Державна служба України з надзвичайних ситуацій

Національний університет цивільного захисту України

**Черкаський інститут пожежної безпеки
ім. Героїв Чорнобиля**

МЕТОДИЧНІ ВКАЗІВКИ ТА ЗАВДАННЯ ДЛЯ ВИКОНАННЯ КОНТРОЛЬНОЇ РОБОТИ З ДИСЦИПЛІНИ МАТЕМАТИЧНІ МЕТОДИ В ПСИХОЛОГІЇ СЛУХАЧАМИ ІІ КУРСУ ЗАОЧНОГО НАВЧАННЯ

Кафедра вищої математики та інформаційних технологій

Черкаси, 2017 рік

Загальні методичні вказівки

Виконувати роботу необхідно за варіантом, який видає кафедра. Робота, виконана не за своїм варіантом, не перевіряється, не зараховується і повертається слухачеві.

Титульна сторінка оформлюється згідно встановленого зразка.

Умови усіх задач повинні записуватися повністю, графіки виконуватись охайно.

Розв’язування прикладів і задач повинні супроводжуватись усіма обчисленнями та формулами, які використовуються, а також короткими поясненнями.

Якщо робота не зараховується, слухач виправляє її і представляє ще раз.

Контрольна робота містить 10 варіантів. Варіант вибирають за останньою цифрою шифру залікової книжки.

Контрольна робота виконується на друкованих аркушах формату А4 і здається в паперовому та електронному форматах.

ВСТУП

Крім аудиторних занять, навчальні плани з навчальної дисципліни «Математичні методи в психології» передбачають самостійну роботу здобувачів вищої освіти, яка має на меті формування пізнавальної активності, засвоєння основних вмінь та навичок роботи з навчальним матеріалом, поглиблення та розширення вже здобутих знань, підвищення рівня організованості студентів.

Особливої уваги при самостійному опрацюванні потребують розділи, за якими не передбачено лекції.

У процесі самостійної роботи студенти мають оволодіти вміннями та навичками:

- організації самостійної навчальної діяльності;
- самостійної роботи в бібліотеці з каталогами;
- праці з навчальною, навчально-методичною, науковою, науково-популярною літературою;
- обчислювати числові характеристики вибіркової сукупності;
- перевіряти відповідність емпіричного розподілу з теоретичним;
- застосовувати параметричні і непараметричні критерії для обробки експериментальних даних;
- аналізувати номінативні дані;
- досліджувати кореляційний зв'язок між ознаками, що вивчаються;

Кожен здобувач вищої освіти повинен вміти раціонально організовувати свою навчальну діяльність. Важливим є вміння скласти план своєї роботи, чітко визначити її послідовність. Необхідно, щоб план самостійного навчання був реальним і його виконання давало плідні наслідки у навчальному процесі.

Для успішної самостійної роботи значну частину часу здобувач вищої освіти присвячує роботі в бібліотеці. Треба розуміти сутність складання алфавітного й тематичного каталогів, вміти швидко знаходити в них необхідну літературу, знати особливості бібліографічного шифрування. Для плідної роботи з літературними джерелами здобувачу вищої освіти корисно скласти бібліографію, заповнити бібліографічні картки на необхідні книги, брошури або статті. Для роботи у бібліотеці треба знати її структуру, спеціалізацію окремих підрозділів, вміти користуватися різноманітними каталогами, правильно заповнювати бланки вимог на літературу тощо.

Для виконання завдань для самостійної роботи здобувачу вищої освіти доцільно скористатися електронними версіями конспектів лекцій та методичних вказівок з практичних занять з дисципліни.

Теоретичні відомості

Випадкові величини.

Вище розглядалися випадкові події, які є якісною характеристикою випадкового результату досліду. Для отримання кількісної характеристики вводиться поняття випадкової величини.

Означення. *Випадковою величиною* називається величина, яка в результаті досліду може приймати те чи інше значення, причому заздалегідь відомо яке саме.

Випадкові величини можна розділити на дві категорії.

Означення. *Дискретною випадковою величиною* називається така величина, яка в результаті досліду може приймати певні значення з певною ймовірністю, що утворюють нумеровану множину.

Ця множина може бути як скінченною, так і нескінченною.

Наприклад, кількість пострілів до першого влучання в ціль є дискретною випадковою величиною, тому що ця величина може приймати і нескінченне, хоча і обраховану кількість значень.

Означення. *Неперервною випадковою величиною* називається така величина, яка може приймати будь-які значення з деякого скінченного або нескінченного проміжку.

Очевидно, що число можливих значень неперервної випадкової величини нескінченне.

Для задання випадкової величини недостатньо просто вказати її значення, необхідно також вказати ймовірність цього значення.

Закон розподілу дискретної випадкової величини.

Означення. Співвідношення між можливими значеннями випадкової величини та їх ймовірностями називається *законом розподілу дискретної випадкової величини*.

Закон розподілу може бути заданий аналітично, у вигляді таблиці або графічно.

Таблиця відповідностей значень випадкової величини та їх імовірностей називається *рядом розподілу.*

Графічне представлення цієї таблиці називається *многокутником розподілу.*

При цьому сума усіх ординат многокутника розподілу представляє собою імовірність усіх можливих значень випадкової величини, а, отже, рівна одиниці.

По цілі проводиться 5 пострілів. Імовірність влучення для кожного пострілу рівна 0,4. Знайти імовірність числа влучень і побудувати многокутник розподілу.

• Імовірності п’яти влучень з п’яти можливих, чотирьох з п;яти і трьох з п’яти були знайдені вище по формулі Бернуллі і рівні відповілно:

$$P_{5,5} = 0{,}01024, \quad P_{4,5} = 0{,}0768, \quad P_{3,5} = 0{,}2304.$$

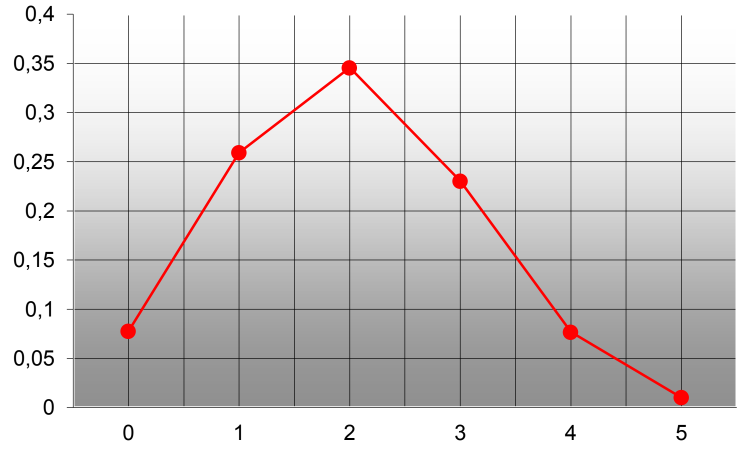
Аналогічно знайдемо:

$$P_{2,5} = \frac{5!}{2!\cdot 3!} 0{,}4^2 \cdot 0{,}6^3 = 0{,}3456;$$

$$P_{1,5} = \frac{5!}{1!\cdot 4!} 0{,}4^1 \cdot 0{,}6^4 = 0{,}2592;$$

$$P_{0,5} = \frac{5!}{0!\cdot 5!} 0{,}4^0 \cdot 0{,}6^5 = 0{,}6^5 = 0{,}0778.$$

Представимо графічно залежність числа влучень від їх імовірностей.



При побудові многокутника розподілу потрібно пам'ятати, що з'єднання отриманих точок носить умовний характер. У проміжках між значеннями випадкової величини імовірність не приймає ніякого значення. Точки є'єднані лише для наочності.

Біноміальний розподіл.

Якщо проводиться n незалежних випробувань, в кожному з яких подія A може з'явитись з однаковою імовірністю p у кожному з випробувань, то імовірність того, що подія не з'явиться, рівна $q=1-p$.

Приймемо число появ подій у кожному випробуванні за деяку випадкову величину X. Щоб знайти закон розподілу цієї випадкової величини, необхідно визначити значення цієї величини та їх імовірності.

Значення знайти досить просто. У результаті n випробувань подія може не з'явитись зовсім, з'явитись один раз, двічі і т.д. до n разів.

Імовірність кожного значення цієї випадкової величини можна знайти по формулі Бернуллі:

$$P_n(k) = C_n^k p^k q^{n-k}, \qquad k = 0,1,2,... \tag{9}$$

Ця формула аналітично виражає шуканий закон розподілу. Цей закон розподілу називається *біноміальним*.

У партії 10% нестандартних деталей. Навмання відібрані 4 деталі. Записати біноміальний закон розподілу дискретної випадкової величини X – числа нестандартниї деталей серед чотирьох відібраних і побудувати многокутник цього розподілу.

● Імовірність появи нстандартної деталі у кожному випадку рівна 0,1. Знайдемо імовірності того, що серед відібраних деталей:

1) взагалі немає нестандартних: $P_4(0) = \frac{4!}{0! \cdot 4!} 0{,}1^0 \cdot 0{,}9^4 = 0{,}6561.$

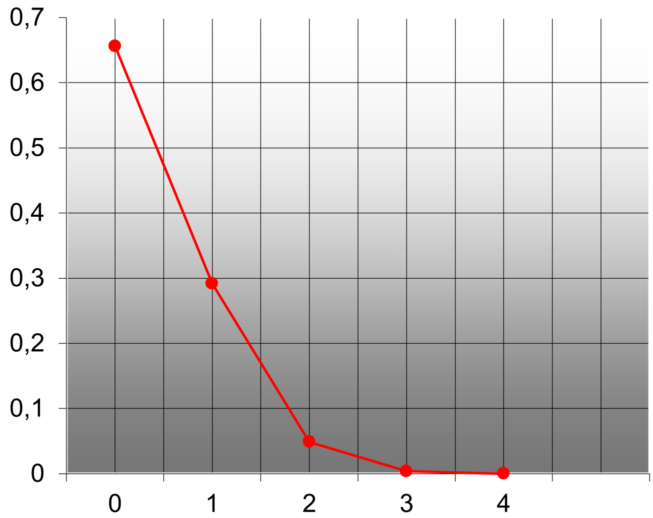
2) одна нестандартна: $P_4(1) = \frac{4!}{1! \cdot 3!} 0{,}1^1 \cdot 0{,}9^3 = 0{,}2916.$

3) дві нестандартні деталі: $P_4(2) = \frac{4!}{2! \cdot 2!} 0{,}1^2 \cdot 0{,}9^2 = 0{,}0486.$

4) три нестандартні деталі: $P_4(3) = \frac{4!}{3! \cdot 1!} 0{,}1^3 \cdot 0{,}9^1 = 0{,}0036.$

5) чотири нестандартні: $P_4(4)=\dfrac{4!}{4!\cdot 0!}0{,}1^4\cdot 0{,}9^0=0{,}0001.$

Побудуємо многокутник розподілу:



Дві гральні кістки одночасно кидають двічі. Написати біноміальний закон розподілу дискретної випадкової величини X – числа випадань парної кількості очок на двох гральних кістках.

● Кожна гральна кістка має три варіанти парних очок – 2, 4 і 6 з шести можливих. Таким, чином, імовірність випадання раного числа очок на одній кістці рівна 0,5.

Імовірність одночасного випадання парних очок на двох кістках рівна 0,25.

Імовірність того, що при двох випробуваннях обидва рази випали парні очки на обох кістках, рівна:

$$P_2(2)=\frac{2!}{0!\cdot 2!}0{,}25^2\cdot 0{,}75^0=0{,}0625.$$

Імовірність того, що при двох випробуваннях один раз випали парні очки на обох кістках:

$$P_2(1)=\frac{2!}{1!\cdot 1!}0{,}25^1\cdot 0{,}75^1=0{,}375.$$

Імовірність того, що при двох випробуваннях жодного разу не випаде парного числа очок на обох кістках:

$$P_2(0)=\frac{2!}{0!\cdot 2!}0{,}25^0\cdot 0{,}75^2=0{,}5625.$$

Розподіл Пуасона.

Нехай проводиться n незалежних випробувань, у яких поява події A має імовірність p. Якщо випробувань n достатньо велике, а імовірність появи події A у кожному досліді мале ($p\le 0{,}1$), то знаходження імовірності появи події A k разів проводиться таким чином.

Зробимо важливе припущення: добуток np зберігає стале значення:

$$np=\lambda.$$

Практично це припущення означає, що середнє число появи події в різних серіях випробувань (при різному n) є незмінним.

По формулі Бернуллі отримаємо:

$$P_n(k)=\frac{n(n-1)(n-2)\ldots(n-(k-1))}{k!}\,p^k(1-p)^{n-k};$$

$$P_n(k)=\frac{n(n-1)(n-2)\ldots(n-(k-1))}{k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}.$$

Знайдемо границю цієї імовірності при $n \to \infty$.

$$P_n(k) \cong \lim_{n\to\infty} \frac{n(n-1)(n-2)...(n-(k-1))}{k!} \frac{\lambda^k}{n^k}\left(1-\frac{\lambda}{n}\right)^{n-k} =$$

$$= \frac{\lambda^k}{k!}\lim_{n\to\infty}\left[\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)...\left(1-\frac{k-1}{n}\right)\left(1-\frac{\lambda}{n}\right)^{n-k}\right] = \frac{\lambda^k}{k!}\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n}\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k} = \frac{\lambda^k}{k!}e^{-\lambda}.$$

Отримаємо формулу *розподілу Пуасона*:

$$P_n(k) = \frac{\lambda^k e^{-\lambda}}{k!}. \qquad (10)$$

Якщо відомі числа λ і k, то значення імовірності можна знайти по відповідним таблицям розподілу Пуасона.

Числові характеристики дискретних випадкових величин.

Закон розподілу повністю характеризує випадкову величину. Однак, коли неможливо знайти закон розподілу, або цього не потрібно, можна обмежитись знаходженням значень, які називаються числовими характеристиками випадкової величини. Ці величини визначають деяке середнє значення, навколо якого групуються значення випадкової величини, і ступінь їх розкиданості навколо цього середнього значення.

Означення. *Математичним очікуваням* дискретної випадкової величини називається сума добутків усіх можливих значень випадкової величини на їх імовірності:

$$m_x = M(X) = x_1 p_1 + x_2 p_2 + ... + x_n p_n = \sum_{i=1}^{n} x_i p_i. \qquad (11)$$

Математичне очікування існує, якщо ряд, що стоїть у правій частині рівності, збіжний абсолютно.

З точки зору імовірності можна сказати, що математичне очікування наближено рівне середньому арифметичному значень випадкової величини, що розглядається.

Властивості математичного очікування.

1) Математичне очікування сталої величини рівне самій сталій: $M(C) = C.$

2) Сталий множник можна виносити за знак математичного очікування: $M(Cx) = CM(x).$

3) Математине очікування добутку двох незалежних випадкових величин рівний добутку їх математичних очікувань:

$$M(XY) = M(X)M(Y).$$

Ця властивість справедлива для довільної кількості випадкових величин.

4) Математичне очікування суми двох випадкових величин рівне сумі математичних очікувань доданків:

$$M(X+Y) = M(X) + M(Y).$$

Ця властивість також справедлива для довільної кількості випадкових величин.

Нехай проводиться n незалежних випробувань, імовірність появи події A у яких рівна p.

Теорема. *Математичне очікування M(X) числа появи події A в n незалежних випробуваннях рівне добутку числа випробувань на імовірність появи події у кожному випробуванні:*

$$M(X) = np.$$

Однак, математичне очікуваня не може повністю характеризувати випадковий процес. Крім математичного очікування потрібно увести величину, яка характеризує відхилення значень випадкової величини від математичног очікування.

Це відхилення рівне різниці між випадковою величиною та її математичним очікуваням. При цьому математичне очікування відхилення рівне нулю. Це пояснюється тим, що одні можливі відхилення додатні, інші від'ємні, і у результаті їх взаємного погашення отримуємо нуль.

Означення. *Дисперсією* (*розсіюванням*) дискретної випадкової величини називається математичне очікування квадрта відхилення випадкової величини від її математичного очікування:

$$D(X) = M\left[X - M(X)\right]^2. \qquad (12)$$

Для розглянутого вище прикладу закон розподілу випадкової величини має вигляд:

X	0	1	2
P	0,0625	0,375	0,5625

Знайти математичне очікування та дисперсію випадкової величини.

● Маємо: $M(X) = 0 \cdot 0{,}0625 + 1 \cdot 0{,}375 + 2 \cdot 0{,}5625 = 1{,}5.$

Можливі значення квадрата відхилення:

$$\left[x_1 - M(X)\right]^2 = (0 - 1{,}5)^2 = 2{,}25;$$
$$\left[x_2 - M(X)\right]^2 = (1 - 1{,}5)^2 = 0{,}25;$$
$$\left[x_3 - M(X)\right]^2 = (2 - 1{,}5)^2 = 0{,}25.$$

Тоді

[*X*–*M*(*X*)]^2	2,25	0,25	0,25
P	0,0625	0,375	0,5625

Дисперсія рівна: $D(X) = 2{,}25 \cdot 0{,}0625 + 0{,}25 \cdot 0{,}375 + 0{,}25 \cdot 0{,}5625 = 0{,}375.$

Однак, на практиці подібний спосіб обчислення дисперсії незручний, так як при великій кількості значень випадкової величини до громіздких обчислень.

Тому використовується інший спосіб.

Обчислення дисперсії.

Теорема. *Дисперсія рівна різниці між математичним очікуваням квадрата випадкової величини X і квадратом її математичного очікування:*

$$D(X) = M(X^2) - \left[M(X)\right]^2. \qquad (13)$$

Доведення. З урахуванням того, що математичне очікування $M(X)$ і квадрат математичного очікування $M^2(X)$ – величини сталі, то можна записати:

$$D(X) = M\left[X - M(X)\right]^2 = M\left[X^2 - 2XM(X) + M^2(X)\right] =$$
$$= M(X^2) - 2M(X)M(X) + M^2(X) = M(X^2) - M^2(X).$$
$$D(X) = M(X^2) - \left[M(X)\right]^2.$$

Застосуємо цю формулу для розглядуваного вище прикладу:

X	0	1	2
X^2	0	1	4
P	0,0625	0,375	0,5625

$$M(X^2) = 0 \cdot 0{,}0625 + 1 \cdot 0{,}375 + 4 \cdot 0{,}5625 = 2{,}625.$$

$$D(X) = 2{,}625 - [1{,}5]^2 = 0{,}375.$$

Властивості дисперсії.

1) Дисперсія сталої величини рівна нулю: $D(C) = 0.$

2) Сталий множник можна виносити за знак дисперсії, підносячи його до квадрату:

$$D(CX) = C^2 D(X).$$

3) Дисперсія суми двох незалежних випадкових величин рівна сумі дисперсій цих величин:

$$D(X+Y) = D(X) + D(Y).$$

4) Дисперсія різниці двох незалежних випадкових величин рівна сумі дисперсій цих величин:

$$D(X-Y) = D(X) + D(Y).$$

Справедливість цієї рівності витікає з властивості 2.

Теорема. *Дисперсія числа появи події A в n незалежних випробувань, у кожному з яких імовірність p появи події стала, рівна добутку числа випробувань на імовірності появи і не появи події в будь-якому випробуванні:*

$$D(X) = npq. \tag{14}$$

Середнє квадратичне відхилення.

Означення. *Середнім квадратичним відхиленням* випадкової величини X називається квадратний корінь з дисперсії:

$$\sigma(X) = \sqrt{D(X)}. \tag{15}$$

Теорема. *Середнє квадратичне відхилення суми скінченної кількості взаємно незалежних випадкових величин рівне квадратному кореню з суми квадратів середніх квадратичних відхилень цих величин:*

$$\sigma(X_1 + X_2 + ... + X_n) = \sqrt{\sigma^2(X_1) + \sigma^2(X_2) + ... + \sigma^2(X_n)}.$$

Завод випускає 96% виробів першого сорту та 4% виробів другого сорту. Навмання вибирають 1000 виробів. Нехай X – число виробів першого сорту у даній вибірці. Знайти закон розподілу, математичне очікування та дисперсію випадкової величини *X*.

● Вибір кожного з 1000 виробів можна вважати незалежним випробуванням, у якому імовірність появи виробів першого сорту однакова і рівна *p*=0,96.

Таким чином, закон розподілу може вважатися біноміальним:

$$m_x = pn = 1000 \cdot 0{,}96 = 960;$$

$$D_x = npq = 1000 \cdot 0{,}96 \cdot 0{,}04 = 38{,}4.$$

Знайти дисперсію дискретної випадкової величини *X* – число появи події *A* у двох незалежних випробуваннях, якщо імовірності появи цієї події у кожного випробуванні рівні і відомо, що $M(X)$=0,9.

● Так як випадкова величина X розподілена по біномінальнгму закону, то

$$M(X) = np = 2p = 0{,}9; \quad \Rightarrow \quad p = 0{,}45;$$

$$D(X) = npq = 2p(1-p) = 2 \cdot 0{,}45 \cdot 0{,}55 = 0{,}495.$$

Проводяться незалежні випробування з однаковою ймовірністю появи події A у кожному випробуванні. Знайти імовірність появи події A, якщо дисперсія числа появи події у трьох незалежних випробуваннях рівна 0,63.

● По формулі дисперсії біноміального закону отримуємо:

$$D(X) = npq = 3p(1-p) = 0{,}63;$$

$$3p^2 - 3p + 0{,}63 = 0;$$

$$p^2 - p + 0{,}21 = 0;$$

$$p_1 = 0{,}7; \quad p_2 = 0{,}3.$$

Досліджується пристрій, що складається з чотирьох незалежно працюючих приладів. Імовірності відмови кожного приладів рівні відповідно $p_1=0{,}3$; $p_2=0{,}4$; $p_3=0{,}5$; $p_4=0{,}6$. Знайти математичне очікування і дисперсію кількості приладів, що відмовили.

● Приймаючи за випадкову величину кількість приладів, що відмовили, бачимо, що ця випадкова величина може приймати значення 0, 1, 2, 3 або 4.

Для складання закону розподілу цієї випадкової величини необхідно визначити відповідні імовірності. Приймемо $q_i = 1 - p_i$.

1) Не відмовив жодний прилад: $p(0) = q_1q_2q_3q_4 = 0{,}7 \cdot 0{,}4 \cdot 0{,}5 \cdot 0{,}4 = 0{,}084.$

2) Відмовив один з приладів:

$$p(1) = p_1q_2q_3q_4 + q_1p_2q_3q_4 + q_1q_2p_3q_4 + q_1q_2q_3p_4 = 0{,}302.$$

3) Відмовили два прилади:

$$p(2) = p_1p_2q_3q_4 + p_1q_2p_3q_4 + p_1q_2q_3p_4 + q_1p_2p_3q_4 + q_1p_2q_3p_4 + q_1q_2p_3p_4 = 0{,}38.$$

4) Відмовили три прилади:

$$p(3) = p_1p_2p_3q_4 + p_1p_2q_3p_4 + p_1q_2p_3p_4 + q_1p_2p_3p_4 = 0{,}198.$$

5) Відмовили усі прилади:

$$p(4) = p_1p_2p_3p_4 = 0{,}036.$$

Отримаємо закон розподілу:

x	0	1	2	3	4
x^2	0	1	4	9	16
p	0,084	0,302	0,38	0,198	0,036

Математичне очікування:

$$M(X) = 0{,}302 + 2 \cdot 0{,}38 + 3 \cdot 0{,}198 + 4 \cdot 0{,}036 = 1{,}8.$$

$$M(X^2) = 0{,}302 + 4 \cdot 0{,}38 + 9 \cdot 0{,}198 + 16 \cdot 0{,}036 = 4{,}18.$$

Дисперсія: $$D(X) = M(X^2) - [M(X)]^2 = 4{,}18 - 3{,}24 = 0{,}94.$$

Функція розподілу.

У всіх розглянутих вище випадках випадкова величина визначалась шляхом задання значень самої величини та імовірностей цих значень.

Однак, такий метод застосовується не завжди. Наприклад, у випадку неперервної випадкової величини, її значення можуть заповнювати деякий довільний інтервал. Очевидно, що в цьому випадку задати усі значення випадкової величини нереально.

Навіть у випадку, коли це зробити можна, задача розв’язується складно. Розглянутий щойно приклад навіть при простій умові приводить до достатньо незручних обчислень, а якщо у задачі буде декілька сотень приладів?

Тому постає задача по можливості відмовитись від індивідуального підходу до кожної задачі і знайти, по можливості, найбільш загальний спосіб задання будь-яких типів випадкових величин.

Нехай x – дійсне число. Імовірність події, що X набуде значення, меншого x, тобто $X < x$, позначимло через $F(x)$.

Означення. *Функцією розподілу* називають функцію $F(x)$, що визначає імовірність того, що випадкова величина X у результаті випробування прийме значення, менше за x.

$$F(x) = P(X < x).$$

Функцію розподілу також називають *інтегральною функцією*.

Функція розподілу існує як для неперервних, так і для дискретних випадкових величин. Вона повністю характеризує випадкову величину і є однією з форм закону розподілу.

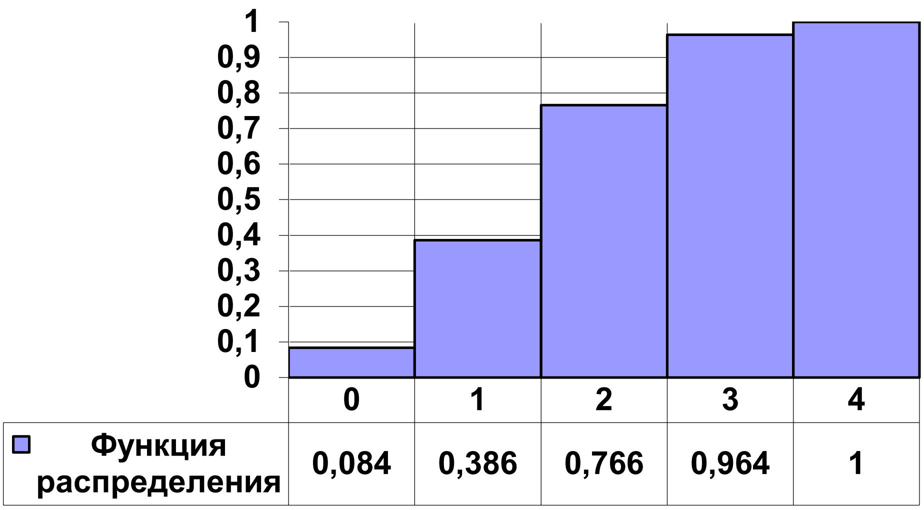
Для дискретної впадкової величини функція розподілу має вигляд:

$$F(x) = \sum_{x_i < x} P(X = x_i).$$

Знак нерівності під знаком суми показує, що сумування розповсюджується на ті можливі значення випадкової величини, які менші за аргумент x.

Функція розподілу дискретної випадкової величини X розривна і зростає стрибками при переході через кожне значення x_i.

Так для прикладу, що розглядався вище, функція розподілу матиме вигляд:



Властивості функції розподілу.

1) значення функції розподілу належать відрізку [0; 1]: $0 \le F(x) \le 1.$

2) $F(x)$ – неспадна функція: $F(x_2) \ge F(x_1)$ при $x_2 \ge x_1$.

3) Імовірність того, що випадкова величина прийме значення з інтегрвала (a; b), рівна приросту функції розподілу на цьому інтервалі:

$$P(a \le X < b) = F(b) - F(a).$$

4) На «мінус нескінченності» функція розподілу рівна нулю; на «плюс нескінченності» функція розподілу рівна одиниці.

5) Імовірність того, що неперервна випадкова величина X набуде одного певного значення, рівна нулю.

Таким чином, немає потреби казати про будь-яке конкретне значення випадкової величини. Цікавість представляє лише імовірність потрапляня випадкової величини у будь-який інтервал, що відповідає більшості практичних задач.

Густина розподілу.

Функція розподілу повністю характеризує випадкову величину, однак, має один недолік. По функції розподілу важко казати про характер розподілу випадкової величини у невеликому околі тій чи іншої точки числової осі.

Означення. *Густиною розподілу* імовірностей неперервної випадкової величини X називається функція $f(x)$ – перша похідна від функції розподілу $F(x)$:

$$f(x) = F'(x).$$

Густину розподілу також називають *диференціальною функцією*.

Зміст густини розподілу полягає у тому, що вона показує, як часто з'являється випадкова величина X у деякому околі точки x при повторенні випробувань.

Після введення функцій та густини розподілу можна дати означення *неперервної в ипадкової величини*.

Означення. Випадкова величина X називається *неперервною*, якщо її функція розподілу $F(x)$ неперервна на усій осі OX, а густина розподілу $f(x)$ існує завжди, за виключенням, можливо, скінченної кількості точок.

Знаючи густину розподілу, можна обчислити імовірність того, що деяка випадкова величина X прийме значення, що належить даному інтервалу.

Теорема. *Імовірність того, що неперервна випадкова величина X набуде значення, що належить інтервалу* $(a;\ b)$, *рівна визначеному інтегралу від густини розподілу, узятому в межах від a до b*:

$$P(a < X < b) = \int_a^b f(x)dx.$$

Доведення цієї теореми основане на означенні густини розподілу та третій властивості функції розподілу, що записані вище.

Геометрично це означає, що імовірність того, що неперервна випадкова величина набуває значення, що належить інтервалу $(a,\ b)$, рівна площі криволінійної трапеції, обмеженої віссю OX, кривою розподілу $f(x)$ і прямими $x=a$ та $x=b$.

Функція розподілу може бути легко знайдена, якщо відома густина розподілу, за формулою:

$$F(x) = \int_{-\infty}^{x} f(x)dx.$$

Властивості густини розподілу.

