end{cases}$$

Графік емпіричної функції розподілу:



Приклад 2. 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 статистичний ряд розподілу оцінок в групі. Дані представити графічно.

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

$[a_i;a_{i+1})$	$[40;60)$	$[60;75)$	$[75;90)$	$[90;100]$
n_i	5	12	7	1

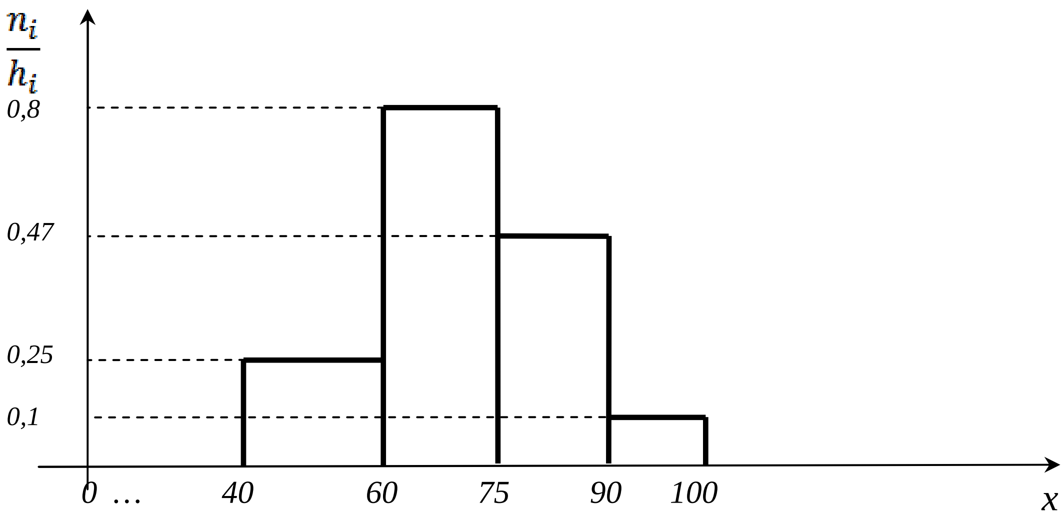
Контроль обчислень: $\sum_{i=1}^{4} n_i = 5+12+7+1=25.$

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

Побудуємо гістограму. Маємо різновеликі інтервали. Їх довжини $h_1 = 60 - 40 = 20, h_2 = 75 - 60 = 15, h_3 = 90 - 75 = 15, h_4 = 100 - 90 = 10$.
Обчислимо щільність частот, $\frac{n_i}{h_i}$, $i = 1,2,3,4$:

$\frac{n_i}{h_i}$	$\frac{5}{20} = 0{,}25$	$\frac{12}{15} = 0{,}8$	$\frac{7}{15} = 0{,}47$	$\frac{1}{10} = 0{,}1$

Побудуємо гістограму, що складається з прямокутників із основами h_i і висотами $\frac{n_i}{h_i}$.



Приклад обробки вибірки невеликого об'єму за допомогою програми MathCAD розглянуто в документі 1. В цьому прикладі елементи вибірки задаються вручну. Потім за допомогою вбудованої функції sort отримаємо варіаційний ряд для заданої вибірки. Для побудови статистичного ряду потрібно вручну вибрати з варіаційного ряду всі різні значення вибіркової величини $(x_i \neq x_j, i \neq j)$ і занести їх до масиву. Подальші обчислення легко автоматизуються за допомогою функцій додавання.

Приклад обробки вибірки великого об'єму розглянуто в документі 2. Також в цих документах наведені варіанти графічного зображення вибірок.

1.2. Статистичні оцінки параметрів розподілу

У багатьох випадках треба дослідити кількісну ознаку X генеральної сукупності, використовуючи результати вибірки. Іноді, з деяких міркувань, вдається встановити закон розподілу X. Тоді виникає задача оцінювання параметрів цього закону розподілу.

За даними вибірки розраховують числові характеристики, які називають *статистиками*. Статистики, що одержані з різних вибірок, як правило, відрізняються одна від одної. Тому статистика, отримана з вибірки, є тільки оцінкою невідомого параметра генеральної сукупності. Оцінка параметра – певна числова характеристика, отримана з вибірки. Коли оцінка визначається одним числом, її називають точкової оцінкою.

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

- **незсуненості** (математичне сподівання такої оцінки повинно бути рівним оцінюваному параметру за будь-якого об'єму вибірки. Незсуненость означає її правильність «в середньому», відсутність систематичної похибки);
- **ефективності** (тобто при заданому об'єму вибірки n така оцінка

повинна мати найменшу з можливих дисперсію);

- **спроможності** (тобто якщо $n \to \infty$, статистична оцінка за ймовірністю повинна прямувати до оцінюваного параметра. Спроможність оцінки - це можливість з її допомогою визначити шуканий параметр з будь-якою точністю і як завгодно великою вірогідністю за рахунок використання вибірки досить великого об'єму).

Якщо досліджувану ознаку X генеральної сукупності розглядати як випадкову величину, то її математичне сподівання $M(X)$ дорівнює **генеральній середній** $\bar{x}_\text{г}$:

$$\bar{x}_\text{г} = \frac{1}{N}\sum_{i=1}^{l} N_i x_i = M(X) = a, \qquad (2)$$

де N – об'єм генеральної сукупності, $x_1, x_2, ..., x_l$ – значення ознаки X генеральної сукупності, $N_1, N_2, ..., N_l$ – відповідні частоти, $\sum_{i=1}^{l} N_i = N$.

Незсуненою оцінкою генеральної середньої (математичного сподівання) є **вибіркова середня**

$$\bar{x}_\text{в} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i, \qquad (3)$$

де x_i – варіанта вибірки; n_i – частота варіанти; $n = \sum_{i=1}^{k} n_i$ – об'єм вибірки.

Якщо потрібно знайти середню для інтервального ряду, то вважать що всі елементи з *i*-го інтервалу приймають значення xs_i (середина *i*-го інтервалу), тобто

$$\bar{x}_\text{в} = \frac{1}{n}\sum_{i=1}^{k} n_i xs_i \qquad (3')$$

Для всіх подальших оцінок формули для інтервального ряду записуються аналогічно.

Степінь розсіювання значень ознаки X генеральної сукупності навколо середнього значення $\bar{x}_\text{г}$ визначається величиною **генеральної дисперсії**, що обчислюється за формулою:

$$D_\text{г} = \frac{1}{N}\sum_{i=1}^{l} N_i (x_i - \bar{x}_\text{г})^2. \qquad (4)$$

Вибіркова дисперсія є зсуненою оцінкою генеральної дисперсії:

$$D_\text{в} = \frac{1}{n}\sum_{i=1}^{k} n_i \left(x_i - \bar{x}_\text{в}\right)^2 \text{ або} \qquad (5)$$

$$D_\text{в} = \frac{1}{n}\sum_{i=1}^{k} n_i {x_i}^2 - \left(\bar{x}_\text{в}\right)^2 \qquad (5')$$

Незсуненою оцінкою генеральної дисперсії є **виправлена вибіркова дисперсія**:

$$S^2 = \frac{1}{n-1}\sum_i^k n_i\left(x_i - \bar{x}_\text{в}\right)^2 \text{ або} \tag{6}$$

$$S^2 = \frac{n}{n-1} D_\text{в} \tag{6′}$$

Для оцінки середнього квадратичного відхилення генеральної сукупності використовують **вибіркове середнє квадратичне відхилення**, що дорівнює квадратному кореню із вибіркової дисперсії

$$\sigma_B = \sqrt{D_B} \tag{7}$$

.

або **«виправлене» середнє квадратичне відхилення**, яке дорівнює квадратному кореню з виправленої дисперсії:

$$S = \sqrt{\frac{1}{n-1}\sum_i^k n_i (x_i - \bar{x}_\text{в})^2}\ . \tag{8}$$

Коефіцієнт варіації визначається формулою

$$V_B = \frac{\sigma_B}{\bar{x}_B} \tag{9}$$

Якщо $|V_B| < 0{,}35$ можна зробити висновок про однорідність даних.

Зауваження. Якщо дані подані у вигляді варіаційного ряду, то формули 3 та5 набувають більш простого вигляду:

$$\bar{x}_\text{в} = \frac{1}{n}\sum_{i=1}^n x_i \tag{3″}$$

$$D_\text{в} = \frac{1}{n}\sum_{i=1}^n \left(x_i - \bar{x}_\text{в}\right)^2 \tag{5″}$$

Приклад 3. За даними приклада 1 визначте середній розмір (середнє число членів) сім’ї. Охарактеризуйте коливання розміру сім’ї за допомогою показників варіації (дисперсії, середнього квадратичного відхилення, коефіцієнта варіації). Поясніть отримані результати, зробіть висновки.

Розв’язання. Розрахуємо середній розмір (середнє число членів) сім’ї за формулою (3). Об’єм вибірки $n = 50$, тому

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

$$\bar{x}_B = \frac{1}{50}\left(1\cdot 2 + 2\cdot 4 + 3\cdot 6 + 4\cdot 8 + 5\cdot 10 + 6\cdot 9 + 7\cdot 6 + 8\cdot 5\right) = \frac{246}{50} = 4{,}92\,.$$

Середній розмір сім’ї – 4,92 особи.

Вибіркову дисперсію знайдемо за формулою (5′);

$$D_B = \frac{1}{50}\left(1^2\cdot 2 + 2^2\cdot 4 + 3^2\cdot 6 + 4^2\cdot 8 + 5^2\cdot 10 + 6^2\cdot 9 + 7^2\cdot 6 + 8^2\cdot 5\right) -$$
$$- 4{,}92^2 = 3{,}554\left(\textit{особ}^2\right)$$

«Виправлену» вибіркову дисперсію знайдемо за формулою (6 ′):

$$S^2 = \frac{50}{50-1} 3{,}554 = 3{,}627\left(ос.^2\right).$$

Вибіркове середнє квадратичне відхилення:

$$\sigma_B = \sqrt{D_B} = \sqrt{3{,}554} = 1{,}885(особ);$$

«Виправлене» вибіркове середнє квадратичне відхилення:

$$S = \sqrt{D_{BB}} = \sqrt{3{,}627} = 1{,}904(особ).$$

Знайдемо коефіцієнт варіації розміру сім'ї за формулою (9)

$$V = \frac{1{,}885}{4{,}92} = 0{,}38.$$

Коефіцієнт варіації становить 38%. Так як коефіцієнт варіації більше ніж 35%, можна зробити висновок про те, що досліджувана сукупність сімей є неоднорідною, чим і пояснюється високе коливання розміру сім'ї в даній сукупності.

Приклад 4. Наведені результати зважування (в г) випадковим чином відібраних 100 страв.

вага (г)	154–158	158–162	162–166	166–170	170–174	174–178	178–182
Число страв	10	14	26	28	12	8	2

Знайти вибіркову середню і вибіркову дисперсію ваги страви.

Розв'язання. Запишемо дані у вигляді статистичного ряду, прийнявши за варіанти середини інтервалів:

xs_i	156	160	164	168	172	176	180
n_i	10	14	26	28	12	8	2

Визначимо вибіркову середню і вибіркову дисперсію:

$$\bar{x}_в = \frac{1}{n}\sum_{i=1}^{k} n_i xs_i =$$

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

$$D_в = \frac{1}{n}\sum_{i=1}^{k} n_i\left(xs_i - \bar{x}_в\right)^2 = \frac{1}{100}((156-166)^2 \cdot 10 + (160-166)^2 \cdot 14 +$$
$$+(164-166)^2 \cdot 26 + (168-166)^2 \cdot 28 + (172-166)^2 \cdot 12 +$$
$$+(176-166)^2 \cdot 8 + (180-166)^2 \cdot 2) = 33{,}44(г^2).$$

Середня вага страви 166г, середня вибіркова дисперсія 33,44 $г^2$.

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

Медіаною називається таке значення ознаки, що поділяє ранжируваний ряд розподілу на 2 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. Якщо в ряді парна кількість (2m) елементів, медіану визначають як середню арифметичну величину з двох серединних значень.

Формули для обчислення медіани з парною та непарною кількістю варіант:

$$M_e = \frac{x_m + x_{m+1}}{2}, \; M_e = x_{m+1}. \qquad (10)$$

Варіанти знаходження числових характеристик варіаційного ряду в середовищі MathCAD дивись в документах 1 та 2.

1.3. Інтервальні статистичні оцінки генеральних характеристик

Середня вибіркова, вибіркова дисперсія і середнє квадратичне відхилення є оцінками параметрів генеральної сукупності, що виражаються одним числом. Очевидно, що точкова оцінка залежить від об'єму вибірки і може відрізнятися від дійсного значення параметру, що оцінюється, тобто може привести до грубих помилок. Це викликає необхідність оцінювати точність і надійність точкових оцінок, отриманих за даною вибіркою, за допомогою так званих інтервальних оцінок.

Нехай оцінкою $\tilde{X}$ деякої кількісної ознаки X генеральної сукупності є значення $\bar{\theta}$. Очевидно, що $\bar{\theta}$ тим точніше, чим менша величина **відхилення** $|\tilde{X} - \bar{\theta}| < \Delta$. У такому випадку число Δ можна вважати **точністю оцінки**. Внаслідок випадковості того, що та чи інша варіанта потрапляє до вибірки, говорити про виконання нерівності $|\tilde{X} - \bar{\theta}| < \Delta$ можна лише з деякою ймовірністю γ, яка називається **надійністю (надійною ймовірністю)** оцінки, тобто

$$P(|\tilde{X} - \bar{\theta}| < \Delta) = \gamma \text{ або } P(\bar{\theta} - \Delta < \tilde{X} < \bar{\theta} + \Delta) = \gamma.$$

Це означає, що ймовірність того, що інтервал $(\bar{\theta} - \Delta, \bar{\theta} + \Delta)$ містить (покриває) невідомий параметр $\tilde{X}$ дорівнює γ; інтервал $(\bar{\theta} - \Delta, \bar{\theta} + \Delta)$ називається **надійним інтервалом**.

Особливий інтерес викликає надійний інтервал для оцінки математичного сподівання кількісної ознаки X генеральної сукупності, розподіленої за **нормальним** законом. Тут можливі такі випадки:

- Якщо заздалегідь **відома** величина **середнього квадратичного відхилення** σ або **у випадку великої вибірки** ($n > 30$), то надійний інтервал для оцінки математичного сподівання a $(\bar{x}_в - \Delta; \bar{x}_в + \Delta)$ має вигляд:

$$\left(\bar{x}_в - t\frac{\sigma}{\sqrt{n}}; \bar{x}_в + t\frac{\sigma}{\sqrt{n}}\right), \qquad (11)$$

де $\bar{x}_в$ – середня вибіркова; n – об'єм вибірки; σ – відоме середнє квадратичне відхилення генеральної сукупності; t – величина, що визначається за таблицею значень функції Лапласа (дод.2) $\Phi(t)=\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}dx$ із співвідношення $2\Phi(t)=\gamma$, де γ – заздалегідь вибрана надійна ймовірність.

Слід пам'ятати, що надійна ймовірність пов'язана тут не з величиною параметра a, а лише з межами інтервалу, які змінюються при зміні вибірки. Надійність γ показує, що якщо зроблено достатньо велике число вибірок, то γ% з них визначають такі інтервали, в яких параметр a дійсно міститься, і лише в (1 – γ)% випадків він може вийти за межі надійного інтервалу.

Величину α=1–γ називають **рівнем значущості**.

- 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:

$$\left(\bar{x}_{\text{в}}-t_{\gamma}\frac{S}{\sqrt{n}};\bar{x}_{\text{в}}+t_{\gamma}\frac{S}{\sqrt{n}}\right), \tag{12}$$

де $\bar{x}_{\text{в}}$ – вибіркова середня; S – «виправлене» середнє квадратичне відхилення; n – об'єм вибірки; t_{γ} – величина, що визначається за таблицею розподілу Стьюдента за рівнем значущості α=1–γ та числом степенів свободи $\boldsymbol{k=n-1}$ (дод.4).

Для оцінки **генеральної частки *p*** нормально розподіленого кількісної ознаки за вибірковою часткою *ω= m/n (об'єм вибірки n* ≥ 30) використовують формулу:

$$\left(\omega-t\sqrt{\frac{\omega(1-\omega)}{n}};\omega+t\sqrt{\frac{\omega(1-\omega)}{n}}\right). \tag{13}$$

де t визначається за таблицями функції Лапласа (дод. 2) із співвідношення $2\Phi(t)=\gamma$; ω – вибіркова частка; *n* – об'єм вибірки (число обстежених одиниць).

Для оцінки **генеральної частки *p*** нормально розподіленого кількісної ознаки за вибірковою часткою *ω= m/n (для малого об'єму вибірки n* < 30) надійний інтервал $(\omega-\Delta;\omega+\Delta)$знаходять за формулою

$$\left(\omega-t_{\gamma}\sqrt{\frac{\omega(1-\omega)}{n}};\omega+t_{\gamma}\sqrt{\frac{\omega(1-\omega)}{n}}\right). \tag{14}$$

где t_{γ} – величина, що визначається за таблицею розподілу Стьюдента за рівнем значущості α=1–γ та числом степенів свободи $\boldsymbol{k=n-1}$.

Приклад 5. За допомогою випадкового відбору керівництво фірми провело вибіркове обстеження 900 своїх службовців. Середній стаж їхньої роботи у фірмі дорівнює 8,70 років, а середнє квадратичне (стандартне) відхилення – 2,70 років. Серед обстежених виявилося 270 жінок. Вважаючи стаж роботи службовців фірми розподіленим за нормальним законом, визначте:

а) з імовірністю 0,95 довірчий інтервал, в якому буде середній стаж роботи всіх службовців фірми;

б) з імовірністю 0,90 довірчий інтервал, що покриває невідому частку жінок у всьому колективі фірми.

Розв'язання. За умовами задачі обсяг вибірки n = 900 одиниць, тобто вибірка велика.

а) Знайдемо межі довірчого інтервалу середнього стажу роботи всього колективу фірми, тобто межі довірчого інтервалу для генеральної середньої. За умовою $\bar{x}_В = 8{,}70$; $\sigma = 2{,}70$; $n = 900$; $\gamma = 0{,}95$. Використовуємо формулу (11)

Знайдемо t із співвідношення $2\Phi(t) = \gamma$: $2\Phi(t) = 0{,}95$; $\Phi(t) = 0{,}95/2 = 0{,}475$.

Із таблиці функції Лапласа *(дод.2)* знайдемо $t = 1{,}96$.

Знайдемо граничну похибку вибірки Δ.

$$\Delta = t\frac{\sigma}{\sqrt{n}} = 1{,}96\frac{2{,}70}{\sqrt{900}} = 0{,}176.$$

Знайдемо довірчий інтервал

$$(\bar{x}_B - \Delta, \bar{x}_B + \Delta) = (8{,}70 - 0{,}176; 8{,}70 + 0{,}176) = (8{,}524; 8{,}876).$$

Отже, з імовірністю 0,95 можна очікувати, що середній стаж роботи всього колективу фірми знаходиться в інтервалі від 8,524 до 8,876 років.

б) Оцінимо справжнє значення частки жінок у всьому колективі фірми. За умовою $m = 270$; $n = 900$; $\gamma = 0{,}90$.

Вибіркова частка $\omega = 270/900 = 0{,}3$.

Застосуємо формулу (2.13). Знайдемо t із співвідношення $2\Phi(t) = \gamma$: $2\Phi(t) = 0{,}90$; $\Phi(t) = 0{,}90/2 = 0{,}45$.

Із таблиці функції Лапласа *(дод.2)* знайдемо $t = 1{,}64$.

Знайдемо граничну похибку вибірки Δ.

$$\Delta = t\sqrt{\frac{\omega(1-\omega)}{n}} = 1{,}64\sqrt{\frac{0{,}3(1-0{,}3)}{900}} = 0{,}025.$$

$$\omega - \Delta < p < \omega + \Delta;$$
$$0{,}3 - 0{,}025 < p < 0{,}3 + 0{,}025;$$
$$0{,}275 < p < 0{,}325.$$

Отже, з імовірністю 0,90 можна очікувати, що частка жінок у всьому колективі фірми знаходиться в інтервалі від 0,275 до 0,325.

Приклад 6. Яким повинен бути об'єм випадкової вибірки, щоб з надійністю 0,90 можна було стверджувати, що максимальне відхилення вибіркової частки жінок у вибірці від частки жінок у всьому колективі фірми не перевищує 0,05, якщо в минулому аналогічному обстеженні вибіркова частка жінок виявилася рівною 0,30?

Розв'язання. У цій задачі треба знайти необхідну чисельність вибірки. Її розрахунок дає відповідь на питання: «Скільки потрібно обстежити одиниць сукупності, щоб з наперед заданою ймовірністю не перевищити заздалегідь задану похибку?»

В задачі дано: $\Delta = 0{,}05$; $\omega = 0{,}30$; $\gamma = 0{,}90$.

Із формули $\Delta = t\sqrt{\frac{\omega(1-\omega)}{n}}$ знайдемо n:

$$n = \frac{t^2\omega(1-\omega)}{\Delta^2}$$

Знайдемо *t* із співвідношення $2\Phi(t) = \gamma$: $2\Phi(t) = 0{,}90$; $\Phi(t) = 0{,}90/2 = 0{,}45$. Із таблиці функції Лапласа *(дод.2)* знайдемо $t = 1{,}64$.

Розрахуємо необхідну чисельність вибірки

$$n = \frac{1{,}64^2 0{,}3(1-0{,}3)}{0{,}05^2} = 225{,}93$$

Оскільки n - ціле число, округлимо отриманий результат до більшого цілого, враховуючи, що необхідно не перевищувати задану похибку. Отже, n=226.

Відповідь. Щоб з ймовірністю 0,90 і похибкою 0,05 з допомогою випадкового відбору визначити частку жінок у всьому колективі фірми, необхідно обстежити не менше ніж 226 службовців.

Приклади розв'язання задач з теми в середовищі MathCAD містяться в документах 3 та 4.

1.4. Статистична перевірка гіпотез

Статистичною гіпотезою називають будь-яке висловлювання про вигляд розподілу генеральної сукупності або про параметри відомих розподілів, що перевіряється за вибіркою. Процес використовування вибірки 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 похибками двох родів. Якщо за висновком буде відкинута правильна гіпотеза H_0, то кажуть, що це **похибка першого роду**. Якщо за висновком буде прийнята неправильна гіпотеза H_0, то кажуть, що це **похибка другого роду**.

	Прийнята гіпотеза H_1	Прийнята гіпотеза H_0
H_0 – правильна	Похибка 1-го роду, її ймовірність $P(H_1/H_0) = \alpha$	Правильне рішення, його ймовірність $P(H_0/H_0) = 1 - \alpha$
H_0 – неправильна	Правильне рішення, його ймовірність	Похибка 2-го роду, її ймовірність

	$P(H_1/H_1) = 1 - \beta$	$P(H_0/H_1) = \beta$

Правило, за яким приймають рішення правильна чи ні гіпотеза 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.

Якщо статистична характеристика розподілена нормально, то критерій позначають не літерою *K*, а літерами *U* або *Z*. Якщо статистична характеристика розподілена за законом Фішера-Снедекора, то її позначають *F*. У разі розподілу статистичної характеристики за законом Стьюдента її позначають *T*, а у випадку закону «Хі - квадрат» – χ^2.

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.

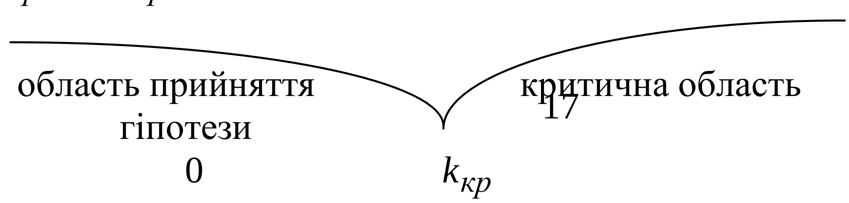
Критичною областю називається сукупність значень критерію, за яких основна гіпотеза відхиляється.

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

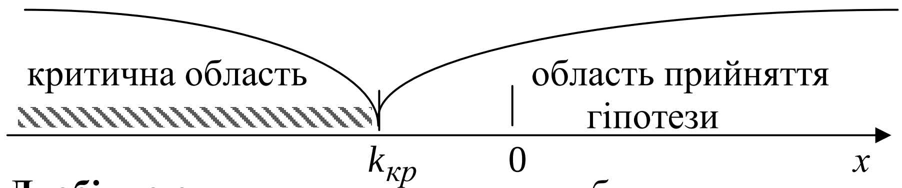
Критерій узгодження *K* – одновимірна випадкова величина, усі її можливі значення належать деякому інтервалу. Тому критична область та область прийняття гіпотези також будуть інтервалами, а це означає, що існують точки, які ці інтервали відокремлюють.

Критичними точками (межами) критерію К 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 нерівністю $K > k_{кр}$, де $k_{кр}$ – додатне число.



Лівобічною називається критична область, що визначається нерівністю $K < k_{кр}$, де $k_{кр}$ – від'ємне число.



Двобічною називається критична область, що визначається нерівностями $K < k_1, \quad K > k_2 \quad k_2 > k_1$.

Щоб знайти однобічну критичну область, треба знайти критичну точку $k_{кр}$. Для цього задають достатньо малу ймовірність – **рівень значущості** $\boldsymbol{\alpha}$**,** а потім шукають критичну точку із врахуванням вимоги $P(K > k_{кр}) = \alpha$ у випадку правобічної критичної області, або $P(K < k_{кр}) = \alpha$ у випадку лівобічної критичної області, у випадку двобічної критичної області має виконуватись тотожність $P(K < k_1) + P(K > k_2) = \alpha$.

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

Якщо рівень значущості α вже вибраний, 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 <u>необхідно</u>:

1) визначити гіпотезу H_1 альтернативну до гіпотези H_0;

2) вибрати статистичну характеристику перевірки;

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

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

Деякі критерії перевірки статистичних гіпотез

Таблиця 1.

Порівняння двох дисперсій нормальних генеральних сукупностей X та Y

Гіпотези	Значення критерію та правило знаходження критичної точки
H_0: D(X) = D(Y)	$F_{сп} = \dfrac{S_{б}{}^2}{S_{м}{}^2}$
H_1: D(X) > D(Y)	$F_{кр}(\alpha, k_1, k_2)$
H_1: D(X) ≠ D(Y)	$F_{кр}(\alpha/2, k_1, k_2)$

$S_{б}^2$ – більша «виправлена» дисперсія; $S_{м}^2$ – менша «виправлена» дисперсія;
α – рівень значущості;
$k_1 = n_1 - 1$ – число степенів свободи для вибірки з більшою дисперсією;
$k_2 = n_2 - 1$ – число степенів свободи для вибірки з меншою дисперсією;
$F_{кр}(\alpha, k_1, k_2)$ знаходять в таблиці критичних точок розподілу Фішера-Снедекора (дод.5).

Таблиця 2

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

Гіпотези	Значення критерію та правило знаходження критичної точки
H_0: $D(X) = \sigma_0^2$	$\chi^2_{сп} = \dfrac{(n-1)S^2}{\sigma_0{}^2}$
H_1: $D(X) > \sigma_0^2$	$\chi^2_{кр}(\alpha, k)$
H_1: $D(X) \neq \sigma_0^2$	$\chi^2_{лів\ кр}(1-\alpha/2, k)$ $\chi^2_{пр\ кр}(\alpha/2, k)$
H_1: $D(X) < \sigma_0^2$	$\chi^2_{кр}(1-\alpha, k)$

σ_0^2 – значення дисперсії, що припускають;
n – об’єм вибірки; S^2 – «виправлена» дисперсія; α – рівень значущості;
$k = n - 1$ – число степенів свободи;

$\chi^2_{кр}(\alpha, k)$ знаходять в таблиці критичних точок розподілу «Хі-квадрат» (дод. 3)

Таблиця 3

Порівняння двох середніх нормальних генеральних сукупностей X та Y, дисперсії яких відомі (незалежні вибірки)

Гіпотези	Значення критерію та правило знаходження критичної точки
H_0: M(X) = M(Y)	$Z_{сп} = \dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{\dfrac{D(x_1)}{n_1} + \dfrac{D(x_2)}{n_2}}}$
H_1: M(X) > M(Y)	$\Phi(z_{kr}) = \dfrac{1-2\alpha}{2}$
H_1: M(X) <M(Y)	$\Phi(z_{kr}) = \dfrac{1-2\alpha}{2}$
H_1: M(X) ≠M(Y)	$\Phi(z_{kr}) = \dfrac{1-\alpha}{2}$
$\overline{x_i}$ – вибіркова середня; $D(x_i)$– відома дисперсія; – об’єм вибірки; α – рівень значущості; $z_{кр}$ знаходять в таблиці Лапласа (дод. 2).	

Таблиця 4

Порівняння двох середніх нормальних(або ні) генеральних сукупностей X та Y, дисперсії яких невідомі (незалежні великі вибірки $n > 30$)

Гіпотези	Значення критерію та правило знаходження критичної точки
H_0: M(X) = M(Y)	$Z_{сп} = \dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{\dfrac{D_B(x_1)}{n_1} + \dfrac{D_B(x_2)}{n_2}}}$
H_1: M(X) > M(Y)	$\Phi(z_{kr}) = \dfrac{1-2\alpha}{2}$
H_1: M(X) <M(Y)	$\Phi(z_{kr}) = \dfrac{1-2\alpha}{2}$
H_1: M(X) ≠M(Y)	$\Phi(z_{kr}) = \dfrac{1-\alpha}{2}$
$\overline{x_i}$ – вибіркова середня; $D_B(x_i)$– вибіркова дисперсія; – об’єм вибірки; α – рівень значущості; $z_{кр}$ знаходять в таблиці Лапласа (дод. 2).	

Таблиця 5

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

Гіпотези	Значення критерію та правило знаходження критичної точки
H_0: M(*X*) = M(*Y*)	$T_{\text{сп}} = \dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{(n_1-1)\cdot S_1{}^2 + (n_2-1)\cdot S_2{}^2}} \cdot \sqrt{\dfrac{n_1 n_2 (n_1+n_2-2)}{n_1+n_2}}$
H_1: M(*X*) > M(*Y*)	$t_{кр\_прав}(\alpha, k)$
H_1: M(*X*) <M(*Y*)	$t_{кр\_лів}(\alpha, k) = -t_{кр\_прав}(\alpha, k)$
H_1: M(*X*) ≠M(*Y*)	$t_{кр\_двоб}(\alpha, k)$
$\overline{x_i}$ – вибіркова середня; S_i^2 – виправлена дисперсія; $k = n_1 + n_2 - 2$ – число степенів свободи; $t_{кр}$ знаходять в таблиці критичних точок розподілу Стьюдента (дод. 4).	

Таблиця 6

Порівняння вибіркової середньої з гіпотетичною генеральною середньою a_0 нормальної сукупності X, генеральна дисперсія якої σ^2 відома, або вибірки великі

Гіпотези	Значення критерію та правило знаходження критичної точки
H_0: $M(X) = a_0$	$U_{сп} = \dfrac{(\bar{x} - a_0)\sqrt{n}}{\sigma}$
H_1: $M(X) > a_0$	$\Phi(u_{кр}) = (1-2\alpha)/2$
H_1: $M(X) \neq a_0$	$\Phi(u_{кр}) = (1-\alpha)/2$
H_1: $M(X) < a_0$	$\Phi(u_{кр}) = (1-2\alpha)/2$
a_0 – значення середньої, що припускають; n – об'єм вибірки; $u_{кр}$ знаходять в таблиці значень функції $\Phi(x)$ (дод. 2);	

Таблиця 7

Порівняння вибіркової середньої з гіпотетичною генеральною середньою a_0 *нормальної сукупності X, генеральна дисперсія якої* σ^2 *невідома, (невеликі незалежні вибірки* $n < 30$*)*

Гіпотези	Значення критерію та правило знаходження критичної точки
H₀: $M(X) = a_0$	$T_{сп} = \dfrac{(\overline{x} - a_0)\sqrt{n}}{S}$
H₁: $M(X) > a_0$	$t_{кр\_прав}(\alpha, k)$
H₁: $M(X) \neq a_0$	$t_{кр\_двоб}(\alpha, k)$
H₁: $M(X) < a_0$	$t_{кр\_лів}(\alpha, k) = -t_{кр\_прав}(\alpha, k)$
a_0 – значення середньої, що припускають; n – об’єм вибірки; $k = n - 1$ – число степенів свободи; $t_{кр}(\alpha, k)$ знаходять в таблиці критичних точок розподілу Стьюдента (дод. 4).	

Таблиця 8

Порівняння відносної частоти з гіпотетичною ймовірністю появи події

Гіпотези	Значення критерію та правило знаходження критичної точки
H₀: $p = p_0$	$U_{сп} = \dfrac{(\omega - p_0)\sqrt{n}}{\sqrt{p_0 q_0}}$
H₁: $p > p_0$	$\Phi(u_{кр}) = (1 - 2\alpha)/2$
H₁: $p \neq p_0$	$\Phi(u_{кр}) = (1 - \alpha)/2$
H₁: $p < p_0$	$\Phi(u_{кр}) = (1 - 2\alpha)/2$
ω - відносна частота, що спостерігається p_0 – значення ймовірності, що припускають; $q_0 = 1 - p_0$ n – об’єм вибірки; $k = n - 1$ – число степенів свободи; $u_{кр}$ знаходять в таблиці значень функції $\Phi(x)$ (дод. 2);	

Таблиця 9

Порівняння двох ймовірностей біноміальних розподілів

Гіпотези	Значення критерію та правило знаходження критичної точки
H_0: $p_1 = p_2$	$U_{сп} = \dfrac{(p_1 - p_2)}{\sqrt{pq\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$
H_1: $p_1 > p_2$	$\Phi(u_{кр}) = (1 - 2\alpha)/2$
H_1: $p_1 \neq p_2$	$\Phi(u_{кр}) = (1 - \alpha)/2$
H_1: $p_1 < p_2$	$\Phi(u_{кр}) = (1 - 2\alpha)/2$
$p_i = \dfrac{m_i}{n_i}$ - відносна частота появи події в i-тій вибірці $p = \dfrac{m_1 + m_2}{n_1 + n_2}$ – узагальнене значення ймовірності; $q = 1 - p$ n_i – об’єм i-тої вибірки; $u_{кр}$ знаходять в таблиці значень функції $\Phi(x)$ (дод. 2);	

Приклад 7. Технічна норма передбачає в середньому 40 с на виконання певної технологічної операції на конвеєрі по виробництву годинника. Від працівників на цій операції надійшли скарги, що вони насправді витрачають на неї більше часу. Для перевірки цієї скарги зроблені хронометричні вимірювання часу виконання цієї технологічної операції у 16 робітниць, зайнятих на ній, та отримано середній час виконання операції $\overline{X}$ = 42 с. Чи можна за наявними хронометричними даними на рівні значущості α = 0,01 відхилити гіпотезу про те, що середній час виконання цієї операції відповідає нормі, якщо виправлене вибіркове середнє квадратичне відхилення s – 3,5 с.

Розв’язання. Для розв’язання цієї задачі необхідно перевірити гіпотезу про те, що невідома генеральна середня нормальної сукупності точно дорівнює певному числу, коли дисперсія генеральної сукупності невідома (вибірка мала, n = 16).

Сформулюємо нульову і конкуруючу гіпотези згідно з умовою задачі.

H_0: $M(X)$ = a_0 = 40 – невідоме математичне сподівання $M(X)$ (нормально розподіленої генеральної сукупності з невідомою дисперсією) дорівнює гіпотетично передбачуваному числовому значенню a_0. Тобто нульова гіпотеза: час виконання технологічної операції відповідає нормі.

H_1: $M(X) > 40$ – невідоме математичне сподівання (нормально розподіленої генеральної сукупності з невідомою дисперсією) більше числового значення a_0 (час виконання технологічної операції більше за встановлену норму).

Оскільки конкуруюча гіпотеза – правобічна, то і критична область – правобічна.

Використаємо правило з таблиці 6.

$T_{сп}$ знайдемо за формулою:

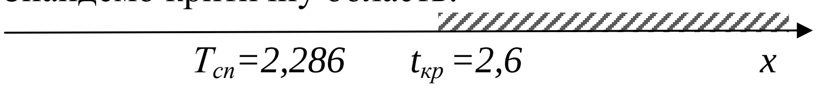
$$T_{сп} = \frac{(\bar{x} - a_0)\sqrt{n}}{S} = \frac{42-40}{3{,}5}\sqrt{16} = 2{,}286.$$

Критичне значення знайдемо за допомогою таблиць розподілу Стьюдента (дод. 4) за рівнем значущості α і числом степенів свободи k.

За умовою $\alpha = 0{,}01$; число степенів свободи $k = 16 - 1 = 15$.

Знайдемо $t_{кр}$ за рівнем значущості $\alpha = 0{,}01$ (для однобічної критичної області) и числу степенів свободи $k = 15$: $t_{кр}(0{,}01;15) = 2{,}6$.

Знайдемо критичну область:



$T_{сп}$ не належить критичній області, отже, на даному рівні значущості немає підстав відхилити нульову гіпотезу.

Відповідь. За наявними хронометричними даними на рівні значущості 0,01 не має підстав відхилити гіпотезу про те, що середній час виконання цієї операції відповідає нормі. Отже, скарги робітниць - необґрунтовані**.**

Приклад 8. Передбачається, що вживання нового типу змішувача скоротить час вимішування тіста. Хронометраж часу вимішування 9 порцій тіста старим апаратом, дав наступні результати: середній час вимішування 57 хв., виправлена вибіркова дисперсія 186,2 (хв.^{2}). Середній час вимішування 15 порцій на новому обладнанні за даними хронометражних вимірів – 52 хв., а виправлена вибіркова дисперсія 166,4 (хв.^{2}). На рівні значущості 0,01 дайте відповідь на питання, чи дозволило використання нового змішувача скоротити час приготування тіста?

Розв'язання. Для розв'язання цього завдання необхідно порівняти 2 середні нормально розподілених генеральних сукупностей, генеральні дисперсії яких невідомі, але передбачаються однаковими (малі незалежні вибірки $n_x = 9$ і $n_y = 15$).

Сформулюємо нульову і конкуруючу гіпотези, згідно з умовою завдання.

H_0: $\overline{X} = \overline{Y}$ – генеральні середні 2 нормально розподілених сукупностей з невідомими дисперсіями (але передбачуваними однаковими) рівні. Тобто нульова гіпотеза: середній час, що витрачається на вимішування тіста, – однаковий, тобто використання нового змішувача не дозволяє скоротити час на приготування тіста).

H_1: $\overline{X} > \overline{Y}$ — генеральна середня для X більше, ніж генеральна середня для Y (тобто середній час, що витрачається на вимішування старим апаратом, більше, ніж у нового, тобто використання змішувача дозволяє працювати швидше).

Оскільки конкуруюча гіпотеза – правобічна, то і критична область – правобічна.

Приступати до перевірки гіпотези про рівність генеральних середніх двох нормально розподілених сукупностей з невідомими дисперсіями можна лише в тому випадку, якщо генеральні дисперсії рівні. Інакше, таке завдання в теорії розв’язати неможливо. Тому, перш ніж перевіряти цю гіпотезу, перевіримо гіпотезу про рівність генеральних дисперсій нормальних сукупностей.

Сформулюємо нульову і конкуруючу гіпотези, згідно з умовою задачі.

H_0: $D(X) = D(Y)$ – генеральні дисперсії 2 нормально розподілених сукупностей рівні.