1) Густина розподілу – невід'ємна функція: $f(x) \geq 0.$

2) Невласний інтеграл по густині розподілу в межах від $-\infty$ до $+\infty$ рівний одиниці:

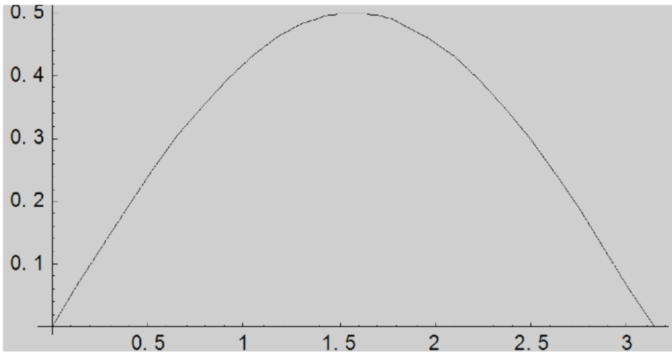
$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

Випадкова величина підкорена закону розподілу з густиною:

$$f(x) = \begin{cases} a\sin x, & при \quad 0 \leq x \leq \pi \\ 0, & при \quad x < 0 \quad або \quad x > \pi \end{cases}.$$

Потрібно знайти коефіцієнт a, побудувати графік функції густини розподілу, визначити імовірність того, що випадкова величина потрапить у інтервал від 0 до $\dfrac{\pi}{4}$.

● Побудуємо графік густини розподілу:



Для знаходження коефіцієнта a скористуємось властивістю: $\int\limits_{-\infty}^{\infty} f(x)dx = 1$.

$$\int\limits_{-\infty}^{\infty} f(x)dx = \int\limits_{-\infty}^{0} 0dx + \int\limits_{0}^{\pi} a\sin x dx + \int\limits_{\pi}^{\infty} 0dx = a\int\limits_{0}^{\pi} \sin x dx = -a\cos x \Big|_{0}^{\pi} = 2a = 1;$$

$$a = \frac{1}{2}.$$

Задана неперервна випадкова величина x своєю функцією розподілу *f(x)*.

$$f(x) = \begin{cases} A\cos 2x, \quad при \quad -\dfrac{\pi}{4} \le x \le \dfrac{\pi}{4} \\ 0, \qquad при \quad |x| > \dfrac{\pi}{4}. \end{cases}$$

Потрібно визначити коефіцієнт A, знайти функцію розподілу, побудувати графіки функції розподілу і густини розподілу, визначити імовірність того, що випадкова величина x потрапить в інтервал $\left(\dfrac{\pi}{6}; 2\right)$.

● Знайдемо коефіцієнт A.

$$\int\limits_{-\infty}^{\infty} f(x)dx = \int\limits_{-\infty}^{-\pi/4} 0dx + \int\limits_{-\pi/4}^{\pi/4} A\cos 2x dx + \int\limits_{\pi/4}^{\infty} 0dx = \frac{A\sin 2x}{2}\Big|_{-\pi/4}^{\pi/4} = A = 1.$$

Знайдемо функцію розподілу:

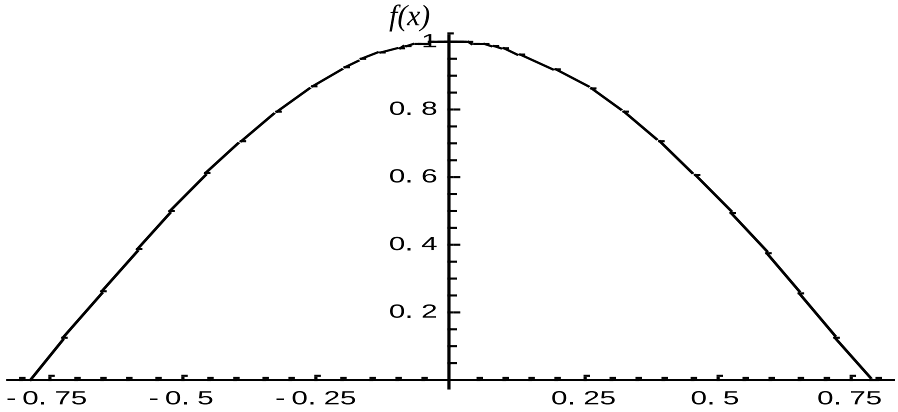
1) на $x < -\dfrac{\pi}{4}$: $F(x) = \int\limits_{-\infty}^{x} f(x)dx = \int\limits_{-\infty}^{x} 0dx = 0;$

2) на $-\dfrac{\pi}{4} \le x \le \dfrac{\pi}{4}$: $F(x) = \int\limits_{-\infty}^{-\pi/4} 0dx + \int\limits_{-\pi/4}^{x} \cos 2x dx = \dfrac{\sin 2x}{2}\Big|_{-\pi/4}^{x} = \dfrac{\sin 2x}{2} + \dfrac{1}{2};$

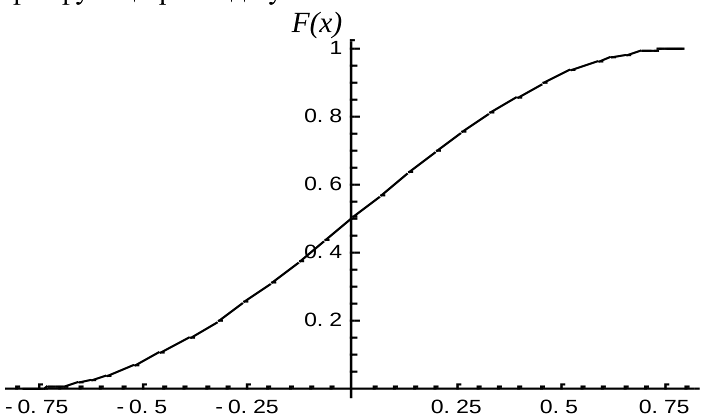
3) на $x > \dfrac{\pi}{4}$: $F(x) = \int\limits_{-\infty}^{-\pi/4} 0dx + \int\limits_{-\pi/4}^{\pi/4} \cos 2x dx + \int\limits_{\pi/4}^{x} 0dx = \dfrac{\sin 2x}{2}\Big|_{-\pi/4}^{\pi/4} = 1.$

Отже: $f(x)=\begin{cases}\cos 2x, & при \quad -\dfrac{\pi}{4}\le x\le \dfrac{\pi}{4};\\ 0, & при \quad |x|>\dfrac{\pi}{4}\end{cases}$; $\quad F(x)=\begin{cases}0, & при \quad x<-\dfrac{\pi}{4};\\ \dfrac{\sin 2x+1}{2}, & при \quad -\dfrac{\pi}{4}\le x\le \dfrac{\pi}{4};\\ 1, & при \quad x>\dfrac{\pi}{4}.\end{cases}$

Побудуємо графік густини розподілу:



Побудуємо графік функції розподілу:



Знайдемо імовірність потрапляння випадкової величини у інтегрвал $\left(\dfrac{\pi}{6};2\right)$.

$$P\left(\frac{\pi}{6}<x<2\right)=\int_{\pi/6}^{2} f(x)dx=\int_{\pi/6}^{\pi/4}\cos 2xdx+\int_{\pi/4}^{2}0dx=\frac{\sin 2x}{2}\Big|_{\pi/6}^{\pi/4}=\frac{1}{2}-\frac{\sqrt{3}}{4}=0{,}067.$$

Ту ж саму імовірність можна шукати і іншим способом:

$$P\left(\frac{\pi}{6}<x<2\right)=F(2)-F(\pi/6)=1-\frac{\sin(\pi/3)+1}{2}=\frac{1}{2}-\frac{\sqrt{3}}{4}=0{,}067.$$

Числові характеристики неперервних випадкових величин.

Нехай неперервна випадкова величина X задана функцією розподілу $f(x)$. Нехай усі можливі значення випадкової величини належать відрізку $[a;b]$.

Означення. *Математичним очікуванням* неперервної випадкової величини X, можливі значення якої належать відрізку $[a;b]$, називається визначений інтеграл

$$M(X)=\int_a^b xf(x)dx.$$

Якщо можливі значення випадкової величини розглядаються на усій числовій осі, то математичне очікування знаходиться за формулою:

$$M(X)=\int_{-\infty}^{\infty} xf(x)dx.$$

При цьому невласний інтеграл має збігатися.

Означення. *Дисперсією* неперервної випадкової величини називається математичне очікування квадрату її відхилення:

$$D(X)=\int_{-\infty}^{\infty}[x-M(X)]^2 f(x)dx.$$

По аналогії з дисперсією дискретної випадкової величини, для практичного обчислення дисперсії використовується формула:

$$D(X)=\int_{-\infty}^{\infty}x^2 f(x)dx-[M(X)]^2.$$

Означення. *Середнім квадратичним відхиленням* називається квадратний корінь з дисперсії:

$$\sigma(X)=\sqrt{D(X)}.$$

Означення. *Модою* M_0 дискретної випадкової величини називається її найбільш імовірнісне значення. Для неперервної випадкової величини мода – таке значення випадкової величини, при якій густина розподілу має максимум:

$$f(M_0)=\max.$$

Якщо многокутник розподілу для дискретної випадкової величини або крива розподілу для неперервної випадкової величини має два або більше максимумів, то такий розподіл називається *двохмодальним* або *багатомодальним*.

Якщо розподіл має мінімум, але не має максимуму, то він називається *антимодальним*.

Означення. *Медіаною* M_D випадкової величини X називається таке її значення, відносно якого рівно ймовірне отримати більше або менше значення випадкової величини:

$$P(X<M_D)=P(X>M_D).$$

Геометрично медіана – абсциса точки, у якій площа, обмежена кривою розподілу ділиться навпіл.

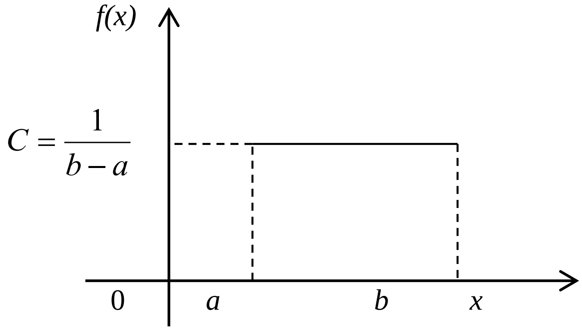
Відмітимо, що якщо розподіл одномодальний, то мода і медіна співпадають з математичним очікуванням.

Рівномірний розподіл.

Означення. Неперервна величина має *рівномірний розподіл* на відрізку [a, b], якщо на цьому відрізку густина розподілу випадкової величини стала, а поза ним рівна нулю:

$$f(x)=\begin{cases}0, & x<a;\\ C, & a\le x\le b;\\ 0, & x>b.\end{cases}$$

Стала величина C може бути визначена з умови рівності одиниці площі, обмеженої кривою розподілу.

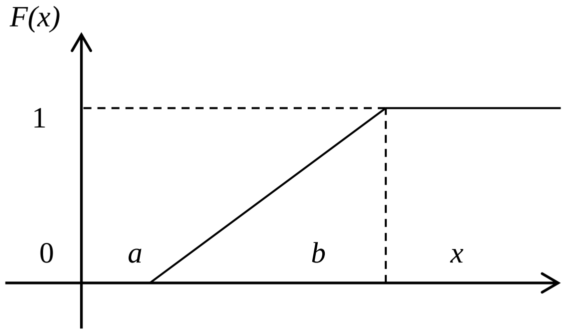


Отримаємо: $C = \frac{1}{b-a}$.

Знайдемо функцію розподілу *F(x)* на відрізку [*a*,*b*].

$$F(x) = \int_{-\infty}^{x} f(x)dx = \int_{a}^{x} \frac{1}{b-a} dx = \frac{x}{b-a}\Big|_{a}^{x} = \frac{x-a}{b-a}.$$

$$F(x) = \begin{cases} 0, & при \quad x < a; \\ \frac{x-a}{b-a}, & при \quad a \le x \le b; \\ 1, & при \quad x > b. \end{cases}$$



Для того, щоб випадкова величина підкорювалась закону рівномірного розподілу необхідно, щоб її значення лежали всередині деякого визначеного інтервала, і всередині цього інтервала значення цієї величини були б рівноймовірнісні.

Визначимо математичне очікування і дисперсію випадкової величини, що підкорена рівномірному закону розподілу.

$$m_x = \int_a^b xf(x)dx = \int_a^b \frac{x}{b-a} dx = \frac{x^2}{2(b-a)}\Big|_a^b = \frac{a+b}{2}.$$

$$m_{x^2} = \int_a^b x^2 f(x)dx = \int_a^b \frac{x^2}{b-a} dx = \frac{x^3}{3(b-a)}\Big|_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{b^2+ab+a^2}{3}.$$

$$D_x = m_{x^2} - m_x^2 = \frac{b^2+ab+a^2}{3} - \frac{a^2+2ab+b^2}{4} = \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12}.$$

$$\sigma_x = \sqrt{D_x} = \frac{b-a}{2\sqrt{3}}.$$

Імовірність потрапляння випадкової величини у заданий інтервал:

$$P(\alpha < X < \beta) = \int_\alpha^\beta \frac{dx}{b-a} = \frac{\beta-\alpha}{b-a}.$$

Показниковий розподіл.

Означення. *Показниковим* (*експоненціальним*) називається розподіл імовірностей неперервної випадкової величини Х, яке описується густиною:

$$f(x)=\begin{cases}0, & при \quad x<0;\\ \lambda e^{-\lambda x}, & при \quad x\ge 0.\end{cases}$$

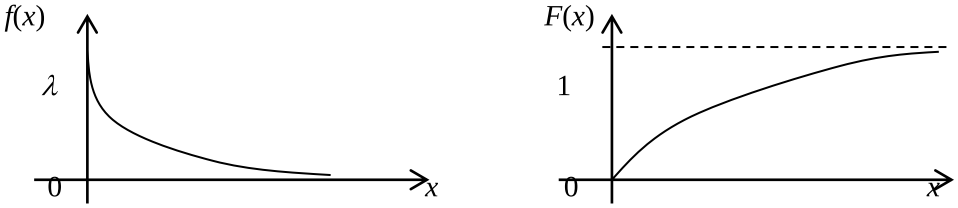
де λ – додатнє число.

Знайдемо закон розподілу.

$$F(x)=\int_{-\infty}^{x} f(x)dx=\int_{-\infty}^{0} 0dx+\lambda\int_{0}^{x} e^{-\lambda x}dx=1-e^{-\lambda x}.$$

$$F(x)=\begin{cases}0, & при \quad x<0;\\ 1-e^{-\lambda x} & при \quad x\ge 0.\end{cases}$$

Графіки функції розподілу і густини розподілу:



Знайдемо математичне очікування випадкової величини, що підкорена показниковому розподілу.

$$m_x=\int_{-\infty}^{\infty} xf(x)dx=\int_{0}^{\infty} x\lambda e^{-\lambda x}dx=\begin{Bmatrix}u=x; e^{-\lambda x}dx=dv;\\ du=dx; -\dfrac{e^{-\lambda x}}{\lambda}=v;\end{Bmatrix}=\lambda\left(-\frac{xe^{-\lambda x}}{\lambda}\Big|_0^{\infty}+\int_0^{\infty}\frac{e^{-\lambda x}}{\lambda}dx\right)$$

$$=\int_0^{\infty} e^{-\lambda x}dx=-\frac{e^{-\lambda x}}{\lambda}\Big|_0^{\infty}=\frac{1}{\lambda}.$$

Результат отриманий з використанням факту, що

$$xe^{-\lambda x}\Big|_0^{\infty}=\lim_{x\to\infty}\frac{x}{e^{\lambda x}}-0=\begin{Bmatrix}По \quad правилу\\ Лопіталя\end{Bmatrix}=\lim_{x\to\infty}\frac{1}{-\lambda e^{\lambda x}}=0.$$

Для знаходження дисперсії знайдемо величину $M(X^2)$.

$$M(X^2)=\int_{-\infty}^{\infty} x^2 f(x)dx=\int_0^{\infty}\lambda x^2 e^{-\lambda x}dx.$$

Двічі інтегруючи по частинам, отримаємо: $M(X^2)=\dfrac{2}{\lambda^2}.$

Тоді $D(X)=M(X^2)-[M(X)]^2=\dfrac{1}{\lambda^2}.$

Отже: $M(X)=\dfrac{1}{\lambda};\quad D(X)=\dfrac{1}{\lambda^2};\quad \sigma_x=\dfrac{1}{\lambda}.$

Видно, що у випадку показникового розподілу математичне очікуваня і середнє квадратичне відхилення рівні.

Також легко визначити і ймовірність потрапляння випадкової величини, що підкорена показниковому закону розподілу, у заданий інтервал:

$$P(a<x<b)=F(b)-F(a)=e^{-\lambda a}-e^{-\lambda b}.$$

Нормальний закон розподілу.

Означення. *Нормальним* називається розподіл імовірностей неперервної випадкової величини, яке описується густиною імовірності:

$$f(x)=\frac{1}{\sigma_x\sqrt{2\pi}}e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}.$$

Нормальний закон розподілу також називається *законом Гаусса*.

Нормальний закон розподілу займає центральне місце у теорії ймовірностей. Це обумовлено тим, що цей закон проявляється у всіх випадках, коли випадкова величина є результатом дії великої кількості різних факторів. До нормального закону наближаються усі інші закони розподілу.

Можна легко показати, що параметри m_x і σ_x, які входять у густину розподілу є відповідно математичним очікуванням і середнім квадратичним відхиленням величини *X*.

Знайдемо функцію розподілу *F*(*x*).

$$F(x)=\frac{1}{\sigma_x\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}dx.$$

Графік густини нормального розподілу називається *нормальною кривою* або *кривою Гаусса*.

Нормальна крива має наступні *властивості*:

1) функція визначена на усій числовій осі;

2) при усіх *x* функція розподілу набуває лише додатніх значень;

3) вісь *OX* є горизонтальною асимптотою графіка густини імовірності, так як пр. необмеженому зростанні по абсолютній величині аргумента *x*, значення функції прямує до нуля;

4) знайдемо екстремум функції:

$$y'=-\frac{x-m}{\sigma^3\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}=0; \qquad x=m.$$

Так як при $y'>0$ при $x<m$ та $y'<0$ при $x>m$, то в точці $x=m$ функція має максимум, рівний $\frac{1}{\sigma\sqrt{2\pi}}$.

5) функція є симетричною відносно прямої *x*=*a*, так як різниця (*x*–*a*) входить у функцію густини розподілу у квадраті;

6) для знаходження точок перегину графіка знайдемо другу похідну функції густини:

$$y''=-\frac{1}{\sigma^3\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}\left[1-\frac{(x-m)^2}{\sigma^2}\right].$$

При $x=m+\sigma$ і $x=m-\sigma$ друга похідна рівна нулю, а при переході через ці точки змінює знак, тобто у цих точках функція має перегин.

У цих точках значення функції рівне $\frac{1}{\sigma e\sqrt{2\pi}}$.

Побудуємо графік функції густини розподілу:

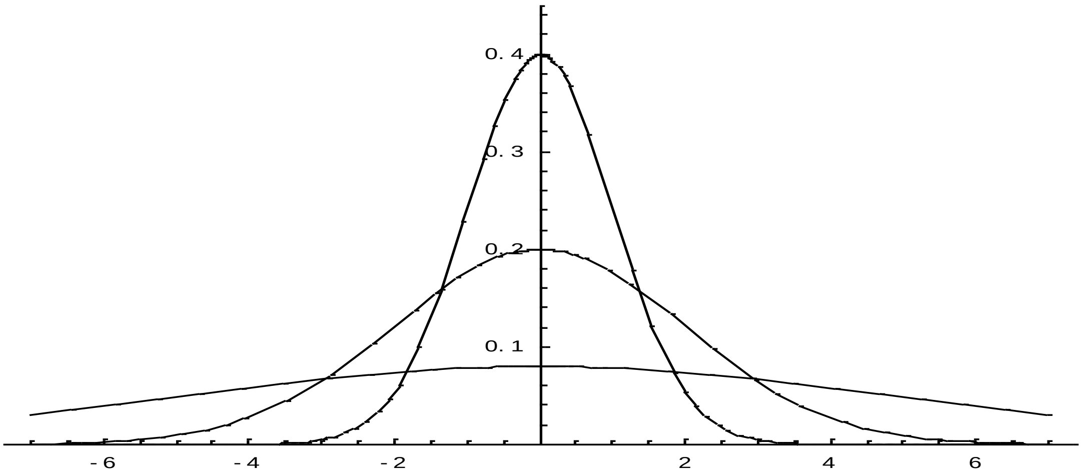


Рис. 8.1

Побудовані графіки при m=0 та трьох можливих значеннях середнього квадратичного відхилення σ=1, σ=2 та σ=7. Як видно, пр. збільшенні значення середнього квадратичного відхилення графік стає більш пологим, а максимальне значення зменшується.

Якщо a>0, то графік зміститься у додатному напрямку, якщо a<0 – у від'ємному.

При a=0 та σ=1 крива називається *нормованою*. Рівняння нормованої кривої:

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

Функція Лапласа.

Знайдемо імовірність потрапляння випадкової величини, розподіленої по нормальному закону, у заданий інтервал:

$$P(a < X < b) = \int_a^b f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{(x-m)^2}{2\sigma^2}} dx.$$

Позначимо $\dfrac{x-m}{\sigma\sqrt{2}} = t;\quad \dfrac{a-m}{\sigma\sqrt{2}} = \alpha;\quad \dfrac{b-m}{\sigma\sqrt{2}} = \beta.$

Тоді $P(a < X < b) = \dfrac{1}{\sigma\sqrt{2\pi}} \displaystyle\int_\alpha^\beta e^{-t^2} \sigma\sqrt{2}dt = \dfrac{1}{\sqrt{\pi}} \int_\alpha^\beta e^{-t^2} dt = \dfrac{1}{2}\left[\Phi(\beta) - \Phi(\alpha)\right].$

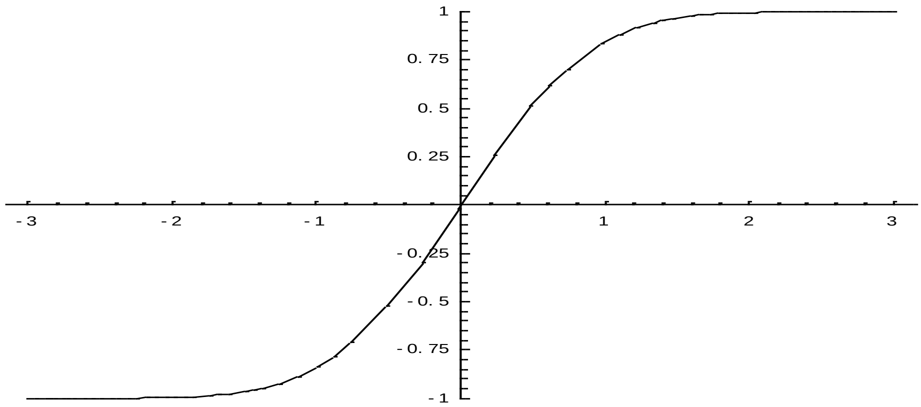
Так як інтеграл $\int e^{-t^2} dt$ не виражається через елементарні функції, то розглянемо функцію:

$$\Phi(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt,$$

яка називається *функцією Лапласа* або *інтегралом імовірностей*.

Значення цієї функції при різних значення x пораховані і приводяться у спеціальних таблицях.

Нижче показано графік функції Лапласа.



Функція Лапласа має наступні властивості:

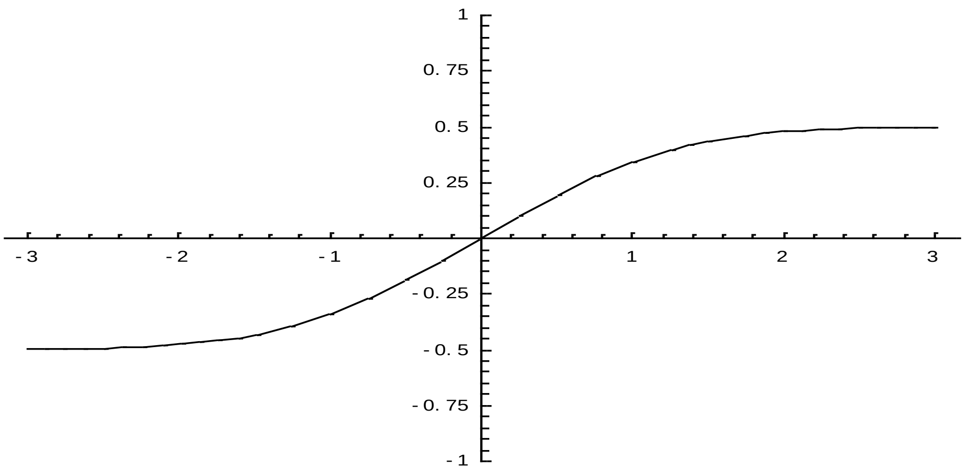
1) $\Phi(0) = 0$;
2) $\Phi(-x) = -\Phi(x)$;
3) $\Phi(\infty) = 1$.

Функцію Лапласа також називають *функцією помилок* і позначають *erf x*.

Ще використовується *нормована функція Лапласа*, яка пов'язана з функцією Лапласа співвідношенням:

$$\overline{\Phi}(x) = \frac{1}{2}\Phi\left(\frac{x}{\sqrt{2}}\right) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-t^2/2}dt.$$

Нижче показано графік нормованої функції Лапласа.



При розгляді нормального закону розподілу виділяється важливий частинний випадок, відомий як *правило трьох сигм*.

Запишемо імовірність того, що відхилення нормально розподіленій випадковій величині від математичного очікуваня менше заданої величини Δ:

$$P(|X - m| < \Delta) = \overline{\Phi}\left[\frac{m + \Delta - m}{\sigma}\right] - \overline{\Phi}\left[\frac{m - \Delta - m}{\sigma}\right] = \overline{\Phi}\left[\frac{\Delta}{\sigma}\right] - \overline{\Phi}\left[-\frac{\Delta}{\sigma}\right] = 2\overline{\Phi}\left[\frac{\Delta}{\sigma}\right].$$

Якщо прийняти $\Delta=3\sigma$, то отримаємо з використанням таблиць значення функції Лапласа:

$$P(|X - m| < 3\sigma) = 2\overline{\Phi}(3) = 2 \cdot 0{,}49865 = 0{,}9973.$$

Тобто імовірність того, що випадкова величина відхилиться від свого математичного очікування на величину, більшу, ніж потроєне середнє квадратичне відхилення, практично рівне нулю.

Це правило називається *правилом трьох сигм*.

На практиці вважається, що якщо для будь-якої випадкової величини виконується правило трьох сигм, то ця випадкова величина має нормальний розподіл.

Потяг складається зі 100 вагонів. Маса кожного вагона – випадкова величина, розподілена по нормальному закону з математичним очікуванням a=65 т та середнім квадратичним відхиленням σ=0,9 т. Локомотив може везти склад масою не більше 6600 т, в противному випадку необхідно чіпляти другий локомотив. Знайти імовірність того, другий локомотив не знадобиться.

● Другий локомотив не знадобиться, якщо відхилення маси складу від очікуваного (100·65 = 6500) не перевищує 6600 – 6500 = 100 т.

Так як маса кожного вагона має нормальний розподіл, то й маса усього складу теж буде розподілена нормально.

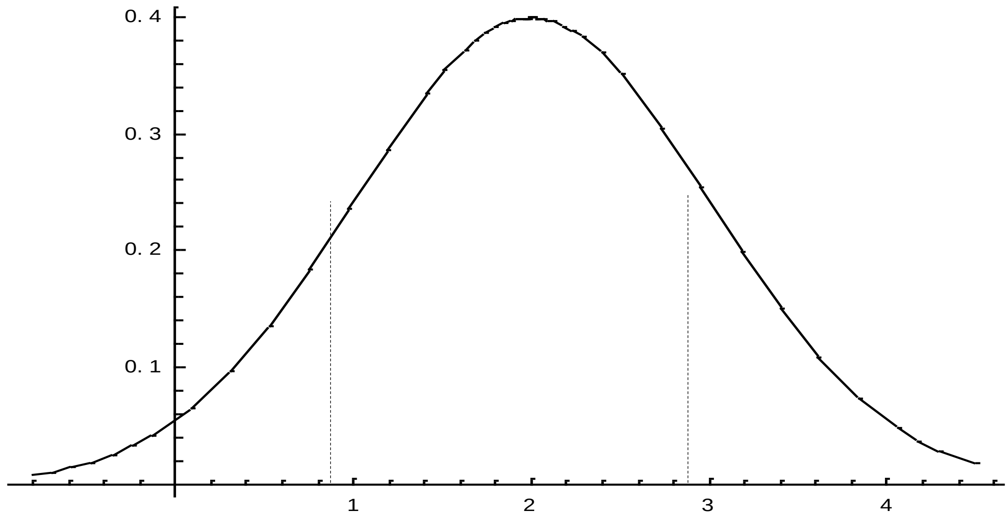
Отримаємо:

$$P(|X - M(X)| < 100 = 2\overline{\Phi}\left[\frac{100}{100\sigma}\right] = 2\overline{\Phi}[1{,}111] = 2 \cdot 0{,}3665 = 0{,}733.$$

Нормально розподілена випадкова величина X задана своїми параметрами – $a=2$ – математичне очікування і σ=1 – середнє квадратичне відхилення. Потрібно написати густину імовірності та побудувати її графік, знайти імовірність того, що X відхилиться (по модулю) від математичного відхилення не більш, ніж на 2.