H_1: $D(X) > D(Y)$ — генеральна дисперсія для X більше за генеральну дисперсію для Y. Висуваємо правобічну конкуруючу гіпотезу, оскільки виправлена вибіркова дисперсія для X значно більше, ніж виправлена вибіркова дисперсія для Y (критична область – правобічна).

Використаємо правило з таблиці 1.

Значення критерію, що спостерігають, знайдемо за формулою

$$F_{сп} = \frac{S_б{}^2}{S_м{}^2} = \frac{186{,}2}{166{,}4} = 1{,}12.$$

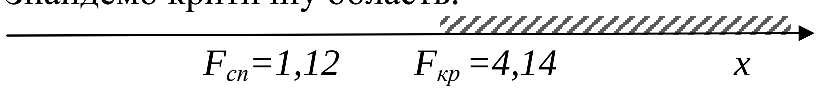
Критичне значення знайдемо за допомогою таблиці розподілу Фішера-Снедекора (додаток 5) по рівню значущості α і числу степенів свободи k_1 і k_2. За умовою α = 0,01; число степенів свободи знайдемо за формулами

$$k_1 = n_1 - 1 = 9 - 1 = 8;\ k_2 = n_2 - 1 = 15 - 1 = 14.$$

Знайдемо $F_{кр}$ за рівнем значущості $\alpha = 0{,}01$ и числу степенів свободи $k_1 = 8$ і $k_2 = 14$:

$$F_{кр}(0{,}01; 8; 14) = 4{,}14.$$

Знайдемо критичну область:



$F_{сп}$ не належить критичній області, отже, на даному рівні значущості немає підстав відхилити нульову гіпотезу.

Отже, можна приступити до перевірки гіпотези про рівність генеральних середніх двох нормально розподілених сукупностей.

Використаємо правило з таблиці 5.

$T_{сп}$ знайдемо за формулою:

$$T_{\text{сп}} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{(n_1 - 1)\cdot S_1{}^2 + (n_2 - 1)\cdot S_2{}^2}} \cdot \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

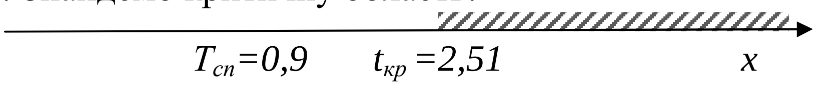
$$T_{\text{сп}} = \frac{57 - 52}{\sqrt{(9-1)\cdot 186{,}2 + (15-1)\cdot 166{,}4}} \cdot \sqrt{\frac{9\cdot 15(9+15-2)}{9+15}} = 0{,}9.$$

Критичне значення знайдемо за допомогою таблиць розподілу Стьюдента (дод. 4) за рівнем значущості α і числом степенів свободи k.

За умовою α = 0,01; число степенів свободи $k = n_1 + n_2 - 2$=9+15-2=22.

Знайдемо $t_{кр}$ за рівнем значущості α = 0,01 (для однобічної критичної області) и числу степенів свободи k = 22: $t_{кр}(0{,}01{,}22) = 2{,}51$.

. Знайдемо критичну область:



$T_{сп}$ не належить критичній області, отже, на даному рівні значущості немає підстав відхилити нульову гіпотезу, тобто середній час, що витрачається на вимішування тіста старим і новим змішувачем, відрізняється незначущо, розбіжності між середніми є випадковими, використання нового обладнання не дозволяє скоротити час вимішування..

Відповідь. На рівні значущості 0,01 не можна стверджувати, що використання нового типу різців дозволило скоротити час обробки деталі**.**

Приклад 9. Два заводи виготовляють однотипні деталі. Для оцінки їх якості зроблені вибірки із продукції цих заводів і отримані наступні результати:

Вибірки	Завод №1	Завод №2
Об'єм вибірки	*200*	300
Число бракованих деталей	*20*	15

На рівні значущості α = 0,025 визначите, чи є суттєва різниця у якості деталей, що виготовляються цими заводами?

Розв'язання. Для розв'язку цієї задачі необхідно порівняти дві ймовірності біноміальних розподілів.

Сформулюємо нульову й конкуруючу гіпотези згідно з умовою задачі.

H_0: $p_1 = p_2$ – імовірності появи події в 2 генеральних сукупностях, що мають біноміальний розподіл, рівні (імовірність браку на двох заводах однакова)

H_1: $p_1 \neq p_2$ – імовірності появи події в 2 генеральних сукупностях, що мають біноміальний розподіл, не рівні (заводи виготовляють деталі різної якості. Оскільки за умовою задачі не потрібно перевірити, на якому заводі якість деталей вище, висуваємо двобічну конкуруючу гіпотезу.)

Оскільки конкуруюча гіпотеза двобічна, то критична область теж двобічна.

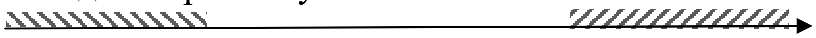
За умовою m_1=20; n_1=200; m_2=15; n_2=300; α= 0,025. Тоді відносна частота браку в 1-ій вибірці – $p_1 = \frac{m_1}{n_1} = \frac{20}{200} = 0{,}1$, в 2-ій $p_2 = \frac{m_2}{n_2} = \frac{15}{300} = 0{,}05$, узагальнене значення ймовірності $p = \frac{m_1 + m_2}{n_1 + n_2} = \frac{20 + 15}{200 + 300} = 0{,}07$; $q = 1 - p = 1 - 0{,}07 = 0{,}93$.

Значення критерію, що спостерігають, знайдемо за формулою з таблиці 9.

$$U_{сп} = \frac{(0{,}1 - 0{,}07)}{\sqrt{0{,}07 \cdot 0{,}93\left(\frac{1}{200} + \frac{1}{300}\right)}} = 1{,}288$$

Критичне значення ($u_{кр}$) знайдемо із рівності $\Phi_0(u_{кр}) = (1 - \alpha)/2$. За умовою $\alpha = 0{,}025$. Тоді $\Phi(u_{кр}) = (1 - 0{,}025)/2 = 0{,}4875$. З таблиці функції Лапласа (дод.2) $u_{кр} = 2{,}24$.

Знайдемо критичну область:



$-u_{кр}$ $\quad$ $U_{сп} = 1{,}288$ $\quad$ $u_{кр} = 2{,}24$ $\quad$ x

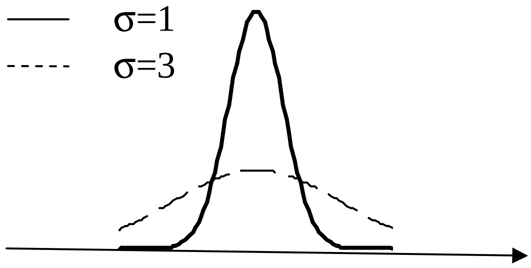
$U_{сп}$ не належить критичній області, отже, на даному рівні значущості немає підстав відхилити нульову гіпотезу.

Отже, заводи виготовляють деталі однакової якості (різниця в якості випадкова, незначна).

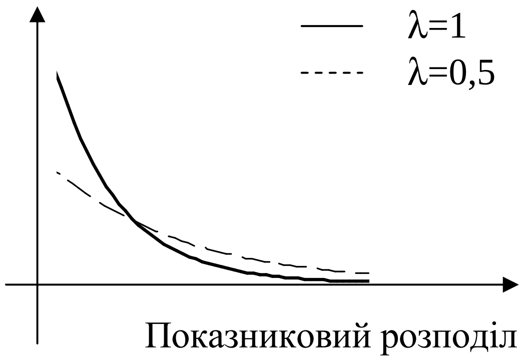
Перевірка гіпотези про розподіл

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.

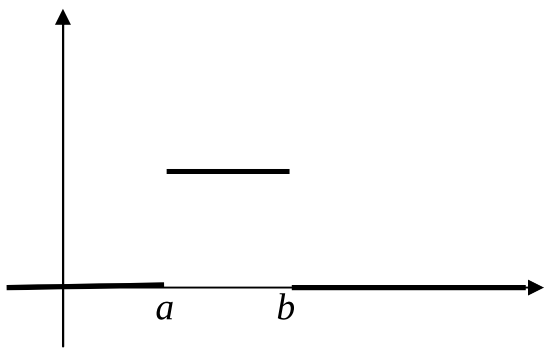
Законами розподілу випадкової величини, які найчастіше зустрічаються під час дослідженні проблем галузі, є нормальний, показниковий, пуассонівський.



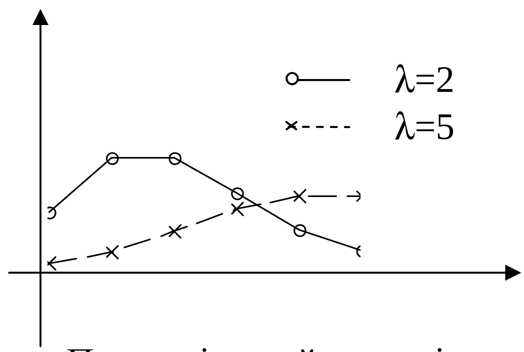
Нормальний розподіл



Показниковий розподіл



Рівномірний розподіл



Пуассонівський розподіл

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

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

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

$$\chi_{cn}^2 = \sum_{i=1}^{s} \frac{(n_i - n_i')^2}{n_i'} \qquad (15)$$

3) знайти степені свободи χ^2 за формулою $k = s - 1 - r$, де s – число груп (інтервалів групування) вибірки; r – число невідомих параметрів передбаченого закону розподілу (якщо всі параметри відомі, то r =0).

4) знайти в додатку 3 критичну точку $\chi_{kp}^2(\alpha, k)$, яка відповідає заданому рівню значущості α та степені свободи k;

5) порівняти χ_{kp}^2 та χ_{cn}^2 і зробити висновок; якщо $\chi_{cn}^2 < \chi_{kp}^2$, то гіпотеза про розподіл генеральної сукупності приймається з ймовірністю $\gamma = 1 - \alpha$, якщо $\chi_{cn}^2 > \chi_{kp}^2$, гіпотезу відкидають з тією ж ймовірністю.

Правила для знаходження теоретичних частот

Нормальний розподіл		
Формула обчислення ймовірності	Степені свободи	Значення аргументів
$P_i = \Phi(z_{i+1}) - \Phi(z_i)$	$k = s - 3$	$z_i = \frac{x_i - \bar{x}_в}{\sigma_в}$ $z_1 = -\infty; \quad z_{s+1} = \infty$

Показниковий розподіл		
$P_i = e^{-\lambda x_i} - e^{-\lambda x_{i+1}}$	$k = s - 2$	$\lambda = \dfrac{1}{\bar{x}_в}$
Рівномірний розподіл		
$P_1 = \dfrac{1}{b^* - a^*}(x_2 - a^*)$ $P_i = \dfrac{1}{b^* - a^*}(x_{i+1} - x_i)$ $P_s = \dfrac{1}{b^* - a^*}(b^* - x_s)$	$k = s - 3$	$a^* = \bar{x}_в - \sqrt{3}\sigma_в$ $b^* = \bar{x}_в + \sqrt{3}\sigma_в$
Закон Пуассона		
$P_i = \dfrac{\lambda^i \cdot e^{-\lambda}}{i!}$	$k = s - 2$	$\lambda = \bar{x}_в$ $i = 1...r$ – кількість появ події в n випробуваннях (r – максимальне число події, що спостерігали)

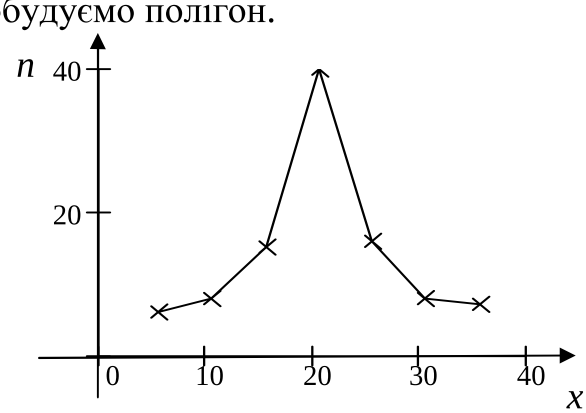
Приклад 10. Висунути гіпотезу про розподіл. Використовуючи критерій Пірсона, на рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про розподіл.

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: $xs_i = \dfrac{x_i + x_{i+1}}{2}$. Маємо розподіл

xs_i	5,5	10,5	15,5	20,5	25,5	30,5	35,5
n_i	6	8	15	40	16	8	7

Побудуємо полігон.



За графіком можна висунути гіпотезу про нормальний розподіл.

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

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

$$\bar{x}_{\text{в}} = \frac{1}{n}\sum_{i=1}^{k} n_i xs_i = \frac{1}{100}(5{,}5\cdot 6+10{,}5\cdot 8+15{,}5\cdot 15+20{,}5\cdot 40+25{,}5\cdot 16+30{,}5\cdot 8+35{,}5\cdot 7)=20{,}7;$$

$$\sigma_{\text{в}} = \sqrt{\frac{1}{n}\sum_{i=1}^{k} n_i xs_i^2 - \left(\bar{x}_{\text{в}}\right)^2} =$$

$$=\sqrt{\frac{1}{100}(5{,}5^2\cdot 6+10{,}5^2\cdot 8+15{,}5^2\cdot 15+20{,}5^2\cdot 40+25{,}5^2\cdot 16+30{,}5^2\cdot 8+35{,}5^2\cdot 7)-20{,}7^2}=7{,}28.$$

Обчислимо нормовані значення кінців інтервалів за формулою $z_i = \dfrac{x_i - \bar{x}_B}{\sigma_B}$,

вважаючи, що $z_1=-\infty$, $z_7=\infty$, та побудуємо таблицю для обчислення теоретичних частот (в таблиці Φ(z) – функція Лапласа з *додатку 2*).

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 (15)

$$\chi^2_{cn} = \frac{(6-4{,}09)^2}{4{,}09}+\frac{(8-10{,}37)^2}{10{,}37}+\frac{(15-21{,}11)^2}{21{,}11}+\frac{(40-26{,}98)^2}{26{,}98}+\frac{(16-21{,}58)^2}{21{,}58}+\frac{(8-11{,}32)^2}{11{,}32}+$$
$$+\frac{(7-4{,}55)^2}{4{,}55}=13{,}22.$$

З *додатку 3* критичних точок розподілу χ^2, за рівнем значущості $\alpha=0{,}05$ та числом степенів свободи $k=s-3=7-3=4$ знаходимо критичну точку $\chi^2_{кр}(0{,}05;4)=9{,}5$. Оскільки $\chi^2_{cn}>\chi^2_{кр}$, з ймовірністю 0,95 гіпотеза про нормальний розподіл генеральної сукупності відкидається.

Приклад 11. Власник фірми вважає, що домогтися більш високих фінансових результатів йому завадила нерівномірність поставок комплектуючих по місяцях року, незважаючи на те, що постачальник у повному обсязі виконав свої зобов’язання за півроку. Постачальник стверджує, що постачання були не так вже нерівномірними. Розподіл поставок по місяцях півріччя має наступний вигляд:

Місяць	1	2	3	4	5	6

Обсяг поставок, од.	19	23	26	18	20	20

На рівні значущості $\alpha = 0{,}05$ визначте хто правий: власник фірми чи постачальник?

Розв'язання. Сформулюємо нульову і конкуруючу гіпотези згідно з умовою задачі.

$H_0 : X \sim R(a;\, b)$ — випадкова величина X підпорядковується рівномірному розподілу з параметрами *(a; b)* (поставки були рівномірними, правий постачальник);

H_1 : випадкова величина X не підпорядковується рівномірному розподілу (правий власник фірми).

Знайдемо обсяг поставок: $n = \sum_{i=1}^{6} n_i = 18 + 23 + 24 + 18 + 17 + 20 = 120$.

За умови рівномірності поставок фірма повинна була отримати щомісяця $\frac{120}{6} = 20$ одиниць товару.

Складемо таблицю розподілу емпіричних і теоретичних частот.

$m_{(емп)i}$	18	23	24	18	17	20
$m_{(теор)i}$	20	20	20	20	20	20

Обчислимо спостережене значення критерію Пірсона за формулою

.

$$\chi^2_{сп} = \frac{(18-20)^2}{20} + \frac{(23-20)^2}{20} + \frac{(24-20)^2}{20} + \frac{(18-20)^2}{20} + \frac{(17-20)^2}{20} + \frac{(20-20)^2}{20} = 2{,}032$$

З *додатку 3* критичних точок розподілу χ^2, за рівнем значущості $\alpha = 0{,}05$ та числом степенів свободи $k = s - 1 - r = 6 - 1 - 0 = 5$ ($s = 6$ кількість поставок, $r = 0$ оскільки всі параметри розподілу відомі) знаходимо критичну точку $\chi^2_{кр}(0{,}05;5) = 11{,}1$. Оскільки $\chi^2_{сп} < \chi^2_{кр}$, на рівні значущості 0,05 немає підстав відкидати гіпотезу про рівномірний розподіл. Отже поставки були рівномірними.

Приклади розв'язання задач теми в середовищі MathCAD знаходяться в документах 5, 6.

1.5. Елементи теорії кореляції

Функціональною залежністю називається такий зв'язок між змінними величинами, коли залежна величина (функція) повністю визначається значеннями (впливаючих) 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 змінювання однієї з величин викликає змінювання розподілу іншої. Якщо при змінюванні однієї з величин змінюється середнє значення іншої, має місце **кореляційна залежність**.

Кореляційна залежність найчастіше характеризується **коефіцієнтом кореляції** $r_{\text{в}}$, який характеризує степінь лінійної функціональної залежності між X та Y. Його величина визначається за формулою

$$r_{\text{в}} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} \text{ або } r_{\text{в}} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sigma_x \sigma_y}. \qquad (16)$$

Слід пам'ятати **властивості коефіцієнта кореляції**:

1. $-1 \le r_{\text{в}} \le 1$.

2. Якщо $r_{\text{в}} = 0$, то лінійний кореляційний зв'язок відсутній, при цьому нелінійний кореляційний зв'язок є можливим.

3. Якщо абсолютна величина коефіцієнта кореляції дорівнює одиниці, тобто $|r_{\text{в}}| = 1 \Rightarrow r_{\text{в}} = \pm 1$, то між X і Y існує функціональний лінійний зв'язок.

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

Діапазон змін $\lvert r_{\text{в}} \rvert$	0,1 ÷ 0,3	0,3 ÷ 0,5	0,5 ÷ 0,7	0,7 ÷ 0,9	0,9 ÷ 0,99
Характер тісноти зв'язку	слабкий	помірний	помітний	високий	надто високий

Коефіцієнт парної кореляції, обчислений за вибірковими даними, є випадковою величиною. Зі зменшенням кількості спостережень надійність коефіцієнта кореляції падає. Тому має сенс перевірити гіпотезу H$_0$: $r_{\Gamma} = 0$ (коефіцієнт кореляції генеральної сукупності дорівнює 0) при конкуруючій гіпотезі H$_1$: $r_{\Gamma} \neq 0$. За критерій приймають випадкову величину T:

$$T_{\text{сп}} = \frac{r_B\sqrt{n-2}}{\sqrt{1-r_B^2}}, \qquad (17)$$

а критичну точку $t_{кр}(\alpha, k)$ знаходять в таблиці критичних точок розподілу Стьюдента для двобічної області за заданим рівнем значущості α та числом степенів свободи $k = n - 2$.

Якщо за умовну середню прийняти середнє арифметичне значень ознаки Y, що відповідають значенню x ознаки X $\bar{y}_x = \dfrac{y_1 + y_2 + \ldots + y_k}{k}$, 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 регресією 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 $\bar{y}_x = f(x)$.

Кореляція встановлює силу зв'язку між ознаками, а *регресія* – форму цього зв'язку.

Під час розв'язання задач найчастіше зустрічаються лінійні кореляційні залежності. 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} ($\rho_{yx} = r_{\text{в}} \dfrac{\sigma_y}{\sigma_x}$).

Класичним методом знаходження параметрів k, b є метод найменших квадратів (МНК), що дозволяє отримати формули для обчислення параметрів лінійного рівняння регресії:

$$\rho_{yx} = \frac{n\sum xy - \sum x \cdot \sum y}{n\sum x^2 - \left(\sum x\right)^2} \text{ або } \rho_{yx} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x^2}, \qquad (18)$$

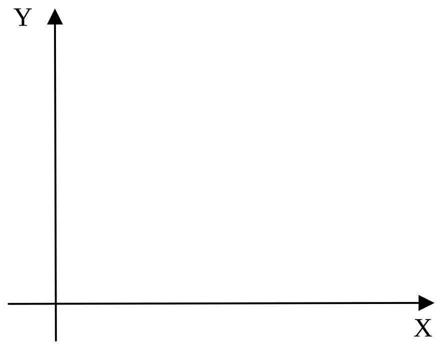
$$b = \frac{\sum x^2 \sum y - \sum x \cdot \sum xy}{n\sum x^2 - \left(\sum x\right)^2} \text{ або } b = \bar{y} - \rho_{yx}\bar{x}. \qquad (19)$$

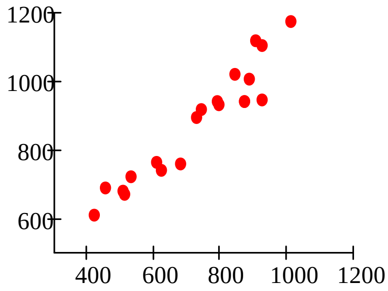
Приклад 12. Провели дослідження 20 випадково відібраних магазинів і отримали такі дані про виручку від продажу пива у банках в магазинах міста протягом дня.

номер магазину	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Число покупців	907	926	506	741	789	889	874	510	529	420	679	872	924	607	452	729	794	844	1010	621
Виручка, ум.од	1120	1105	684	921	942	1008	945	673	724	612	763	943	946	764	692	895	933	1023	1177	741

Дослідити чи є лінійна залежність між числом відвідувачів та кількістю проданого пива.

Розв'язання. В нашому випадку доцільно досліджувати залежність проданого пива (позначимо Y) від кількості відвідувачів (X). Обернена залежність не має сенсу. Представимо дані у вигляді точкової діаграми:





Діаграма показує наявність досить гарної лінійної залежності виручки від продажу пива від числа відвідувачів магазину. Видно, що зі збільшенням числа покупців зростає і виторг.

Допоміжні розрахунки можна оформити за допомогою таблиці:

Магазин	Число покупців X	Виручка Y	X^2	Y^2	XY
1	*2*	*3*	*4*	*5*	*6*
1	907	1120	822649	1254400	1015840
2	926	1105	857476	1221025	1023230
3	506	684	256036	467856	346104
4	741	921	549081	848241	682461
5	789	942	622521	887364	743238
6	889	1008	790321	1016064	896112
7	874	945	763876	893025	825930
8	510	673	260100	452929	343230
9	529	724	279841	524176	382996
10	420	612	176400	374544	257040
11	679	763	461041	582169	518077
12	872	943	760384	889249	822296
13	924	946	853776	894916	874104
14	607	764	368449	583696	463748
15	452	692	204304	478864	312784
16	729	895	531441	801025	652455
17	794	933	630436	870489	740802
18	844	1023	712336	1046529	863412
19	1010	1177	1020100	1385329	1188770
20	621	741	385641	549081	460161
Разом	14623	17611	11306209	16020971	13412790

Використовуючи дані таблиці знайдемо:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i = \frac{14623}{20} = 731{,}15; \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{k} y_i = \frac{17611}{20} = 880{,}55;$$

$$\overline{x \cdot y} = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot y_i = \frac{13412793}{20} = 670639{,}5;$$

$$\sigma_x^2 = D_x = \frac{1}{n}\sum_{i=1}^{k} x_i^2 - \left(\overline{x}\right)^2 = \frac{11306209}{20} - 731{,}15^2 = 30730{,}13;$$

$$\sigma_y^2 = D_y = \frac{1}{n}\sum_{i=1}^{k} y_i^2 - \left(\overline{y}\right)^2 = \frac{160209719}{20} - 880{,}55^2 = 25680{,}25.$$

Знайдемо коефіцієнт кореляції за формулою 16:

$$r_{в} = \frac{\overline{xy} - \overline{x}\cdot\overline{y}}{\sigma_x \cdot \sigma_y} = \frac{670639{,}5 - 731{,}15\cdot 880{,}55}{\sqrt{30730{,}13}\cdot\sqrt{25680{,}25}} = 0{,}955.$$

Значення коефіцієнта свідчить про дуже тісний лінійний зв'язок.

Знайдемо параметри рівняння регресії $\overline{y}_x = b + kx$

Тоді з формул 18 та 19 знайдемо:

$$k = \rho_{yx} = \frac{\overline{xy} - \overline{x}\cdot\overline{y}}{\sigma_x^2} = \frac{670639{,}5 - 731{,}15\cdot 880{,}55}{30730{,}13} = 0{,}873;$$

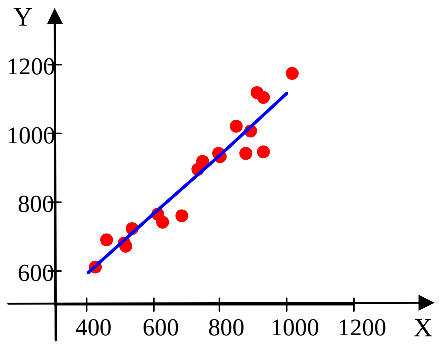
$$b = \overline{y} - \rho_{yx}\overline{x} = 880{,}55 - 0{,}873\cdot 731{,}15 = 242{,}3.$$

Отже рівняння регресії має вид

$$\overline{y}_x = 242{,}3 + 0{,}873x.$$

Коефіцієнт $k=0{,}873$ характеризує нахил лінії регресії. Це означає, що при збільшенні X на одиницю очікуване значення Y зросте на 0,873. Тобто регресійна модель указує на те, що кожний новий відвідувач магазину в середньому збільшує тижневий виторг магазину на 0,873 ум.од.

Побудуємо лінію регресії: вона проходить через точки (400, 591,5) та (1000,1115,3).



За великої кількості спостережень одне й те саме значення X може зустрічатися n_X разів, одне й те саме значення Y – n_Y разів, пара чисел (X,Y) – n_{XY} разів. Тому дані має сенс угрупувати і записати у вигляді кореляційної таблиці:

Y \ X	X_1	X_2	…	X_m	n_Y
Y_1	$n_{X_1Y_1}$	$n_{X_2Y_1}$	…	$n_{X_mY_1}$	n_{Y_1}

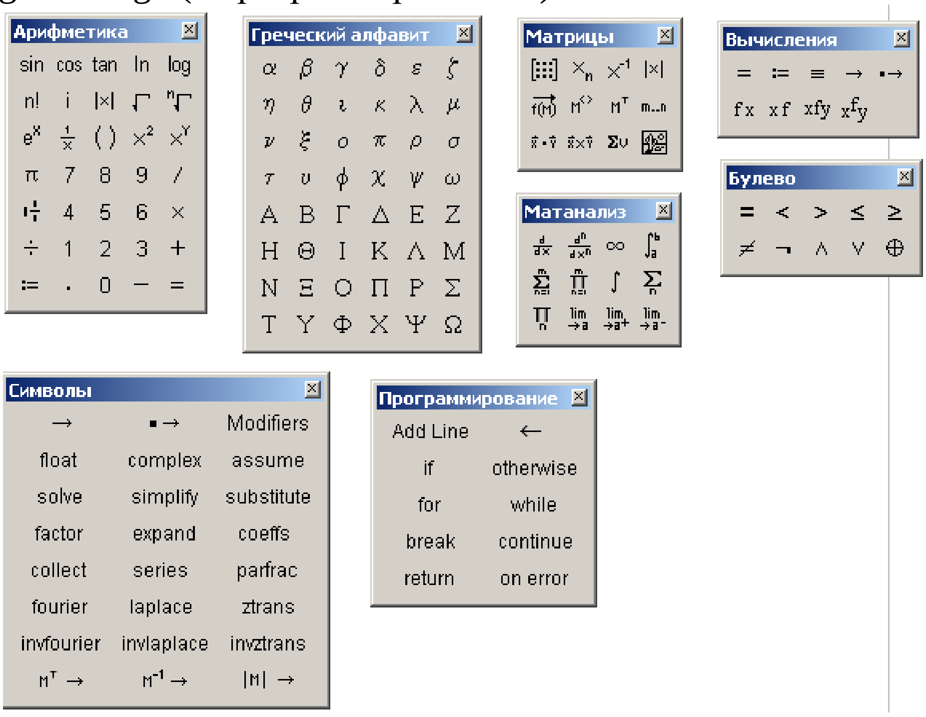
Y_2	$n_{X_1Y_2}$	$n_{X_2Y_2}$	…	$n_{X_mY_2}$	n_{Y_2}
…	…	…	…	…	…
Y_l	$n_{X_1Y_l}$	$n_{X_2Y_l}$	…	$n_{X_mY_l}$	n_{Y_l}
n_X	n_{X_1}	n_{X_2}	…	n_{X_m}	$n=\sum_{i=1}^{m} n_{X_i}=\sum_{j=1}^{l} n_{Y_j}$

Приклад побудови лінії регресії наведено в документі 7.

РОЗДІЛ 2. ЕЛЕМЕНТИ MATHCAD

Для того, щоб увійти в MathCAD, виберіть Пуск – Програми – MathSoft Apps – MathCAD [Version] Professional та натисніть Enter. На екрані з’явиться заставка MathCAD Plus, а потім саме вікно MathCAD.

MathCAD має 8 основних панелей: «Calculator» («Арифметика»), «Greek» «Греческий алфавит», «Matrix» («Матрицы»), «Calculus» («Матанализ»), «Evaluation» («Вычисления»), «Boolean» («Булево»), «Symbolic» («Символы»), «Programming» («Программирование»).



Якщо до знаку на панелі підвести курсор, то з’явиться підказка: що означає цей знак та як його набрати з клавіатури.

2.1. Робота з текстом

Для того, щоб ввести текст, необхідно виконати пункт «Text Region» («Создать текстовую область») з меню «Insert» («Вставка»), або натиснути клавішу з подвійними лапками (“), або натиснути кнопку тексту на панелі інструментів. Після того, як виконаєте одну з цих операцій, MathCAD замінить хрестик на вертикальну лінію – маркер вводу. Цей маркер буде обрамлений рамкою і позначатиме текстову область. Текст буде вводитися після маркеру, текстова область буде розширюватися до розмірів тексту. Також текст можна вставляти з інших текстових редакторів, наприклад, Word. Це можна зробити двома способами: або безпосередньо вставити в MathCAD-документ скопійований текст (Ctrl + V або Shift + Insert), або вставити його в попередньо відкриту текстову область. В першому випадку текст буде вставлений в вигляді об’єкта OLE, і для його редагування буде викликатися та програма, в якій він був створений, в другому випадку редагування можливе і в MathCAD. Перший спосіб зручніше, коли текст повинен бути добре відформотований – наприклад, в MathCAD немає системи автоматичних переносів.

2.2. Приклади простих обчислень

Весь аркуш MathCAD призначений для розрахунків. Місце, в якому буде вводитись формула позначається червоним хрестиком. Після початку вводу формули хрестик змінюється на блакитний кут, нижня лінія якого показує до якої частини виразу відноситься дія. Змінити довжину лінії можна за допомогою клавіші пробіл.

Приклад. Обчислити $\frac{1+2}{5}$ та $1+\frac{2}{5}$. В першому випадку потрібно набрати 1+2 і виділити обидві цифри $\boxed{1+2}$ після чого набрати « / ». В другому випадку виділити потрібно тільки останню 2.

Числа задаються за допомогою арабських цифр, десяткової **точки** (а не коми) та знака « – » (мінус). Наприклад, дві цілі одна десята – 2.1. Дуже великі або маленькі числа можна записувати за допомогою мантиси та порядку. Наприклад, 12.3*10^25. Знак множення * при введенні числа на екран змінюється на крапку, а операція піднесення до степені (символ « ^ » з клавіатури) відображається у вигляді надрядкового елемента. Десятинні числа мають основу 10. Діапазон їх можливих значень лежить в межах від 10^{-15} до 10^{307} (це машинний нуль та машинна нескінченність).

Імена змінних та функцій можуть складатися з великих та маленьких літер, грецьких символів, знаку підкреслювання (_), символів штрих, відсоток, нескінченність та нижніх символів (потрібно відрізняти індекс (вводиться як « [» з клавіатури) – посилання на елемент матриці або вектора та нижній регістр (вводиться як « . » з клавіатури) – варіант написання імені), але починатись повинні з букви.

Потрібно пам'ятати, що MathCAD ***відрізняє*** букви, які виглядають *однаково,* але набрані в різних мовах, наприклад, англійська A та російська або українська А, а також великі та маленькі літери.

Для того, щоб визначити змінну *x* необхідно ввести символ *x* та « := » або « ≡ » з панелі «Evaluation» («Вычисления») (з клавіатури знак « : » або « ~ » відповідно). Перший знак означає локальне, другий глобальне присвоювання. Локально визначену змінну *x* можна використовувати в обчисленнях всюди **нижче і праворуч** рівності, де вона визначена, глобальне присвоювання діє по всьому документу, але використати її в іншому глобальному присвоюванні можна тільки нижче та праворуч рівності. Щоб подивитись, чому дорівнює значення змінної, потрібно набрати *x* =.

Для того щоб обчислити будь-який вираз в діапазоні значень *x*, то необхідно обрати на панелі «Matrix» («Матрицы») команду « m..n » – задати як дискретну змінну (з клавіатури « ; »). Потрібно вказати перше значення, що приймає *x*, набрати кому та наступне значення, кінцеве значення. Наприклад, *x* := 1,1.2..2 – *x* приймає значення 1, 1.2, 1.4, 1.6, 1.8, 2. Якщо наступне значення не вказується, за замовчанням MathCAD бере крок, що дорівнює 1. Наприклад, *x* := 1..5, тоді *x* приймає значення 1, 2, 3, 4, 5.

2.3. Робота з матрицями та векторами

Вектором в MathCAD називається матриця, що має один <u>стовпець</u>. Та якщо вектор має велику кількість елементів, відображати на екрані його зручніше у вигляді рядка за допомогою функції транспонування (^{T}).

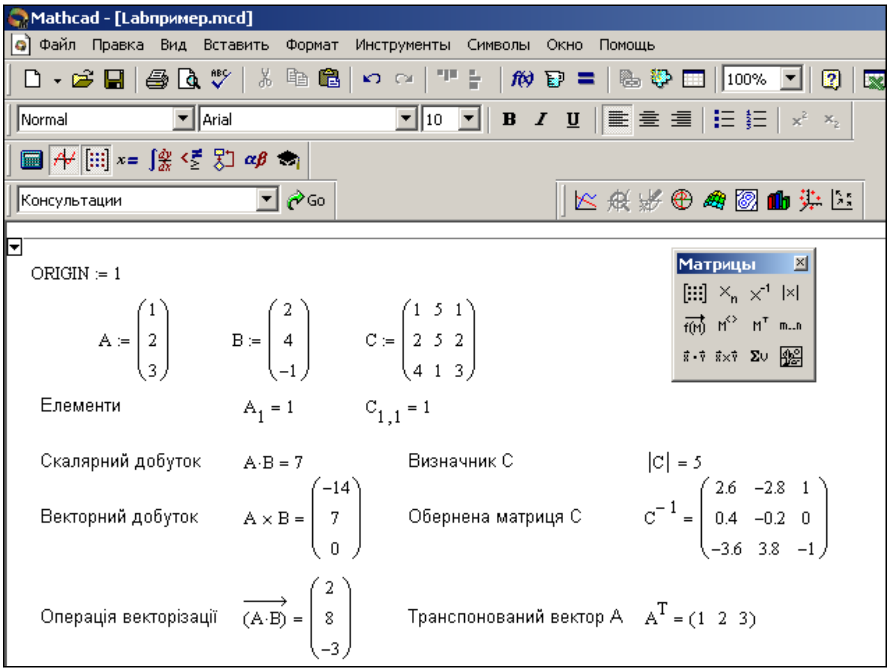
Потрібно пам'ятати, що перший елемент матриці має індекс 0. Для того щоб змінити початковий індекс на одиницю, використовують функцію ORIGIN.

Оператори для векторів та матриць		
Назва оператору	набір з клавіатури	позначення на панелі «Матрицы»
Векторний добуток	Ctrl + 8	$\vec{X} \times \vec{Y}$
Визначник	|	$\lvert X \rvert$
Скалярний добуток	*	$\vec{X} \cdot \vec{Y}$
Обернена матриця	^–1	A^{-1}
Множення матриць	*	$A \cdot B$
Нижній індекс вектор(матриця)	[	X_n
Верхній індекс (виводить вказаний стовпчик)	Ctrl + 6	$M^{\langle n \rangle}$
Транспонування	Ctrl + 1	A^T
Оператор векторизації*	Ctrl + –	$\overrightarrow{f(M)}$
Утворити матрицю	Ctrl + M	$\begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}$
Сума елементів вектору	Ctrl + 4	$\sum V$

* Операція векторізації працює так: вказана операція виконується для відповідних елементів векторів (матриць). Наприклад $\overrightarrow{A + BC}$, виконується $a_i + b_i c_i$ для всіх *i*. Всі аргументи повинні мати однакову розмірність.

Вектор (матрицю) можна задати як дискретну величину, якщо відома формула для його елементів (наприклад, вектор n=(0,1,2,3). Набираємо i:=0..3, та n_i:=i.) або за допомогою команди «Matrix or Vector» («Вставить матрицу») з панелі «Matrix» («Матрицы»): з'являється форма, в якій потрібно вказати кількість рядків та стовпців, а потім ввести елементи вектора (матриці).

Нижче приведені результати кількох операцій.



Функція	Повертає
cols(A)	число стовпців матриці A
rows (A)	число рядків матриці A
last(v)	індекс останнього елементу вектора v
length(v)	число елементів вектора v
max(A)	найбільший елемент матриці A
min(A)	найменший елемент матриці A
csort(A, n)	сортує стовпці за зростанням елементів у рядку n
rsort(A, n)	сортує рядки за зростанням елементів у стовпці n
sort (v)	сортує за зростанням елементи вектора v
reverse(A)	вектор (матрицю), елементи (рядки) якого записані у оберненому порядку
diag (v)	діагональна матриця, що має на діагоналі елементи з v
identify(n)	одиничну матрицю розміру $n \times n$ (n – натуральне)
matrix (m, n, f)	створює матрицю розміру $m \times n$, у якій (i, j)–й елемент має значення $f(i, j)$, де $i = 0..m-1$, $j = 0..n-1$.
augment(A, B)	до матриці A справа приєднується матриця B (матриці повинні мати однакову кількість рядків)
stack (A, B)	до матриці A знизу приєднується матриця B (матриці повинні мати однакову кількість стовпців)
submatrix(A,a,b,c,d)	матриця, яка складається з елементів A, що знаходяться на перетині рядків з a по b та стовпців з c по d (щоб зберегти порядок рядків та стовпців $a \geq b$, $d \geq c$)
lsolve(M,v)	вектор x, який є розв'язком рівняння Mx=v
rank(A)	ранг матриці A
lookup(z, A, B)	в матриці A шукає значення $a_{i,j} = z$, повертає

	елементи В, які мають такі саме індекси

Результати деяких функцій приведені нижче.