● Густина розподілу має вигляд: $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-2)^2}{2}}$.

Побудуємо графік:



Знайдемо імовірність потрапляння випадкової величини у інтервал (1; 3).

$$P(1 < X < 3) = \frac{1}{\sqrt{\pi}} \int_{\alpha}^{\beta} e^{-t^2} dt = \frac{1}{2}\left[\Phi\left(\frac{3-2}{\sqrt{2}}\right) - \Phi\left(\frac{1-2}{\sqrt{2}}\right)\right] = \frac{1}{2}[\Phi(0{,}7071) + \Phi(0{,}7071)] = 0{,}6778.$$

Знайдемо імовірність відхилення випадкової величини від математичного очікуваня на величину, не більшу, ніж 2.

$$P(|X - 2| < 2) = \Phi\left(\frac{\Delta}{\sigma\sqrt{2}}\right) = \Phi\left(\frac{2}{\sqrt{2}}\right) = \Phi(\sqrt{2}) \approx 0{,}95.$$

Той же результат може бути отриманий з використанням нормованої функції Лапласа:

$$P(|X - 2| < 2) = 2\overline{\Phi}\left(\frac{\Delta}{\sigma}\right) = 2\overline{\Phi}(2) = 2 \cdot 0{,}4772 \approx 0{,}95.$$

Теорема Бернуллі.

Нехай проводиться n незалежних випробувань, у кожному з яких імовірність появи події A рівна p.

Потрібно визначити відносну частоту появи події A.

Теорема. *Якщо у кожному з n незалежних випробувань імовірність p появи події A стала, то як завгодно близька одиниці імовірність того, що відхилення відносної частоти від імовірності p по абсолютній величині буде як завгодно малим, якщо число випробувань достаньо велике*:

$$\lim_{n\to\infty} P(|m/n - p| < \varepsilon) = 1.$$

Тут m – число появ події A. З усього сказаного вище не слідує, що із збільшенням числа випробувань відносна частота неухильно прямує до імовірності p, тобто $\lim\limits_{n\to\infty}\frac{m}{n} = p$. У теоремі мається на увазі лише імовірність наближення відносної частоти до імовірності появи події A у кожному випробуванні.

У випадку, якщо імовірності появи події A у кожному випробуванні різні, то справедлива теорема, відома як *теорема Пуассона*.

Теорема Пуассона. *Якщо проводиться n незалежних випробувань і ймовірність появи події A у кожному досліді рівна* p_i, *то при збільшенні n частота події A збігається по імовірності до середнього арифметичного імовірностей* p_i.

Теорема Муавра–Лапласа.

Теорема Муавра–Лапласа. *Якщо проводиться n незалежних випробувань, у кожному з яких подія A з'являється з імовірністю p, то для будь-якого інтервала (α, β) справедливе співвідноешння:*

$$P\left(\alpha < \frac{Y - np}{\sqrt{npq}} < \beta\right) = \frac{1}{2}\left[\Phi\left(\frac{\beta}{\sqrt{2}}\right) - \Phi\left(\frac{\alpha}{\sqrt{2}}\right)\right] = \overline{\Phi}(\beta) - \overline{\Phi}(\alpha).$$

де Y – число появ події A в n випробуваннях, q=1–p, Φ(x) – функція Лапласа, $\overline{\Phi}(x)$*– нормована функція Лапласа.*

Теорема Муавра–Лапласа описує поведінку біноміального розподілу при великих значеннях n.

Дана теорема дозволяє суттєво спростити обчислення по формулі біноміального розподілу.

Розрахунок імовірності потрапляння значення випадкової величини у заданий інтервал $P(\alpha < Y < \beta) = \sum\limits_{\alpha<k<\beta} C_n^k p^k q^{n-k}$ при великих значеннях n складний. Простіше скористатися формулою:

$$P(\alpha < Y < \beta) = \frac{1}{2}\left[\Phi\left(\frac{\beta - np}{\sqrt{2npq}}\right) - \Phi\left(\frac{\alpha - np}{\sqrt{2npq}}\right)\right].$$

Теорема Муавра–Лапласа широко використовується при розв'язуванні практичних задач.

Математична статистика

Термін *“статистика”* походить від латинського слова *status* (*статус*), що означає “визначене положення речей”. У даний час термін “*статистика*” вживається в трьох значеннях:

1) як *галузь* практичної діяльності зі збору, обробки та аналізу даних соціально-економічного та іншого масового характеру;

2) як *наука*, що містить теоретичні положення і методи розв’зання практичних задач статистики (*математ ична статистика*);

3) як *підсумкові* статистичні дані, тобто як результати застосування статистичних методів до початкової статистичної інформації.

Математична статистика – це наука про методи реєстрації, опису й аналізу статистичних експериментальних даних, які одержуються у результаті спостереження масових випадкових явищ.

Задачами математичної статистики, які найчастіше зустрічаються на практиці, є:

1) визначення закону розподілу випадкової величини (або системи випадкових величин) за статистичними початковими даними;

2) перевірка правдоподібності гіпотез, наприклад – чи узгодяться результати експерименту з гіпотезою про те, що дана випадкова величина підпорядкована закону *розподілу* $F(x)$;

3) відшукання невідомих параметрів розподілу випадкової величини або системи випадкових величин (або оцінок цих параметрів).

У математичній статистиці використовується властива цій предметній галузі система категорій і понять. Розглянемо основні з цих понять.

Варіаційний ряд, гістограма.

Припустимо, що виникла ситуація, у якій потрібно вивчити сукупність об’єктів (наприклад покупців продукції фірми), оцінюючи деяку ознаку, яка властива кожному з цих об’єктів (наприклад, вік). Якщо аналізована ознака має кількісне значення, то частіше за все цю ознаку ототожнюють із випадковою величиною, а конкретне значення ознаки сприймається як значення випадкової величини.

Означення. *Вибірковою сукупністю* (синоніми: вибірка, простий статистичний ряд) називається сукупність значень x_i тієї самої ознаки (ξ) у випадково відібраних n об’єктів.

Звичайно вибіркова сукупність оформляється у вигляді таблиці в два рядки (див. табл. 4. 1). У першому рядку вказують номер об’єкта (або спроби), у другому – значення ознаки цього об’єкта або в цій спробі (значення випадкової величини ξ). Така вибіркова сукупність далі підлягає обробці, наприклад – побудові статистичної функції розподілу. Вибіркова сукупність є елементом більш загальної – генеральної сукупності.

Таблиця 9.1

i	1	2	…	k	…	N
ξ	x_1	x_2	…	x_k	…	x_n

Означення. *Генеральною сукупністю* називається сукупність значень тієї самої ознаки (ξ) в *усіх* об’єктах, зі складу яких проводиться вибірка.

Означення. *Обсягом сукупності* (вибіркової або генеральної) називають *кількість* (n) об’єктів цієї сукупності.

Серед отриманих n значень ознаки можуть бути однакові значення. Так, значення x_i може зустрічатися n_i разів. У результаті, *різні* значення ознаки ξ можуть виявитися рівними $x_1, x_2, \ldots, x_m$, а кількість таких *різних* значень ознаки ξ у вибірці може дорівнювати деякій величині m ($m \le n$). Для опису зазначеної ситуації використовують поняття “варіант”.

Означення. *Варіантом* називають конкретне, відмінне від інших, отримане в спробі значення x_i ознаки ξ .

Означення. *Частотою варіанту* x_i у вибірці називають кількість n_i *однакових* значень (x_i) ознаки у вибірці.

Сума частот усіх варіантів дорівнює обсягу вибірки:

$$\sum_{i=1}^{m} n_i = n .$$

Означення. *Відносною частотою* (n_i^*) значення x_i ознаки ξ називається відношен-ня частоти n_i варіанта x_i до обсягу вибірки n:

$$n_i^* = \frac{n_i}{n} \le 1; \qquad \sum_{i=1}^{m} n_i^* = 1.$$

Означення. *Дискретним варіаційним* рядом (*варіацією*) називають упорядковану за зростанням значень x_i послідовність варіантів із вказівкою їх абсолютних або відносних частот і подану у вигляді таблиці (див., наприклад, табл. 9.2).

Означення. *Полігоном відносних частот* (див. рис. 9.1) називають ламану лінію, яка з’єднує точки з координатами (x_i, n_i^*).

Означення. *Розмахом варіації* (R) називають *різницю* між максимальним (x_{max}) і мінімальним (x_{min}) значеннями варіантів у вибірці:

$$R = x_{max} - x_{min}.$$

Таблиця 9.2

Дискретний варіаційний ряд

Варіант x_i	x_1	x_2	x_3	…	x_m
Частота n_i	n_1	n_2	n_3		n_m
Відносна частота n_i^*	n_1^*	n_2^*	n_3^*		n_m^*

Записати у вигляді варіаційного ряду вибірку:

3; 9; 9; 5; 3; 6; 9; 5; 5; 5; 6; 3; 6; 5; 6; 3; 3; 5; 6; 5.

Визначити розмах вибірки, побудувати полігон частот і полігон відносних частот.

● Обсяг вибірки (число її елементів) дорівнює $\boldsymbol{n} = 20$. Упорядкуємо варіанти за зростанням і підрахуємо кількість повторень значень x_i у кожному варіанті, одержимо:

$$x_1 = 3; \quad x_2 = 5; \quad x_3 = 6; \quad x_4 = 9;$$
$$n_1 = 5; \quad n_2 = 7; \quad n_3 = 5; \quad n_4 = 3.$$

Контроль обчислень виконаємо шляхом підсумовування частот варіантів:

$$\Sigma\, n_i = 20.$$

Для кожного варіанту $\boldsymbol{x}_i$ знаходимо відносну частоту: $\boldsymbol{n}_i^* = \boldsymbol{n}_i / \boldsymbol{n}$.

Отриманий варіаційний ряд набуде вигляду (див. табл. 9.3).

Таблиця 9.3

Дискретний варіаційний ряд

x_i	3	5	6	9
n_i	5	7	5	3
n_i^*	0,25	0,35	0,25	0,15

Контроль обчислень виконаємо шляхом підсумовування відносних частот варіантів:

$$\Sigma\, n_i^* = 1.$$

Розмах вибірки одержимо:

$$R = x_{max} - x_{min} = 9 - 3 = 6.$$

Для побудови полігону час-тот нанесемо отримані точки (x_i, n_i^*) із табл. 9.3 на графік (див. рис. 9.2).

Якщо обсяг вибірки великий (сотні варіантів), то дискретний варіаційний ряд стає незручною формою запису. Тоді увесь діапазон значень ознаки ξ розбивають на k часткових інтервалів, які не перетинаються, (синонім – розрядів) $[a_{i-1}; a_i)$, $(i = 1, 2, \ldots, k)$.

Для кожного i-го інтервалу підраховують кількість значень x_i ознаки ξ, що потрапили в цей інтервал, – частоту інтервалу (n_i). Елемент x_i, який *співпав* із границею інтервалу, відносять до наступного інтервалу, а не до попереднього.

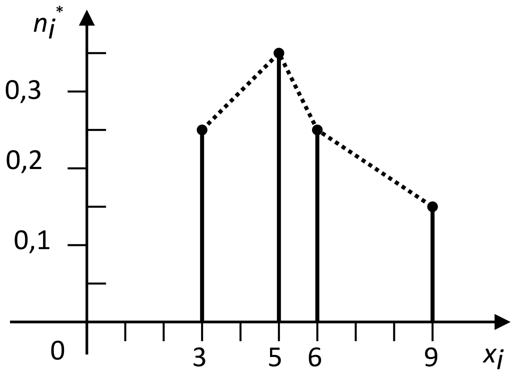


Рис. 9.1. Полігон відносних частот (див. табл. 9.3)

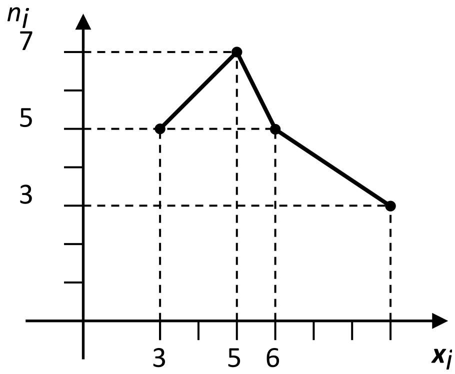


Рис. 9.2. Полігон частот (див. табл. 9.3)

Довжина розрядів може бути як однаковою, так і різною, що визначається необхідністю наявності в кожному інтервалі (розряді) *не менше* 5–10 значень випадкової ознаки ξ. Для ділянок із найбільшою щільністю значень ознаки довжина інтервалу (розряду) може бути меншою, із малою щільністю – більшою.

Раціональне число інтервалів (розрядів) складає 10–20.

Якщо до точності розрахунків немає дуже високих вимог, то частіше використовують розряди рівної довжини. Результати групування значень ознаки ξ за інтервалами записують у вигляді таблиці.

Означення. *Інтервальним варіаційним рядом* називається таблиця відповідності інтервалів значень (розрядів) випадкової величини ознаки ξ і частот n_i або/і відносних частот n_i^* входження значень ознаки в ці розряди, яка (таблиця) отримана за результатами спостережень або спроб (див. табл. 9.4). Такий ряд іноді називають *безперервним* або *статистичним* рядом.

Таблиця 9.4

Інтервальний варіаційний ряд

Інтервал (розряд)	$[a_1; a_2)$	$[a_2; a_3)$	$[a_3; a_4)$	…	$[a_{m-1}; a_m]$
Частота n_i	n_1	n_2	n_3		n_m
Відносна частота n_i^*	n_1^*	n_2^*	n_3^*		n_m^*

Якщо в кожному інтервалі як представницьке значення ознаки ξ узяти середнє значення інтервалу: $x_i = \dfrac{a_{i-1} + a_i}{2}$, ($i$=1, 2, …, k), то *інтервальний варіаційний* ряд можна умовно подати вже розглянутим *дискретним варіаційним* рядом. У цьому випадку в першому рядку табл. 9.4 указуються *середні* значення ознак для кожного інтервалу.

Графічно варіаційний ряд зображають у вигляді полігону частот або полігону відносних частот.

Використати 6 інтервалів рівної довжини і побудувати інтервальний варіаційний ряд розподілу за даною вибіркою:

15	18	22	26	15	10	23	27	19	18	14	15	28	6	29	7	11
26	24	19	14	15	7	27	14	19	8	20	5	29	16	10	16	5
18	20	12	16	22	23	20	21	11	16	22	22	6	18	14	11	

Потім перейти до дискретного варіаційного ряду.

● Обсяг вибірки $\boldsymbol{n} = 50$.

Знайдемо розмах вибірки: $R = x_{max} - x_{min} = 29 - 5 = 24$.

За умовою кількість інтервалів $m = 6$, тому довжина кожного часткового інтервалу $h = R / m = 24 / 6 = 4$. Перший інтервал починається з $x_{min} = 5$

і закінчується точкою $x_{min} + h = 5 + 4 = 9$, значення якої до складу інтервалу *не входить*. Ця ж точка є початком другого інтервалу, до складу якого вона *входить*. Виконуючи послідовність таких же розрахунків, одержимо границі 6 інтервалів: [5;9), [9;13), [13;17), [17; 21), [21; 25) [25; 29).

Підраховуємо кількість елементів вибірки, що потрапили в кожний із знайдених інтервалів. Елемент вибірки 21 є границею інтервалів [17; 21) і [21; 25). При підрахунку частоти відносимо його до інтервалу [21; 25). У результаті одержимо такий інтервальний варіаційний ряд (див. табл. 9.5).

Таблиця 9.5

Інтервальний варіаційний ряд

$a_{i-1} \div a_i$	[5÷9)	[9÷13)	[13÷17)	[17÷21)	[21÷25)	[25÷29]
n_i	7	6	12	10	7	8
n_i^*	0,14	0,12	0,24	0,20	0,14	0,16
$n_i^{(n)}$	7	13	25	35	42	50
$F_n(x)$	0,00	0,14	0,26	0,50	0,70	0,84 ($F_n(x>29)=1$

Контроль обчислень виконаємо шляхом підсумовування частот варіантів, тоді одержимо: Σ $n_i = 50$.

Для кожного часткового інтервалу обчисляємо відносну частоту за формулою: $n_i^* = n_i / n$ і результат заносимо в нижній рядок таблиці.

Контроль обчислень виконаємо шляхом підсумовування відносних частот варіантів, одержимо: $\Sigma\, n_i^* = 1$.

Для переходу від інтервального варіаційного ряду до дискретного варіаційного ряду візьмемо як варіанти середнє значення в кожному інтервалі: $x_i = (a_{i-1} + a_i) / 2$. Одержимо наступний дискретний варіаційний ряд:

x_i	7	11	15	19	23	27
n_i	7	6	12	10	8	7
n_i^*	0,14	0,12	0,24	0,20	0,16	0,14

Означення. *Накопиченою частотою* називається число $n_i^{(n)}$, що дорівнює сумі частот усіх варіантів від x_1 до x_i включно: $n_i^{(n)} = \sum_{q=1}^{i} n_q$.

Означення. *Кривою зростаючих результатів* (кумулятою) називається лінія в координатах $(x_i,\ n_i^{(n)})$.

Для зображення інтервального варіаційного ряду використовують гістограму.

Іноді для порівняльного аналізу гістограму подають у вигляді прямокутників із площею, яка дорівнює частоті відповідних розрядів, і з основою, яка дорівнює довжині цих розрядів (див. рис. 9.3).

Означення. *Гістограмою* називається східчаста фігура, складена з прямокутників, які мають за основи часткові інтервали $[a_{i-1}; a_i)$, а за висоти – *відносні* частоти (іноді – частоти). Приклад гістограми поданий на рис. 9.3 і 9.4.

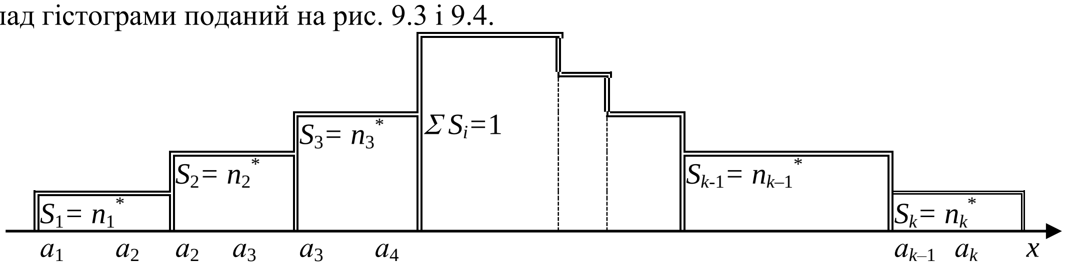


Рис. 9.3. Гістограма інтервального варіаційного ряду

Емпірична функція розподілу.

Означення. *Емпіричною функцією розподілу* називається функція $F_n(x)$, яка визначає для кожного значення x відносну частоту події ($\xi < x$) і має наступні властивості:

1) значення функції $F_n(x)$ належать інтервалу [0; 1]: $0 \le F_n(x) \le 1$;
2) $F_n(x)$ – функція неспадна;
3) $F_n(x \le x_1) = 0$, якщо x_1 – найменший варіант;
4) $F_n(x > x_m) = 1$, якщо x_m – найбільший варіант.

Розрахунковий вираз для емпіричної функції розподілу можна записати таким способом:

$$F_n(x) = \sum_{x_i < x} n_i^* \; ; \quad \to \quad F_n(x) = \begin{cases} 0 & \text{при} \quad x \le x_1, \\ n_1^* & \text{при} \quad x_1 < x \le x_2, \\ n_1^* + n_2^* & \text{при} \quad x_2 < x \le x_3, \\ n_1^* + n_2^* + n_3^* & \text{при} \quad x_3 < x \le x_4, \\ \dots \;\; \dots \;\; \dots \;\; \dots & \dots \;\; \dots \;\; \dots \;\; \dots \\ n_1^* + n_2^* + \dots + n_m^* = 1 & \text{при} \quad x > x_m. \end{cases}$$

Вигляд емпіричної функції розподілу поданий на рис. 9.5.

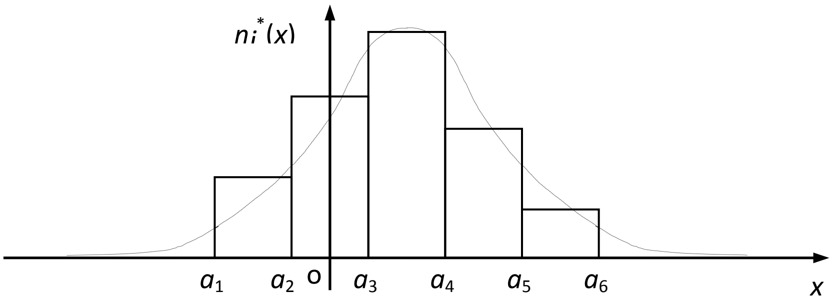


Рис. 9.4. Гістограма, загальний випадок

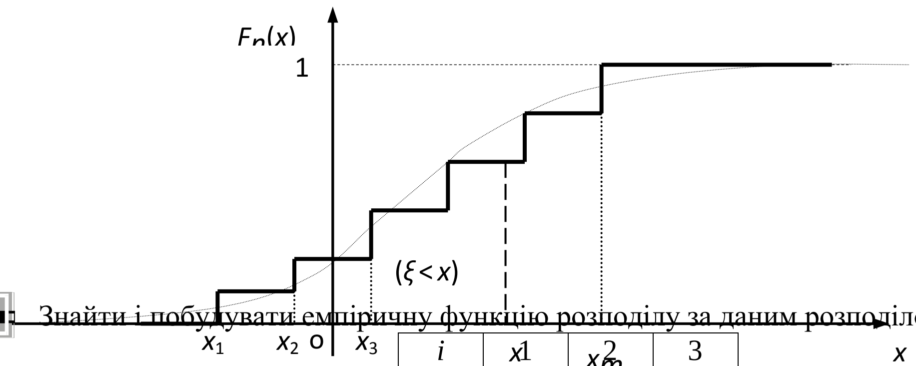


Рис. 9.5. Емпірична функція розподілу, загальний випадок

Приклад. Знайти і побудувати емпіричну функцію розподілу за даним розподілом вибірки:

i	1	2	3
x_i	1	4	7
n_i	5	20	25

● Знаходимо обсяг вибірки: $n = \Sigma\, n_i = 5 + 20 + 25 = 50$.

Для кожного варіанту x_i обчисляємо відносну частоту за формулою: $n_i^* = n_i / n$ і результат заносимо в таблицю:

i	1	2	3
x_i	1	4	7
n_i^*	0,1	0,4	0,5

Контроль обчислень виконуємо шляхом підсумовування відносних частот варіантів, тоді одержимо: $\Sigma\, n_i^* = 1$.

1. Знаходимо емпіричну функцію розподілу $F_n(x)$:

якщо $x \le x_1 = 1$, то $F_n(x) = 0$;
якщо $x_1 < x \le x_2$, тобто $1 < x \le 4$, то $F_n(x) = n_1^* = 0{,}1$;
якщо $x_2 < x \le x_3$, тобто $4 < x \le 7$, то $F_n(x) = n_1^* + n_2^* = 0{,}1 + 0{,}4 = 0{,}5$;
якщо $x > x_3$, тобто $x > 7$, то $F_n(x) = n_1^* + n_2^* + n_3^* = 0{,}1 + 0{,}4 + 0{,}5 = 1$.

Отже, одержимо:

$$F_n(x) = \begin{cases} 0 & \text{при } x \le 1, \\ 0{,}1 & \text{при } 1 < x \le 4, \\ 0{,}5 & \text{при } 4 < x \le 7, \\ 1 & \text{при } x > 7. \end{cases}$$

Використовуючи отримані значення функції розподілу $F_n(x)$, будуємо графік (див. рис. 4.6).

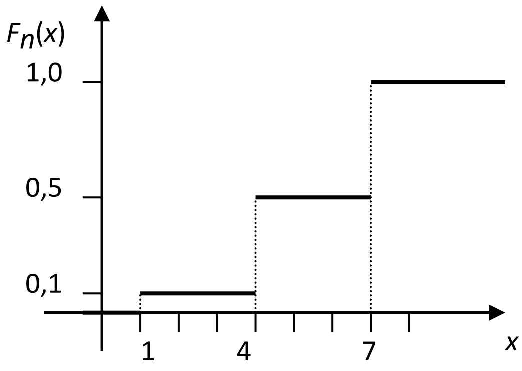


Рис. 9.6. Емпірична функція розподілу

Числові характеристики вибірки.

Означення. *Вибірковим середнім* (середнім арифметичним, позначається $\overline{x}$) називається середнє арифметичне значення ознаки ξ у вибірці x_i.

Коли всі значення ознаки *різні*, розрахункова формула має вигляд:

$$\overline{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

Коли у вибірці є *частоти* значень ознаки або обчислення виконуються для інтервального варіаційного ряду, тоді варто визначити: x_i – середнє значення ознаки $\boldsymbol{\xi}$ в i-му інтервалі, $\boldsymbol{n_i}$ – частоту значень ознаки в i-му інтервалі. Тоді розрахункова формула набуде вигляду:

$$\overline{x} = \frac{x_1 n_1 + x_2 n_2 + ... + x_m n_m}{n} = \frac{1}{n}\sum_{i=1}^{m} x_i n_i .$$

Означення. *Початковим моментом* k-го порядку (ν_k^*) варіаційного ряду називається середнє арифметичне від k-их ступенів варіантів x_i:

$$\nu_k^* = \frac{1}{n}\cdot\sum_{i=1}^{m} x_i^{\,k}\cdot n_i = \sum_{i=1}^{m} x_i^{\,k}\cdot n_i^* , \qquad \nu_1^* = \overline{x} .$$

Означення. *Центральним моментом* k-го порядку (μ_k^*) варіаційного ряду називається середнє арифметичне від k-их ступенів центрованих варіантів x_i:

$$\mu_k^* = \frac{1}{n}\cdot\sum_{i=1}^{m}\left(x_i - \overline{x}\right)^k \cdot n_i = \sum_{i=1}^{m}\left(x_i - \overline{x}\right)^k \cdot n_i^* .$$

Означення. *Вибірковою дисперсією* (дисперсією варіаційного ряду) називається середнє арифметичне квадратів відхилення варіантів x_i від свого вибіркового середнього значення, при цьому зберігаються відомі з теорії ймовірностей співвідношення:

$$D_в = S_x^2 = \frac{1}{n}\cdot\sum_{i=1}^{m}\left(x_i - \overline{x}\right)^2 \cdot n_i ; \ \rightarrow\ D_в = S_x^2 = \mu_2^* = \nu_2^* - (\nu_1^*)^2 = \nu_2^* - (\overline{x})^2 .$$

Тут S_x – вибіркове значення середнього квадратичного відхилення варіантів x_i.

Властивості вибіркового середнього і вибіркової дисперсії збігаються з розглянутими раніше в теорії ймовірностей властивостями математичного сподівання і дисперсії випадкових величин відповідно.

Умовні варіанти.

Якщо значення x_i ознаки ξ виражаються багатозначними (великими) числами, то обчислення характеристик вибірки стає трудомістким. У такому випадку переходять до використання *умовних* варіантів, для чого від реальних значень варіантів x_i переходять до їх "центрованих" і потім "нормованих" еквівалентів u_i. Як "центр" нормування вибирають варіант із *найбільшою частотою* і позначають його величину символом a. Якщо вибірка задана у вигляді розподілу *рівновіддалених* варіантів і їх частот (x_i, n_i), то, використовуючи різницю $(h = x_i - x_{i-1})$ між двома сусідніми варіантами, умовні варіанти u_i знаходять за формулою:

$$u_i = \frac{x_i - a}{h}.$$

Далі, після здійснення розрахунків з урахуванням відомих властивостей вибіркової середньої і дисперсії:

1) $\left(\overline{k \cdot x}\right) = k \cdot \overline{x}; \quad D_в(k \cdot x) = k^2 \cdot D_в(x);$

2) $\left(\overline{x - a}\right) = \overline{x} - a; \quad D_в(x - a) = D_в(x),$

повертаються до початкової вибірки, для якої вирази початкової вибіркової середньої, дисперсії і середнього квадратичного відхилення через відповідні характеристики умовних варіантів мають вигляд:

$$\overline{x} = h \cdot \overline{u} + a; \quad D_в(x) = h^2 \cdot D_в(u); \quad S_u = \sqrt{D_в(u)}; \quad S_x = h \cdot S_u.$$

Закон великих чисел. Центральна гранична теорема.

Випадковість виходу одиничної спроби, випробування або експерименту полягає в тому, що заздалегідь передбачати її вихід і оцінити характеристики цього виходу поки не можна. Проте при великій кількості випробувань *характеристики* випадкових подій і випадкових величин втрачають випадковий характер. Ця властивість дозволяє використовувати результати спостережень за масовими явищами і *прогнозувати* можливі результати майбутніх випробувань.

Умови, за яких можна використовувати статистичні оцінки характеристик випадкових величин для оцінки їх параметрів (ймовірності, математичного сподівання, дисперсії, законів розподілу), визначаються *системою граничних теорем*, яка називається *законом великих чисел*. Розглянемо деякі з цих теорем без доведення.