ORIGIN := 1

$i := 1..3$

$A_i := \sin(2 \cdot i) + 1$

$B := \begin{pmatrix} 1 & 5 \\ 5 & 2 \\ 3 & 1 \end{pmatrix}$ $\quad A = \begin{pmatrix} 1.909 \\ 0.243 \\ 0.721 \end{pmatrix}$

Елементи $\quad A_1 = 1.909 \quad B_{1,1} = 1$

Число стовпців В

$\mathrm{cols}(B) = 2$

Число рядків В

$\mathrm{rows}(B) = 3$

Індекс останнього елемента А

$\mathrm{last}(A) = 3$

Сортування В по першому стовпцю

$\mathrm{csort}(B,1) = \begin{pmatrix} 1 & 5 \\ 3 & 1 \\ 5 & 2 \end{pmatrix}$

Сортування В по другому стовпцю

$\mathrm{csort}(B,2) = \begin{pmatrix} 3 & 1 \\ 5 & 2 \\ 1 & 5 \end{pmatrix}$

Сортування А

$\mathrm{sort}(A) = \begin{pmatrix} 0.243 \\ 0.721 \\ 1.909 \end{pmatrix}$

Виділення мінору матриці В, що стоїть на перехресті другого-третього рядка та першого-другого стовпця

$\mathrm{submatrix}(B,3,2,1,2) = \begin{pmatrix} 3 & 1 \\ 5 & 2 \end{pmatrix}$

Приєднання до А матриці В справа

$\mathrm{augment}(A,B) = \begin{pmatrix} 1.909 & 1 & 5 \\ 0.243 & 5 & 2 \\ 0.721 & 3 & 1 \end{pmatrix}$

Приєднання до траспонованої матриці А транспонованої матриці В знизу

$\mathrm{stack}\left(A^T, B^T\right) = \begin{pmatrix} 1.909 & 0.243 & 0.721 \\ 1 & 5 & 3 \\ 5 & 2 & 1 \end{pmatrix}$

Виділення другого стовпця В

$B^{\langle 2 \rangle} = \begin{pmatrix} 5 \\ 2 \\ 1 \end{pmatrix}$

Виділення другого рядка В

$\left(\left(B^T\right)^{\langle 2 \rangle}\right)^T = (5 \ \ 2)$ $\quad$ або

$\mathrm{submatrix}(B,2,2,1,2) = (5 \ \ 2)$

2.4. Побудова графіків у MathCAD

Пакет MathCAD має великі графічні можливості. В MathCAD є такі види графіків: декартовий, полярний, поверхні, карта ліній рівня, векторне поле, тривимірний точковий, тривимірна стовпчаста діаграма. Всі графіки є стандартними об’єктами MathCAD: їх можна редагувати, а при перерахунку вихідних даних вони автоматично перерисовуються.

Декартовий графік будується, як правило, в три кроки:

- крок 1: формування вектора значень аргументу;
- крок 2: завдання виду функції однієї змінної;
- крок 3: побудова графіку.

Третій крок в свою чергу можна поділити ще на три кроки:

- крок 3.1: виклик на екран дисплею шаблона графіка – прямокутника з чорними квадратиками ліворуч та знизу; шаблон графіка з’являється у відміченому курсором місці після того, як користувач натисне одну з кнопок панелі «Graph» («Графики»);
- крок 3.2: заповнення користувачем двох чорних квадратів шаблона графіка ім’ям функції та ім’ям аргументу. Якщо функцій більше за одну, то їх імена вводяться через кому. В заготовці є й інші чорні квадрати, які можна не заповнювати. Середовище MathCAD заповнить їх автоматично. Графік

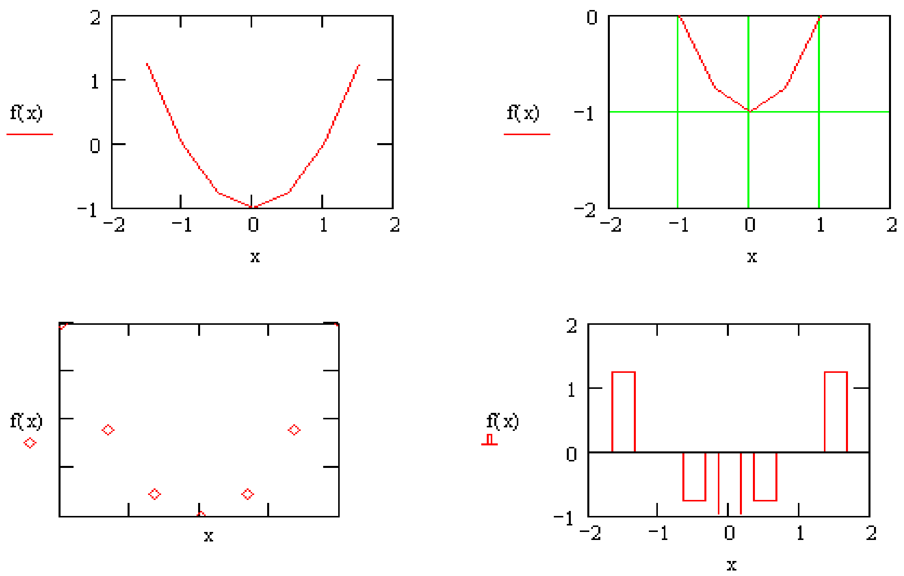
з’являється на дисплеї після виводу курсору з зони графіка. Параметри графіка задаються стандартами за замовчанням;

- крок 3.3 необхідний, якщо параметри графіка, встановлені за замовчанням не задовольняють користувача і він хоче їх змінити.

Нижче приведений графік функції x^2+1 на проміжку [–1,5, 1,5] з різними варіантами завдання сліду (проекспериментуйте самостійно).

$x := -1.5, -1 .. 1.5$

$f(x) := x^2 - 1$



2.5. Елементи програмування

Розглянемо основні команди панелі «Programming» («Программирование»)

Add Line – викликає шаблон для створення блоків виразів (його також можна викликати натиском клавіші «] ».). Шаблон має вид:

▪
▪

← – знак присвоювання змінній конкретного значення. Його можна викликати натиском клавіші « { ».

while – – викликає шаблон оператора while для створення оператору циклу з передумовою. В верхній позначці вводиться логічний (булевий) вираз, в нижній – оператор циклу. Перевірка правильності логічного виразу відбувається перед виконанням нової ітерації. Оператор можна викликати натиском клавіш Ctrl +].

if – викликає шаблон для введення умовного оператора if (якщо) в програму. Його можна викликати також натиском клавіші « } ». В другу позначку вводиться умова, а в першу – результат виконаної дії.

otherwise – дозволяє перетворити неповну альтернативу в повну (його можна викликати натиском клавіш Ctrl + }):

C ← D if A > B

E ← F otherwise

for – кнопка для вводу в програми циклу з параметром. В циклі задається початкове, кінцеве значення параметру, крок, на який змінюється параметр циклу. Введення оператора for можна також здійснити комбінацією клавіш Ctrl + ".

Приклад. Побудувати функцію розподілу для вибірки

x_i	2	3
n_i	0,3	0,7

Наберіть «$f(x)$:=» і в позначку введіть оператор Add Line двічі.

У всіх позначках введемо умовний оператор if**,** тобто натиснемо if на панелі управління. На екрані будемо мати

$$f(x) := \left|\begin{array}{l} \blacksquare \text{ if } \blacksquare \\ \blacksquare \text{ if } \blacksquare \\ \blacksquare \text{ if } \blacksquare \end{array}\right.$$

Задамо відповідні умови та вирази для обчислення функції. Для введення символів <, >, ≤ використаємо панель управління «Boolean» («Булевы»). Маємо

$$f(x) := \left|\begin{array}{l} 0 \text{ if } x \le 2 \\ 0.3 \text{ if } 2 < x \le 3 \\ 1 \text{ if } x > 3 \end{array}\right.$$

2.6. Статистичні функції

MathCAD має велику кількість вбудованих функцій. Зверніть увагу, як пишуть ім'я функції, наприклад, var і Var – це різні функції.

Функція	Повертає
hist(intvls, A)	вектор, i-ий елемент якого показує кількість елементів A що попали в інтервали $(\text{intrvl}_i, \text{intrvl}_{i+1})$, де intvls – вектор, елементи якого задають межі інтервалів у порядку зростання; (якщо довжина вектору intvls дорівнює n, довжина вектору hist дорівнює n–1) або ціле число – кількість підінтервалів однакової довжини
histogram(n,A)	матрицю розміром $n \times 2$. У першому стовпцю відображаються середини n підінтервалів однакової довжини. Елементи другого стовпця співпадають з елементами hist(n, A). Перший аргумент також може бути вектором інтервалів
mean(A)	вибіркове середнє вибірки A
var(A)	вибіркову дисперсію вибірки A
Var(A)	«виправлену» вибіркову дисперсію вибірки A
stdev(A)	середнє квадратичне відхилення вибірки A
Stdev(A)	«виправлене» середнє квадратичне відхилення вибірки A
mode(A)	моду вибірки A

Функція	Повертає
median(A)	медіану вибірки A
cvar(A, B)	коваріацію елементів вибірок A та B однакового розміру
corr(A, B)	коефіцієнт кореляції для вибірок A і B
slope(vx, vy)	скаляр: нахилення лінії регресії для даних вибірок vx та vy
intercept(vx,vy)	скаляр: зміщення по осі ординат лінії регресії для даних вибірок vx та vy

Зауважимо, що всі ***вибірки*** – аргументи статистичних функцій задаються як ***варіаційний ряд***. Якщо вибірка задана ***статистичним рядом,*** відповідні формули потрібно задавати ***вручну***.

Приклад:

	$x := (1 \quad 2 \quad 2 \quad 2)^T$	$X := (1 \quad 2)^T \quad n := (1 \quad 3)^T$
об'єм вибірки	$n := length(x) \quad n = 4$	$N := \sum_{i=1}^{2} n_i \quad N = 4$
середня вибіркова	$xv := mean(x) \quad xv = 1.75$	$Xv := \frac{1}{N}\sum_{i=1}^{2} X_i \cdot n_i \quad X = 1.75$

Розглянемо найбільш важливі функції розподілу:

Функція	Повертає
cnorm(x)	функцію щільності стандартного нормального розподілу $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ (математичне сподівання a = 0, середнє квадратичне відхилення σ = 1)
dnorm(x, a, σ)	функцію щільності нормального розподілу $\varphi_{a,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}$
pnorm(x, a, σ)	інтегральну функцію нормального розподілу $\Phi_{a,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}} dt$
qnorm(p, a, σ)	обернену функцію для pnorm(x, a, σ)
dbinom (k, n, p)	функцію щільності біноміального розподілу $C_n^k p^k q^{n-k}$ k успіхів в n випробуваннях, p – ймовірність успіху, $0 \le p \le 1$
pbinom (k, n, p)	інтегральну функцію біноміального розподілу
dexp (x, r)	функцію щільності експоненціального розподілу $r \cdot e^{-rx}$
pexp (x, r)	інтегральну функцію експоненціального розподілу

Функція	Повертає
dgeom (k, p)	функцію щільності геометричного розподілу $p^k(1-p)^{n-k}$, p – ймовірність успіху, $0 \le p \le 1$
pgeom(k, p)	інтегральну функцію геометричного розподілу
dpois(k, λ)	функцію щільності розподілу Пуассона $\frac{\lambda^k}{k!}e^{-\lambda}$, $\lambda > 0$
ppois(k, λ)	інтегральну функцію розподілу Пуассона
dunif(x, a, b)	функцію щільності рівномірного розподілу $\frac{1}{b-a}$, a та b – кінці відрізку ($a < b$)
punif(x, a, b)	інтегральну функцію рівномірного розподілу
dt(x, d)	функцію щільності t – розподілу Ст'юдента з d степенями свободи
pt(x ,d)	інтегральну функцію t – розподілу Ст'юдента
qt(p, d)	обернену функція для pt(x, d)
dF(p, $d1$, $d2$)	функцію щільності F – розподілу Фішера-Снедекора з $d1$ та $d2$ степенями свободи
pF(p, $d1$, $d2$)	інтегральну функцію F – розподілу Фішера-Снедекора
qF(p, $d1$, $d2$)	обернену функцію для pF(p, $d1$, $d2$)
dchisq(x ,d)	функцію щільності розподілу χ^2 з d степенями свободи
pchisq(x ,d)	інтегральну функцію розподілу χ^2
qchisq(x ,d)	обернену функцію для pchisq(x ,d)
rnd(x)	випадкове число, що рівномірно розподілене на інтервалі (0, x)
rnorm(m, a, σ)	вектор з m випадковими числами, що розподілені за нормальним законом

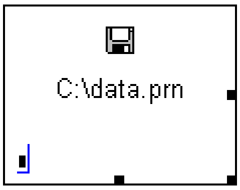
Функції розподілу можна використовувати для знаходження табличних значень

Додаток 1 $\varphi(x)$	dnorm(x, 0, 1) або cnorm(x)
Додаток 2 $\Phi(x)$	pnorm(x, 0, 1) – 0,5 для знаходження ймовірності p qnorm(p + 0.5, 0, 1) для знаходження x
Додаток 3 $\chi^2(\alpha,k)$	$\mathrm{qchisq}(1-\alpha,\mathrm{k})$
Додаток 4 $t_\gamma(\gamma,n)$	$qt\left(\frac{1+\gamma}{2},n-1\right)$
Додаток 5 $F(\alpha,k_1,k_2)$	$qF(1-\alpha,k_1,k_2)$, k_1 – число ступенів свободи більшої дисперсії,k_2 – число ступенів свободи меншої дисперсії
Додаток 6 $t(\alpha,k)$	$qt(1-\alpha,k)$, якщо область однобічна, $qt\left(1-\frac{\alpha}{2},k\right)$, якщо область двобічна

Приклади використання цих функцій дивись в документах 1–7.

2.7. Робота з даними

Якщо під час роботи ви отримали матрицю даних, з якими потрібно працювати далі, їх можна зберегти в окремому файлі. Простіше за все це зробити за допомогою команди меню: Insert (Вставка)/ Data (Данные)/ File Output (Файл вывод). З’явиться вікно, в якому потрібно обрати у якому вигляді зберегти дані (текстовий – Formatted Text (*.prn), Tab Delimited Text (*.dat, *.txt), Comma Separated Values (*.cvs); Microsoft Excel (*.xls) тощо) та вказати ім’я файла та шлях до потрібної директорії (наприклад, дані потрібно зберегти на диску С з ім’ям data.prn). Також можна обрати вид роздільника точка або кома. Після цього вибрати «Готово», якщо потрібно зберегти всі дані, або команду «Далее», якщо бажаєте зберегти окремі стовпці або рядки. З’явиться прямокутник, в якому потрібно вказати ім’я матриці з даними:



Дані (у вигляді з точкою або комою) у форматі *.txt відокремлюються одне від одного за допомогою табуляції. Для даних у форматі Comma Separated Values зберігається тільки ціла частина, числа відокремлюються крапкою з комою.

Якщо потрібно зчитати дані, збережені раніше в MathCAD або в інших програмах, можна у вільне місце поставити курсор і застосувати команду Insert (Вставка)/ Data (Данные)/ File Input (Файл ввод). Подальші дії як і під час збереження даних. Після цього потрібно вказати ім’я змінної, якій будуть присвоєні ці дані.

В MathCAD є вбудовані функції readprn(file) для зчитування та writeprn(file) для збереження даних. Аргумент цих функцій містить або повний опис шляху, наприклад, C:\Мои документи\data.prn або ім’я файлу, наприклад, data.prn, якщо дані зберігають (знаходяться) в поточному каталозі. Але деякі версії програми не підтримують російські або українські літери в імені файлу, тому звернутись до наведеного вище файлу не завжди можливо.

Можна використовувати і дані, які набрані в програмі Microsoft Word. Числа можуть бути набрані з роздільником точка або кома (2.1 або 2,1) в таблиці або відокремлені одне від одного табуляцією. В документі потрібно виділити та скопіювати необхідні дані, після чого в MathCAD набрати ім’я змінної, команду « := » та операцію «Вставка» (Ctrl + V). Програма відобразить дані у вигляді матриці.

2.8. Збереження робочого документу

Щоб зберегти робочий документ необхідно вибрати пункт «Save»

(«Сохранить») з меню «File» («Файл»), або натиснути кнопку з дискетою на панелі інструментів. Якщо файл до цього часу не зберігався, то з'явиться діалогове вікно «Save as» («Сохранить как»). Введіть ім'я файлу, вказавши повний опис шляху. Інакше MathCAD збереже файл в каталозі, в якому він встановлений, або з якого зчитаний останній робочий документ. Зберігати документ рекомендується як можна частіше, щоб не довелося у випадку зависання програми набирати текст ще раз.

2.9. Вихід з MathCAD

Для того, щоб вийти з системи MathCAD по закінченню роботи, необхідно вибрати пункт «Exit» («Выход») з меню «File» («Файл») або натиснути Alt+F4. MathCAD закриє всі відкриті ним вікна і вийде в диспетчер програм. Якщо в робочому документі будуть не збережені раніше зміни, то MathCAD виведе діалогове вікно з запитанням зберегти чи ні ці зміни.

Документ 1

Обробка вибірки невеликого об'єму

Вибіркова величина приймає наступні значення 10, 10, 10, 30, 20, 12, 10, 12, 20, 10.

1. Задамо вибіркову величину як вектора x

$x_0 := 10 \quad x_1 := 10 \quad x_2 := 10 \quad x_3 := 30 \quad x_4 := 20$

$x_5 := 12 \quad x_6 := 10 \quad x_7 := 12 \quad x_8 := 20 \quad x_9 := 10$

$x^T =$

	0	1	2	3	4	5	6	7	8	9
0	10	10	10	30	20	12	10	12	20	10

2. Знайдемо об'єм вибірки N

$N := \text{length}(x) \quad N = 10$

3. Знайдемо вариаційний ряд Y для вихідної вибірки x за допомогою вбудованої функції **sort**

$Y := \text{sort}(x) \quad Y^T =$

	0	1	2	3	4	5	6	7	8	9
0	10	10	10	10	10	12	12	20	20	30

4. Побудуємо статистичний ряд.

В масив X запишемо по 1 разу кожне значення, що приймає вибіркова величина x

$X_0 := 10 \quad X_1 := 12 \quad X_2 := 20 \quad X_3 := 30$

$X^T = (10 \;\; 12 \;\; 20 \;\; 30)$

n - кількість різних елементів

$n := \text{length}(X) \quad n = 4$

Знайдемо частоти m для всіх елементів X

$j := 0..n-1 \quad k := 0..N-1$

$m_j := \sum_k \text{if}(x_k = X_j, 1, 0) \qquad m^T = (5 \;\; 2 \;\; 2 \;\; 1)$

Знайдемо відносні частоти p для всіх елементів X

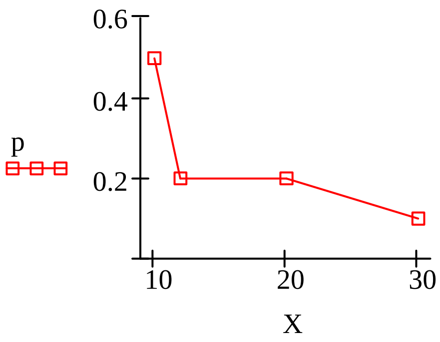
$p := \frac{m}{N}$ Перевірка $\sum_j p_j = 1$

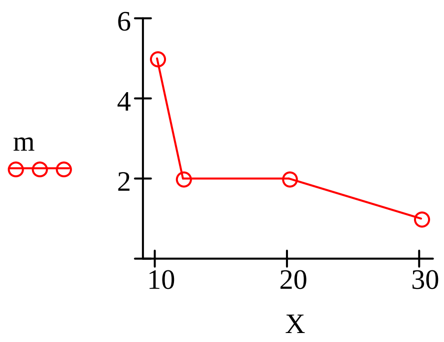
$p^T = (0.5 \;\; 0.2 \;\; 0.2 \;\; 0.1)$

Статистичний ряд

$C := \text{augment}(X, m) \qquad C^T = \begin{pmatrix} 10 & 12 & 20 & 30 \\ 5 & 2 & 2 & 1 \end{pmatrix}$

5. Побудуємо полігон частот та полігон відносних частот

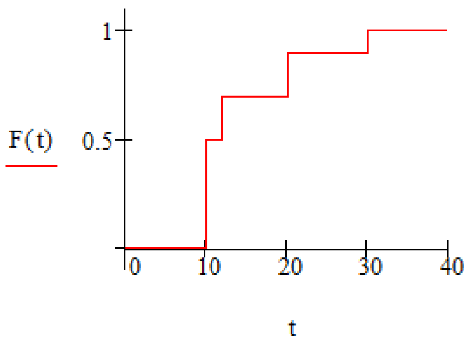




6. Побудуємо статистичну функцію розподілу

$$F(t) := \begin{vmatrix} 0 \ \text{ if } \ t \le 10 \\ 0.5 \ \text{ if } \ 10 < t \le 12 \\ 0.7 \ \text{ if } \ 12 < t \le 20 \\ 0.9 \ \text{ if } \ 20 < t \le 30 \\ 1 \ \text{ otherwise} \end{vmatrix}$$

6.1. Графік статистичної функції розподілу



Закінчення документа 1

7. Знайдемо числові характеристики X

Вибіркове середнє	$Xv := \sum_j X_j \cdot p_j$	$Xv = 14.4$
Вибіркова дисперсія	$Dv := \sum_j (X_j - Xv)^2 \cdot p_j$	$Dv = 41.44$
"Виправлена" вибіркова дисперсія	$Dvv := \frac{N}{N-1}\left[\sum_j (X_j - Xv)^2 \cdot p_j\right]$	$Dvv = 46.04$
Вибіркове середнє квадратичне відхилення	$\sigma v := \sqrt{Dv}$	$\sigma v = 6.44$
"Виправлене" середнє квадратичне відхилення	$S := \sqrt{Dvv}$	$S = 6.79$
Знайдемо максимальне та мінімальне значення	$x_{max} := \max(x)$	$x_{max} = 30$
	$x_{min} := \min(x)$	$x_{min} = 10$
Знайдемо розмах вибірки	$roz := x_{max} - x_{min}$	$roz = 20$
Медіана (N=10)	$M_e := \frac{Y_4 + Y_5}{2}$	$M_e = 11$
Мода $\max(m) = 5$	$M_0 := \mathrm{lookup}\left(\max(m), C^{\langle 1\rangle}, C^{\langle 0\rangle}\right)$	$M_0 = (10)$

Документ 2

Обробка вибірки великого об'єму.