Нерівність Чебишова. Нехай є випадкова величина ξ з математичним сподіванням m і дисперсією D. Нерівність Чебишева стверджує, що яке б не було позитивне число α, ймовірність того, що величина ξ відхилиться від свого математичного сподівання m *не менше* ніж на α (тобто більше або дорівнює α), обмежена зверху величиною $\dfrac{D}{\alpha^2}$, що формально можна записати так:

$$P(|X - m| \ge \alpha) \le \frac{D}{\alpha^2}. \tag{1}$$

Зокрема, якщо взяти $\alpha = 3\sigma$, то можна знайти оцінку понад ймовір-ності відхилення величини ξ від свого математичного сподівання не менше ніж на 3σ:

$$P(|X - m| \ge 3\sigma) \le \frac{D}{(3\sigma)^2} = \frac{1}{9} \approx 0{,}111. \tag{2}$$

Відзначимо, що на практиці в більшості випадків ймовірність того, що величина ξ вийде за границі ділянки $m \pm 3\sigma$, значно менше 1/9. Наприклад, для нормального закону ця ймовірність дорівнює 0,003. У випадку, коли закон розподілу випадкової величини невідомий, а відомі тільки m і σ, то ділянкою практично можливих значень випадкової величини вважають відрізок $m \pm 3\sigma$ (так зване "*правило трьох сигм*").

Теорема Чебишова. Якщо в n незалежних випробуваннях спостеріга-ються $x_1, \ldots, x_n$ значень випадкової величини ξ, то при $n \to \infty$ середнє арифметичне випадкової величини ξ

збігається за ймовірністю до її математич-ного сподівання *m*. Тобто при будь-якому $\varepsilon > 0$ і $\delta > 0$ виконується нерівність:

$$\lim_{n\to\infty} P\left(\left| m - \frac{1}{n}\sum_{i=1}^{n} x_i \right| \geq \varepsilon \right) \leq 1-\delta. \tag{3}$$

Для уточнення змісту поняття "*збігається за ймовірністю*" з'ясуємо, як змінюється середнє арифметичне ($\overline{x}_n$) випадкової величини ξ *при збільшенні числа спроб* – числа доданків під знаком суми у виразі (9.3), від значення $n-1$ до значення n, одержимо:

$$\overline{x}_{n-1} = \frac{1}{n-1}\sum_{i=1}^{n-1} x_i; \qquad \to \qquad \sum_{i=1}^{n-1} \boldsymbol{x_i} = (\boldsymbol{n}-1)\cdot \overline{\boldsymbol{x}}_{n-1}. \tag{4}$$

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\left(\sum_{i=1}^{n-1} \boldsymbol{x_i} + x_n\right) = \frac{1}{n}\left((\boldsymbol{n}-1)\cdot \overline{\boldsymbol{x}}_{n-1} + x_n\right) = \overline{\boldsymbol{x}}_{n-1} + \frac{x_n - \overline{x}_{n-1}}{n}. \tag{5}$$

Уведемо позначення:

$$\Delta\overline{x}_n = \frac{x_n - \overline{x}_{n-1}}{n}. \tag{6}$$

Тоді з формули (9.5) одержимо:

$$\overline{x}_n = \overline{x}_{n-1} + \Delta\overline{x}_n. \tag{7}$$

У підсумку, нове (наступне) значення середнього арифметичного ($\overline{x}_n$) випадкової величини $\boldsymbol{\xi}$ можна рекурентно висловити *через попереднє значення* її середнього арифметичного (формула (9.7)) *з урахуванням* виникаючого збільшення $\Delta\overline{x}_n$ (формули (9.5, 9.6)). Під знаком суми (формули (9.4, 9.5)) складаються випадкові значення (x_i), тому *середнє арифметичне* $\overline{x}_n$ випадкової величини ξ також *буде* величиною *випадковою*.

Змінні, що є в чисельнику в правій частині виразу (9.6) для збільшення $\Delta\overline{x}_n$, є величинами *випадковими*. Врахуємо діапазон ($x_{max} \leq x_i \leq x_{min}$) можливих значень (x_i) випадкової величини ξ і спробуємо з'ясувати, як змінюється абсолютна величина приросту $\Delta\overline{x}_n$ при збільшенні числа спроб n. Для цього замінимо вираз в чисельнику у формулі (9.6) на явно більшу (не меншу) величину, одержимо:

$$\left|\Delta\overline{x}_n\right| \leq \left|\frac{x_{\max} - x_{\min}}{n}\right|. \tag{8}$$

Виявляється, що після першої спроби (при $n = 1$) величина приросту може дорівнювати величині всього діапазону можливих значень (x_i), після другої спроби (при $n = 2$) – половині діапазону, після третьої ($n = 3$) – третій частині діапазону можливих значень і так далі відбувається ділення діапазону можливих значень на усе більш дрібні частини. Отже, можливий *приріст оцінок* середнього арифметичного (формула (9.7)) при збільшенні числа спроб n буде зменшуватися за відзначеним законом (9.8). Знак різниці випадкових величин, що стоять у дужках (формула (9.6)), може бути як позитивним, так і негативним, тому можна стверджувати, що *абсолютне значення приросту* оцінок не перевищує величину, яка стоїть у правій частині нерівності (9.8).

При необмеженому збільшенні числа спроб n значення величини приросту $\Delta\overline{x}_n$ (формула (9.6)), залишаючись *випадковою величиною*, буде наближатися до нуля:

$$\lim_{n\to\infty} \Delta\overline{x}_n = \lim_{n\to\infty} \frac{x_n - \overline{x}_{n-1}}{n} = 0. \tag{9}$$

Отже, *середнє арифметичне* $\overline{x}_n$ випадкової величини ξ буде *величи-ною випадковою* при будь-якій *кінцевій* кількості доданків ($n < \infty$). Проте при необмеженому збільшенні числа доданків ($n \to \infty$) величина $\Delta\overline{x}_n$ додаткової *випадкової* складової (формула (9.7)) зменшується аж *до нуля* (формула (9.9)). Тому середнє арифметичне ($\overline{x}_n$) *перестає залежати від* значення *випадкової* добавки $\Delta\overline{x}_n$, набуваючи нову якість, тобто перетворюючись із випадкової величини $\overline{x}_n$ *в невипадкову величину*, яка називається математичне сподівання (m) випадкової величини ξ. До цього процесу і застосовують термін – "*збігається за ймовірністю*".

Із формули (9.8) шляхом диференціювання за змінною n можна одержати оцінку швидкості (V) "*збіжності за ймовірністю*":

$$V = \left| \frac{d\Delta\overline{x}}{dn} \right| \le \frac{-|x_{\max} - x_{\min}|}{n^2}. \tag{10}$$

Отже, швидкість "*збіжності за ймовірністю*" (V) негативна і обернено пропорційна квадрату обсягу n вибірки.

Приклад зміни величини приросту $\Delta\overline{u}_{n.відн}$ оцінок середнього арифметичного $\overline{u}_n$ випадкової величини, яка відповідає початковій величині ξ, рівномірно розподіленої у границях $x_i \in [0;\ 1]$, при зміні числа значень n (числа елементів у вибірці) для $x_{\min} = 0$; $x_{\max} = 1$; $C = 1$.

Теорема Бернуллі. *Якщо проводиться n незалежних випробувань, у кожному з який подія A може з'явитися з ймовірністю p, то частота m/n появи події A при нескінченному збільшенні числа n спроб збігається за ймовірністю до величини p, тобто при будь-якому* $\varepsilon > 0$ і $\delta > 0$:

$$\lim_{n\to\infty} P\left(\left| \frac{m}{n} - p \right| \ge \varepsilon \right) \le 1 - \delta,$$

де m – число позитивних виходів (число появ події A) у n випробуваннях.

Центральна гранична теорема. *Якщо* $\xi_1, \ldots, \xi_n$ *– незалежні, однаково розподілені випадкові величини, що мають математичне сподівання* m_x *і дисперсію* σ_x^2*, то при* $n\to\infty$ *закон розподілу їх суми* $\sum_{i=1}^{n} \xi_i$ *нескінченно наближається до нормального.*

Дана теорема може бути справедливою і для суми неоднаково розподілених, але порівнянних за своїм розкидом, доданків. На практиці її застосовують і для випадків суми від десятьох доданків, а іноді і менше. При цьому не обов'язково знати закони розподілу окремих доданків, достатньо знати їх математичні сподівання і дисперсії.

Для дискретних випадкових величин окремим випадком центральної граничної теореми є теорема Лапласа.

Теорема Лапласа. *Якщо в кожному з n незалежних випробувань подія A з'являється з ймовірністю p (і не з'являється з ймовірністю q=1–p), то границя ймовірності влучення нормованої випадкової величини m – числа появ події A у n спробах має вигляд:*

$$\lim_{n\to\infty} P\left(a < \frac{\boldsymbol{m} - n\cdot p}{\sqrt{n\cdot p\cdot q}} < b \right) = \Phi^*(b) - \Phi^*(a), \quad \Phi^*(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{v^2}{2}} dv,$$

де $\Phi^*(t)$ – функція Лапласа.

Статистичні оцінки параметрів.

Розглянуті раніше формули для статистичних оцінок фактично реалізують ідею усереднення значень оцінюваних характеристик, які (характеристики, параметри) позначимо символом E. При цьому розрізняють реальне (істинне) значення параметра E, що поки невідомо, і одержуване зі спроб значення оцінки $\widetilde{E}$ цього параметра.

Означення. *Оцінки*, отримані в результаті статистичної обробки результатів ***n*** спостережень деякої випадкової величини ξ (у результаті n спроб), повинні задовольняти *вимогам*: *незсуненності*, *ефективності* та *обгрунтованості*.

1. *Незсуненою* називається оцінка $\widetilde{E}$, якщо її математичне сподівання $M[\widetilde{E}]$ дорівнює *самому* оцінюваному параметру E при будь-якому обсязі вибірки, тобто: $M[\widetilde{E}] = E$.

2. *Ефективною* називається оцінка $\widetilde{E}$, якщо вона серед усіх інших *незсунених* оцінок цього параметра має *найменшу дисперсію* даної оцінки $\widetilde{E}$ щодо істинного значення E, у порівнянні з будь-якою іншою оцінкою $\widetilde{E}_k$: $M[(\widetilde{E} - E)^2] \le M[(\widetilde{E}_k - E)^2]$.

3. *Обгрунтованою* називається оцінка $\widetilde{E}$, якщо в міру зростання n – числа спостережень (спроб) оцінка *збігається за ймовірністю* до самого оцінюваного параметра E:

$$\lim_{n\to\infty} P(|\widetilde{E} - E| < \varepsilon) = 1, \qquad \varepsilon > 0$$

або

$$\lim_{n\to\infty} M\left[(\widetilde{E} - E)^2\right] = 0.$$

Можна строго показати, що оцінка математичного очікування:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{11}$$

є незсуненою, обгрунтованою, має дисперсію оцінки:

$$D[\bar{x}] = D[\widetilde{m}] = \frac{1}{n}\cdot D_x, \tag{12}$$

а якщо випадкова величина ξ розподілена за нормальним законом, то дисперсія (9.12) буде мінімально можливою, тобто оцінка (9.11) – ефективна.

Дійсно, у виразі (9.11) складаються незалежні значення однієї і тієї ж випадкової величини x, що мають однакову дисперсію:

$$D_{x1} = D_{x2} = \ldots = D_{xi} = D_x.$$

Застосовуючи операцію дисперсії до лівої і правої частин рівняння (9.11) і з огляду на властивості дисперсії, одержимо:

$$D[\bar{x}] = D\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2}\cdot D\left[\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2}\cdot\sum_{i=1}^{n} D_x = \frac{1}{n^2}\cdot nD_x = \frac{1}{n}\cdot D_x.$$

Оцінка вибіркової *дисперсії* $D_в$ випадкової величини ξ:

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

є ефективною, обгрунтованою, але *зсуненою* убік менших значень у $\frac{n-1}{n}$ разів.

Для усунення цього зсуву праву частину рівності (9.12) достатньо *розділити* на зазначений множник, у підсумку одержимо так називану “виправлену” дисперсію:

$$\widetilde{D}_{в.\,випр} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \widetilde{m}_x)^2\cdot n_i = \left[\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \widetilde{m}_x^2\right]\cdot\frac{n}{n-1} = S_x^2\cdot\frac{n}{n-1} \tag{13}$$

У випадку *нормального* розподілу випадкової величини ξ *дисперсію* самої *оцінки* виправленої дисперсії (9.13) і середнє квадратичне відхилення знайдемо за формулами:

$$D[\widetilde{D}_x] = \frac{2}{n-1}\widetilde{D}^2, \qquad \sigma_x = \widetilde{D}_x\cdot\sqrt{\frac{2}{n-1}}.$$

Одержувані статистичні (точкові) оцінки сходяться до істинних оцінок лише при *нескінченному* збільшенні числа n спроб. Отже, заміна *істинних* значень шуканих параметрів їх статистичними оцінками завжди супроводжується *помилками*. Якщо такі помилки занадто великі, то наслідки заміни можуть бути істотними. Тому потрібно вміти оцінювати точність і надійність одержуваних оцінок, особливо при малій кількості спостережень n (при малій кількості спроб).

Для визначення точності і надійності статистичних оцінок використовують поняття “довірчий інтервал” і “довірча ймовірність”.

Довірчий інтервал, довірча ймовірність.

Припустимо, що в результаті n спроб потрібно одержати статистичну оцінку $\tilde{E}$ параметра, що має реальне значення E, із похибкою не більш ε, причому надійність такої оцінки повинна характеризуватися деякою ймовірністю β.

Розглянемо подію, яка полягає в тому, що *реальне* (істинне) значення параметра E, яке поки невідоме (наприклад, математичне сподівання випадкової величини ξ), буде відрізнятися *від одержуваного* зі спроб значення *оцінки* $\tilde{E}$ цього параметра не більш, ніж на величину ε. Ймовірність такої події можна записати у такий спосіб:

$$P(\tilde{E}-\varepsilon < E < \tilde{E}+\varepsilon) = P(|\,\tilde{E}-E\,| < \varepsilon) = \beta\,.$$

Така рівність означає, що з ймовірністю β невідоме значення параметра E потрапляє в інтервал I_β (див. рис. 9.7):

$$I_\beta = (\tilde{E}-\varepsilon\,;\; \tilde{E}+\varepsilon).$$

Означення. *Довірчим інтервалом* називається встановлений інтервал I_β можливих значень оцінюваного параметра E за рівнем довірчої ймовірності β.

Означення. *Довірчою ймовірністю* називається встановлена ймовірність β, з якою значення невідомого параметра E потрапляє в довірчий інтервал $\pm\varepsilon$ в околі отриманої оцінки $\tilde{E}$ цього параметра.

Тоді діапазон практично можливих значень *помилки*, що виникає при заміні реального значення параметра E на його оцінку $\tilde{E}$, із ймовірністю β *дорівнюватиме* $\pm\varepsilon$. Ще більші помилки будуть з'являтися тільки з малою ймовірністю $\alpha = 1 - \beta$, значення якої (α) називається *рівнем значущості*.

Як значення довірчої ймовірності у статистиці беруть числа, близькі до одиниці: 0,95; 0,99.

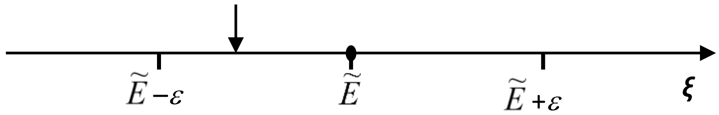


Рис. 9.7. Поняття “довірчий інтервал”.

Знайти довірчі оцінки для величини $\bar{x}$ – оцінки математичного очікування випадкової величини ξ.

● Із формули для оцінки математичного сподівання:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

випливає, що ця характеристика є сумою n *незалежних*, однаково розподілених випадкових величин x_i. При будь-якому кінцевому значенні n сама *величина* $\bar{x}$ також *є випадковою*. Відповідно до центральної граничної теореми теорії ймовірностей (див. п. 9.5), при достатньо великому n закон розподілу суми випадкових величин близький до нормального. На практиці ця умова починає виконуватися, починаючи з $n \geq 10$. Тому можна вважати, що величина $\overline{x}$ має нормальний закон розподілу із математичним сподіванням *m*, дисперсією $D[\overline{x}] = \sigma_m^2 = \frac{D_x}{n} = \frac{\sigma_x^2}{n}$ і СКВ, що дорівнюють (див. п. 9.6):

$$\sigma_m = \sqrt{\frac{\sigma_x^2}{n}} = \frac{\sigma_x}{\sqrt{n}}, \quad \text{де} \quad \sigma_x^2 = \left[\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\overline{x}\right)^2\right] \cdot \frac{n}{n-1} = \sigma_{x.\,випр}^2.$$

Знайдемо таку величину *ε*, для якої ймовірність відхилення оцінки $\overline{x}$ від реального значення математичного сподівання *m* дорівнює довірчій ймовірності β , тобто: $P(|\,\overline{x} - m\,| < \varepsilon) = \beta$.

Раніше відзначалося, що ймовірність влучення деякої нормально розподіленої випадкової величини η на інтервал ($\pm\varepsilon$), симетричний щодо її математичного сподівання *m*, знаходиться як ймовірність відхилення центрованої випадкової величини (η–*m*) не більш ніж на ($\pm\varepsilon$):

$$P(|\,\eta - m\,| \le \varepsilon) = \Phi\left(\frac{\varepsilon}{\sigma}\right) - \Phi\left(\frac{-\varepsilon}{\sigma}\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right). \qquad (14)$$

У даній ситуації випадковою є величина $\overline{x}$, яка має нормальний розподіл із математичним сподіванням *m* і середнім квадратичним відхиленням: $\sigma = \sigma_m = \frac{\sigma_x}{\sqrt{n}}$, тому формула (9.14), після підстановки цих величин, набуде вигляду:

$$P(|\,\overline{x} - m\,| < \varepsilon) = \beta = 2\Phi\left(\frac{\varepsilon}{\sigma}\right) = 2\Phi\left(\frac{\varepsilon \cdot \sqrt{n}}{\sigma_x}\right), \;\rightarrow\; 2\Phi\left(\frac{\varepsilon \cdot \sqrt{n}}{\sigma_x}\right) = \beta .$$

Із цієї рівності знайдемо величину довірчого інтервалу *ε*:

$$\Phi\left(\frac{\varepsilon \cdot \sqrt{n}}{\sigma_x}\right) = \frac{\beta}{2}; \;\rightarrow\; \frac{\varepsilon \cdot \sqrt{n}}{\sigma_x} = \arg\Phi\left(\frac{\beta}{2}\right); \;\rightarrow\; \varepsilon = \frac{\sigma_x}{\sqrt{n}} \cdot \arg\Phi\left(\frac{\beta}{2}\right) = \frac{\sigma_x}{\sqrt{n}} \cdot t_\beta ,$$

де $arg\,\Phi\,(x)$ – функція, обернена до функції $\Phi\,(x)$; $\quad t_\beta = \arg\Phi\left(\frac{\beta}{2}\right)$.

Нагадаємо, що поняття оберненої функції передбачає необхідність відшукання такого значення аргументу ($x = arg\,\Phi(x)$) функції $\Phi(x)$, при якому ця функція дорівнює відомому значенню, наприклад величині β/2. Як приклад скористаємося таблицею в додатку 9 і для $2\Phi\,(x)$=0,95 знайдемо значення аргументу *x*. Для цього на першому кроці з попередньо-го рівняння знайдемо значення функції $\Phi\,(x) = 0{,}95/2$=0,475. На другому кроці за таблицею в додатку 9 знаходимо, що значенню функції $\Phi(x) = 0{,}475$ відповідає значення аргументу x = 1,960, тобто $\Phi\,(x = 1{,}96) = 0{,}475$.

Відповідно до використаних позначень виявилося наступне:

$$x = t_\beta = \arg\Phi\left(\frac{\beta}{2}\right) = \arg\Phi\left(\frac{0{,}95}{2}\right) = \arg\Phi\left(0{,}475\right) = 1{,}960 .$$

Дані розглянутого прикладу, а також дані для інших значень β подані у табл. 9.6.

Отже, величина довірчого інтервалу має вигляд:

$$\varepsilon = \frac{\sigma_x}{\sqrt{n}} \cdot t_\beta . \qquad (15)$$

У підсумку, шукані границі інтервалу набудуть вигляду:

$$\overline{x} - \varepsilon < m < \overline{x} + \varepsilon ,$$

де, залежно від значення довірчої ймовірності β, величина t_β табульована і приймає значення, подані в табл. 9.6.

Таблиця 9.6

Параметри взаємозв'язку довірчої ймовірності

β	0,8	0,85	0,9	0,95	0,99	0,999
$t_\beta = \arg \Phi \left(\frac{\beta}{2}\right)$	1,282	1,439	1,643	1,960	2,576	3,290
$(t_\beta)^2$	1,644	2,071	2,699	3,842	6,636	10,824

Із формули (9.15) можна знайти оцінку необхідного числа спроб ***n***, яка дозволяє одержати необхідний розмір довірчого інтервалу ε при заданому значенні довірчої ймовірності β:

$$n = \frac{t_\beta^2 \cdot \sigma_x^2}{\varepsilon^2}. \qquad (16)$$

Для практичної оцінки інтервалу можливих значень реального середнього значення $\bar{x}$ ознаки ξ за інформацією наявної вибірки необхідно виконати такі дії.

1. Знайти оцінку математичного сподівання $\bar{x}$ значення ознаки ξ:

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

2. Знайти незсунену (виправлену) оцінку СКВ:

$$D_{x.\,випр} = \left[\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\bar{x}\right)^2\right] \cdot \frac{n}{n-1}; \qquad \sigma_x = \sqrt{D_{x.\,випр}} \; .$$

3. За значенням довірчої ймовірності β із таблиці 9.6 або з доданку 9 знайти значення величини t_β.

4. Розмір довірчого інтервалу ε розрахувати за формулою:

$$\varepsilon = \frac{\sigma_x}{\sqrt{n}} \cdot t_\beta \, .$$

5. Знайти шукані границі інтервалу, що будуть мати вигляд:

$$\bar{x} - \varepsilon < m_x < \bar{x} + \varepsilon.$$

6. При необхідності, потрібну кількість спроб (обсяг вибірки n) можна оцінити за формулою (9.16):

$$n = \frac{t_\beta^2 \cdot \sigma_x^2}{\varepsilon^2}.$$

Розрахунки можливі як в абсолютних, так і у відносних величинах:

$$\delta\varepsilon = \frac{\varepsilon}{m_x} = \frac{\sigma_x}{m_x\sqrt{n}} \cdot t_\beta ; \qquad n = \frac{t_\beta^2}{(\delta\varepsilon)^2} \cdot \left(\frac{\sigma_x}{m_x}\right)^2 .$$

При виконанні аналогічних оцінок *ймовірності p* події A за результатами n спроб спочатку знаходять статистичні оцінки ймовірності p^* події A як середнього значення випадкової величини $\xi = x_1 = 1$ (подія A з'явилася) і $\xi = x_2 = 0$ (подія A *не* з'явилася). Оцінки характеристик випадкової величини (ξ) в одній спробі дорівнюють:

$$M[x] = 0 \cdot q + 1 \cdot p = p; \qquad D[x] = D_x = p \cdot q = \sigma_x^2.$$

Для серії з n спроб статистична оцінка p^* ймовірності події A може бути виконана за відомими формулами оцінки математичного сподівання і дисперсії випадкової величини (ξ):

$$p^* = \frac{1}{n}\sum_{i=1}^{n} x_i; \quad q^* = 1 - p^*; \quad D[p^*] = \frac{D_x}{n} = \frac{p^* \cdot q^*}{n}; \quad \sigma_{p^*} = \sqrt{\frac{p^* \cdot q^*}{n}} \; .$$

Тоді далі можна задатися величиною довірчої ймовірності β (і відпо-відно t_β) і *для загального випадку* знайти границі довірчого інтервалу, використовуючи формули (4.40, 4.41):

$$\varepsilon = \frac{\sigma_{p^*}}{\sqrt{n}} \cdot t_\beta , \rightarrow \; p^* - \varepsilon < m_x < p^* + \varepsilon \; .$$

Проте якщо розмір добутку np^* і nq^* більший чотирьох, то розподіл випадкової величини p^* виявляється близьким до нормального, що дозволяє визначити довірчий інтервал оцінки ймовірності p^* події A з урахуванням допустимості застосування нормального закону розподілу:

$$\varepsilon = \sqrt{\frac{p^* \cdot q^*}{n}} \cdot t_\beta = \sigma_{p^*} \cdot t_\beta ; \qquad (17)$$

$$p^* - \sigma_{p^*} \cdot t_\beta < p < p^* + \sigma_{p^*} \cdot t_\beta .$$

Значення t_β вибираються із табл. 9.6.

Із формули (9.17) можна знайти оцінку числа спроб n, необхідного для одержання результатів із довірчою ймовірністю β (і відповідно t_β) і з довірчим інтервалом ε:

$$n \cdot \varepsilon^2 = p^* \cdot q^* \cdot (t_\beta)^2 \to n = \frac{t_\beta^2 \cdot p^* \cdot (1 - p^*)}{\varepsilon^2} . \qquad (18)$$

Частота появи події A (успішного складання іспиту студентом) у серії з n = 100 спроб (число студентів на курсі) виявилася рівною n^* = 0,78. Визначити 90% довірчий інтервал для ймовірності p події A.

● Перевіряємо придатність використання нормального закону, для цього оцінимо величини np і nq. Уважаючи $p \approx n^*$, одержимо:

$$np \approx n \cdot n^* = 78 ; \quad nq \approx n \cdot (1 - p^*) = 22 .$$

Обидві величини значно більше чотирьох, отже, нормальний закон можна застосовувати. Із табл. 9.6 для значення β = 0,9 знаходимо:

$$t_\beta = 1{,}643; \; D[p^*] = (0{,}78 \cdot 0{,}22)/(100 - 1) = 0{,}00173;$$

$$\sigma_{p^*} = \sqrt{D[p^*]} = 0{,}041; \; \varepsilon = t_\beta \cdot \sigma_{p^*} = 1{,}643 \cdot 0{,}041 = 0{,}068 .$$

Із формули (9.18) одержуємо:

$$p_1 = p^* - \sigma_{p^*} \cdot t_\beta = 0{,}712; \qquad p_2 = p^* + \sigma_{p^*} \cdot t_\beta = 0{,}848 .$$

Тоді шуканий довірчий інтервал I_β для ймовірності p події A (успішного складання іспиту) за рівнем довірчої ймовірності β = 0,9 виявиться таким, що дорівнює:

$$I_\beta = (0{,}712; \; 0{,}848) .$$

Чи відповідають дійсності рекламні дані про параметри того або іншого товару? Таке питання може виникнути в ході прийняття рішень щодо управління діяльністю підприємства. Відповідь на це питання можна одержати, розв'язавши задачу порівняння вибіркової середньої із значенням параметра, що анонсується. Розглянемо такий приклад.

Фірма-постачальник комплектуючих виробів у рекламному буклеті стверджує, що середній термін безвідмовної роботи запропонованого виробу знаходиться у границях від 2600 до 3100 годин і в серед-ньому складає – 2900 годин. Перед закупівлею великої партії виробів була зроблена вибірка з 50 виробів і проведені випробування. Середній $\left(\overline{\boldsymbol{T}}_{50}\right)$ термін безвідмовної роботи виявився рівним 2 720 годинам при "виправле-ному" середньому квадратичному відхиленні 361 годин. Необхідно при 5% рівні значущості:

1) перевірити гіпотезу про те, що значення 2900 годин дійсно є математичним сподіванням часу безвідмовної роботи виробів;

2) у випадку незадовільного результату перевірки оцінити:

– яким повинно бути мінімальне значення середнього часу роботи виробів в отриманій вибірці, при якому можна було б уважати рекламні дані дійсними;

– яким варто уважати *реальне* значення максимально можливого часу безвідмовної роботи виробу.

● Наявне число спроб n = 50 достатньо для того, щоб використовувати нормальний закон розподілу випадкового відхилення середнього значення від свого математичного сподівання.

1. Знайдемо розмір довірчого інтервалу середнього часу безвідмовної роботи виробу. Для цього з урахуванням 5% рівня значущості знаходимо спочатку величину $\beta = 1 - 0{,}05 = 0{,}95$, потім із табл. 4.6 знаходимо величину t_β (0,95) = 1,96. Далі, за наявними початковими даними n = 50, σ_x =

361 і $t_\beta\,(0{,}95) = 1{,}96$, знаходимо величину довірчого інтервалу для оцінки середнього значення часу справної роботи виробів за формулою:

$$\varepsilon = \frac{\sigma_x}{\sqrt{n}} \cdot t_\beta = \frac{361 \cdot 1{,}96}{\sqrt{50}} = \frac{707{,}56}{7{,}07} = 100 \text{ годин.}$$

У спробі відхилення склало $\Delta = 2900 - 2720 = 180$ годин, що перевершує розмір довірчого інтервалу ($180 > 100$). Отже, *фірма* в рекламі *завищує* термін безвідмовної роботи виробів.

2. Мінімальним значенням часу (T) безвідмовної роботи виробів у вибірці, при якому можна було б вважати рекламні дані дійсними, є час:

$$T = 2900 - \varepsilon = 2900 - 100 = 2800 \text{ годин.}$$

Реальне значення максимально можливого часу (T) безвідмовної роботи виробу за даними вибірки є:

$$T = 2720 + \varepsilon = 2720 + 100 = 2820 \text{ годин.}$$

У практиці страхового бізнесу і в практиці менеджменту іноді виникає необхідність *оцінити ймовірність* порівняно *рідкісних подій*, таких як аварії на виробництві, смерть застрахованої особи у результаті добре відпрацьованої і практично безпечної хірургічної операції, нещасний (або страховий) випадок природного характеру і т. ін. Причому може виявитися, що можливість розглянутого випадку теоретично існує, а реально такий випадок або не мав місця зовсім, або зафіксований один-два рази, що не дає підстав для статистичних оцінок. Як у такій ситуації можна одержати кількісні оцінки, корисні для прийняття розв'язань, розглянемо на прикладі.