Вибіркова величина – вектор довжини 500, елементи якого дорівнюють значення генератора випадкових чисел rnd(x), що приймає значення від 0 до 100 (x=100) і які округлені до цілого значення.

1. Отримаємо вибірку заданого об'єма N за допомогою вбудованого генератора випадкових чисел rnd та функції округлення

$N := 500 \qquad k := 0..N-1$

В масиві X - рівномірно розподілені натуральні випадкові числа

$X_k := \mathrm{round}(\mathrm{rnd}(100))$ $\qquad X^T =$

	0	1	2	3	4	5	6	7	8	9
0	81	81	90	27	23	87	84	4	18	45

2.Знайдемо Y варіаційний ряд для X

$Y := \mathrm{sort}(X)$ $\qquad Y^T =$

	0	1	2	3	4	5	6	7	8	9
0	0	0	0	1	1	1	1	1	2	2

3. Побудуємо інтервальний статистичний ряд

3.1. Знайдемо розмах вибірки $\quad R := Y_{N-1} - Y_0 \qquad R = 100$

3.2. Знайдемо число інтервалів $\quad k := \mathrm{round}(1 + 3.322 \cdot \log(N)) \qquad k = 10$

3.3. Знайдемо довжину інтервалу $\quad h := \frac{1.01 \cdot R}{k} \qquad h = 10.1$

3.4. Знайдемо межі інтервалів (масив int) та середини інтервалів (масив xs)

$i := 0..k-1$

$int_0 := Y_0 \quad int_{i+1} := int_i + h \quad xs_i := \dfrac{int_{i+1} + int_i}{2}$

$int^T =$

	0	1	2	3	4	5	6	7	8	9
0	0	10.1	20.2	30.3	40.4	50.5	60.6	70.7	80.8	90.9

$xs^T =$

	0	1	2	3	4	5	6	7	8	9
0	5.05	15.15	25.25	35.35	45.45	55.55	65.65	75.75	85.85	95.95

3.5. Знайдемо частоти для інтервалів (массив n) за допомогою функції hist

$n := hist(int, X) \quad n^T =$

	0	1	2	3	4	5	6	7	8	9
0	49	46	60	54	52	46	41	48	59	45

Перевіремо правильність обчислення частот $\sum_i n_i = 500$

3.6. Знайдемо відносні частоти

$P := \dfrac{n}{N} \quad P^T =$

	0	1	2	3	4	5	6	7	8	9
0	0.098	0.092	0.12	0.108	0.104	0.092	0.082	0.096	0.118	0.09

Перевіремо правильність обчислення відносних частот $\sum_i P_i = 1$

Побудувати інтервальний ряд можна за допомогою вбудованої функції histogram, указавши кількість інтервалів (якщо всі інтервали однакової довжини) або межі інтервалів

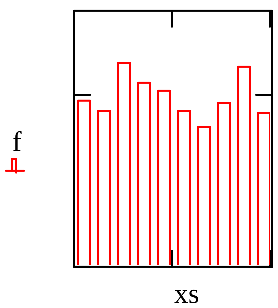
3.а. $histogram(k, Y)^T =$

	0	1	2	3	4	5	6	7	8	9
0	5.05	15.15	25.25	35.35	45.45	55.55	65.65	75.75	85.85	95.95
1	49	46	60	54	52	46	41	48	59	45

3.б $histogram(int, Y)^T =$

	0	1	2	3	4	5	6	7	8	9
0	5.05	15.15	25.25	35.35	45.45	55.55	65.65	75.75	85.85	95.95
1	49	46	60	54	52	46	41	48	59	45

4. Побудуємо гістограму щільності відносних частот (масив f) $f := \dfrac{P}{h}$



Закінчення документа 2

7. Знайдемо числові характеристики X	обчислені за інтервальним рядом		за варіаційним рядом
Вибіркове середнє	$Xv := \sum_i xs_i \cdot P_i$	$Xv = 49.147$	$mean(Y) = 49.198$
Вибіркова дисперсія	$Dv := \sum_i (xs_i - Xv)^2 \cdot P_i$	$Dv = 828.326$	$var(Y) = 827.259$
"Виправлена" вибіркова дисперсія	$Dvv := \frac{N}{N-1} Dv$	$Dvv = 829.986$	$Var(Y) = 828.917$
Вибіркове середнє квадратичне відхилення	$\sigma v := \sqrt{Dv}$	$\sigma v = 28.781$	$stdev(Y) = 28.762$
"Виправлене" середнє квадратичне відхилення	$S := \sqrt{Dvv}$	$S = 28.809$	$Stdev(Y) = 28.791$

Документ 3

Вибіркове обстеження бюджету 36 сімей виявило середньомісячний доход на одну сім'ю 1186 грн. Знайти надійний інтервал для оцінки математичного сподівання a – середньомісячного доходу всіх 10 тисяч обстежених сімей, якщо відомо, що він розподілений за нормальним законом з середнім квадратичним відхиленням σ = 25 грн. і задана надійність оцінки γ = 0,9.

Запишемо дані задачі

Об'єм вибірки $N := 36$

Середньомісячний доход $Xv := 1186$

Середнє квадратичне відхилення $\sigma := 25$

Надійність оцінки $\gamma := 0.9$

Математичне сподівання належить інтервалу $\left(\bar{x}_в - t\frac{\sigma}{\sqrt{n}}; \bar{x}_в + t\frac{\sigma}{\sqrt{n}}\right)$

Знайдемо t $\quad t := qnorm\left(\frac{\gamma}{2} + 0.5, 0, 1\right) \qquad t = 1.645$

Знайдемо межі надійного інтервалу (x0,x1)

$$x0 := Xv - \frac{t\cdot\sigma}{\sqrt{N}} \qquad x1 := Xv + \frac{t\cdot\sigma}{\sqrt{N}}$$

$$x0 = 1179.15 \qquad x1 = 1192.85$$

Документ 4

За даними вибіркового обстеження 16 студентських їдалень середня виробка на одного працівника кухні складає 10,6 блюд на годину при вибірковому середньому квадратичному відхиленні S = 1,4 бл./год. Знайти надійний інтервал для оцінки математичного сподівання *a* нормально розподіленої ознаки X – виробки на одного працівника, якщо надійна

Запишемо дані задачі

Об'єм вибірки $N := 16$

Середня виробка $Xv := 10.6$

Вибіркове середнє квадратичне відхилення $S := 1.4$

Надійність оцінки $\gamma := 0.95$

Математичне сподівання належить інтервалу $\left(\bar{x}_в - t_\gamma \frac{S}{\sqrt{n}}; \bar{x}_в + t_\gamma \frac{S}{\sqrt{n}}\right)$

Знайдемо t_γ $t\gamma := qt\left(\frac{1+\gamma}{2}, N-1\right)$ $t\gamma = 2.131$

Знайдемо межі надійного інтервалу (x0,x1)

$$x0 := Xv - \frac{t\gamma \cdot S}{\sqrt{N}} \quad x1 := Xv + \frac{t\gamma \cdot S}{\sqrt{N}}$$

$$x0 = 9.85 \quad x1 = 11.35$$

ймовірність $\gamma = 0{,}95$.

Документ 5

Для порівняння точності двох верстатів взяли дві вибірки. Отримані наступні результати вимірювання контрольного розміру:

перший верстат	1,08	1,12	1,14	1,15	1,25	1,36	1,38	1,39	1,40	1,42
другий верстат	1,11	1,13	1,15	1,26	1,26	1,27	1,30	1,39		

Чи можна вважати, що верстати мають однакову точність H_0: D(X) = D(Y), якщо прийняти рівень значущості α=0,01 та альтернативну гіпотезу H_1: D(X) ≠D(Y)

1. Занесемо дані задачі

X - результати вимірювання контрольного розміру першого верстата

$X := (1.08\ \ 1.12\ \ 1.14\ \ 1.15\ \ 1.25\ \ 1.36\ \ 1.38\ \ 1.39\ \ 1.4\ \ 1.42)^T$

Y - результати вимірювання контрольного розміру другого верстата

$Y := (1.11\ \ 1.13\ \ 1.15\ \ 1.26\ \ 1.26\ \ 1.27\ \ 1.3\ \ 1.39)^T$

Рівень значущості $\alpha := 0.1$

Нульова гіпотеза H0: D(X) = D(Y),

Альтернативна гіпотеза H1: D(X)<> D(Y)

2. Знайдемо об'єм вибірок

$N1 := \mathrm{length}(X) \qquad N1 = 10 \qquad N2 := \mathrm{length}(Y) \qquad N2 = 8$

3. Знайдемо вибіркові дисперсії

$DX := \mathrm{Var}(X) \qquad DX = 0.018$

$DY := \mathrm{Var}(Y) \qquad DY = 0.009$

4. Знайдемо Fсп

$F_{сп} := \mathrm{if}\left(DX > DY, \frac{DX}{DY}, \frac{DY}{DX}\right) \qquad F_{сп} = 1.978$

5. Знайдемо Fкр

$F_{кр} := \mathrm{if}\left(DX > DY, qF\left(1 - \frac{\alpha}{2}, N1-1, N2-1\right), qF\left(1 - \frac{\alpha}{2}, N2-1, N1-1\right)\right)$

$F_{кр} = 3.677$

6. Зробимо висновок

$висновок := \mathrm{if}\left(F_{сп} < F_{кр}, \text{"da"}, \text{"net"}\right) \qquad висновок = \text{"da"}$

Документ 6

Результати дослідження міцності на стиснення 200 зразків бетону представлені у вигляді згрупованого статистичного ряду.

інтервали міцності	частота
190–200	10
200–210	26
210–220	56
220–230	64
230–240	30
240–250	14

Потрібно перевірити нульову гіпотезу про нормальний розподіл міцності. Рівень значущості прийняти α=0,001.

1. Занесемо дані задачі

int - масив меж інтервалів

$i := 1..6 \quad int_0 := 190 \quad int_i := int_{i-1} + 10$

$int^T = (190\ \ 200\ \ 210\ \ 220\ \ 230\ \ 240\ \ 250)$

n - масив відповідних частот $n := (10\ \ 26\ \ 56\ \ 64\ \ 30\ \ 14)^T$

Рівень значущості $\alpha := 0.001$

Нульова гіпотеза H0: розподіл нормальний

2. Знайдемо об'єм вибірки $N := \sum n \qquad N = 200$

3. Знайдемо середини інтервалів - xs

$xs_{i-1} := \frac{int_i + int_{i-1}}{2} \quad xs^T = (195\ \ 205\ \ 215\ \ 225\ \ 235\ \ 245)$

4. Знайдемо характеристики ряду

4.1. Знайдемо кількість інтервалів $k := length(n) \qquad k = 6$

4.2. Знайдемо вибіркову середню $Xv := \frac{1}{N} \cdot \sum_{i=0}^{k-1} xs_i \cdot n_i \qquad Xv = 221$

4.3. Знайдемо вибіркову дисперсію $Dv := \frac{1}{N} \cdot \sum_{i=0}^{k-1} \left[\left(xs_i - Xv\right)^2 \cdot n_i\right] \qquad Dv = 152$

4.4. Знайдемо середнє квадратичне відхилення $\sigma v := \sqrt{Dv} \qquad \sigma v = 12.329$

5. Знайдемо нормовані інтервали

$z_0 := -\infty \qquad z_i := \dfrac{int_i - Xv}{\sigma v} \qquad z_6 := \infty$

$z^T = \left(-1\times10^{307}\ \ -1.703\ \ -0.892\ \ -0.081\ \ 0.73\ \ 1.541\ \ 1\times10^{307}\right)$

6. Знаходимо ймовірності P влучення в інтервал (zi-1,zi)

$P_{i-1} := pnorm(z_i, 0, 1) - pnorm(z_{i-1}, 0, 1)$

$P^T = (0.04425\ \ 0.14188\ \ 0.28154\ \ 0.29963\ \ 0.17105\ \ 0.06165)$

7. Знайдемо теоретичні частоти

$n1_{i-1} := N\cdot P_{i-1} \qquad n1^T = (8.851\ \ 28.377\ \ 56.308\ \ 59.925\ \ 34.21\ \ 12.329)$

8. Знайдемо хі-квадрат спостережене

$\chi sp := \sum_{i=0}^{k-1} \dfrac{(n_i - n1_i)^2}{n1_i} \qquad \chi sp = 1.372$

9. Знайдемо хі-квадрат критичне

$\chi kr := qchisq(1-\alpha, k-3) \qquad \chi kr = 16.266$

10. Зробимо висновок

$висновок := if(\chi sp < \chi kr, "da", "net") \qquad висновок = "da"$

Можна не знаходити нормовані інтервали (п. 5), а в функцію pnorm підставити замість 0 вибіркову середню, замість 1 – середнє квадратичне відхилення (п. 6).

$int_0 := -\infty \qquad int_6 := \infty$

$P_{i-1} := pnorm(int_i, Xv, \sigma v) - pnorm(int_{i-1}, Xv, \sigma v)$

$P^T = (0.044\ \ 0.142\ \ 0.282\ \ 0.3\ \ 0.171\ \ 0.062)$

Документ 7

Дати кількісну характеристику залежності між забезпеченням робітниками та виробництвом продукції на 100 га сільськогосподарських ділянок на підставі 15 господарств харківської області (дані наведені у таблиці).

№ з/п	Валова продукція на 100 га сільськогосподарських ділянок, тис. грн..	Чисельність робітників на 100 га сільськогосподарських ділянок, люд.(х)
1	199	7,2
2	513	16,9
3	178	10,7
4	212	6,9
5	271	9,5
6	215	11,6
7	145	8,9
8	336	10,2
9	251	7,8
10	195	4,8
11	275	7,4
12	319	10
13	375	9,6
14	232	6,6
15	242	5,8

Середньорічна чисельність робітників на 100 га сільськогосподарських ділянок є факторною ознакою (X)

Результативною ознакою (Y) є вартість валової продукції на 100 га сільськогосподарських ділянок

Занесемо дані задачі

$$Y := (199\ \ 513\ \ 178\ \ 212\ \ 271\ \ 215\ \ 145\ \ 336\ \ 251\ \ 195\ \ 275\ \ 319\ \ 375\ \ 232\ \ 242)^T$$

$$X := (7.2\ \ 16.9\ \ 10.7\ \ 6.9\ \ 9.5\ \ 11.6\ \ 8.9\ \ 10.2\ \ 7.8\ \ 4.8\ \ 7.4\ \ 10\ \ 9.6\ \ 6.6\ \ 5.8)^T$$

Знайдемо: за формулами за допомогою вбудованих функцій

Вибірковий коефіцієнт регресії

$$k := \frac{\mathrm{mean}\left(\overrightarrow{(X\cdot Y)}\right) - \mathrm{mean}(X)\cdot\mathrm{mean}(Y)}{\mathrm{var}(X)} \qquad k = 21.424 \qquad \mathrm{slope}(X, Y) = 21.424$$

Вільний доданок рівняння регресії

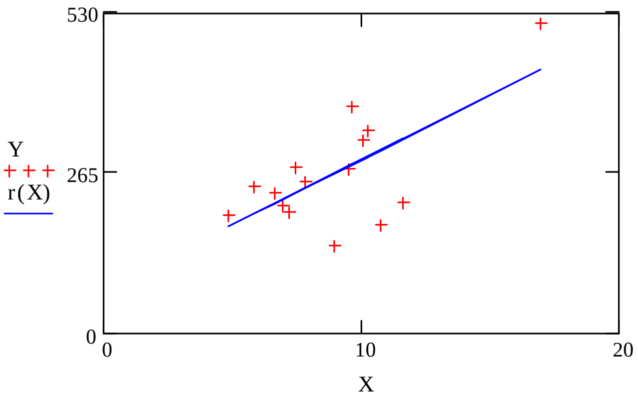
$$b := \mathrm{mean}(Y) - k\cdot\mathrm{mean}(X) \qquad b = 72.62 \qquad \mathrm{intercept}(X, Y) = 72.62$$

Коефіцієнт кореляції

$$rv := \frac{\mathrm{mean}\left(\overrightarrow{(X\cdot Y)}\right) - \mathrm{mean}(X)\cdot\mathrm{mean}(Y)}{\mathrm{stdev}(X)\cdot\mathrm{stdev}(Y)} \qquad rv = 0.678 \qquad \mathrm{corr}(X, Y) = 0.6783$$

Рівняння регресії Y на X

$r(x) := b + k \cdot x$



ЗАВДАННЯ ДЛЯ САМОСТІЙНОГО РОЗВ'ЯЗУВАННЯ

Завдання до лабораторної роботи 2:

Задача 1. Отримали дані різних соціологічних досліджень. Ранжирувати дані. Записати дані у вигляді статистичного ряду (знайти частоти та відносні частоти). Записати дані у вигляді інтервального ряду (кількість інтервалів 3, всі інтервали однакової довжини).