Фірмою, що спеціалізується на транспортуванні озброєння, зроблено 100 перевірок саме довільного спрацьовування упакованого детонатора при падінні з вантажівки (подія A). Детонатор не спрацював.

1. Знайти 90% довірчу оцінку ймовірності p саме довільного *неспра-цьовування* упакованого детонатора при падінні з вантажівки.

2. Скільки разів потрібно переконатися у відсутності саме довільного спрацьовування упакованого детонатора при падінні з вантажівки, щоб із гарантією 95% стверджувати, що на практиці самовибух буде не більш ніж у 5% усіх випадків падіння?

● Шукана ймовірність p появи події A при n спробах є малою величиною, що дозволяє припустити її розподіл за законом Пуассона з математичним сподіванням $a = np$.

Довірчу ймовірність *непояви* події A (появи рівно 0 разів) відповідно до першого пункту задачі можна оцінити як $1-\beta$ і одночасно знайти цю ймовірність за формулою Пуассона:

$$P_{k=0} = \mathrm{e}^{-np} = 1-\beta.$$

Звідки для $\beta = 0{,}9$ знаходимо:

$$p = \frac{-\ln(1-\beta)}{n} = \frac{-\ln(0{,}1)}{100} = \frac{2{,}303}{100} = 0{,}023.$$

Для відповіді на друге питання врахуємо, що $\beta = 0{,}95$; $p = 0{,}05$. Із тієї ж формули одержуємо необхідне число спроб n:

$$n = \frac{-\ln(1-\beta)}{p} = \frac{-\ln(0{,}05)}{0{,}05} = \frac{2{,}996}{0{,}05} = 59{,}9 \approx 60.$$

Перевірка статистичних гіпотез. Критерій Пірсона.

Після виконання оцінок статистичних параметрів може виникнути *задача ідентифікації* отриманого в спробі розподілу $F^*(x)$ випадкової величини ξ з *відомими* теоретичними функціями розподілу. Таку задачу називають *задачею вирівнювання*.

Означення. *Задачею вирівнювання статистичних рядів* називається задача добору підхожої теоретичної функції розподілу ймовірностей або щільності ймовірностей при обробці статистичних даних. Задача може бути виконана як за методом найменших квадратів (див. п. 9.10), так і з інших міркувань, наприклад із використанням *методу моментів*.

Метод моментів передбачає виконання вимоги збігу статистичних моментів статистичного ряду випадкової величини ξ, наприклад математичного сподівання і дисперсії, із їх значеннями в теоретичній кривій розподілу ймовірностей $f(x)$. Якщо крива лінія $f(x)$ залежить від двох параметрів, то можна підібрати їх так, щоб співпали перші два моменти і так далі – до чотирьох моментів. *Точність* розрахунків моментів вище четвертого порядку різко спадає, тому їх застосування уважається недоцільним.

Після вирівнювання статистичного розподілу $F^*(x)$ за допомогою теоретичної кривої $F(x)$ можна помітити, що між цими кривими є роз-біжності.

Тому виникає таке питання: чи є отримані розбіжності випадковими, пов'язаними з недостатньою кількістю спроб, або вони вказують на невідповідність підібраної теоретичної кривої лінії $F(x)$ реальному розподілу $F^*(x)$.

Формалізуємо цю ситуацію у вигляді такої гіпотези H_0: випадкова величина ξ має інтегральний закон розподілу $F(x)$.

Для відповіді на питання "Дану гіпотезу H_0 варто прийняти або спростувати?" служать так називані "критерії згоди".

Критерій Пірсона (критерій χ^2).

Нехай за вибіркою обсягу n побудований дискретний (див. табл. 9.7) або інтервальний (див. табл. 9.8) варіаційний ряд,

Таблиця 9.7

x_i	x_1	x_2	…	x_m
n_i^*	n_1^*	n_2^*	…	n_m^*

Таблиця 9.8

$a_{i-1}\div a_i$	$a_1\div a_2$	$a_2\div a_3$	…	$a_{m-1}\div a_m$
n_i^*	n_1^*	n_2^*	…	n_m^*

де n_i^* ($i = 1, 2, \ldots, m$) – відносні частоти, причому $\Sigma\, n_i^* = 1$.

Перевіряється гіпотеза H_0, яка підтверджує, що вибіркова сукупність *має* конкретний *закон* $F(x)$ розподілу. Таким законом може бути нормальний закон, закон Пуассона і т. ін.

Процедура застосування критерію Пірсона (критерію χ^2) складається з таких етапів.

1. За варіантним рядом знаходяться "*реальні*" оцінки невідомих параметрів *припущеного* теоретичного закону розподілу $F(x)$.

2. Визначаються *теоретичні* частоти на основі припущеного закону $F(x)$ розподілу в такий спосіб.

Якщо ξ – дискретна випадкова величина, то обчислюються ймовірності *кожного значення* цієї величини $p_i = p(\xi = x_i)$.

Якщо ξ – безперервна випадкова величина, то обчислюються *ймовірності влучення* значень x_i випадкової величини ξ *у* кожний i-й *інтервал*, тобто: $p_i = P(a_{i-1} < x_i < a_i) = F(a_i) - F(a_{i-1})$.

Потім за знайденими ймовірностями p_i і обсягом вибірки n розраховуються значення *теоретичних* частот n^T_i за формулою:

$$n_i^T = n \cdot p_i .$$

Значення ($n \cdot p_i$) може бути дробовим числом, а частота n^T_i повинна виражатися цілим числом, тому за теоретичну частоту n^T_i можна взяти найближче до ($n \cdot p_i$) ціле число, при цьому $\sum n_i^T = n$.

3. За *кількісну міру* ступеня *збігу* теоретичного закону розподілу з законом розподілу у вибірковій сукупності (у якості "критерію" перевірки істинності нульової гіпотези H_0) приймається випадкова величина χ^2:

$$\chi^2 = \sum_{i=1}^{m} \frac{(n_i - n_i^T)^2}{n_i^T} .$$

Підставляючи в цю формулу *отримані* з вибірки емпіричні частоти n_i і обчислені *теоретичні* частоти $n^T{}_i$, знаходять значення випадкової величини χ^2, яку можна позначити як $\chi^2{}_{спост}$.

4. Розраховується число ступенів свободи r. Виявляється, що закон розподілу випадкової величини χ^2 визначається *одним* параметром r – *числом ступенів свободи*, яке обчислюється за формулою:

$$r = m - k - 1,$$

де m – кількість *варіантів* або інтервалів заданого варіаційного ряду,

k – кількість *параметрів* припущеного теоретичного розподілу, *які* (*параметри*) *оцінюються* за вибіркою. Число k є кількістю умов, що “обмежують свободу”.

Ще однією умовою (обмеженням) *завжди* є умова нормування, тому в правій частині відзначеної формули є величина “–1”.

Зокрема, якщо припущений розподіл – *нормальний*, то оцінюють *два* параметри (математичне сподівання a випадкової величини і середнє квадратичне відхилення σ), тому $k = 2$ і число ступенів свободи виявляється рівним $r = m - 3$.

Якщо припускають, що генеральна сукупність розподілена за законом Пуассона, то оцінюють один параметр λ , тому $k = 1$ і $r = m - 2$.

5. За заданим рівнем значущості α і числом ступенів свободи r, використовуючи таблицю критичних точок розподілу χ^2, знаходять критичну точку $\chi^2_{кр}(\alpha, r)$.

6. Якщо $\chi^2_{спост} < \chi^2_{кр}(\alpha, r)$ – то уважається, що розбіжності в спробі виявилися навіть меншими теоретичних і тому підстав відкинути гіпотезу H_0 *немає*.

7. Якщо $\chi^2_{спост} > \chi^2_{кр}(\alpha, r)$ – то гіпотезу H_0 відхиляють.

Зауваження. Для застосування критерію Пірсона необхідно, щоб обсяг вибірки n був достатньо великий ($n \geq 50$). Крім того, частоти n_i повинні бути $n_i \geq 5$. Тому частоти $n_i < 5$ варто об’єднати в більш численний інтервал. У цьому випадку і відповідні їм теоретичні частоти $n^T{}_i$ варто об’єднати.

Для упорядкування графіка роботи співробітників досліджу-вана кількість клієнтів, що звернулися у фірму в першій половині робочого дня. Результати дослідження числа клієнтів залежно від часу роботи подані в табл. 9.9.

Таблиця 9.9

Номер інтервалу часу, i	1	2	3	4
Часи роботи (a_i; a_{i+1})	9 – 10	10 – 11	11 – 12	12 – 13
Кількість клієнтів (n_i)	6	20	35	15

При рівні значущості α = 0,05 потрібно перевірити гіпотезу H_0 про те, що кількість клієнтів, які звернулися у відзначені часи роботи, має нормальний розподіл.

● У даному випадку значеннями безперервної випадкової величини ξ є відліки моментів часу (x_i) *приходу клієнтів* в офіс фірми, які (відліки) можуть потрапити в різні часи роботи і сформувати в такий спосіб кількість значень випадкової величини ξ, що потрапили у від-повідний інтервал часу. Ці значення вже згруповані за $m = 4$ інтервалами, що дозволяє вважати середину кожного інтервалу – представницьким значенням усіх значень випадкової величини, що потрапили в цей інтервал. Процес розв’зання оформимо у вигляді таблиці (див. табл. 9.10).

Відзначимо, що застосування критерію Пірсона потребує, щоб частоти були не менше 5, тобто $n > 5$. В іншому випадку необхідно об’єднати інтервали до одержання необхідних частот. У випадку даного прикладу такої необхідності немає.

1. За варіантним рядом (див. табл. 9.9) знаходимо “*реальні*” оцінки невідомих параметрів *припущеного* теоретичного, у даному випадку – нормального закону розподілу $F(x)$. Для цього спочатку знайдемо загальну кількість клієнтів, що відвідали офіс фірми (див. табл. 9.10, рядок 3):

$$n = \sum_{i=1}^{4} n_i = 76.$$

Далі знайдемо середні (представницькі) значення ($x_{iср}$) для кожного інтервалу (див. табл. 9.10, рядок 4). Потім домножимо кожне з цих середніх значень ($x_{iср}$) на кількість випадкових величин (n_i), що потрапили в цей інтервал (див. табл. 9.10, рядок 5), і знайдемо їх суму, яка дозволяє розрахувати оцінку математичного сподівання моментів часу появи клієнтів в офісі фірми:

$$\bar{x} = \tilde{m} = \frac{1}{n}\sum_{i=1}^{4} x_{icp} \cdot n_i = \frac{857}{76} = 11{,}276.$$

Для оцінки середнього *квадратичного* відхилення (S_x) моментів часу появи клієнтів спочатку (див. табл. 9.10, рядок 6) домножимо *квадрат* кожного із середніх значень інтервалів на кількість випадкових величин, які потрапили в цей інтервал ($x^2_{icp} \cdot n_i$), і знайдемо їх суму, що дозволяє розрахувати оцінку другого початкового моменту випадкового часу появи клієнтів в офісі фірми:

$$\tilde{M}[x^2] = \frac{1}{n}\sum_{i=1}^{4} x_{icp}^2 \cdot n_i = \frac{9719}{76} = 127{,}882.$$

Потім знайдемо оцінки дисперсії зсунутої і виправленої:

$$\tilde{D}_в = \tilde{M}[x^2] - \tilde{m}_x^2 = 127{,}882 - 127{,}148 = 0{,}726;$$

$$\tilde{D}_{в.\,випр} = \frac{n}{n-1} \cdot \tilde{D}_{в.} = \frac{4}{3} \cdot 0{,}726 = 0{,}968,$$

що дозволяє одержати оцінку виправленого (незсуненого) середнього квадратичного відхилення:

$$S_x = \sigma_{x.\,випр} = \sqrt{\tilde{D}_{в.\,випр}} = \sqrt{0{,}968} = 0{,}984.$$

2. Визначаємо *теоретичні* частоти на основі припущеного закону $F(x)$ розподілу в такий спосіб.

В даному випадку ξ – моменти часу приходу клієнтів в офіс фірми – безперервна випадкова величина, тому обчисляємо *ймовірності влучення* значень x_i випадкової величини ξ *у* кожний *i*-й *інтервал*, тобто $p_i = P(a_i < x_i < a_{i+1}) = F(a_{i+1}) - F(a_i)$, для чого скористаємося функцією Лапласа і відомим виразом оцінки ймовірності влучення випадкової величини на інтервал (a; b):

$$P(a < x < b) = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right),$$

одержимо формулу:

$$n \cdot \varepsilon^2 = p^* \cdot q^* \cdot (t_\beta)^2 \rightarrow n = \frac{t_\beta^2 \cdot p^* \cdot (1-p^*)}{\varepsilon^2}. \qquad (19)$$

У (9.19) аргументами є нормовані границі інтервалів. Тому далі знайдемо значення нормованих границь цих інтервалів (див. табл. 9.10, рядки 7, 8) і, скориставшись таблицею $\Phi(x)$ – значень функції Лапласа (додаток 9), запишемо відповідні значення (див. табл. 9.10, рядки 9 і 10). Відзначимо, що для невід'ємного значення аргументу варто використовувати властивість непарності функції: $\Phi(-x) = -\Phi(x)$. У наступному рядку знайдемо різницю значень функції Лапласа для кожного інтервалу, що і є теоретичною оцінкою ймовірності (p_i) влучення значень випадкової величини в кожний інтервал (див. табл. 9.10, рядок 11).

Таблиця 9.10

Схема оцінки правдоподібності гіпотези про нормальний розподіл безперервної випадкової величини

№ п/п	Параметри, що розраховуються	*i*				Σ	M[*]
		1	2	3	4		
1	a_i	9	10	11	12		
2	a_{i+1}	10	11	12	13		
3	n_i	6	20	35	15	76	
4	x_{icp}	9,5	10,5	11,5	12,5		
5	$x_{icp} \cdot n_i$	57	210	402,5	187,5	857	11,276

6	$x^2_{icp} \cdot n_i$	541,5	2205	4629	2344	9719	127,882
7	$x^{н}_{min.i} = (a_i - \overline{x})/S_x$	–2,313	–1,297	–0,281	0,735	$\widetilde{\boldsymbol{D}}_{\boldsymbol{в}} =$	0,726
8	$x^{н}_{max.i} = (a_{i+1} - \overline{x})/S_x$	–1,297	–0,281	0,735	1,752	$\widetilde{\boldsymbol{D}}_{\boldsymbol{в.випр}}$	0,968
9	$\Phi(x^{н}_{min.i})$	–0,490	–0,403	–0,111	0,269	$S_x =$	0,984
10	$\Phi(x^{н}_{max.i})$	–0,403	–0,111	0,269	0,460		
11	p_i	0,087	0,292	0,380	0,191	0,950	
12	$n_i^{Т}$	6,609	22,201	28,843	14,526		
13	$n_i - n_i^{Т}$	–0,609	–2,201	6,157	0,474		
14	$(n_i - n_i^{Т})^2 / n_i^{Т}$	0,056	0,218	1,314	0,016	1,604 =	$\chi^2_{спост}$

Потім за знайденими ймовірностями p_i і обсягом вибірки $n = 71$ розрахуємо (див. табл. 9.10, рядок 12) значення *теоретичних* частот n_i^T за формулою: $n_i^T = n \cdot p_i$.

3. За заданими в умові емпіричними частотами n_i (див. табл. 9.10, рядок 3) і отриманими теоретичними частотами n_i^T (див. табл. 9.10, рядок 12) обчисляємо величину $\chi^2_{спост}$. Для цього попередньо для кожного інтервалу знайдемо (див. табл. 9.10, рядок 13) різницю між емпіричною і теоретичною частотою влучення випадкової величини в цей інтервал: $(n_i - n_i^T)$. Потім знайдемо *окремі доданки* (див. табл. 9.10, рядок 14) для визначення значення показника "Хі-квадрат":

$$\chi_i^2 = \frac{(n_i - n_i^T)^2}{n_i^T}$$

і потім розрахуємо суму цих значень, що і є шуканою величиною для оцінки правдоподібності початкової гіпотези:

$$\chi^2_{спост} = \sum_{i=1}^{4} \frac{(n_i - n_i^T)^2}{n_i^T} = 1,604.$$

У вибірці з m=4 інтервалів оцінювалися *два* ($k = 2$) параметри – $\overline{x} = \widetilde{m}$ і S_x, тому кількість ступенів свободи r дорівнюватиме:

$$r = m - k - 1 = 4 - 2 - 1 = 1.$$

За таблицею знаходимо критичне значення:

$$\chi^2_{кр}(\alpha, r) = \chi^2_{кр}(0{,}05; 1) = 3{,}84; \qquad \chi^2_{спост} = 1{,}604.$$

Значення $\chi^2_{спост} = 1{,}604$, тобто $\chi^2_{спост} < \chi^2_{кр}(\alpha, r)$, тому *гіпотеза* H_0 про те, що кількість клієнтів (випадкова величина ξ) має нормальний розподіл – *приймається*.

При $n = 757$ випробувань блоків радіоелектронної апаратури отримані і наведені в табл. 9.11 дані про кількість відмов на один блок.

Таблиця 9.11

Кількість відмов (x_i)	0	1	2	3	4	5	≥6
Кількість випадків (n_i)	427	235	72	21	1	1	0

За рівнем значущості $\alpha = 0{,}05$ потрібно перевірити гіпотезу H_0 про те, що кількість відмов має розподіл Пуассона:

$$P_k = \frac{\lambda^k}{k\,!} \cdot e^{-\lambda}.$$

• 1. Знайдемо оцінку параметра λ розподілу Пуассона, яка дорівнює середньому числу відмов. Процес розв'язання оформимо у вигляді розрахункових таблиць 9.12–9.14.

Із табл. 9.12 знаходимо середнє число відмов: $\bar{x} = \dfrac{\sum_{i=1}^{7} x_i n_i}{\sum_{i=1}^{7} n_i} = \dfrac{451}{757} = 0{,}6$.

Отже, теоретичний закон розподілу числа відмов блоків апаратури має вигляд:

$$p_i = p(\xi = x_i) = \frac{(0{,}6)^{x_i}}{x_i\,!} e^{-0{,}6}, \qquad x_i = 0,\ 1, \ldots,\ 5.$$

Таблиця 9.12 *Таблиця* 9.13 *Таблиця* 9.14

i	x_i	n_i	$x_i\, n_i$	p_i	np_i	$n^{Т}_i$	n_i	$n^{Т}_i$	$(n_i – n^{Т}_i)^2$	$(n_i – n^{Т}_i)^2/n^{Т}_i$
1	0	427	0	0,5488	415,4504	416	427	416	121	0,291
2	1	235	235	0,3293	249,2702	249	235	249	196	0,787
3	2	72	144	0,0988	74,78107	75	72	75	9	0,120
4	3	21	63	0,0198	14,95621	15	23	17	36	2,118
5	4	1	4	0,0030	2,243432	2				
6	5	1	5	0,0004	0,269212	0				
7	6	0	0							
Σ		757	451			757			$\chi^2_{спост} = 3{,}316$	

2. На основі отриманого теоретичного закону розподілу знаходимо теоретичні частоти $n^{Т}_i$, для чого складемо розрахункову табл. 9.13. Ймовірність p_i знаходимо або за таблицею, або обчисляємо за формулою Пуассона для всіх значень x_i.

3. За заданими в умові емпіричними частотами $\boldsymbol{n}_i$ і отриманими теоретичними частотами $\boldsymbol{n}^{Т}_i$ обчисляємо величину $\chi^2_{спост}$:

$$\chi^2_{спост.} = \sum_{i=1}^{m} \frac{(n_i - n_i^T)^2}{n_i^T}.$$

Для цього складемо розрахункову таблицю (див. табл. 9.14), об’єднавши останні три рядки з рядком для $x_i = 3$, тому що застосування критерію Пірсона потребує, щоб частоти не були малими.

У вибірці оцінювався *один* параметр λ, тому число ступенів свободи r дорівнюватиме:

$$r = m - k - 1 = 4 - 1 - 1 = 2.$$

За таблицею знаходимо критичне значення:

$$\chi^2_{кр}(\alpha, r) = \chi^2_{кр}(0{,}05;\ 2) = 5{,}99.$$

Отримане в спробі $\chi^2_{спост}$.= 3,316, і це значення виявляється меншим критичного, тобто $\chi^2_{спост}. < \chi^2_{кр}(\alpha, r)$, тому *гіпотеза* H_0 про розподіл кількості відмов блоків апаратури за законом Пуассона *приймається*.

Вибіркове рівняння прямої лінії регресії.

Математична формалізація реальних процесів у вигляді функціональ-ної залежності однієї змінної (функції y) від іншої змінної (аргументу x) $y = f(x)$ дозволяє кожному значенню аргументу однозначно поставити у відповідність одне заздалегідь відоме значення функції. Така залежність, як правило, є оборотною, тобто за відомим значенням функції завжди можна знайти відповідне значення її аргументу: $x = f^{-1}(y)$.

Проте на практиці реальні процеси демонструють іншу відповідність змінних. Так за один і той же робочий час тим самим підприємством може бути зроблена різна кількість продукції, та сама продукція може мати різну собівартість, та сама кількість продукції може бути продана за різну ціну, люди того самого зросту можуть мати різну вагу.

У підсумку, *тому* самому *значенню* аргументу x можуть відповідати *різні* значення функції y. Часто значення функції носять характер випадкових величин, які при зміні значень аргументу можуть мати різні центри групування, дисперсію і закон розподілу. У цьому випадку зворотне перетворення з метою відшукання значення аргументу x за відомим значенням функції y не є однозначним.

Для встановлення залежності між такими неоднозначними змінними використовують теорію кореляційно-регресійного аналізу, у якій дослід-жуються зміни середніх значень функції при зміні одного або багатьох аргументів. Проте формальні методи кореляційного аналізу не дають відпо-віді на питання "що є причиною, а що є наслідком або яка змінна є аргументом, а яка – функцією?". Відповідь на це питання має дати дослідник.

Метою побудови регресійних моделей може бути встановлення залежності між *середніми* значеннями двох змінних (параметрів), одну з яких дослідник *призначає* функцією, а другу – її аргументом.

У основі регресійного аналізу лежать дві гіпотези.

1. Передбачається, що досліджувана сукупність параметрів *має* внутрішній статистичний зв'язок, який може бути виявлений і формалізований у вигляді кореляційної (отже лінійної) залежності одного параметра від іншого або від інших. Тобто вважається, що існує внутрішній лінійний зв'язок *середніх* значень цих параметрів.

2. Передбачається, що випадковий розкид (дисперсія) значень (кожного) параметру має регулярну компоненту, яка залежить від деякого аргументу ("сигналу"), і випадкову компоненту ("шум"). Випадкова компонента ("шум") розподілена за нормальним законом.

Початковою інформацією для побудови лінійної однофакторної регресійної моделі є сукупність із n двовимірних точок (x_i, y_i), де кожна координата точки, як правило, має свій фізичний сенс, наприклад x_i – зріст людини в сантиметрах, y_i – її вага в кілограмах.

Під час формалізації постановки задачі розглянемо двовимірну випадкову величину (X, Y), над якою проведено n незалежних випробувань і в результаті випробувань отримана вибірка – n пар чисел (координат точки):

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n),$$

де x_i – значення випадкової величини X у i-му випробуванні;
y_i – значення випадкової величини Y у i-му випробуванні.

Необхідно знайти *наближене* зображення значень однієї з випадкових величин як функції значень другої випадкової величини.

Означення. *Вибірковим рівнянням регресії Y* на X $(y \to x)$ називається рівняння, яке встановлює залежність змінної y від змінної x, тобто коли змінна y вважається *функцією*, а змінна x – *аргументом*: $y = f(x)$, при цьому початковою інформацією є вибірка з n пар чисел.

Означення. *Вибірковим* рівнянням регресії *X на Y* $(x \to y)$ називається рівняння $x = \varphi(y)$, у якому при тій же початковій інформації вже змінна x уважається *функцією*, а змінна **y** – її *аргументом*.

Означення. *Лінійною* називається регресія у випадку, коли залежності $f(x)$ і $\varphi(y)$ є *лінійними функціями*. Тоді рівняння регресії мають вигляд:

$$y = a \cdot x + b; \qquad x = c \cdot y + d.$$

Порядок розрахунку параметрів рівняння регресії розглянемо без виведення і для двох варіантів умов: при відсутності і при наявності збіжних точок у вибірці.

1. Нехай серед точок (x_i, y_i) вибірки збіжних точок немає. Для того щоб скласти вибіркове рівняння прямої лінії регресії, виконуються наступні розрахунки:

а) обчислюються середні значення величин: $\bar{x}$, $\bar{y}$, $\overline{x \cdot y}$, $\overline{x^2}$, $\overline{y^2}$ і знаходяться середні квадратичні відхилення $S_x = \sigma_x$, $S_y = \sigma_y$ із використанням формул, перерахованих з обліком доцільної послідовності їх застосування:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \,; \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \,; \quad \overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \,; \quad \overline{y^2} = \frac{1}{n}\sum_{i=1}^{n} y_i^2 \,;$$

$$\overline{x \cdot y} = \frac{1}{n}\sum_{i=1}^{n} x_i \cdot y_i \,; \quad S_x^2 = \overline{x^2} - \left(\bar{x}\right)^2 ; \quad S_x = \sqrt{S_x^2} \,; \quad S_y^2 = \overline{y^2} - \left(\bar{y}\right)^2 ; \quad S_y = \sqrt{S_y^2} \,;$$

б) обчислюється значення вибіркового коефіцієнту кореляції r:

$$r = \frac{\overline{x \cdot y} - \overline{x} \cdot \overline{y}}{S_x \cdot S_y}.$$

Нагадаємо, що вибірковий коефіцієнт кореляції характеризує рівень лінійного кореляційного зв’язку двох випадкових величин. Чим ближче $|r|$ до одиниці, тим більш сильним є зв’язок двох величин, чим ближче $|r|$ до нуля, тим зв’язок слабше;

в) для одержання рівняння регресії Y на X: $y = a \cdot x + b$ обчисляємо коефіцієнти a і b даного рівняння за формулами:

$$a = r\frac{S_y}{S_x}; \quad b = \overline{y} - a \cdot \overline{x}.$$

Для одержання рівняння регресії X на Y: $x = c \cdot y + d$ обчисляємо коефіцієнти c і d даного рівняння за формулами:

$$c = r\frac{S_x}{S_y}; \quad d = \overline{x} - c \cdot \overline{y}.$$

Обидві прямі лінії $y = a \cdot x + b$ і $x = c \cdot y + d$ проходять через точку $(\overline{x}, \overline{y})$. Для зображення обох прямих ліній на одному графіку друге рівняння варто подати у вигляді: $y = x/c - d/c$.

2. При великій кількості точок $\boldsymbol{n}$ у вибірці значення x_i може зустрітися $\boldsymbol{m}_i$ разів, значення y_j може зустрітися n_j разів. Та сама пара чисел (x_i, y_j) може зустрітися n_{ij} разів. У цьому випадку вибірку зручно подати у вигляді кореляційної таблиці (див. табл. 9.15) так, що кількість повторень m_i значень координат x_i, кількість повторень $\boldsymbol{n}_j$ значень координат y_j і обсяг вибірки $\boldsymbol{n}$ дорівнюватимуть:

$$m_i = \sum_{j=1}^{\ell} n_{ij}, \quad n_j = \sum_{i=1}^{k} n_{ij}; \quad n = \sum_{i=1}^{k}\sum_{j=1}^{\ell} n_{ij} = \sum_{j=1}^{\ell} n_j = \sum_{i=1}^{k} m_i = n.$$

Таблиця 9.15

Кореляційна таблиця

x_i	Y_j				
	y_1	y_2	…	y_ℓ	m_i
x_1	n_{11}	n_{12}	…	$n_{1\ell}$	m_1
x_2	n_{21}	n_{22}	…	$n_{2\ell}$	m_2
..	…	…	…	…	…
x_k	n_{k1}	n_{k2}	…	$n_{k\ell}$	m_k
n_j	n_1	n_2	…	n_ℓ	n

Вибіркове рівняння прямої лінії регресії знаходимо аналогічно першому випадку, розрахунки величин $\overline{x},\ \overline{y},\ \overline{x \cdot y},\ \overline{x^2},\ \overline{y^2}$ виконуємо з урахуванням наявності повто-рюваних значень змінних за форму-лами:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot m_i; \quad \overline{y} = \frac{1}{n}\sum_{i=1}^{\ell} y_i \cdot n_j;$$

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{k} x_i^2 \cdot m_i; \quad \overline{y^2} = \frac{1}{n}\sum_{i=1}^{\ell} y_i^2 \cdot n_j; \quad \overline{x \cdot y} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{\ell} x_i \cdot y_i \cdot n_{ij}.$$

Якщо розглядається вибірка з генеральної сукупності безперервних випадкових величин X і Y, то кореляційна таблиця буде містити інтервали $[a_{i-1}, a_i)$ і $[b_{j-1}, b_j)$.

У цьому випадку для обчислення величин $\overline{x},\ \overline{y},\ \overline{x \cdot y},\ \overline{x^2},\ \overline{y^2}$ спочатку потрібно перейти до дискретних рядів, а потім виконати обчислення за розглянутими вище формулами.

Зауваження. Якщо значення координат точок (x_i, y_j) є дуже великими або дуже маленькими числами, то при обчисленні значень величин $\overline{x},\ \overline{y},\ \overline{x \cdot y},\ \overline{x^2},\ \overline{y^2}$ можна спочатку перейти до розглянутих раніше (див. п. 9.4) *умовних варіантів* u_i і v_j:

$$u_i = \frac{x_i - \alpha}{p}, \quad v_j = \frac{y_j - \beta}{q}$$

і знайти їх середні значення $\overline{u}$, $\overline{v}$ і S_u, S_v, а потім знайти *середні* значення початкових координат $\overline{x}$, $\overline{y}$ та їх середніх квадратичних відхилень S_x, S_y:

$$\overline{x} = p \cdot \overline{u} + \alpha, \quad \overline{y} = q \cdot \overline{v} + \beta; \quad S_x = p \cdot S_u, \quad S_y = q \cdot S_v.$$

Перехід до умовних варіантів не змінює величину вибіркового коефіцієнту кореляції, тому $r_{xy} = r_{uv} = r$.

Знайти вибіркові рівняння прямих ліній регресії Y на X і X на Y за даними таблиці спостережень (див. табл. 9.16)

Таблиця 9.16

Таблиця спостережень

x	1,00	1,50	3,00	4,50	5,00
y	1,25	1,40	1,50	1,75	2,25

- Складемо розрахункову таблицю (див. табл. 9.17).

Таблиця 9.17

Розрахункова таблиця

i	x_i	y_j	x_i^2	y_j^2	$x_i\ y_j$
1	1,00	1,25	1,00	1,56	1,25
2	1,50	1,40	2,25	1,96	2,10
3	3,00	1,50	9,00	2,25	4,50
4	4,50	1,75	20,25	3,06	7,88
5	5,00	2,25	25,00	5,06	11,25
$\sum$	15,00	8,15	57,50	13,90	26,98
$\sum/n$	3	1,63	11,5	2,779	5,395

За допомогою таблиці знаходимо оцінки середніх значень:

$$\overline{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = \frac{15}{5} = 3; \quad \overline{y} = \frac{1}{5}\sum_{j=1}^{5} y_j = \frac{8{,}15}{5} = 1{,}63; \quad \overline{xy} = \frac{1}{5}\sum_{i=1}^{5} x_i \cdot y_i = \frac{26{,}975}{5} = 5{,}395;$$

$$\overline{x^2} = \frac{1}{5}\sum_{i=1}^{5} x_i^2 = \frac{57{,}5}{5} = 11{,}5; \quad \overline{y^2} = \frac{1}{5}\sum_{i=1}^{5} y_i^2 = \frac{13{,}897}{5} = 2{,}779.$$

Потім визначаємо значення середніх квадратичних відхилень:

$$S_x = \sqrt{\overline{x^2} - \left(\overline{x}\right)^2} = \sqrt{11{,}5 - 9} = \sqrt{2{,}5} = 1{,}58;$$

$$S_y = \sqrt{\overline{y^2} - \left(\overline{y}\right)^2} = \sqrt{2{,}779 - 2{,}657} = \sqrt{0{,}122} = 0{,}35.$$

Знайдемо вибірковий коефіцієнт кореляції:

$$r = \frac{\overline{x \cdot y} - \overline{x} \cdot \overline{y}}{S_x \cdot S_y} = \frac{5{,}395 - 3 \cdot 1{,}63}{1{,}58 \cdot 0{,}35} = \frac{0{,}505}{0{,}553} = 0{,}913.$$

Для вибіркового рівняння прямої лінії регресії Y на X, що має вигляд $y = a \cdot x + b$, обчислимо значення коефіцієнтів a і b за відомими формулами, одержимо:

$$a = r\frac{S_y}{S_x} = 0{,}913 \cdot \frac{0{,}35}{1{,}58} = 0{,}202;$$

$$b = \overline{y} - a \cdot \overline{x} = 1{,}63 - 0{,}202 \cdot 3 = 1{,}63 - 0{,}606 = 1{,}024.$$

Остаточно вибіркове рівняння прямої лінії регресії Y на X набуде вигляду:

$$y = 0{,}202 \cdot x + 1{,}024.$$

Для одержання рівняння регресії X на Y: $x = c \cdot y + d$ обчисляємо коефіцієнти c і d за формулами:

$$c = r\frac{S_x}{S_y} = 0{,}913\cdot\frac{1{,}58}{0{,}35} = 4{,}122\,; \quad d = \bar{x} - c\cdot\bar{y} = 3 - 4{,}122\cdot 1{,}63 = -3{,}719\,.$$

Вибіркове рівняння прямої лінії регресії X на Y набуде вигляду:

$$x = 4{,}122\cdot y - 3{,}719.$$

Для побудови даного рівняння в тій же системі координат, що і рівняння регресії Y на X, скористаємося варіантом перерахунку $y = x/c - d/c$ і виразимо змінну y через змінну x, одержимо:

$$y = 0{,}242\cdot x + 0{,}902.$$

Метод найменших квадратів.

У ряді практичних задач обробки результатів вимірів, у тому числі значень випадкових величин, виникає необхідність згладженого подання значень однієї величини (наприклад, величини *y)*, як функції іншої величини (наприклад, величини *x)*. Одним із найбільш поширених способів такого наближеного зображення є метод найменших квадратів.

За допомогою методу найменших квадратів розв'язується задача добору такої аналітичної залежності $y(x) = \Psi_m(x, a_0, a_1, \ldots, a_m)$, графік якої *не обов'язково* проходив би через *усі* задані точки, але максимально "згладжував" би випадкові похибки вимірюваних ординат функції $y_i = f(x_i)$ $(i = 0, 1, 2, \ldots, n)$, тобто щоб сума квадратів відхилень значень аналітичної залежності $\Psi_m(x, a_0, a_1, \ldots, a_m)$ від значень вимірюваних ординат $y_i = f(x_i)$ у цих точках була мінімальною:

$$S = \sum_{i=1}^{n}[y_i - \Psi_m(x_i)]^2 \Rightarrow \min.\, \Psi_m(x_i). \qquad (20)$$

Апроксимація за методом найменших квадратів виконується у два етапи:

– *на першому* етапі вибирають *вигляд* $\Psi_m(x, a_0, \ldots, a_m)$ шуканої формули;

– *на другому* етапі для формули обраного вигляду "підбирають" *значення параметрів* $a_0, a_1, \ldots, a_m$, виходячи з вимоги (9.20).

Процес добору полягає в одержанні системи з (m + 1)-го рівняння для визначення значень усіх (m + 1) параметрів $a_0, a_1, \ldots, a_m$.

Виходячи з вимоги мінімізації відхилення значень $\Psi_m(x_i)$ від значень $y_i = f(x_i)$, система рівнянь формується шляхом відшукання похідних за кожним параметром (a_k) від рівняння (9.20) і прирівнювання їх до нуля:

$$\frac{\partial S}{\partial a_k} = -2\sum_{i=1}^{n}[y_i - \Psi_m(x_i, a_0, \ldots, a_k, \ldots, a_m)]\cdot\left(\frac{d\,\Psi_m}{da_k}\right)_{x_i} = 0;\ \ k = 0, 1, \ldots, m.$$

Розділивши ліву і праву частини на множник (–2), який "заважає", одержимо:

$$\sum_{i=1}^{n}[y_i - \Psi_m(x_i, a_0, \ldots, a_k, \ldots, a_m)]\cdot\left(\frac{d\,\Psi_m}{da_k}\right)_{x_i} = 0; \quad k = 0, 1, \ldots, m, \qquad (21)$$

де $\left(\frac{d\,\Psi_m}{da_k}\right)_{x_i} = \Psi'_{a_k}(x_i, a_0, \ldots, a_k, \ldots, a_m)$ – значення часткової похідної від функції Ψ_m за параметром a_k у точці x_i.

Для розв’зання системи рівнянь (9.21) потрібно задатися конкретним видом апроксимуючої функції Ψ_m.

Лінійна апроксимація.

У цьому випадку $\Psi'_{a_0} = 1$; $\Psi'_{a_1} = x$. Тоді рівняння (9.21) набуде вигляду:

$$\sum_{i=1}^{n}[y_i - (a_0 + a_1\, x_i)] = 0\,; \qquad \sum_{i=1}^{n}[y_i - (a_0 + a_1\, x_i)]\, x_i = 0\,.$$

У цьому рівнянні розкриємо дужки і зробимо підсумовування, що доз-воляє одержати:

$$\sum_{i=1}^{n} y_i - n \cdot a_0 - a_1 \sum_{i=1}^{n} x_i = 0\,; \qquad \sum_{i=1}^{n} x_i y_i - a_0 \sum_{i=1}^{n} x_i - a_1 \sum_{i=1}^{n} x_i^2 = 0.$$

Розділивши ліву і праву частини кожного рівняння на кількість точок *n*, одержимо:

$$\frac{1}{n}\sum_{i=1}^{n} y_i - a_0 - a_1 \frac{1}{n}\sum_{i=1}^{n} x_i = 0\,; \qquad \frac{1}{n}\sum_{i=1}^{n} x_i y_i - a_0 \frac{1}{n}\sum_{i=1}^{n} x_i - a_1 \frac{1}{n}\sum_{i=1}^{n} x_i^2 = 0.$$

Доданки в цих виразах дорівнюють:

$$\frac{1}{n}\sum_{i=1}^{n} y_i = m_y = \overline{y}\,; \quad \frac{1}{n}\sum_{i=1}^{n} x_i = m_x = \overline{x}\,; \qquad \frac{1}{n}\sum_{i=1}^{n} x_i y_i = \alpha_{11} = \overline{xy}\,; \quad \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \alpha_{2x} = \overline{x^2}$$

і мають сенс оцінок математичних сподівань (m_x, m_y), змішаного початкового другого моменту (α_{11}) і початкового другого моменту (α_{2x}), тому система рівнянь для пошуку значень коефіцієнтів a_0 і a_1 може бути записана більш компактно:

$$\left.\begin{aligned} m_y - a_0 - a_1 \cdot m_x &= 0, \\ \alpha_{11} - a_0 \cdot m_x - a_1 \cdot \alpha_{2x} &= 0, \end{aligned}\right\} \qquad \to \qquad \left.\begin{aligned} \overline{y} - a_0 - a_1 \cdot \overline{x} &= 0, \\ \overline{xy} - a_0 \cdot \overline{x} - a_1 \cdot \overline{x^2} &= 0. \end{aligned}\right\}.$$

Звідки випливає, що в апроксимуючому поліномі $\Psi_1 = a_0 + a_1 x$ коефіцієнти a_0 і a_1 можна визначити за формулами:

$$a_1 = \frac{\alpha_{11} - m_x \cdot m_y}{\alpha_{2x} - m_x^2} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2}\,; \qquad a_0 = m_y - a_1 \cdot m_x = \overline{y} - a_1 \cdot \overline{x}\,.$$

Аналогічно знаходяться параметри і для апроксимуючих функцій будь-якого іншого вигляду. Проте загального правила для вибору відповідного вигляду емпіричної формули не існує. Водночас, для найбільш поширених семи видів функцій існує процедура формальної оцінки їх придатності.

Для умов, розглянутих у попередньому прикладі (див. табл. 9.16), знайти лінійну апроксимацію залежності змінної *y* від змінної *x* за даними тієї ж таблиці спостережень, що повторюємо (див. табл. 9.18).

Таблиця 9.18

x	1,00	1,50	3,00	4,50	5,00
y	1,25	1,40	1,50	1,75	2,25

- Отримана раніше розрахункова таблиця має вигляд (див. табл. 9.19):

Таблиця 9.19

i	x_i	y_j	x_i^2	y_j^2	$x_i\, y_j$
1	1,00	1,25	1,00	1,56	1,25
2	1,50	1,40	2,25	1,96	2,10
3	3,00	1,50	9,00	2,25	4,50

4	4,50	1,75	20,25	3,06	7,88
5	5,00	2,25	25,00	5,06	11,25
Σ	15,00	8,15	57,50	13,90	26,98
Σ/n	3	1,63	11,5	2,779	5,395

Необхідні оцінки з таблиці були знайдені і мають такі значення:

$$\overline{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = \frac{15}{5} = 3\,;\quad \overline{y} = \frac{1}{5}\sum_{j=1}^{5} y_j = \frac{8{,}15}{5} = 1{,}63\,;\quad \overline{xy} = \frac{1}{5}\sum_{i=1}^{5} x_i y_i = \frac{26{,}975}{5} = 5{,}395;$$

$$\overline{x^2} = \frac{1}{5}\sum_{i=1}^{5} x_i^2 = \frac{57{,}5}{5} = 11{,}5\,;\quad \overline{y^2} = \frac{1}{5}\sum_{i=1}^{5} y_i^2 = \frac{13{,}897}{5} = 2{,}779\,.$$

Отже, для лінійної апроксимації $y = a_0 + a_1 x$ знаходимо:

$$a_1 = \frac{\overline{xy} - \overline{x}\cdot\overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2} = \frac{5{,}395 - 3\cdot 1{,}63}{11{,}5 - (3)^2} = \frac{5{,}395 - 4{,}89}{11{,}5 - 9} = \frac{5{,}395 - 4{,}89}{11{,}5 - 9} = \frac{0{,}505}{2{,}5} = 0{,}202;$$

$$a_0 = \overline{y} - a_1 \cdot \overline{x} = 1{,}63 - 0{,}202\cdot 3 = 1{,}63 - 0{,}606 = 1{,}024\,.$$

Результуючий вираз має вигляд: $y = 1{,}024 + 0{,}202\cdot x$ і цілком збігається з отриманим раніше рівнянням регресії Y на X: $y = 0{,}202\cdot x + 1{,}024$.

Графічне зображення цих ліній наведене на рис. 9.8.

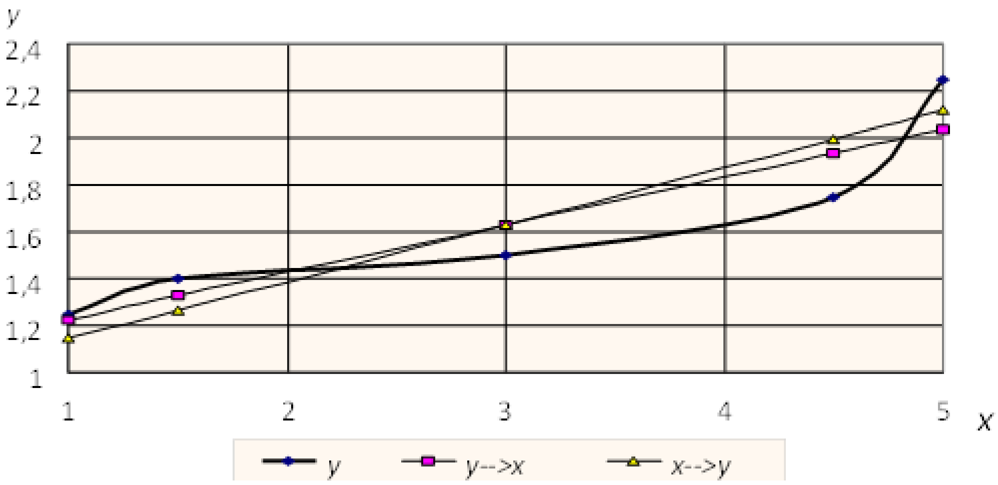


Рис 9.8. Лінії регресії

Відзначимо, що лінія регресії $y \to x$ у даному випадку має менший кут нахилу.

Таблиця значення функції $\varphi = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3652	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920

0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

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

X	Ф(x)	X	Ф(x)	X	Ф(x)	X	Ф(x)
0,00	0,0000	**0,24**	0,0948	**0,48**	0,1844	**0,72**	0,2642
0,01	0,0040	**0,25**	0,0987	**0,49**	0,1879	**0,73**	0,2673
0,02	0,0080	**0,26**	0,1026	**0,50**	0,1915	**0,74**	0,2703
0,03	0,0120	**0,27**	0,1064	**0,51**	0,1950	**0,75**	0,2734
0,04	0,0160	**0,28**	0,1103	**0,52**	0,1985	**0,76**	0,2764
0,05	0,0199	**0,29**	0,1141	**0,53**	0,2019	**0,77**	0,2794

0,06	0,0239	0,30	0,1179	0,54	0,2054	0,78	0,2823
0,07	0,0279	0,31	0,1217	0,55	0,2088	0,79	0,2852
0,08	0,0319	0,32	0,1255	0,56	0,2123	0,80	0,2881
0,09	0,0359	0,33	0,1293	0,57	0,2157	0,81	0,2910
0,10	0,0398	0,34	0,1331	0,58	0,2190	0,82	0,2939
0,11	0,0438	0,35	0,1368	0,59	0,2224	0,83	0,2967
0,12	0,0478	0,36	0,1406	0,60	0,2257	0,84	0,2995
0,13	0,0517	0,37	0,1443	0,61	0,2291	0,85	0,3023
0,14	0,0557	0,38	0,1480	0,62	0,2324	0,86	0,3051
0,15	0,0596	0,39	0,1517	0,63	0,2357	0,87	0,3078
0,16	0,0636	0,40	0,1554	0,64	0,2389	0,88	0,3106
0,17	0,0675	0,41	0,1591	0,65	0,2422	0,89	0,3133
0,18	0,0714	0,42	0,1628	0,66	0,2454	0,90	0,3159
0,19	0,0753	0,43	0,1664	0,67	0,2486	0,91	0,3186
0,20	0,0793	0,44	0,1700	0,68	0,2517	0,92	0,3212
0,21	0,0832	0,45	0,1736	0,69	0,2549	0,93	0,3238
0,22	0,0871	0,46	0,1772	0,70	0,2580	0,94	0,3264
0,23	0,0910	0,47	0,1808	0,71	0,2611	0,95	0,3289
0,96	0,3315	1,37	0,4147	1,78	0,4625	2,36	0,4909
0,97	0,3340	1,38	0,4162	1,79	0,4633	2,28	0,4913
0,98	0,3365	1,39	0,4177	1,80	0,4641	2,40	0,4918
0,99	0,3389	1,40	0,4192	1,81	0,4649	2,42	0,4922
1,00	0,3413	1,41	0,4207	1,82	0,4656	2,44	0,4927
1,01	0,3438	1,42	0,4222	1,83	0,4664	2,46	0,4931
1,02	0,3461	1,43	0,4236	1,84	0,4671	2,48	0,4934
1,03	0,3485	1,44	0,4251	1,85	0,4678	2,50	0,4938
1,04	0,3508	1,45	0,4265	1,86	0,4686	2,52	0,4941
1,05	0,3531	1,46	0,4279	1,87	0,4693	2,54	0,4945
1,06	0,3554	1,47	0,4292	1,88	0,4699	2,56	0,4948
1,07	0,3577	1,48	0,4306	1,89	0,4706	2,58	0,4951
1,08	0,3599	1,49	0,4319	1,90	0,4713	2,60	0,4953
1,09	0,3621	1,50	0,4332	1,91	0,4719	2,62	0,4956
1,10	0,3643	1,51	0,4345	1,92	0,4726	2,64	0,4959
1,11	0,3665	1,52	0,4357	1,93	0,4732	2,66	0,4961
1,12	0,3686	1,53	0,4370	1,94	0,4738	2,68	0,4963
1,13	0,3708	1,54	0,4382	1,95	0,4744	2,70	0,4965
1,14	0,3729	1,55	0,4394	1,96	0,4750	2,72	0,4967
1,15	0,3749	1,56	0,4406	1,97	0,4756	2,74	0,4969
1,16	0,3770	1,57	0,4418	1,98	0,4761	2,76	0,4971
1,17	0,3790	1,58	0,4429	1,99	0,4767	2,78	0,4973
1,18	0,3810	1,59	0,4441	2,00	0,4772	2,80	0,4974
1,19	0,3830	1,60	0,4452	2,02	0,4783	2,82	0,4976

1,20	0,3849	**1,61**	0,4463	**2,04**	0,4793	**2,84**	0,4977
1,21	0,3869	**1,62**	0,4474	**2,06**	0,4803	**2,86**	0,4979
1,22	0,3883	**1,63**	0,4484	**2,08**	0,4812	**2,88**	0,4980
1,23	0,3907	**1,64**	0,4495	**2,10**	0,4821	**2,90**	0,4981
1,24	0,3925	**1,65**	0,4505	**2,12**	0,4830	**2,92**	0,4982
1,25	0,3944	**1,66**	0,4515	**2,14**	0,4838	**2,94**	0,4984
1,26	0,3962	**1,67**	0,4525	**2,16**	0,4846	**2,96**	0,4985
1,27	0,3980	**1,68**	0,4535	**2,18**	0,4854	**2,98**	0,4986
1,28	0,3997	**1,69**	0,4545	**2,20**	0,4861	**3,00**	0,4986
1,29	0,4015	**1,70**	0,4554	**2,22**	0,4868	**3,20**	5
1,30	0,4032	**1,71**	0,4564	**2,24**	0,4875	**3,40**	0,49931
1,31	0,4049	**1,72**	0,4573	**2,26**	0,4881	**3,60**	0,49966
1,32	0,4066	**1,73**	0,4582	**2,28**	0,4887	**3,80**	0,49984
1,33	0,4082	**1,74**	0,4591	**2,30**	0,4893	**4,00**	0,49992
1,34	0,4099	**1,75**	0,4599	**2,32**	0,4898	**4,50**	0,49996
1,35	0,4115	**1,76**	0,4608	**2,34**	0,4904	**5,00**	0,49999
1,36	0,4131	**1,77**	0,4616				

Таблиця значень $t_v = t(\gamma, n)$

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

Таблиця значень $q = q(\gamma, n)$

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

Критичні точки розподілу χ^2

Число степенів свободи *k*	**Рівень значимості α**					
	0,01	**0,025**	**0,05**	**0,95**	**0,975**	**0,89**

1	6,6	5,0	3,8	0,0039	0,00098	0,0001
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

Критичні точки розподілу Стьюдента

Число степенів свободи *k*	**Рівень значимості α (двостороння 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	4,60	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,40
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,99
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,96
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,51	3,79
23	1,71	2,07	2,50	2,81	3,49	3,77
24	1,71	2,06	2,49	2,80	3,47	3,74
25	1,71	2,06	2,49	2,79	3,45	3,72
26	1,71	2,06	2,48	2,78	3,44	3,71
27	1,71	2,05	2,47	2,77	3,42	3,69
28	1,70	2,05	2,46	2,76	3,40	,3,66
29	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,29
	0,05	**0,025**	**0,01**	**0,005**	**0,001**	**0,0005**
	Рівень значимості α (одностороння критична область)					

Приклади

Приклад № 1

Вибірка обсягом 30 осіб, розбита на дві рівні групи за ознакою статі, пройшла функціональну діагностику мозкової активності, в результаті якої у 13 жінок і 4 чоловіків було виявлено домінування правої півкулі, а у 2 жінок і 11 чоловіків - домінування лівої півкулі. Перевірте гіпотезу про зв'язокстаті з функціональною асиметрією головного мозку.

Рішення:
Оскільки в обох вибірках n_1 і $n_{2>}$ 11 і діапазони розкиду значень у двох вибірках не збігаються між собою, ми можемо скористатися самим простим критерієм для зіставлення двох вибірок - критерієм Q Розенбаума. Обсяги вибірок різняться менш ніж на 10 чоловік, так, що обмеження про примірному рівність вибірок також не перешкоджає нам.
Таблиця 1. Показники вираженості функціональної асиметрії у чоловіків і жінок

	Група 1 - чоловіки (N = 15 осіб)	Група 2 - жінки (n = 15 осіб)
Домінування правового півкулі	4	13
Домінування лівого півкулі	11	2

Дані в таблиці 1 розташовані за ступенем домінування тієї чи іншої півкулі в чоловічій або жіночої вибірці. Першим вищим є ряд значень в жіночій вибірці.
Середня величина в чоловічій і жіночій вибірці ідентична і дорівнює 7,5.
Сформулюємо гіпотези.
Формулювання гіпотез систематизує припущення дослідника й представляє їх у чіткому й лаконічному вигляді [5; с. 24]. Статистичні гіпотези підрозділяються на нульові і альтернативні.
Нульова гіпотеза - це гіпотеза про відсутність розходжень. Вона позначається як H_0 і називається нульовою тому, що містить число 0:
$X_{1\text{-}X\,2} = 0$, де $X_{1,}$ X_2 - зіставлення значення ознак. Таким чином, нульова гіпотеза - це те, що ми хочемо спростувати, якщо перед нами стоїть завдання довести значимість розходжень.
Альтернативна гіпотеза - це гіпотеза про значимість розходжень. Вона позначається як $H_{1.}$
Альтернативна гіпотеза - це те, що ми хочемо довести, тому іноді її називають експериментальної гіпотезою.
Сформулюємо основні гіпотези:
$H_{0:}$ Функціональна асиметрія головного мозку у чоловіків не виражена більшою мірою, ніж у жінок.
$H_{1:}$ Функціональна асиметрія головного мозку у чоловіків виражена більшою мірою, ніж у жінок.
Зіставимо ряди значень для визначення S_1 і $S_{2.}$
max 2 = 13
$S_1 = 0$
min 1 = 4
$S_2 = 1$
Виробляємо підрахунок емпіричного значення $Q_{ЕМП} = S_1 + S_2 = 0 +1 = 1$
По таблиці визначаємо критичне значення Q для даних n_1 і $n_{2.}$ Якщо $Q_{ЕМП}$ одно $Q_{0,05}$ або перевищує його, H_0 відкидається.
У даному випадку $Q_{кр} = 6$
6 ($p \leq 0{,}01$)
$Q_{ЕМП} < Q_{кр}$
Отже приймається гіпотеза H_0 і відкидається гіпотеза $H_{1.}$
Функціональна асиметрія головного мозку у чоловіків не виражена більшою мірою, ніж у жінок, отже, функціональна асиметрія головного мозку не залежить від ознаки статі.

Приклад 3
Психолог припустив, що в результаті навчання час вирішення еквівалентних завдань "гри в 5" (тобто мають один і той же алгоритм рішення) буде значимо зменшуватися. Для перевірки гіпотези у восьми досліджуваних порівнювався час рішення (у хвилинах) першої і третьої задач. Встановіть, чи правильно припущення дослідника?

№ випробуваного	1	2	3	4	5	6	7	8
1 завдання	4	3,5	4,1	5,5	4,6	6	5,1	4,3
2 завдання	3	3	3,8	2,1	4,9	5,3	3,1	2,7

Щоб встановити чи вірно припущення дослідника про скорочення часу при вирішенні еквівалентних (тобто мають один і той же алгоритм рішення) завдань застосуємо Т - критерій Вілкоксона.

Таблиця № 1

№ випробуваного	Час виконання завдання № 1 $f_{до}$	Час виконання завдання № 2 $f_{після}$	Різниця ($F_{після} - f_{до}$)	Абсолютне значення різниці	Ранговий номер різниці
1	4	3	-1	1	5
2	3,5	3	0,5	0,5	3
3	4,1	3,8	0,3	0,3	1,5
4	5,5	2,1	3,4	3,4	8
5	4,6	4,9	**0,3**	**0,3**	**1,5**
6	6	5,3	0,7	0,7	4
7	5,1	3,1	2	2	7
8	4,3	2,7	1,6	1,6	6
Сума					36

Сформулюємо гіпотезу

$H_{0:}$ Інтенсивність зрушень у бік зменшення тривалості виконання еквівалентних завдань значно перевищує інтенсивність зрушень у бік збільшення часу рішення.

С умма рангів дорівнює 36, що відповідає розрахунковій:

$\Sigma R = N (N +1) / 2 = 72 / 2 = 36$

Тепер відзначимо ті зрушення, які є нетиповими, в даному випадку - позитивними. У табл. № 1 ці зрушення і відповідні їм ранги виділені кольором.

Сума рангів цих "рідкісних" зрушень і становить емпіричне значення критерію Т:

де R_r - рангові значення зрушень з більш рідкісним знаком.

Т = 1,5

З таблиці VI додатка 1 визначаємо критичні значення Т для $n = 8$

Т кр {5 (р <0,05) | 1 (р <0,01)

Отримуємо, що $T_{ЕМП} < T_{кр}$ (0,05)

Відповідь: H_0 підтверджується *(p* <0,05). на 5% рівні.

Приклад № 4

У двох школах району психолог з'ясовував думки вчителів про організацію психологічної служби в школі. У першій школі було опитано 20 вчителів, у другій 15. Психолога цікавило питання: в якій школі психологічна служба поставлена краще? Вчителі давали відповіді по номинативной шкалою - подобається (та), не подобається - (немає). Результати опитування представлені у вигляді четирехпольние таблиці:

	1 школа	2 школа
Число вчителів відповіли на питання ствердно	15	7
Число вчителів, які відповіли на запитання негативно	5	8

Для з'ясування питання про кращу організацію психологічної служби в обох школах за результатами опитування вчителів доцільно отримані дані перевести у відсотки, таким чином, ми отримаємо відсоткове співвідношення відповідей "Так" "Ні". І так, в першій школі з 100% вчителів задоволеними психологічною службою виявилися - 75%, незадоволеними - 25%. У другій школі відсоток позитивних відповідей склав 47% від числа всіх опитаних, негативних - 53%

Застосуємо Критерій φ* - кутове перетворення Фішера.