вар.	Дані соціологічних досліджень
1.	20 15 13 17 17 14 11 13 16 12 14 13 14 11 13 15 12 13 14 14 13 13 12 15 13 14 14 17 18 16
2.	17 18 18 19 19 20 22 20 21 17 18 21 12 13 12 17 19 20 11 19 20 22 20 21 17 18
3.	14 13 15 12 13 13 14 14 13 12 15 13 11 14 13 14 12 16 13 11 14 13 15 12
4.	190 192 192 194 196 198 200 202 204 200 192 192 190 200 190 190 192 202 200 215 220 192 200 196 202
5.	45 47 54 45 58 54 47 49 54 52 54 52 54 54 49 64 58 54 49 47 45 60 54 60 60 60
6.	11 21 25 21 15 11 35 15 35 11 35 21 15 17 21 21 21 16 21 29 11 35 11
7.	29 32 32 33 33 34 31 32 35 34 36 31 32 35 33 31 33 34 35 31 32 34 31 34 33 35 32 32 33 34 36 31 32 34 33 34 34 32 33 31 34 34
8.	23 13 13 15 18 16 17 14 18 18 19 15 16 14 14 13 15 15 13 16 14 13 16 17 14 15 13 13 18 14 17 17 19 19 13 18 14 18 15 16 16 15
9.	15 18 16 13 11 19 26 21 20 24 15 15 10 20 2 26 11 16 19 11 20 10 15 20 21 24 10 18 13 19 21 24 15 20 26 24 26
10.	25 36 16 27 22 32 25 32 21 35 23 23 22 32 25 32 21 18 17 28 24 30 30 37 26 23 25 29 30 24 32 25 32 21
11.	45 42 44 43 44 46 44 45 41 42 42 43 44 45 43 44 45 42 41 44 45 45 44 43 44 41 42 44 44 43 45 46 44 43 43 41 43 44 43 41 42 44
12.	39 41 56 47 54 69 58 60 61 45 50 47 53 64 35 56 45 63 47 49 58 60 34 45 50 39 34 44 39 63 40 58

13.	94 94 99 93 93 91 92 90 94 92 93 95 97 910 96 95 97 93 92 99 98 91 94 96 95 94 92 91 90 98
14.	12 12 13 11 12 10 12 11 12 12 12 12 13 13 12 11 11 12 11 12 13 12 12 12 12 13 13 13 13 12
15.	32 32 33 34 35 33 34 35 32 31 34 33 34 36 31 32 34 34 33 35 36 34 34 35 36 31 33 34 35 33 33 33 32 31 32 31 36 35 34 33 32 33
16.	54 69 58 60 61 45 50 47 53 64 58 54 45 50 50 50 60 34 45 50 39 34 44 49 58 60 34
17.	75 69 60 26 80 17 87 99 58 18 76 62 35 91 27 58 48 32 23 62 58 56 82 85 20 62 15 38 67 79 32 27 94 75 11 86
18.	21 78 78 32 45 59 12 53 53 56 70 65 76 79 47 64 71 36 32 33 27 19 77 76 61 33 53 52 24 15 54 60 34 46 67 66
19.	1 7 2 7 2 3 6 8 8 7 3 6 7 3 2 6 7 10 2 5 7 6 7 5 2 7 0 9 8 1 6 4 6 6 6 2 3 5 6 2
20.	14 1 3 0 2 12 12 12 9 8 11 11 4 1 1 2 11 5 2 4 12 5 4 10 1 6 9 3 12 4 5 7 13 3 10 5 3 1 5 8
21.	11 10 11 10 11 13 19 13 18 10 16 13 9 23 8 9 14 9 8 7 13 13 12 10 13 12 13 18 16 13 17 16 17 12 14 15 21 14
22.	10 8 6 6 2 6 4 3 7 5 5 11 12 11 9 7 7 1 11 10 3 10 12 13 9 3 12 4 7 3 2 4 3 11 2 7 14 4 7 11
23.	11 5 2 4 3 4 6 4 5 1 2 2 3 14 5 3 4 5 2 1 14 5 5 4 3 4 6 1 2 14 4 3 5 6 4 3 3 11 3 1 4
24.	4 4 4 2 12 3 3 4 3 14 4 5 6 2, 6 6 6 4 13 5 4 3 3 4 1 1 3 16 15 2 3 4 15 4 4 5 4, 2 4 5 4 13 3 3 3 2
25.	70 73 63 65 39 69 42 70 73 63 76 50 62 74 75 64 75 78 68 81 82 70 84 45 92 96 72 64 65 65 66 39 68 75 69 70
26.	68 67 84 90 70 73 63 73 63 76 50 62 74 75 64 75 78 68 81 82 70 84 45 92 96 80 82 92 68 72 64 69 70 64 61 74
27.	131 121 123 129 121 133 123 123 121 129 131 131 127 139 139 131 123 131 127 133 125 121 123 121 133 133 133 127
28.	120 125 125 115 135 125 125 135 130 130 130 120 120 125 120 125 135 140 140 140 125 135
29.	12 12 14 16 18 12 14 18 16 12 14 16 12 14 12 14 16 12 14 14 18 16 16 14 12 16
30.	21 25 26 24 25 25 26 21 21 25 24 26 24 25 25 21 24 26 25 25 24 26 27 25 26

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

Завдання до лабораторної роботи 3:

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

Завдання до лабораторної роботи 4:

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

Завдання до лабораторної роботи 5:

Задача 5. За допомогою випадкової вибірки оцінюється середній час щоденного перегляду телепередач абонентами кабельного телебачення в період з 18 до 22 год. Яким повинен бути обсяг вибірки в цьому випадку, якщо в попередніх вибіркових обстеженнях стандартне відхилення часу перегляду передач склало m хв, а відхилення вибіркової середньої від генеральної середньої по абсолютній величині не повинне перевищувати n хв із імовірністю γ?

вар.	m	n	γ	вар.	m	n	γ
1.	120	5	0,95	16.	100	6	0,99
2.	130	7	0,99	17.	120	5	0,99
3.	140	10	0,95	18.	135	10	0,95
4.	125	5	0,95	19.	125	10	0,95
5.	135	4	0,95	20.	125	5	0,99
6.	145	3	0,95	21.	200	15	0,999
7.	90	10	0,999	22.	220	15	0,9
8.	95	5	0,9	23.	40	5	0,99
9.	92	5	0,95	24.	45	8	0,95
10.	45	2	0,99	25.	80	5	0,95
11.	40	5	0,95	26.	85	3	0,9
12.	30	2	0,95	27.	75	2	0,95
13.	80	5	0,98	28.	125	10	0,95
14.	75	5	0,99	29.	135	10	0,95
15.	100	7	0,99	30.	130	5	0,99

Задача 6. З генеральної сукупності витягли вибірку

вар.	Дані соціологічних досліджень
1.	55 53 52 54 53 54 53 54 54 53 53 52 55 53 54
2.	37 38 23 37 38 23 32 33 33 39 20 22 20 23
3.	73 73 74 74 73 72 75 73 77 74 73 74
4.	90 92 95 94 96 98 92 92 90 90 90 95 92 96 93
5.	45 47 49 54 52 64 52 54 49 47 45 60 54 60 62 60
6.	21 32 36 41 45 18 32 32 34 43 21 29 45 43
7.	65 64 66 61 62 65 66 61 66 64 65 64 66 61 62 64 66 64 64 62
8.	23 47 44 48 48 49 45 43 45 45 43 46 44 43 46 47 44 45 44 47
9.	33 33 39 26 23 20 24 35 35 30 20 26 33 36 39 33 20 30 35 20
10.	41 35 43 43 44 34 45 34 41 30 30 37 46 44 34 45 34 41

вар.	Дані соціологічних досліджень
11.	52 52 53 55 55 53 55 55 52 55 51 52 55 55 53 55
12.	60 61 45 50 47 53 64 35 56 45 63 47 49 58
13.	14 14 12 13 15 17 110 16 15 17 15 14 12 11 10 18
14.	99 93 92 95 92 93 93 92 95 92 92 93 93 94 93 92
15.	33 34 35 32 31 34 33 33 35 36 31 33 34 35 33
16.	60 61 45 50 47 53 45 50 39 34 44 49 58 60 34
17.	17 17 22 11 11 17 12 32 27 24 15 11 16
18.	21 59 42 53 53 44 34 41 33 32 43 31 33 53 54
19.	17 16 17 15 12 17 10 19 18 11 16 14 16 16 16
20.	14 21 23 20 22 12 12 12 29 28 11 11 24 21
21.	68 60 66 63 23 63 63 62 60 63 62 63 68 66 63
22.	16 11 12 11 19 17 17 11 11 10 16 10 12 16 12 16 11
23.	14 25 25 24 23 24 26 21 22 14 24 23 25 26
24.	34 34 34 32 12 33 33 34 33 14 34 35 36
25.	69 72 70 73 63 75 78 68 81 82 70 84 75 92
26.	73 63 76 74 68 81 82 70 84 45 82 92 68 72 64 69
27.	18 23 20 16 23 28 21 29 30 31 27 39 14 24 27 34 25 28
28.	17 17 17 12 13 25 22 12 12 12 51 15 11 11
29.	13 13 27 47 57 77 92 12 18 18 18 28
30.	13 53 33 23 13 27 27 37 17 37 17 12 12 12 28 28 18 28

Оцінити з надійністю 0,95 математичне сподівання *a* нормально розподіленої ознаки генеральної сукупності за допомогою надійного інтервалу.

Завдання до лабораторної роботи 6:

Задача 7. 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? (Спочатку перевірити гіпотезу про рівність дисперсій).

вар.	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)≠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)≠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)≠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)≠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)≠M(Y)

вар.	n_x	m_x	s_x^2	n_y	m_y	s_y^2	α	H_1
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)≠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)≠M(Y)
14	8	62	130,2	11	42	129,5	0,02	M(X)>M(Y)
15	12	65	95,2	10	53	120	0,01	M(X)≠M(Y)
16	12	65	95,2	10	53	120	0,01	M(X)>M(Y)
17	5	92	140,2	10	75	123	0,01	M(X)≠M(Y)
18	5	92	140,2	10	75	123	0,01	M(X)>M(Y)
19	7	105	111,2	6	40	120,5	0,02	M(X)≠M(Y)
20	7	105	111,2	6	40	120,5	0,02	M(X)>M(Y)
21	5	52,6	93,7	8	37	112,3	0,02	M(X)≠M(Y)
22	5	52,6	93,7	8	37	112,3	0,02	M(X)>M(Y)
23	4	67,2	145,4	12	58,1	192,4	0,02	M(X)≠M(Y)
24	4	67,2	145,4	12	58,1	192,4	0,02	M(X)>M(Y)
25	11	63	95,1	12	51,2	143,2	0,02	M(X)≠M(Y)
26	11	63	95,1	12	51,2	143,2	0,02	M(X)>M(Y)
27	10	75	103,4	7	71	133,2	0,02	M(X)≠M(Y)
28	10	75	103,4	7	71	133,2	0,02	M(X)>M(Y)
29	11	65	95	8	45	141,2	0,01	M(X)≠M(Y)
30	11	65	95	8	45	141,2	0,01	M(X)>M(Y)

Завдання до лабораторної роботи 7:

Задача 8. Використовуючи критерій Пірсона, при рівні значущості α перевірити, чи узгоджується гіпотеза H_0 про розподіл генеральної сукупності з емпіричним розподілом вибірки?

вар.	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]	

вар.	H_0: нормальний розподіл сукупності						α
	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
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	[200;207]	[207;214]	[214;221]	[221;228]	[228;235]	[235;242]	
	n_i	2	24	51	47	30	8	0,025
13.	x_i	[0;3]	[3;6]	[6;9]	[9;12]	[12;15]	[15;18]	
	n_i	10	50	48	69	30	12	0,025
14.	x_i	[40;42]	[42;44]	[44;46]	[46;48]	[48;50]	[50;52]	
	n_i	8	20	36	47	21	10	0,05
15.	x_i	[120;124]	[124;128]	[128;132]	[132;136]	[136;140]	[140;144]	
	n_i	7	37	40	30	21	10	0,05
16.	x_i	[10;12]	[12;14]	[14;16]	[16;18]	[18;20]	[20;22]	
	n_i	7	10	42	55	31	9	0,01
17.	x_i	[5;7]	[7;9]	[9;11]	[11;13]	[13;15]	[15;17]	
	n_i	14	32	60	51	26	9	0,05
18.	x_i	[500;510]	[510;520]	[520;530]	[530;540]	[540;550]	[550;560]	
	n_i	20	25	35	51	31	2	0,01
19.	x_i	[300;320]	[320;340]	[340;360]	[360;380]	[380;400]	[400;420]	
	n_i	15	37	40	51	31	18	0,05
20.	x_i	[700;705]	[705;710]	[710;715]	[715;720]	[720;725]	[725;730]	
	n_i	9	32	44	49	31	5	0,05
21.	x_i	[190;195]	[195;200]	[200;205]	[205;210]	[210;215]	[215;220]	
	n_i	15	23	56	55	20	16	0,025
22.	x_i	[101;107]	[107;113]	[113;119]	[119;125]	[125;131]	[131;137]	
	n_i	4	26	50	71	30	13	0,025
23.	x_i	[50;53]	[53;56]	[56;59]	[59;62]	[62;65]	[65;68]	
	n_i	3	50	45	45	21	12	0,025
24.	x_i	[110;112]	[112;114]	[114;116]	[116;118]	[118;120]	[120;122]	
	n_i	8	20	46	46	20	10	0,05
25.	x_i	[120;124]	[124;128]	[128;132]	[132;136]	[136;140]	[140;144]	
	n_i	7	37	40	30	21	10	0,05
26.	x_i	[30;32]	[32;34]	[34;36]	[36;38]	[38;40]	[40;42]	
	n_i	5	12	41	33	21	8	0,01

вар.	H_0: рівномірний розподіл сукупності						α
27	x_i	[20;21]	[21;22]	[22;23]	[23;24]	[24;25]	[25;26]	
	n_i	24	30	35	36	31	29	0,01
28	x_i	[–10;0]	[0;10]	[10;20]	[20;30]	[30;40]	[40;50]	

	n_i	20	40	37	42	41	35	0,05

вар.	H_0: показниковий розподіл сукупності							α
29	x_i	[0;5]	[5;10]	[10;15]	[15;20]	[20;25]	[25;30]	
	n_i	250	90	40	10	5	5	0,025
30	x_i	[0;10]	[10;20]	[20;30]	[30;40]	[40;50]	[50;60]	
	n_i	460	154	50	20	10	6	0,05

Завдання до лабораторної роботи 8:

1. Туристична компанія пропонує місця в готелях приморського курорту. Менеджера компанії цікавить, наскільки зростає привабливість готелю залежно від її відстані до пляжу. З цією метою по 14 готелях міста була з'ясована середньорічна наповнюваність номерів і відстань у кілометрах від пляжу

Відстань, км	0,1	0,2	0,2	0,3	0,4	0,4	0,5	0,6	0,7	0,7	0,8	0,8	0,9	0,9
Наповнюваність, %	90	95	96	90	89	86	90	83	85	80	78	76	72	75

Побудуйте графік вихідних даних і визначите по ньому характер залежності. Розрахуйте вибірковий коефіцієнт лінійної кореляції Пирсона, перевірте його значущість при $\alpha = 0,05$. Побудуйте рівняння регресії і дайте інтерпретацію отриманих результатів.

2. Компанію з прокату автомобілів цікавить залежність між пробігом автомобілів (X) і вартістю щомісячного технічного обслуговування (Y). Для з'ясування характеру цьому зв'язку було відібрано 15 автомобілів.

X	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Y	13	16	15	20	19	21	26	24	30	32	30	35	34	40	39

Побудуйте графік вихідних даних і визначите по ньому характер залежності. Розрахуйте вибірковий коефіцієнт лінійної кореляції Пирсона, перевірте його значущість при $\alpha = 0,05$. Побудуйте рівняння регресії і дайте інтерпретацію отриманих результатів.

3. Лікар-дослідник з'ясовує залежність площі ураженої частини легенів у людей, що занедужали емфіземою легенів, від числа років паління. Статистичні дані, зібрані їм у деякій області, мають такий вигляд:

Число років паління	25	36	22	15	48	39	42	31	28	33
Уражена частина легенів, %	55	60	50	30	75	70	70	55	30	35

Побудуйте графік вихідних даних і визначите по ньому характер залежності. Розрахуйте вибірковий коефіцієнт лінійної кореляції Пирсона, перевірте його значущість при $\alpha = 0,05$. Побудуйте рівняння регресії і дайте інтерпретацію отриманих результатів. Зробіть прогноз про ступінь поразки легенів у випадково вибраного пацієнта, хворого емфіземою, що паліть 30 років.

4. Компанія, що займається продажем радіоапаратури, встановила на відеомагнітофон певної моделі ціну, диференційовану по регіонах. Наступні дані показують ціни на відеомагнітофон в 8 різних регіонах і відповідне їм число продажів.

Число продаж, шт.	420	380	350	400	440	380	450	420

Ціна, грн	550	600	650	600	500	580	450	500

Побудуйте графік вихідних даних і визначите по ньому характер залежності. Розрахуйте вибірковий коефіцієнт лінійної кореляції Пирсона, перевірте його значущість при $\alpha = 0{,}05$. Побудуйте рівняння регресії і дайте інтерпретацію отриманих результатів.

5. Опитування випадково вибраних 10 студентів, що проживають у гуртожитку університету, дозволяє виявити залежність між середнім балом за результатами попередньої сесії й числом годин у тиждень, витрачених студентом на самостійну підготовку..

Середній бал	4,6	4,3	3,8	3,8	4,2	4,3	3,8	4,0	3,1	3,9
Число годин	25	22	9	15	15	30	20	30	10	17

Побудуйте графік вихідних даних і визначите по ньому характер залежності. Розрахуйте вибірковий коефіцієнт лінійної кореляції Пирсона, перевірте його значущість при $\alpha = 0{,}05$. Побудуйте рівняння регресії і дайте інтерпретацію отриманих результатів. Якщо студент займається самостійно по 12 год. у тиждень, то який прогноз його успішності?

6. Деяка компанія нещодавно провела рекламну кампанію в магазинах з демонстрацією антисептичних якостей свого нового мийного засобу. Через 10 тижнів компанія вирішила проаналізувати ефективність цього виду реклами, зіставивши щотижневі обсяги продажів з витратами на рекламу (тис. грн).

обсяг продаж, тис. грн	72	76	78	70	68	80	82	65	62	90
витрати на рекламу, тис. грн	5	8	6	5	3	9	12	4	3	10

Побудуйте графік вихідних даних і визначите по ньому характер залежності. Розрахуйте вибірковий коефіцієнт лінійної кореляції Пирсона, перевірте його значущість при $\alpha = 0{,}05$. Побудуйте рівняння регресії і дайте інтерпретацію отриманих результатів

7. Припустимо, що ми маємо випадкову вибірку з 10 домогосподарств для вивчення зв'язку між числом холодильників у домогосподарстві й кількістю членів домогосподарства. X - число членів домогосподарства; Y - число холодильників.

X	6	2	4	3	4	4	6	3	2	2
Y	4	1	3	2	2	3	4	1	2	2

Побудуйте графік вихідних даних і визначите по ньому характер залежності. Розрахуйте вибірковий коефіцієнт лінійної кореляції Пирсона, перевірте його значущість при $\alpha = 0{,}05$. Побудуйте рівняння регресії і дайте інтерпретацію отриманих результатів.

8. Є вибіркові дані про стаж роботи (X, років) і виробітку одного робітника за зміну (Y, шт.)

X	1	3	4	5	6	7
Y	14	15	18	20	22	25

Побудуйте графік вихідних даних і визначите по ньому характер залежності. Розрахуйте вибірковий коефіцієнт лінійної кореляції Пирсона, перевірте його значущість при $\alpha = 0{,}05$. Побудуйте рівняння регресії і дайте інтерпретацію отриманих результатів

9. Вивчається залежність собівартості одиниці виробу (Y, тис. грн) від величини випуску продукції (X, тис. шт.) по групах підприємств за звітний період. Економіст обстежив 5 підприємств і одержав наступні дані:

X	*2*	3	4	5	6
Y	1,9	1,7	1,8	1,6	1,4

Вважаючи, що між Y і X має місце лінійна залежність, визначите вибіркове рівняння лінійної регресії й поясніть зміст отриманих коефіцієнтів. Які значимість коефіцієнта кореляції, напрямок і тіснота зв'язку між показниками Y і X, якщо рівень значущості принять рівним 0,05?

10. Є вибіркові дані про глибину оранки полів під озимі культури (X, див) і їхньої врожайності (Y, ц/га):

X	10	15	20	25	30
Y	5	10	16	20	24

При α = 0,05 установити значимість статистичного зв'язку між ознаками X и Y. Якщо ознаки корелюють, побудуйте рівняння регресії й поясніть його зміст. Зробіть прогноз урожайності пшениці при глибині оранки 22 см.