Групи	Стверджуючі відповіді	Негативні відповіді	Суми
Перший школа	15 (75%)	5 (25%)	20
Друга-школа	7 (46,6%)	8 (53,3%)	15
Суми	22	13	35

$H_{0:}$ Частка вчителів задоволених психологічною службою у першій школі не більше, ніж у другій.
$H_{1:}$ Частка вчителів задоволених психологічною службою у першій школі більше ніж у другій.
За табл. XII визначимо показники φ:
$\varphi_{1\,(75\%)} = 2{,}094$
$\varphi_{2\,(46{,}6\%)} = 1{,}503$
Тепер підрахуємо емпіричне значення φ * за формулою:
З умови задачі $n_1 = 20$; $n_2 = 15$
$\varphi^*_{\text{ЕМП}} = 0{,}591 \text{ X } 2{,}93 = 1{,}73$
За Табл. XIII Додатка 1 визначаємо, якому рівню значущості відповідає $\varphi^*_{\text{ЕМП}} = 1{,}73$:
Р = 0,04
φ^* кр = {1, 64 (ρ ≤ 0,05) | 2,31 (ρ ≤ 0,01)
$\varphi^*_{\text{ЕМП}>}$ φ^* кр (ρ ≤ 0,05) |
Відповідь: H_0 відкидається. Частка вчителів задоволених психологічною службою у першій школі більше ніж у другій, тобто в першій школі психологічна служба поставлена краще.

Приклад № 5

При вимірі просторових порогів відчуття дотику отримані такі величини порогів для жінок і чоловіків (у мм):
Жінки - 32, 30, 28, 30, 33, 37, 28, 27.
Чоловіки - 39, 36, 31, 35, 29, 34, 38.
Дослідника цікавить питання: чи відрізняються між собою за величиною пороги жінок і чоловіків?
Для вирішення даної задачі застосовуємо критерій Манна - Уітні

Жінки n_1		Чоловіки n_2	
Показник просторового порогу тактильної чутливості	Ранг	Показник просторового порогу тактильної чутливості	Ранг
37 33 32 30 30 28 28 27	13 9 8 5,5 5,5 2,5 2,5 1	39 38 36 35 34 31 29	15 14 12 11 10 7 4
Суми 245	47	242	73
Середні 30,6		34,5	

Загальна сума рангів: 47 +73 = 120. Розрахункова сума:

$\Sigma R = N (N+1) / 2 = 15X 16 / 2 = 120$
Рівність реальною і розрахункової сум дотримано.
Ми бачимо, що за рівнем просторового порогу тактильної чутливості більш високим поруч опиняється група чоловіків. Саме на цю вибірку доводиться вищий сумарний ранг: 73
$U_{ЕМП} = 56 + 28 - 73 = 11$
Для другої рангової суми (47) величина $U_{ЕМП} = 47$
Вибираємо меншу величину U: $U_{ЕМП} = 11$
За Табл. II Додатка 1 визначаємо критичні значення для $n_1 = 8$, $n_2 = 7$.
$U_{до\ р} = \{13\ (p < 0,05) \mid 7\ (p < 0,01)$
$U_{ЕМП} < U_{до\ р,}$ тобто ми можемо констатувати достовірні відмінності.
Нульова гіпотеза про те, що пороги чутливості у чоловіків відрізняються від жінок за величиною підтверджується на 5% рівні значимості.

Приклад № 6
Припустимо, що на 8 піддослідних проведено тест коректурної проби у звичайних умовах (А) і в умовах емоційної напруженості (В). Фіксувалося кількість помилок.

випробовувані	1	2	3	4	5	6	7	8
Умова А	3	5	6	8	10	12	13	14
Умова У	10	10	12	5	8	11	20	23

Для вирішення даного завдання, на мій погляд, доцільно використовувати Т - критерій Вілкоксона.
Таблиця № 1

Код випробуваного	Корректурная проба Звичайні умови (А) Умови емоційної напруженості (В) Різниця ($F_В - f_А$)	Абсолютне значення різниці	Ранговий номер різниці	
1 2 3 4 5 6 7 8	3 5 6 8 10 12 13 14 10 10 12 5 8 11 20 23 +7 +5 +6 **3** **2**	7 5 6 **3** **2** **1** 7 9	6,5 4 5 **3** **2** **1** 6,5 8	

	1 +7 +9			
Сума				

36

Сума рангів дорівнює 36, що відповідає розрахунковій:

$\Sigma R = N (N +1) / 2 = 72 / 2 = 36$

Відповідно висуваємо гіпотезу. $H_{0:}$ Кількість помилок при коректурної пробі в умовах емоційної напруженості перевершує кількість помилок у звичайних умовах.

Відзначимо нетипові зрушення (у даному випадку негативні). У таблиці № 1 вони виділені жирним шрифтом.

Сума рангів цих "рідкісних" зрушень і становить емпіричне значення критерію Т:

$T = \Sigma_{R r}$

де $_{R r}$ - рангові значення зрушень з більш рідкісним знаком.

$T_{ЕМП} = 3 +2 +1 = 6$

За Таблиці VI Додатка 1 визначаємо критичні значення Т для $n = 8$

$T_{кр} = \{5 (p <0,05) \mid 1 (p <0,01)$

$T_{ЕМП>} T_{кр}$

Рівень достовірності гіпотези не досягає 5% значущості.

Відповідь: H_0 не підтверджується, так як емпіричне значення Т вище критичного значення Т і потрапляє в зону незначущості.

Отже робимо висновок про відсутність значущих статистичних відмінностей між результатами тестування у різних умовах.

Приклад № 7

В експерименті з дослідження інтермодального перенесення отримано, що в одній групі піддослідних (14 осіб) більш ефективним виявилося тактильне ознайомлення з наступним зоровим впізнаванням (8 чоловік з 14), тоді як у другій групі (10 осіб) тільки для трьох досліджуваних цей вид перенесення образу був ефективнішим, ніж перенесення в напрямку зір-дотик.

Значимі відмінності цих двох груп випробовуваних в частині ефективності перенесення дотик-зір?

Для вирішення підходить кутовий φ - критерій Фішера

Групи	Перенесення Дотик-зір	Перенесення Зір-дотик	Всього
№ 1	8 (57,14%)	6 (42,85%)	14
№ 2	3 (30%)	7 (70%)	10
Всього	11	13	

Сформулюємо гіпотези

$H_{0:}$ частка випробовуваних в першій групі, в частині ефективності перенесення дотик-зір, не перевершує частку таких же випробовуваних у другій групі.

$H_{1:}$ частка випробовуваних в першій групі в частині ефективності перенесення дотик-зір переважає частку таких же випробовуваних у другій групі.

Визначаємо величини φ_1 і φ_2 по Таблиці XII додатка 1 (Сидоренко Е. В.) Нагадаємо, що φ_1 - це завжди кут, відповідний більшою процентною часткою.

φ 1 (57,14%)) = 1,713

φ 2 (30%) = 1,159.

φ е. $_{мп.}$ = (Ф1 - φ2) √ n 1 * n 2/n1 + n2 = (1,713 - 1,159) √ 14 * 10/14 +10 = 1,339;

За Табл. XIII Додатка 1 визначаємо, якому рівню значущості відповідає ця величина: p = 0,092

Для практики цей рівень малий, тому слід порівняти

$\varphi_{ЕМП.}$ з $\varphi_{кр}$.= $\{1,64 (p \le 0,05) \mid 2,31 (p \le 0,01)$

Отримане значення φ знаходиться поза зоною значущості, H_0 приймається.

Відповідь: відмінності двох груп випробовуваних в частині ефективності перенесення дотик-зір незначущі.

Приклад № 8

Проведіть порівняння (кореляційний аналіз) паралельних форм деякого опитувальника за результатами, поданими в таблиці:

№ питання	1	2	3	4	5	6	7	8	9	10	11	12
Форма "А"	0	1	0	0	1	1	0	1	0	1	1	1
Форма "А"	0	1	1	0	0	0	1	1	0	0	0	0

Для вирішення даної задачі застосуємо коефіцієнт кореляції φ

№ питання	Форма "А" (X) 0 - розбіжність із "ключем", 1 - збіг з "ключем".	Форма "В" (Y) 0 - розбіжність із "ключем", 1 - збіг з "ключем".
1	0	0
2	1	1
3	0	1
4	0	0
5	1	0
6	1	0
7	0	1
8	1	1
9	0	0
10	1	0
11	1	0
12	1	0

$Ф_{ЕМП}$ = pxy - px py / √ px (1 - px) py (1 - py)
де px - частота або частка ознаки, що має 1 по X,
(1 - px) - частка або частота ознаки, що має 0 по X;
py - частота або частка ознаки, що має 1 по Y,
(1 - py) - частка або частота ознаки, що має 0 по Y,
pxy - частка або частота ознаки, що має 1 одночасно як по X, так і по Y.
$Ф_{ЕМП}$ = - 0,12
Для даного коефіцієнта відсутні таблиці значущості. Значимість розраховується за формулою:
$T_{ф}$ = | - 0,12 | √ 12-2 / 1-0,12 ^{x} - 0,12 = 0,382
Число ступенів свободи для даної вибірки k = n -2 = 12-2 = 10
За табл.16 додатка 1 (Єрмолаєв О. Ю. Математична статистика для психологів) знаходимо критичні значення критерію Стьюдента.
$t_{кр}$ = {2,23 (p <0,05) | 4,59 (p <0,01):
Значення величини $T_{ф}$ потрапило в зону незначущості.
Відповідь: зв'язок між паралельними формами деякого опитувальника не виявлена.

Приклад № 9

За допомогою двох опитувальників (X і Y), що вимагають альтернативних відповідей "так" чи "ні", були отримані первинні результати - відповіді 15 піддослідних. Результати представлена у вигляді сум балів за позитивні відповіді ("так") для кожного випробуваного окремо для опитувальника X і опитування Y. Потрібно визначити, вимірюють чи опитувальники X і Y схожі особистісні якості піддослідних, або не вимірюють. Можна припустити, що якщо опитувальники за змістом і формулювань мало відрізняються один від одного, то сума балів, набрана кожним випробуваним за опитувальником X, буде близька до суми балів, набраних за опитувальником Y. Результати експерименту представлені в таблиці:

№ випробуваного	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	47	71	52	48	35	35	41	82	72	56	59	73	60	55	41
Y	75	79	85	50	49	59	75	91	102	87	70	92	54	75	68

Дану задачу можливо вирішити із застосуванням коефіцієнта кореляції Пірсона.

№ випробуваного	Опитувальник X	Опитувальник Y	XY	XX	YY
1	47	75	3525	2209	5625
2	71	79	5609	5041	6241
3	52	85	4420	2704	7225
4	48	50	2400	2304	2500
5	35	49	1715	1225	2401
6	35	59	2065	1225	3481
7	41	75	3075	1681	5625
8	82	91	7462	6724	8281
9	72	102	7344	5184	10404
10	56	87	4872	3136	7569
11	59	70	4130	3481	4900
12	73	92	6716	5329	8464
13	60	54	3240	3600	2916
14	55	75	4125	3025	5625
15	41	68	2788	1681	4624
Сум а	827	1111	63486	48549	85881

Для розрахунку застосовується формула:
$r_{xy} = (\Sigma_{xy} - (\Sigma_x \Sigma_y / n)) / \sqrt{S_x S_y}$, $r_{xy\ ЕМП} = 0,68$
По таблиці знаходимо критичні значення для обчисленого коефіцієнта Пірсона $r_{xy\ ЕМП}$ з урахуванням числа ступенів свободи розраховуються за формулою $k = n - 2$. У нашому випадку $k = 15 - 2 = 13$
$r_{кр} = \{0,51\ (p < 0,05)\ |\ 0,64\ (p < 0,01)$

Величина розрахованого коефіцієнта потрапляє в зону значущості. Гіпотеза про наявність зв'язку підтверджується. Іншими словами зв'язок між результатами двох опитувальників підтверджується на рівні 1% і є позитивною.
Відповідь: опитувальники X і Y вимірюють схожі особистісні якості піддослідних.

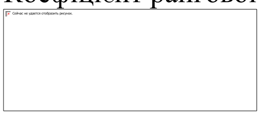
Приклад № 10

Проведіть обробку результатів експерименту: дослідження самооцінки особистості. Мета обробки результатів - визначення зв'язку між ранговими оцінками якостей особистості, що входять до вистави "Я" - ідеальне і "Я" реальне. Міра зв'язку встановлюється за допомогою коефіцієнта рангової кореляції Ч. Спірмена. (Експериментальна робота проводиться під час аудиторного заняття.)
Методика. Дослідження самооцінки особистості
Мета дослідження: визначити рівень самооцінки.
Матеріал й устаткування: список слів або спеціальний бланк зі словами, що характеризують окремі якості особистості, ручка.
Це дослідження має два суттєво різняться за процедурою варіанти визначення самооцінки особистості. В обох випадках можна працювати з одним випробуваним, і з групою. Ми розглянемо перший варіант дослідження. (Див. Психологічні дослідження. Практикум з загальної психології для студентів педагогічних вузів. Укладачі Т. І. Пашукова и др. - М., Видавництво "Інститут практичної психології", Воронеж: НВО "МОДЕК", 1996.)
Перший варіант дослідження
В основі дослідження самооцінки в цьому варіанті методики лежить спосіб ранжирування. Процедура дослідження включає дві серії. Матеріалом, з яким працюють випробовувані, є надрукований на спеціальному бланку список слів, що характеризують окремі якості особистості. Кожен випробовуваний отримує такий бланк на початку дослідження. При роботі з групою піддослідних важливо забезпечити сувору самостійність ранжирування.
Перша серія
Завдання першої серії: виявити уявлення людини про якості свого ідеалу, тобто "Я" ідеальне. Для цього слова, надруковані на бланку, випробовуваний повинен розташувати в порядку пріоритету.
Інструкція випробуваному: "Прочитайте уважно всі слова, що характеризують якості особистості. Розгляньте ці якості з точки зору прісущность їх ідеальної особистості, тобто з точки зору корисності, соціальної значимості і бажаності. Для цього проранжуйте їх, оцінивши кожне в балах від 20 до 1. Оцінку 20 поставте в бланку, в колонці № 1 ліворуч від тієї якості, яке, на Вашу думку, є самим корисним і бажаним для людей.
Оцінку 1 - у тій же колонці № 1 ліворуч від якості, яке найменш корисно, значимо і бажано. Всі інші оцінки від 19 до 2 розташуйте відповідно до Вашім ставленням до всіх інших якостей. Слідкуйте, щоб ні одна оцінка не повторювалася двічі ".
Друга серія
Завдання другої серії: виявити уявлення людини про свої власні якості, то є його "Я" реально. Як і в першій серії випробуваного просять проранжувати надруковані на бланку слова, але вже з точки зору характерності або прісущность охоплюють ними качествлічності собі самому.
Інструкція випробуваному. "Прочитайте знову всі слова, що характеризують якості особистості. Розгляньте ці якості з точки зору прісущность їх Вам. Розставте їх у колонці № 2, оцінивши кожне від 20 до I. Оцінку 20 - поставте праворуч від тієї якості, яке, на Вашу думку, є притаманним Вам найбільшою мірою, оцінку 19 - поставте тому якості, яке характерно для Вас дещо менше, ніж перше, і так далі. Тоді оцінкою 1 у Вас буде позначено то якість, яка властива Вам менше, ніж всі інші. Стежте, що б оцінки - ранги не повторювалися двічі ".
Бланк зі словами, що характеризують якості особистості виглядає наступним чином.

№ 1	Якості особистості	№ 2	d	2 d
19	Поступливість	15	4	16
13	Сміливість	10	3	9

4	Запальність	4	0	0
6	Нервозність	3	3	9
18	Терплячість	16	2	4
12	Увлекаемость	7	5	25
7	Пасивність	13	-6	36
3	Холодність	2	1	1
14	Ентузіазм	8	6	36
11	Обережність	17	-6	36
1	Вередливість	1	0	0
10	Повільність	20	-10	100
20	Гуманність	18	2	4
9	Сором'язливість	14	-5	25
2	Боягузтво	5	-3	9
17	Працьовитість	9	8	64
8	Підозрілість	11	-3	9
16	Педантичність	12	4	16
5	Легковір'я	6	-1	1
15	Акуратність	19	-4	16
210		210	0	Σ d2 = 326

Коефіцієнт рангової кореляції Спірмена підраховується за формулою



де *d* - різниця між рангами за двома змінним для кожного досліджуваного;
N - кількість ранжируваних значень,
rs = 1 - 0,245 = 0,755
За Табл. XVI Додатки 1. Критичні значення вибіркового коефіцієнта кореляції рангів визначаємо критичні значення:
rs кр = {0,45 (p <0,05) | 0,57 (p <0,01) r $_{s\ ЕМП}$> rs кр,
значить досягається статистична значимість.
Отриманий коефіцієнт кореляції свідчить про наявності значимої позитивного зв'язку між Я ідеальним і Я реальним на 1% рівні. Це можна трактувати як прояв (при p від +0,39 до +0,89), тенденції до завищення самооцінки.

Порівняльна характеристика статистичних критеріїв

Параметричні	Непараметричні
1. Дають змогу прямо оцінити різниці середніх, отриманих у двох вибірках (t-критерій Стьюдента)	1. Дають змогу оцінити лише середні тенденції, відповісти на питання, чи частіше у вибірці А трапляються вищі, а у вибірці Б – нижчі значення ознаки (критерії *Q*, *U*, φ* та ін.)
2. Дають змогу прямо оцінити відмінності в дисперсіях (F-критерій Фішера)	2. Уможливлюють оцінювання лише відмінності в діапазонах варіативності ознаки (критерій φ*)
3. Уможливлюють виявлення тенденції зміни ознаки при переході від умови до умови при нормальному розподілі ознаки (однофакторний дисперсійний аналіз)	3. Уможливлюють виявлення тенденції зміни ознаки при переході від умови до умови при будь-якому розподілі ознаки (критерії тенденцій L і S)
4. Уможливлюють оцінювання взаємодії двох і більше факторів у їх впливі на зміни ознаки (двофакторний дисперсійний аналіз)	4. Ця можливість відсутня
5. Експериментальні дані повинні відповідати умовам: а) значення ознаки виміряні за інтервальною шкалою; б) розподіл ознаки є нормальним; в) у дисперсійному аналізі слід дотримуватися вимоги рівності дисперсій у комірках комплексу	5. Експериментальні дані можуть не відповідати жодній із цих умов: а) значення ознаки можуть бути пред'явлені у будь-якій шкалі; б) розподіл ознаки може бути будь-яким, збіг його з теоретичним законом розподілу не потребує перевірки; в) вимога рівності дисперсій відсутня
6. Математичні розрахунки дуже складні	6. Математичні розрахунки здебільшого прості і займають мало часу
7. Якщо умови, перераховані в п. 5, задовольняються, параметричні критерії виявляються дещо потужнішими, ніж непараметричні	7. Якщо умови, перераховані в п. 5, не задовольняються, непараметричні критерії виявляються потужнішими, ніж параметричні, оскільки вони менш чутливі до «засмічень»

Таблиця 5.2.

Класифікація задач і методів їх розв'язання з використанням параметричних статистичних критеріїв

Задача	Умови	Методи (критерії)
1. Оцінка відповідності емпіричного розподілу нормальному закону	зіставлення емпіричною розподілу з теоретичним	Критерії асиметрії і ексцесу, χ^2 – критерій Пірсона
2. Пряма оцінка рівня середньої величини	а) дисперсія відома:	z-критерій
	б) дисперсія невідома.	t-критерій Стьюдента
3. Пряма оцінка рівня дисперсії		χ^2 – критерій Пірсона
4. Оцінка відмінностей у рівні середніх значень ознаки	а) 2 незв'язані вибірки однакової чисельності;	t-критерій Стьюдента
	б)2 незв'язані вибірки різної чисельності;	
	в) 2 зв'язані вибірки:	
	г) 3 і більше вибірок.	Однофакторний дисперсійний аналіз Фішера
5. Оцінка істотності різниць дисперсій ознаки	а) 2 незв'язані вибірки однакової чисельності;	F-критерій Фішера
	б) 2 зв'язані вибірки;	t -критерій Стьюдента
	в) 3 і більше вибірок однакової чисельності;	q-критерій Кохрана
	г) 3 і більше вибірок різної чисельності.	М-критерій Бартлета
6. Аналіз змін ознаки під впливом контрольованих умов	а) під впливом одного фактора;	Однофакторний дисперсійний аналіз Фішера
	б) під впливом двох факторів.	Двохфакторн ий дисперсійний аналіз Фішера

Таблиця 5.3.

Класифікація задач і методів їх розв'язання з використанням непараметричних статистичних критеріїв

Задача	Умови	Методи (критерії)
1. Виявлення відмінностей у рівні досліджуваної ознаки	а) 2 вибірки;	Q-критерій Розенбаума, U-критерій Манна-Уітні, φ-критерій кутового перетворення Фішера
	б) 3 і більше вибірок.	S-критерій Джонкіра, H-критерій Крускала-Уолліса
2. Оцінка зсуву значень досліджуваної ознаки	а) 2 виміри на одній і тій же вибірці;	Т-критерій Вілкоксона, G-критерій знаків, φ-критерій кутового перетворення Фішера
	б) 3 і більше вимірів на одній і тій же вибірці	χ^2-критерій Фрідмана, L-критерій Пейджа
3. Виявлення відмінностей у розподілі досліджуваної ознаки	а) при зіставленні емпіричного розподілу з теоретичним;	χ^2-критерій Пірсона, λ-критерій Колмогорова-Смірнова, m-біноміальний критерій
	б) при зіставленні двох емпіричних розподілів.	χ^2-критерій Пірсона, λ -критерій Колмогорова-Смірнова, φ-критерій кутового перетворення Фішера
4. Аналіз змін ознаки під впливом контрольованих умов	а) під впливом одного фактора;	S-критерій Джонкіра, L-критерій Пейджа, Однофакторний дисперсійний аналіз Фішера
	б) під впливом двох факторів одночасно.	Двохфакторний дисперсійний аналіз Фішера

КОНТРОЛЬНІ ЗАВДАННЯ ДЛЯ ЗДОБУВАЧІВ ВИЩОЇ ОСВІТИ ЗАОЧНОЇ ФОРМИ НАВЧАННЯ

Варіант 1.

1. Порівняйте основні типи вимірювальних шкал у психології.

2. Особливості використання та алгоритм розрахунку критерію t-Стюдента для однієї вибірки.

3. Практичне завдання. Розв'яжіть за допомогою критерію Розенбаума нижченаведену задачу.

Задача. Результати тестування показника відповідальності (фактор Кеттела «у стінах») для підлітків загальноосвітніх шкіл після літніх канікул були такими:

1) 15 учнів, що брали участь у пошуковій діяльності (спеціально організованих загонах):

{8, 7, 8, 9, 7, 9, 6, 8, 9, 8, 8, 9, 7, 6, 8};

2) 14 учнів, що проводили літній відпочинок у звичайний для них спосіб:

{7, 5, 8, 6, 5, 7, 6, 5, 8, 7, 5, 5, 6, 5}.

Чи можна стверджувати, що рівень показника відповідальності вищий у тих учнів, які брали участь у пошуковій діяльності?

4. Умови завдання:

12 учнів були проранжовано одним експертом з їх відкритої неприязні до викладача (х) у і до інших учням (у). Ранг 1 присвоювався учню, який показав себе найбільш ворожим. Результати експертної оцінки наведені в таблиці.

Ранговая оцінка неприязні	Учні											
	1	2	3	4	5	6	7	8	9	10	11	12
До вчителя (x)	2	8	12	3	1	6	7	10	4	9	11	5
До учня (y)	6	5	10	7	3	4	9	8	1	11	12	2

Завдання:

Визначити, чи є зв'язок між відкритою неприязню учнів до викладача і до інших учнів, якщо вірити експерту.

Варіант 2.

1. Дайте порівняльну характеристику номінативним і порядковим шкалам вимірювань, наведіть приклади.

2. Особливості використання та алгоритм розрахунку критерію φ-Фішера.

3. Практичне завдання. Розв'яжіть за допомогою критерію Манна-Уітні нижченаведену задачу.

Задача. Результати тестування показника відповідальності (фактор Кеттела «у стінах») для підлітків загальноосвітніх шкіл після літніх канікул були такими:

1) 15 учнів, що брали участь у пошуковій діяльності (спеціально організованих загонах):

{8, 7, 8, 9, 7, 9, 6, 8, 9, 8, 8, 9, 7, 6, 8};

2) 14 учнів, що проводили літній відпочинок у звичайний для них спосіб:

{7, 5, 8, 6, 5, 7, 6, 5, 8, 7, 5, 5, 6, 5}.

Чи можна стверджувати, що рівень показника відповідальності вищий у тих учнів, які брали участь у пошуковій діяльності?

4. Умови завдання:

В таблиці представлені результати щодо заробітку Y (у тис. грн.) у працівників освіти X (1 — для чоловіків, 0 - для жінок).

i	1	2		4	5	6	7	8	9	10	11	12	13	14	15
X	0	1	0	0	1	1	0	1	0	0	0	1	1	1	0
Y	0,5	1,2	0,8	1,0	2,0	1.1	1,0	2,1	0,8	0,6	1,2	1,2	1,8	2,0	1,1

Завдання:

За даними таблиці визначити коефіцієнт кореляції між статтю працівників освіти X (1 — для чоловіків, 0 - для жінок) та їхнім заробітком Y (у тис. грн.)

Варіант 3.

1. Порівняйте інтервальні і абсолютні шкали вимірювань, наведіть приклади.

2. Особливості використання та алгоритм розрахунку критерію U-Манна-Уітні.

3. Практичне завдання. Завдання . Для досліджуваної ознаки, виміряної за порядковою шкалою, у двох незалежних групах одержано такі значення:
перша група ( ) — 17, 24, 22, 25, 24, 27, 25, 12, 26, 11, 23, 14, 29;
друга група ( ) — 15, 22, 9, 8, 23, 16, 10, 10, 24, 18, 24, 17.
З'ясувати, чи ці групи відрізняються за рівнем досліджуваної ознаки при рівні значущості .

4. Умови завдання:
Досліджувався зв'язок між статтю X (1 - для чоловіків, 0 - для жінок) та активним заняттям спортом студентів 1-го курсу Y (1- регулярно відвідує спортивну секцію, 0 - ні). Результати приведені в таблиці.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	0	1	0	0	1	1	0	1	0	0	0	1	1	1	0
Y	1	1	0	1	0	1	1	1	0	1	0	1	1	0	0

Завдання:
Визначити зв'язок між статтю X (1 - для чоловіків, 0 - для жінок) та активним заняттям спортом студентів 1-го курсу Y (1- регулярно відвідує спортивну секцію, 0 - ні).

Варіант 4.

1. Правила ранжування даних. Наведіть приклади.

2. Охарактеризуйте основні завдання та методи їх розв'язання з використанням параметричних критеріїв.

3. Практичне завдання. Завдання . 14 досліджуваних до впливу фактора і після його впливу за 7-бальною шкалою одержали такі бали:

До впливу фактор а	2	5	3	2	4	5	2	5	4	3	6	2	6	3
Після впливу фактор а	2	4	5	1	5	4	2	6	4	2	7	4	7	4

Для рівня значущості $\alpha = 0{,}05$ визначити, чи відбулися зміни.

4. Умови завдання:
За тестом Айзенка в групі випробуваних виявлено 15 екстравертів, з них 8 з високим рівнем нейротизму (холерики) та 7 — з низьким нейротизмом (сангвініки). Тест Спілбергера виявив у тих і інших наступний рівень особистісної тривожності (РОТ):

Холерики	42	44	40	38	43	37	41	42
Сангвініки	34	36	38	40	35	38	39	

Завдання:
Визначити рівень кореляція та її статистичну достовірність між типом темпераменту та особистісною тривожністю.

Варіант 5.

1. Вкажіть різницю, яка існує між генеральною сукупністю та вибіркою. Наведіть приклади. Які існують рекомендації щодо вибору оптимального обсягу вибірки?
2. Особливості використання та алгоритм розрахунку критерію Н-Крускала-Уоллеса.
3. Практичне завдання. В групі студентів проведено тренінг креативного мислення. Визначити результативність впливу, використовуючи Т- критерій Вілкоксона.

№	1	2	3	4	5	6	7	8	9	10	11	12
До	19	26	18	15	29	21	21	18	21	23	14	10
Після	17	20	20	18	30	25	28	19	20	27	19	13

4. Умови завдання:
30 студентів (14 юнаків та 16 дівчат) під час екзаменаційної сесії були протестовані за опитувальником Спілбергера на рівень реактивної тривожності (РРТ). Отримані наступні результати:

№№ п/п	1	2	3	4	5	6	7	8
Юнаки	32	34	28	43	35	26	41	32
Дівчата	34	30	37	43	42	44	46	36
№№ п/п	9	10	11	12	13	14	15	16
Юнаки	40	39	42	38	44	33		
Дівчата	45	28	34	41	40	35	42	39

Завдання:
Визначити достовірність відмінностей за рівнем реактивної тривожності у юнаків та дівчат.

Варіант 6.

1. Нормальний закон розподілення даних і його застосування. Обґрунтуйте необхідність оцінювання відповідності емпіричного розподілу нормальному законові.

2. Особливості використання та алгоритм розрахунку критерію Т-Вілкоксона.

3. Практичне завдання. В опитуванні, яке проводилося з метою з’ясування позитивного ставлення (у %) студентів до відмови від шкідливих звичок взяли участь 6 груп студентів 1-го курсу, 5 груп 2-го і по чотири групи 3-го і 4-го курсів. Чи відрізняються середні показники позитивного ставлення студентів різних курсів до важливості і необхідності здорового способу життя?

***Середні показники позитивного ставлення студентів різних курсів до відмови від шкідливих звичок* (у %)**

група / курс	1	2	3	4	5	6
1	51,12	28,75	63,28	36,50	45,06	49,88
2	60,17	44,25	68,56	38,23	50,08	
3	75,92	67,13	49,07	77,15		
4	53,44	81,60	79,55	84,16		

Для розв’язування скористайтесь критерієм Крускала-Уоллеса.