11. Зі студентів 4- го курсу одного з факультетів університету відібрані випадковим чином 10 студентів і підраховані середні оцінки, отримані ними на 1- м (X) і 4- м (Y) курсі. Отримані наступні дані:

X	3,5	4,0	3,8	4,6	3,9	3,0	3,5	3,9	4,5	4,1
Y	4,2	3,9	3,8	4,5	4,2	3,4	3,8	3,9	4,6	3,0

Вважаючи, що між Y і X має місце лінійна залежність, визначите вибіркове рівняння лінійної регресії й поясніть зміст отриманих коефіцієнтів. Які значимість коефіцієнта кореляції, напрямок і тіснота зв'язку між показниками Y і X, якщо рівень значущості принять рівним 0,05?

12. Визначте тісноту зв'язку між віком літака (X, років) і вартістю його експлуатації (Y, млн грн) за наступним даними:

X	1	2	3	4	5
Y	2	4	5	8	10

Встановіть значимість коефіцієнта кореляції. Якщо він значимий, то побудуйте рівняння регресії й поясніть його зміст. Яким буде прогноз вартості експлуатації літака, якщо його вік 1,5 року, а рівень значущостіпринять рівним 0,05?

13. Визначите тісноту зв'язку обсягу випуску продукції (X, тис. шт.) і собівартості одиниці виробу (Y, тис. грн) на основі наступних даних:

X	3	4	5	6	7
Y	10	8	7	5	2

Перевірте значимість вибіркового коефіцієнта кореляції при α = 0,05. Побудуйте рівняння лінійної регресії й поясните його.

14. Визначите тісноту зв'язку загальної ваги деякої рослини (X, г) і ваги його насінь (Y, г) на основі наступних вибіркових даних:

X	40	50	60	70	80	90	100
Y	20	25	28	30	35	40	45

Перевірте значимість вибіркового коефіцієнта кореляції при α=0,05. Побудуйте лінійне рівняння регресії й поясните його.

15. При дослідженні залежності часу, витраченого на закріплення деталі на токарському верстаті, від ваги деталі, отримані наступні результати (X - вага деталі, кг, Y - час закріплення деталі, с):

X	7	8	10	12	13	14	15	17	18	20
Y	2,2	2,3	2,4	2,5	2,6	2,7	2,8	3,0	3,1	3,2

Вважаючи, що між Y і X має місце лінійна залежність, визначите вибіркове рівняння лінійної регресії й поясніть зміст отриманих коефіцієнтів. Які значимість коефіцієнта кореляції, напрямок і тіснота зв'язку між показниками Y і X, якщо рівень значущості принять рівним 0,05?

16. На підставі емпіричних даних по 10 підприємствах ЖКГ треба побудувати лінійну модель, яка характеризує залежність між витратами за використані матеріали (x) та обсягом наданих послуг (y);

Витрати на матеріали (тис. грн)	35,2	26,0	12,8	18,4	11,2	18,2	5,0	9,1	27,4	11,2
Обсяг послуг (тис. грн.)	25,4	21,7	9,8	15,7	9,1	16,0	6,3	7,3	22,5	9,2

Перевірте значимість вибіркового коефіцієнта кореляції при α=0,05.

17. Наступні дані отримані з випадкової вибірки по оборотах 8 річних консолідованих балансів. Цифри в таблиці показують обсяг продажу тис. шт., і ціну одиниці товару, грн

Продаж	12,2	18,6	29,2	15,7	25,4	35,2	14,7	11,17
Ціна	29,2	30,5	29,7	31,3	30,8	29,9	27,8	27,0

Розрахуйте вибірковий коефіцієнт кореляції Пирсона між обсягом продаж і ціною товару. Перевірте значимість коефіцієнта кореляції для α = 0,05.

18. Перед здачею іспитів наприкінці семестру в 20 групах студентів університету було проведено опитування про те, яку оцінку по здаваних у сесію курсах вони очікують одержати. Після сесії середні отримані оцінки були зіставлені із середніми очікуваними. Результати наведені в таблиці:

очікувана	3,4	3,1	3,0	2,8	3,7	3,5	2,9	3,7	3,5	3,2
одержана	4,1	3,4	3,3	3,0	4,7	4,6	3,0	4,6	4,6	3,6
очікувана	3,0	3,5	3,3	3,1	3,3	3,9	2,9	3,2	3,4	3,4
одержана	3,5	4,0	3,6	3,1	3,3	4,5	2,8	3,7	3,8	3,9

19. Визначите тісноту зв'язку собівартості одиниці виробу (X, тис. шт.) і обсягу випуску продукції (Y, тис. грн) на основі наступних даних:

X	10	8	7	5	2
Y	3	4	5	6	7

Перевірте значимість вибіркового коефіцієнта кореляції при α = 0,05. Побудуйте рівняння лінійної регресії й поясните його.

20. Є дані по 14 підприємствах про продуктивність праці (Y, шт.) і коефіцієнт механізації робіт (X, %)

X	32	30	36	40	41	47	56	54	60	55	61	67	69	76
Y	20	24	28	30	31	33	34	37	38	40	41	43	45	48

Перевірте значимість вибіркового коефіцієнта кореляції при α = 0,05. Побудуйте рівняння лінійної регресії й поясните його.

21. Менеджером фірми одержано залежність між часом *X* реалізації партії продукції (дні) і величиною партії *Y* (тис. шт.). Результати дослідження наведені в таблиці.

X	13	16	17	19	20	23	25	26
Y	0,52	1,4	1,6	2,2	2,7	3,6	4,1	4,48

Потрібно:
1) встановити форму залежності між ознаками *X* та *Y*;
2) знайти рівняння лінійної регресії *Y* на *X*;
3) оцінити силу лінійного зв'язку і перевірити гіпотезу про значущість коефіцієнта кореляції;

22. За даними аудиторного звіту про діяльність кооператиційних банків

№ банку	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Кредитна ставка, $	59	61	64	66	68	61	64	64	66	67	66	62	59	68	68
Дохідність від кредитних операцій	18	24	35	31	29	25	36	32	30	31	30	28	35	37	34

Визначте функцію, яка описує залежність між розміром кредитної ставки та дохідністю від кредитних операцій, обчисліть параметри рівняння, пояснити їх зміст; перевірте зв'язок на істотність з ймовірністю 0,95. Зробити висновки.

23. Про вік і середній виробіток робітників підприємства відомі такі дані:

Номер групи робітників	1	2	3	4	5	6	7	8
Вік робітників, років	20	25	30	35	40	45	50	55
Середньоденний виробіток, грн.	18	22	25	31	34	38	46	52

Знайти рівняння регресії, яке визначає залежність середньоденного виробітку від віку робітників, і лінійний коефіцієнт кореляції, а також оцінити його істотність (α=0,05).

24. На основі статистичних даних про значення факторного показника *x* – витрати підприємства на придбання нового устаткування, тис. у .о., та результативного показника *y* – прибутку від продажу виготовленої продукції, тис. у.о.:
1) оцінити щільність зв'язку між показниками;
2) побудувати лінійну регресійну модель, оцінити всі її параметри.

Проілюструвати залежність між ознаками графічно.

X	100	150	200	250	300	350	400	450
Y	1050	1000	930	860	700	490	413	333

25. Наведено дані про залежність між видобутком вугілля одним робітником за зміну *Y* (т) і потужністю пласта *X* (м), які характеризують процес видобутку вугілля в n = 17 шахтах.

xi	21	22	3	10	12	15	6	7	17	8	18	10	12	18	15	15	17
yi	2	3	13	6	6	5	3	6	6	5	7	8	7	8	9	8	10

Перевірте значимість вибіркового коефіцієнта кореляції при α = 0,05. Побудуйте рівняння лінійної регресії й поясните його.

26. Досліджується залежність продуктивності праці (у) від рівня механізації робіт (х) по даним 14 промислових підприємств (і - порядковий номер підприємства). Статистичні дані наведені в таблиці.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
x_i	39	43	44	50	59	57	63	58	64	70	72	79	35	33
y_i	31	33	34	36	37	40	41	43	44	46	48	51	23	27

Потрібно знайти оцінки параметрів лінійної регресії у на х. Побудувати діаграму розсіювання й нанести пряму регресії на діаграму розсіювання. На рівні значимості $\alpha = 0{,}05$ перевірити гіпотезу про згоду лінійної регресії з результатами спостережень.

27. По 10 магазинах, що належать одному торговельному підприємству, відомі величина річного товарообігу в (млн. грн) і торговельна площа кожного магазину х (тис. кв. м). Дані представлені в таблиці.

x	0,24	0,31	0,55	0,48	0,78	0,98	0,94	1,21	1,29	1,12
y	19,76	38,09	40,95	41,08	56,29	68,51	75,01	89,05	91,13	91,26

Побудувати рівняння лінійної регресії й пояснити зміст його параметрів. Розрахувати лінійний коефіцієнт кореляції.

28. Визначити зв'язок між виходом цукру з 1 т цукрових буряків та їх цукристістю за такими даними:

Номер заводу	1	2	3	4	5	6	7	8
Цукристість, %	14,8	15,2	17,3	18,4	17,6	16,4	14,3	16,8
Вихід цукру з 1 т буряків, кг	132	134	140	146	138	141	136	144

Оцінити щільність зв'язку за допомогою лінійного коефіцієнта кореляції.

29. Менеджером фірми одержано залежність між часом X реалізації партії продукції (дні) і величиною партії Y (тис. шт.). Результати дослідження наведені в таблиці.

X	13	16	17	19	20	23	25	26
Y	0,52	1,4	1,6	2,2	2,7	3,6	4,1	4,48

Потрібно: встановити форму залежності між ознаками X та Y; знайти рівняння лінійної регресії Y на X; оцінити силу лінійного зв'язку і перевірити гіпотезу про значущість коефіцієнта кореляції;

30. Залежність між вологістю борошна в % і виходом хліба в грамах з 1 кг борошна задана таблицею, де X - вологість борошна (%), Y - випічка хліба (у г з 1 кг борошна):

X	13,5	13,6	13,7	13,8	13,9	14,0	14,1	14,2	14,3	14,4
Y	1362	1368	1357	1363	1360	1346	1354	1347	1359	1348

Визначити тісноту зв'язку між X і Y і скласти рівняння регресії y_x.

Залікове завдання: аналіз продуктів харчування.

Лабораторія проводить аналіз продуктів харчування з метою визначення в них шкідливих речовин. З певним видом продуктів працюють 2 лаборанти, результати аналізу порівнюють. Продукти надходять з двох пунктів. Лабораторія має дати висновок, де продукти більш “чисті”. Крім того керівника лабораторії цікавить: відрізняються за точністю результати експериментів у першого та другого лаборантів. Лаборантам було запропоновано незалежно проаналізувати одні ті самі зразки, для яких потрібно було визначити вміст шкідливої речовини X. В одиниці об’єму продукту кількість X не повинне перевищувати 0,015. Дані вимірювань наведені нижче.

Потрібно за даними спостережень обчислити статистичні оцінки.

Сформулюйте та перевірте статистичні гіпотези, на основі яких можна з’ясувати:

- чи можна обом пунктам видати сертифікат якості;
- однакова чи ні кваліфікація першого та другого лаборантів (тобто чи відрізняються у них значимо результати аналізів);
- скільки зразків достатньо взяти для дослідження на першому та другому пунктах?

Дані досліджень

Лаборант 1, пункт 1

X_i	0,011	0,012	0,0127	0,013	0,0138	0,014	0,015	0,0156	0,017	0,018
n_i	2	2	7	16	30	35	20	5	2	1

Лаборант 1, пункт 2

X_i	0,012	0,0128	0,0135	0,014	0,0147	0,00156	0,016
n_i	1	2	5	10	4	2	1

Лаборант 2, пункт 1

X_i	0,01	0,012	0,0135	0,0142	0,0149	0,0152	0,016	0,0175	0,019
n_i	2	10	17	30	25	17	5	3	1

Лаборант 2, пункт 2

X_i	0,0115	0,0127	0,0136	0,0142	0,015	0,0152
n_i	1	1	3	10	3	1

ДОДАТКИ

Додаток 1 – значення функції $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

X	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3725	0,3712	0,3697
0,4	0,3683	0,3668	0,3653	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0.2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0.1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0,0387	0,0379	0,0371	0,0363
2,2	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001
4,0	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001

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

x	0	1	2	3	4	5	6	7	8	9
0,0	0,00000	0,00399	0,00798	0,01197	0,01595	0,01994	0,02392	0,02790	0,03188	0,03586
0,1	0,03983	0,04380	0,04776	0,05172	0,05567	0,05962	0,06356	0,06749	0,07142	0,07535
0,2	0,07926	0,08317	0,08706	0,09095	0,09483	0,09871	0,10257	0,10642	0,11026	0,11409
0,3	0,11791	0,12172	0,12552	0,12930	0,13307	0,13683	0,14058	0,14431	0,14803	0,15173
0,4	0,15542	0,15910	0,16276	0,16640	0,17003	0,17364	0,17724	0,18082	0,18439	0,18793
0,5	0,19146	0,19497	0,19847	0,20194	0,20540	0,20884	0,21226	0,21566	0,21904	0,22240
0,6	0,22575	0,22907	0,23237	0,23565	0,23891	0,24215	0,24537	0,24857	0,25175	0,25490
0,7	0,25804	0,26115	0,26424	0,26730	0,27035	0,27337	0,27637	0,27935	0,28230	0,28524
0,8	0,28814	0,29103	0,29389	0,29673	0,29955	0,30234	0,30511	0,30785	0,31057	0,31327
0,9	0,31594	0,31859	0,32121	0,32381	0,32639	0,32894	0,33147	0,33398	0,33646	0,33891
1,0	0,34134	0,34375	0,34614	0,34849	0,35083	0,35314	0,35543	0,35769	0,35993	0,36214
1,1	0,36433	0,36650	0,36864	0,37076	0,37286	0,37493	0,37698	0,37900	0,38100	0,38298
1,2	0,38493	0,38686	0,38877	0,39065	0,39251	0,39435	0,39617	0,39796	0,39973	0,40147
1,3	0,40320	0,40490	0,40658	0,40824	0,40988	0,41149	0,41308	0,41466	0,41621	0,41774
1,4	0,41924	0,42073	0,42220	0,42364	0,42507	0,42647	0,42785	0,42922	0,43056	0,43189
1,5	0,43319	0,43448	0,43574	0,43699	0,43822	0,43943	0,44062	0,44179	0,44295	0,44408
1,6	0,44520	0,44630	0,44738	0,44845	0,44950	0,45053	0,45154	0,45254	0,45352	0,45449
1,7	0,45543	0,45637	0,45728	0,45818	0,45907	0,45994	0,46080	0,46164	0,46246	0,46327
1,8	0,46407	0,46485	0,46562	0,46638	0,46712	0,46784	0,46856	0,46926	0,46995	0,47062
1,9	0,47128	0,47193	0,47257	0,47320	0,47381	0,47441	0,47500	0,47558	0,47615	0,47670
2,0	0,47725	0,47778	0,47831	0,47882	0,47932	0,47982	0,48030	0,48077	0,48124	0,48169
2,1	0,48214	0,48257	0,48300	0,48341	0,48382	0,48422	0,48461	0,48500	0,48537	0,48574
2,2	0,48610	0,48645	0,48679	0,48713	0,48745	0,48778	0,48809	0,48840	0,48870	0,48899
2,3	0,48928	0,48956	0,48983	0,49010	0,49036	0,49061	0,49086	0,49111	0,49134	0,49158
2,4	0,49180	0,49202	0,49224	0,49245	0,49266	0,49286	0,49305	0,49324	0,49343	0,49361
2,5	0,49379	0,49396	0,49413	0,49430	0,49446	0,49461	0,49477	0,49492	0,49506	0,49520
2,6	0,49534	0,49547	0,49560	0,49573	0,49585	0,49598	0,49609	0,49621	0,49632	0,49643
2,7	0,49653	0,49664	0,49674	0,49683	0,49693	0,49702	0,49711	0,49720	0,49728	0,49736
2,8	0,49744	0,49752	0,49760	0,49767	0,49774	0,49781	0,49788	0,49795	0,49801	0,49807
2,9	0,49813	0,49819	0,49825	0,49831	0,49836	0,49841	0,49846	0,49851	0,49856	0,49861
3,0	0,49865	0,49869	0,49874	0,49878	0,49882	0,49886	0,49889	0,49893	0,49896	0,49900
3,1	0,49903	0,49906	0,49910	0,49913	0,49916	0,49918	0,49921	0,49924	0,49926	0,49929
3,2	0,49931	0,49934	0,49936	0,49938	0,49940	0,49942	0,49944	0,49946	0,49948	0,49950
3,3	0,49952	0,49953	0,49955	0,49957	0,49958	0,49960	0,49961	0,49962	0,49964	0,49965
3,4	0,49966	0,49968	0,49969	0,49970	0,49971	0,49972	0,49973	0,49974	0,49975	0,49976
3,5	0,49977	0,49978	0,49978	0,49979	0,49980	0,49981	0,49981	0,49982	0,49983	0,49983
3,6	0,49984	0,49985	0,49985	0,49986	0,49986	0,49987	0,49987	0,49988	0,49988	0,49989
3,7	0,49989	0,49990	0,49990	0,49990	0,49991	0,49991	0,49992	0,49992	0,49992	0,49992
3,8	0,49993	0,49993	0,49993	0,49994	0,49994	0,49994	0,49994	0,49995	0,49995	0,49995
3,9	0,49995	0,49995	0,49996	0,49996	0,49996	0,49996	0,49996	0,49996	0,49997	0,49997
4,0	0,499968									
4,5	0,49997									
5,0	0,4999997									

Додаток 3 *Критичні точки розподілу* χ^2

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

Додаток 4 Критичні точки розподілу Стьюдента

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

Додаток 5 *Критичні точки розподілу Фишера-Снедекора (K_1 - число ступенів свободи більшої дисперсії, K_2 - число ступенів свободи меншої дисперсії)*

	Рівень значущості α = 0,01											
	K_1											
K_2	*1*	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6082	6106
2	98,49	99,01	90,17	99,25	99,33	99,30	99,34	99,36	99,36	99,40	99,41	99,42
3	34,12	38,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,96	14,80	14,66	14,54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4.71
11	9,86	7,20	6,22	5,67	5,32	5,07	4,88	4,72	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55
17	8,40	6,11	5,18	4,67	4,44	4,10	3,93	3,79	3,68	3,59	3,52	3,45

	Рівень значущості α = 0,05											
	K_1											
K_2	*1*	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Вища математика : посібник для самостійного вивчення курсу : Посібник / В.Г. Гула, М.С. Синєкоп та ін.; ХДУХТ – Харків, 2007. – 303с.

2. Вища математика у вправах і задачах: Посібник/ В.Г. Гула, М.С. Синкоп та ін.. – Х.: ХДУХТ, 2004.

3. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистики. – М.: Высшая школа, 2003.

4. Жилюк Н.О., Синкоп М.С., Шинкар С.М. Методичні вказівки для самостійної роботи студентів з курсу «Теорія ймовірностей та математична статистика» – Х.: ХДУХТ, 2006.

5. Демченко Т.В., Симоненко В.І. Методичні вказівки для організації самостійної роботи студентів з курсу «Вища математика» розділ «Теорія ймовірностей та математична статистика» – Х.:ХДУХТ, 2005.

6. Полевич В.В. Теорія ймовірностей та математична статистика. Х. : ХДУХТ, 2007.

Навчальне видання

Укладачі: Манжос Наталія Володимирівна,
Півненко Андрій Олександрович,
Пархоменко Лариса Олександрівна

Математичні методи в технології

МЕТОДИЧНІ ВКАЗІВКИ
з розв'язання задач
з використанням ПЕОМ

для студентів технологічних спеціальностей

Підп. до друку , формат 60×84. Папір газ. Друк. офс. Умов. друк. арк. .
Обл.–вид. арк. . Умовн. фарб.–відб. . Тир. прим. Зам № .
Харківський державний університет харчування та торгівлі
61051, Харків–51, вул. Клочківська, 333.

ДОД ХДУХТ Харків–51, вул. Клочківська, 333