4. Умови завдання:

У дослідах на 100 досліджуваних (50 чоловіків і 50 жінок) реєструвалося час простої сенсомоторної реакція (ЧР) на звуковий стимул. Отримано наступне результати:

Час реакції, секунди							
	0,10 ÷ 0,12	0,12 ÷ 0,14	0,14 ÷ 0,16	0,16 ÷ 0,18	0,18 ÷ 0,20	0,20 ÷ 0,22	0,22 ÷ 0,24
Частоти зустрічаємості							
Чоловіки	2	15	26	5	2	0	0
Жінки	0	12	20	8	7	2	1

Завдання:

Користуюся критерієм Стьюдента і Фішера, визначити достовірність відмінностей ЧР у чоловіків до жінок.

Варіант 7.

1. Охарактеризуйте параметричні статистичні критерії.

2. Особливості використання та алгоритм розрахунку критерію χ^2- Пірсона.

3. Практичне завдання. При виборі претендентів на посаду директора однієї зарубіжної фірми будо проведено обстеження. Обстеженням було охоплено 20 чоловіків віком від 25 до 40 років, середній вік – 31,5 р. Наприкінці 8-годинного сеансу діагностичних рольових ігор і вправ було проведено соціометричне опитування учасників групи, під час якого вони були повинні відповісти на запитання: «Якщо представником фірми був би я, то я обрав би на посаду директора: 1) ….2) … 3) … ». В результаті цієї процедури кожен учасник отримав ту чи іншу кількість виборів від інших учасників, яка відображала його соціометричний статус у групі претендентів. (див. таблицю).Чи можна стверджувати, що групи з різним статусом відрізняються і за рівнем авторитетності, що визначалась незалежно від соціометрії за допомогою експрес-відеодіагностики? Для розв’язання скористуйтесь критерієм Джонкіра.

Показники за шкалою «Авторитетність» по групах з різним соціометричним статусом

№№ учасників	Група 1: 0 виборів ($n_1=5$)	Група 2: 1 вибір ($n_2=5$)	Група 3: 2-3 вибори ($n_3=5$)	Група 4: 4 і більше виборів ($n_4=5$)
1	5	5	5	9
2	5	6	6	9
3	2	7	7	8
4	5	6	7	8
5	4	4	5	7
Суми	21	28	30	41
Середні	4,2	5,6	6,0	8,2

4. Умови завдання:

У 100 випробовуваних, протестованих по тесту Айзенка, визначався рівень нейротизму. Отримано такі дані:

Рівень нейротизму	Частота	Рівень нейротизму	Частота	Рівень нейротизму	Частота
1	0	9	6	17	8
2	0	10	8	18	6
3	0	11	9	19	4
4	0	12	7	20	3
5	2	13	10	21	1
6	3	14	8	22	0
7	3	15	9	23	0
8	4	16	9	24	0

Завдання:

1. Побудувати графічне зображення (полігон розподілу або частотну діаграму) експериментальних даних.
2. Визначити відповідність емпіричного розподілу нормальному: а) за умовою хі-квадрат.

Варіант 8.

1. Охарактеризуйте непараметричні статистичні критерії.
2. Дайте визначення і опишіть особливості показників мір центральної тенденції.
3. Практичне завдання. У групі студентів був проведений тренінг креативного мислення. Перед початком роботи тренінгу та по її закінченню були проведені тестові зрізи по паралельним формам тесту Й. Ніссинена і Е. Воутілайнена (методика вивчення творчого потенціалу). Дані зрізів наведені у таблиці. Визначте результативність стимулюючого впливу тренінгу за допомогою критерію Вілкоксона.
Результати визначення творчого потенціалу групи студентів до і після проведення спеціального тренінгу (зрізи 1 і 2)

№	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Зріз 1	19	26	18	15	29	21	21	18	21	23	14	10
Зріз 2	17	20	20	18	30	25	28	19	20	27	19	13

4. Умови завдання:
У 100 випробовуваних, протестованих по тесту Айзенка, визначався рівень нейротизму. Отримано такі дані:

Рівень нейротизму	Частота	Рівень нейротизму	Частота	Рівень нейротизму	Частота
1	0	9	6	17	8
2	0	10	8	18	6
3	0	11	9	19	4
4	0	12	7	20	3
5	2	13	10	21	1
6	3	14	8	22	0
7	3	15	9	23	0
8	4	16	9	24	0

Завдання:
1. Побудувати графічне зображення (полігон розподілу або частотну діаграму) експериментальних даних.
2. Визначити відповідність емпіричного розподілу нормальному: б) за умовою Колмогорова.

Варіант 9.

1. Характеристика рівня статистичної значущості. Традиційна інтерпретація рівнів значущості. Правила знаходження рівня статистичної значущості за таблицею критичних значень.

2. Особливості використання та алгоритм розрахунку коефіцієнта кореляції Спірмена.

3. Практичне завдання. В виборці із 28 чоловіків-керівників перед початком курсу тренінгу партнерського спілкування проводилось дослідження. В Табл. приведені індивідуальні значення по фактору, який відображає життєву досвідченість і проникливість. Дані представлені в "сирих" балах і згруповані по чотирьом віковим групам. Чи можна стверджувати, що є певна тенденція зміни значень фактору при переході від групи до групи?

№ испытуемых	Группа 1: 26-31 год (n_1=7)	Группа 2: 32-37 лет (n_2=7)	Группа 3: 38-42 года (n_3=7)	Группа 4: 46-52 года (n_4=7)
1	2	11	8	11
2	10	7	12	12
3	5	8	14	9
4	8	12	9	9
5	10	12	16	10
6	7	12	14	14
7	12	9	10	13
Суммы	54	71	83	78
Средние	7.71	10,14	11,86	11,14

4. Умови завдання:

У 200 учнів випускних класів визначався коефіцієнт інтелектуальності за стандартними тестовим методиками. Після нормування отриманого розподілу IQ за стандартним відхиленню були отримані наступні результати:

Класовий інтервал	Частоти IQ	Класовий інтервал	Частоти IQ
-4 ÷ -3,5 σ	0	0 ÷ 0,5σ	32
-3,5 ÷ -3 σ	1	0,5 ÷ 1 σ	13
-3 ÷ -2,5 σ	5	1 ÷ 1,5 σ	5
-2,5 ÷ -2 σ	12	1,5 ÷ 2 σ	1
-2 ÷ -1,5 σ	18	2 ÷ 2,5 σ	1
-1,5 ÷ -1 σ	27	2,5 ÷ 3 σ	1
-1 ÷ -0,5 σ	41	3 ÷ 3,5 σ	0
-0,5 σ ÷ 0	43	3.5 ÷4 σ	0

Завдання:

Користуючись критеріями λ і χ^2 визначити, чи відповідає отриманий розподіл IQ нормальному

Варіант 10.

1. Характеристики параметрів нормального закону розподілення.

2. Особливості використання та алгоритм розрахунку коефіцієнта кореляції Кенделла.

3. Практичне завдання. В дослідженні було встановлено, що досліджувані по-різному відносяться до покарань, які здійснюють різні люди по відношенню до їхніх дітей. Наприклад, покарання здійснене одним із батьків вважається більш прийнятним, ніж покарання зі сторони бабусі і тим більше зі сторони вихователя, чи вчителя. Чи можна стверджувати про достовірність тенденції в оцінках?

Оцінка ступеня згоди з твердженнями про допустимість покарань в експериментальній групі. (n=16)

Испытуемые	Условие 1: "Я сам наказываю"	Условие 2: "Бабушка наказывает"	Условие 3: "Учительница наказывает"
1	4	2	1
2	1	1	1
3	5	4	4
4	4	3	2
5	3	3	2
6	4	5	1
7	3	3	1
8	5	5	3
9	6	5	3
10	2	2	2
11	6	3	2
12	5	3	4
13	7	5	4
14	5	5	2
15	5	5	4
16	6	6	4
Суммы	71	60	40

4. Умови завдання:

У 100 школярів випускних класах визначався коефіцієнт інтелекту за тестом Векслера. Отримано наступний варіаційний ряд:

Випробо-вувані	IQ	Випробо-вувані	IQ	Випробо-вувані	IQ	Випробо-вувані	IQ	Випробо-вувані	IQ
1	119	21	117	41	104	61	107	81	111
2	86	22	82	42	88	62	78	82	98
3	100	23	100	43	113	63	110	83	84
4	93	24	86	44	89	64	98	84	102

5	108	25	129	45	103	65	84	85	92
6	88	26	103	46	83	66	107	86	110
7	104	27	88	47	91	67	92	87	101
8	127	28	108	48	97	68	105	88	85
9	103	29	70	49	87	69	89	89	114
10	111	30	111	50	101	70	95	90	102
11	83	31	96	51	95	71	90	91	94
12	114	32	90	52	74	72	81	92	106
13	99	33	118	53	94	73	94	93	87
14	118	34	101	54	93	74	95	94	109
15	92	35	90	55	100	75	91	95	95
16	115	36	116	56	89	76	100	96	105
17	112	37	111	57	106	77	93	97	97
18	90	38	96	58	95	78	91	98	91
19	121	39	99	59	116	79	109	99	102
20	96	40	123	60	91	80	94	100	95

Завдання:

Визначити поду, медіану і середнє арифметичне значення коефіцієнта інтелекту для вибірки в 100 випробовуваних.

Короткий термінологічний словник з курсу „Математичні методи в психології”

Альтернативна гіпотеза –гіпотеза, яка протиставляється нульовій гіпотезі і заперечує її.

Аналіз –метод наукового дослідження об'єкта шляхом розгляду його окремих аспектів, складових частин.

Аномальні спостереження –спостереження, характерні для нестабільних явищ і процесів.

Багатоступінчаста вибірка –типова вибірка, яку проводять кількома стадіями (ступенями). При цьому кожна стадія має свою одиницю відбору. При багатоступінчастому відборі з генеральної сукупності спочатку відбирають укрупнені групи, потім –більш дрібні і так доти, доки не будуть відібрані ті одиниці, що піддаються обстеженню.

Багатофазна вибірка – вибірка, яка припускає збереження однієї й тієї ж одиниці добору на всіх етапах його проведення. При цьому відібрані на кожній стадії одиниці піддаються обстеженню. На кожній наступній стадії добору програма обстеження розширюється.

Безповторна власне випадкова вибірка – вибірка, при якій кожна раніше відібрана одиниця не повертається в генеральну сукупність і в подальшому відборі не бере участі.

Варіанти – окремі значення ознаки, які вона приймає у варіаційному ряді. Варіаційний ряд розподілу – ряд, побудований за кількісною ознакою.

Варіація – коливання значень ознаки. Також: коливання, різноманіття, змінюваність величини ознаки (зміна розміру ознаки) в окремих одиницях сукупності.

Варіативна ознака – ознака, яка має в межах сукупності різні значення.

Величина інтервалу – різниця між верхньою і нижньою межами інтервалу.

Вибіркова сукупність – частина генеральної сукупності, яку вибірково

обстежуватимуть, сукупність відібраних для обстеження одиниць.

Вибіркова частка – питома вага одиниць, що володіють даною ознакою у вибірковій сукупності. Розходження між вибірковою часткою і середнім значенням ознаки у вибірці (вибіркової середньої) визначають особливості обчислення необхідного обсягу, помилок вибірки, інтервалів довіри та ін.

Вибіркове (репрезентативне) спостереження – найбільш поширений вид несуцільного спостереження, при якому закономірності й характеристики, властиві якісь генеральній сукупності, визначають дослідженням деякої її частини. Його основою є випадковий відбір одиниць для обстеження, що гарантує незалежність результатів вибірки від волі осіб, які його проводять.

Вибірковий метод – сукупність математичних способів і обґрунтувань, які використовують при застосуванні вибіркового спостереження.

Власне випадкова вибірка – вибірка, при якій кожна одиниця з генеральноїсукупності відбирається у вибірку випадково, ненавмисно. При цьому генеральна сукупність не розподіляється на складові частини. Відбір одиниць звичайно проводиться жеребкуванням.

Внутрішньогрупова дисперсія – відбиває випадкову варіацію, тобто частину варіації, що відбувається під впливом неврахованих факторів і не залежить від факторної ознаки.

Генеральна сукупність – загальна сукупність одиниць, з якої проводять відбір частини одиниць.

Густина розподілу – відношення частоти до величини інтервалу в

інтервальних рядах розподілу з нерівними інтервалами.

Групувальна ознака – ознака, за якою проводиться розбивка одиниць сукупності на окремі групи.

Групування – розчленовування безлічі одиниць досліджуваної сукупності на групи за певними, суттєвими для них ознакам.

Деревовидна кластеризация – об'єднання об'єктів дослідження в досить великі кластери, використовуючи деяку міру схожості або відстань між об'єктами.

Децилі – додаткові статистичні характеристики рядів розподілу, які поділяють ранжируваний ряд розподілу на 10 рівних частин.

Дискретні варіаційні ряди розподілу – ряди, в яких варіанти виражено цілими числами.

Дисперсія – середній квадрат відхилень індивідуальних значень ознаки від їхньої середньої величини.

Експлікація – опис змісту графіка за допомогою слів (словесний опис).

Екстраполяція – знаходження рівнів за межами досліджуваного ряду, тобто продовження ряду на основі виявленої закономірності зміни рівнів у досліджуваний період.

Загальна середня – середня, що показує типовий розмір ознаки якісно однорідної сукупності в цілому.

Загальні методи кластерного аналізу – об'єднання (деревоподібна кластеризація), метод К середніх.

Закон великих чисел – зі збільшенням кількості спостережень вплив випадкових причин, що визначають величину ознаки окремих одиниць сукупності, в цілому взаємно погашається й у зведених характеристиках виражається дія основних причин, тобто визначається закономірність.

Закон розподілу – це співвідношення між можливими значеннями ознаки величин і відповідними ймовірностями.

Закономірність – повторюваність, послідовність і порядок змін у явищах.

Закономірності розподілу – закономірності зміни частот у варіаційних рядах.

Залежні вибірки – вибірки, коли спостереження в одній вибірці якоюсь мірою пов'язано зі спостереженнями у другій.

Зворотний зв'язок – зменшення чи збільшення значення результативної ознаки із збільшенням чи зменшенням значень факторної ознаки.

Імовірність – міра об'єктивної можливості здійснення певних подій. Кількісно імовірність виражають відношенням кількості сприятливих наслідків до кількості можливих наслідків.

Індекс кореляції (кореляційне відношення) – показник тісноти зв'язку міжфакторною і результативною ознаками при криволінійній формі зв'язку, який показує, яку частку в загальному середньоквадратичному відхиленні результативної ознаки становить середньоквадратичне відхилення факторної ознаки. Змінюється в межах від 0 до +1, тобто завжди є додатною величиною.

Інтерполяція – наближений розрахунок рівнів, що лежать усередині ряду динаміки, але з якихось причин невідомих.

Квартилі – додаткові статистичні характеристики рядів розподілу, що поділяють ряд розподілу за сумою частот на 4 рівні частини.

Кількісні ознаки – ознаки, що реєструються числом. Групувальна ознака може бути виражена числами по-різному. Одні ознаки виражаються тільки цілими числами. Така ознака зветься дискретною, або перервною. Інші ознаки можуть позначатися цілими і дробовими числами. Ці зміни ознаки називають безперервними.

Класифікація – різновид типологічних групувань, систематизований, заздалегідь встановлений поділ явищ і об'єктів на групи, класи, розряди, категорії, за якими проводиться зведення даних. Основою класифікації, як правило, є якісна ознака.

Кластер – однорідні групи, отримані в результаті розподілу сукупності об'єктів, кожний з яких характеризується набором k - ознак.

Кластерний аналіз – метод знаходження кластерів.

Коефіцієнт кореляції рангів – визначення тісноти зв'язку між двома кількісними або якісними ознаками після попереднього ранжирування їх в порядку зростання або спадання.

Коефіцієнт множинної (сукупної) детермінації – показує, яка частка варіації досліджуваного результативного показника зумовлена впливом факторів, включених у рівняння множинної регресії. Він може мати значення від 0 до +1.

Коефіцієнт множинної (сукупної) кореляції – основний показник тісноти зв'язку за множинної кореляції, який може мати значення від 0 до +1.

Коефіцієнт регресії – показує, на скільки в середньому змінюється значення результативної ознаки зі зміною факторної на одиницю власного виміру.

Колінеарність (мультиколінеарність) – кореляційний зв'язок між факторами в рівнянні множинної регресії. Чим вища колінеарність, тим менш надійними будуть показники впливу окремих факторів.

Кореляційна залежність – це функціональне співвідношення лише між середніми значеннями досліджуваних ознак.

Кореляційне відношення – показує зв'язок між двома ознаками.

Кореляційне поле – спосіб графічного зображення взаємозалежності статистичних показників. Дає наочне уявлення про наявність зв'язку між досліджуваними ознаками.

Кореляційний аналіз – це метод визначення і кількісної оцінки взаємозалежностей між статистичними ознаками, що характеризують окремі психологічні явища і процеси.

Кореляційний зв'язок – різновид стохастичного зв'язку, що виявляється в зміні середніх умовних розподілів. При К. з. немає суворої відповідності між значеннями залежних ознак: кожному певному значенню аргументу (факторної ознаки) відповідає кілька різних значень функції (результативної ознаки).

Кореляція – статистична залежність між випадковими величинами, що не має суто функціонального характеру, за якої зміна однієї з випадкових величин приводить до зміни математичного очікування іншої.

Крива розподілу – графічне зображення у вигляді безперервної лінії зміни частот у варіаційному ряді, функціонально зв'язаного зі зміною варіант.

Критерії узгодженості – особливі статистичні показники, що характеризують відповідність емпіричного й теоретичного розподілів. Відомі критерії згоди Д. Пірсона.

Критерій t нормального розподілу – це теоретичне нормоване відхилення для великих вибірок. За законом нормального розподілу варіація індивідуальних значень досліджуваної ознаки перебуває в межах $\chi \pm 3\sigma$ (правило трьох сигм). Числове значення цього критерію залежить від рівня імовірності. Його визначають за спеціальними таблицями «Значення інтеграла імовірностей». Критерій t-Стьюдента використовують для перевірки статистичних гіпотез стосовно середніх при малій вибірці (п < 20). Його застосовують, визначаючи надійні інтервали, інтервально оцінюючи параметри генеральної сукупності. Числове значення критерію залежить від кількості ступенів свободи варіації та рівня ймовірності.

Критерій F-Фішера-Снедекора використовують у оцінці співвідношення дисперсій при малих вибірках, а також суті ступеня варіації ознак і надійності взаємозв'язку між факторами.

Критерій Вілкоксона застосовують для перевірки однорідності розподілів двох генеральних сукупностей.

Критерій $\chi 2$ Пірсона – використовують тоді, коли потрібно визначити ступінь відмінності фактичного розподілу частот від теоретичного. Крім того, його застосовують для оцінки однорідності розподілів, а також як критерій незалежності в розподілі об'єктів сукупності за градаціями досліджуваної ознаки.

Критична область – це ті значення критерію, при яких нульова гіпотеза відхиляється.

Критична оцінка вихідних даних – повнота, якість і вірогідність відповідності емпіричного матеріалу цілям і завданням дослідження.

Критичні точки – точки, які відокремлюють критичну галузь від галузі допустимих значень.

Лаг – проміжок часу відставання одного явища від іншого, пов'язаного з ним.

Лінії регресії – лінії, побудовані на основі рівнянь регресії.

Лінійний зв'язок – статистичний зв'язок між явищами, виражений рівнянням прямої лінії.

Лінійний коефіцієнт парної кореляції – кількісний показник тісноти прямолінійного зв'язку результату з одним фактором. При парній залежності коефіцієнт кореляції коливається від 0 до +1 за прямого зв'язку і від 0 до -1 — за зворотного зв'язку. Якщо г < 0.3, зв'язку немає, якщо г = 0.3 - 0.5 – зв'язок слабкий, якщо г = 0.5-0.7 — зв'язок середній і якщо г > 0.7 — зв'язок тісний.

Мала вибірка – вибіркове спостереження, чисельність одиниць якого не перевищує 20. При малій вибірці діє особливий закон розподілу. Величинаймовірної помилки залежить як від коефіцієнта довіри t, так і від обсягу вибірки за випадку, якщо гранична помилка не перевищує t -кратну середню помилку в малих вибірках.

Матеріали спостереження – це первинна статистична інформація, яка є основою для одержання узагальнених характеристик.

Матриця – прямокутна таблиця числової інформації, що складається з m рядків і n стовпчиків.

Медіана (структурна або розподільна середня) – значення ознаки в одиниці сукупності, що займає середнє положення в ранжируваному ряду розподілу. Вона є центром розподілу сукупності і ділить її на дві рівні за кількістю частини.

Метод головних компонент – засіб зниження розмірності. Використовується і для проведення класифікацій. Суть методу полягає у виділенні лінійних комбінацій вихідних факторних ознак, які мають максимально можливу дисперсію. При цьому, перша головна компонента володіє максимальною дисперсією і є нормованою лінійною комбінацією всіх можливих вихідних ознак, а друга – враховує максимальне значення дисперсії, що залишилась, і кореляційно не пов'язана з першою компонентою.

Метод повного зв'язку – заснований на використанні відстані між одиницями кластерів, які найбільш віддалені від інших пар об'єктів і один від одного.

Механічна вибірка – різновид випадкової вибірки, коли одиниці для вибіркового спостереження відбирають не жеребкуванням, а механічно через відповідний інтервал. Для цього всі одиниці генеральної сукупності розподіляють у певному порядку, але так, щоб порядок не був пов'язаний із розміром досліджуваної ознаки.

Множинна кореляція – кореляція, за допомогою якої вивчається вплив на результативну ознаку двох і більше взаємопов'язаних факторних ознак.

Множинна регресія – модель зв'язку трьох і більш ознак.

Множинний коефіцієнт кореляції – відбиває зв'язок між результативною і декількома факторними ознаками.

Мода (структурна або розподільна середня) – значення ознаки, яке найчастіше повторюється в досліджуваній сукупності. Це варіант, який має найбільшу частоту.

Мода і медіана – структурні середні.

Незалежні вибірки – вибірки, коли кожному спостереженню в одній вибірці не можна протиставити спостереження другої вибірки (варіанти яких змінюються незалежно один від одного).

Незважене попарне арифметичне середнє – відстань між двома різними кластерами визначається як середня відстань між усіма парами об'єктів у них.

Нелінійний зв'язок – статистичний зв'язок між психологічними явищами, аналітично виражений рівнянням кривої лінії (параболи, гіперболи і т. ін.).

Непараметричні критерії – статистичні критерії, використання яких не пов'язане зі знанням закону розподілу випадкової величини, їх можна використовувати і тоді, коли досліджуваний розподіл відрізняється від нормального. До непараметричних належать критерії Колмогорова, Вілкоксона, Уайта тощо.

Нормальний розподіл – це симетричний розподіл, в якому максимумизначень випадкової величини концентруються навколо середньої величини.

Об'єкт статистичного спостереження – сукупність суспільних та психологічних явищ і процесів, про які слід зібрати статистичні відомості.

Обернений кореляційний зв'язок – зв'язок, за якого зі збільшенням факторної ознаки результативна ознака зменшується чи, навпаки, зі зменшенням факторної ознаки результативна зростає.

Об'єднання (деревовидна кластеризація) – об'єднання об'єктів дослідження в досить великі кластери, використовуючи деяку міру схожості або відстань між об'єктами.

Об'єкт спостереження – статистична сукупність, у якій виникають досліджувані соціально-психологічні явища і процеси.

Область допустимих значень – це ті значення критерію, при яких нульова гіпотеза приймається.

Одинокий зв'язок (метод «найближчого сусіда») – відстань між двома кластерами визначається відстанню між двома найбільш близькими об'єктами в різних кластерах.

Ознака – загальна властивість, характерна риса чи інша особливість одиниць сукупності, за якими можна вести статистичне спостереження або виміряти їх. Основна відмітна риса, особливість досліджуваного явища чи процесу.

Опитування – спосіб збирання статистичних даних, при якому відповіді на запитання формуляра записують зі слів опитуваної особи. Розрізняють три способи опитування: усне опитування; самореєстрація; кореспондентський спосіб.

Основна тенденція (тренд) – досить плавна і стійка зміна рівня явища в часі, більш-менш вільна від випадкових коливань. Основну тенденцію можна подати або аналітично — у вигляді рівняння (моделі) тренда, або графічно.

Параметричні критерії – статистичні критерії, які ґрунтуються на припущенні, що розподіл досліджуваної ознаки в сукупності підпорядковується певному відомому закону, наприклад, законам: нормального розподілу, розподілу Стьюдента, Фішера і т. ін. До них належать критерії t, F, $\chi 2$. Особливістю цих критеріїв є те, що їх застосування потребує обчислення оцінок параметрів розподілу.

Парна кореляція – кореляційний зв'язок, при якому аналізують зв'язок між парою показників, один з яких факторний, другий — результативний.

Парна регресія – аналітичне вираження зв'язку двох ознак.

Парний коефіцієнт кореляції – показує ступінь тісноти зв'язку між двома ознаками при фіксованому значенні інших факторних ознак.

Процентилі – значення ознаки, які ділять ряд розподілу на сто частин.

Помилка вибіркового спостереження – різниця між величиною параметра в генеральній сукупності та його величиною, обчисленою за результатами вибіркового спостереження.

Помилка репрезентативності – різниця між показниками вибіркової та генеральної сукупностей.

Помилка спостереження – розбіжність між розрахунковим і дійсним значеннями досліджуваних величин. Вони властиві всім несуцільним спостереженням, оскільки, хоч як правильно проводився відбір одиниць сукупності, узагальнені показники відібраної частини завжди будуть якоюсьмірою відрізнятися від відповідних показників усієї сукупності.

Помилки другого порядку – полягають у тому, що приймається нульова гіпотеза, хоч насправді правильна альтернативна гіпотеза.

Помилки першого порядку – полягають у тому, що відхиляється нульова гіпотеза, хоч насправді вона правильна.

Помилки статистичного спостереження – це розбіжності між розмірами якогось показника, що встановлені за допомогою спостереження, і справжніми його розмірами.

Правила об'єднання: одинокий зв'язок (метод «найближчого сусіда»), повний зв'язок (метод «найбільш віддалених сусідів»), незважене попарне арифметичне середнє, зважене попарне арифметичне середнє, метод Варда.

Прямий зв'язок – із збільшенням чи зменшенням значень факторної ознаки збільшується чи зменшується значення результативної ознаки.

Прямий кореляційний зв'язок – зв'язок, при якому зміна факторної ознаки зумовлює зміну результативної ознаки в тому самому напрямі. Прямолінійний кореляційний зв'язок характеризується рівномірним зростанням або зменшенням результативної ознаки під впливом відповідної зміни факторної ознаки. Аналітично його визначають за рівнянням прямої лінії.

Ранг – порядковий номер відповідної одиниці сукупності у ранжируваному ряду. Також: порядковий номер значення ознаки, розміщеного в порядку зростання чи спадання величин.

Ранжирування – процедура упорядкування об'єктів вивчення, що виконується на основі переваги значень ознаки в порядку зростання чи спадання.

Регресійний аналіз – аналітичне вираження зв'язку, в якому зміна однієї величини –результативної ознаки – зумовлена впливом однієї чи декількох незалежних величин (факторів), а безліч всіх інших факторів, що також чинять вплив на залежну величину, приймається за постійні й середні значення.

Результативна ознака – ознака, яка змінюється під впливом факторної ознаки.

Репрезентативна вибірка – вибіркова сукупність, яка достатньо точно відображує генеральну сукупність.

Рівень значущості – імовірність припуститися помилки першого порядку. Р. з. становить ту мінімальну ймовірність, починаючи з якої можна визнати подію практично неможливою, тобто показує міру, з якою ми ризикуємо, відхиляючи нульову гіпотезу.

Рівень істотності – показує ймовірність, з якою гіпотеза, що перевіряється, може дати помилковий результат.

Рівняння регресії (кореляційне рівняння) – рівняння, за допомогою якого визначають статистичний зв'язок між величинами, що корелюють.

Розподіл $\chi 2$ (хі – квадрат) – це закон розподілу вибіркової дисперсії параметрів, які підпорядковуються закону нормального розподілу при малих вибірках.

Розподіл t-Стьюдента –це закон розподілу нормованого відхилення при малих вибірках ($n < 20$).

Середнє квадратичне відхилення – мірило надійності середньої величини. Характеризує середнє коливання ознаки в сукупності, зумовленеіндивідуальними особливостями одиниць сукупності. Обчислюють добуванням квадратного кореня з дисперсії.

Середнє лінійне відхилення – показник варіації, який становить середню з абсолютних відхилень усіх варіантів від середнього значення варіативної ознаки.

Середня арифметична – найбільш поширений вид середніх величин, який застосовують тоді, коли загальний обсяг варіативної ознаки для всієї сукупності становить суму індивідуальних значень усередненої ознаки. Визначають як відношення суми окремих значень ознаки до кількості одиниць сукупності.

Статистична гіпотеза – це припущення відносно параметрів або форми розподілу генеральної сукупності, яке можна перевірити на основі вибірки. У процесі перевірки статистичної гіпотези потрібно визначити, чи узгоджуються дані спостереження з висунутим припущенням. Унаслідок перевірки гіпотеза приймається або відхиляється.

Статистична оцінка параметра розподілу – наближене значення шуканої величини генеральної сукупності, встановлене на основі вибіркового спостереження.

Статистична сукупність – безліч одиниць, що володіють масовістю, однорідністю, визначеною цілісністю, взаємозалежністю станів окремих одиниць і наявністю варіації. Це сума об'єктів (подій, елементів, явищ тощо), які мають єдину якісну основу, але відрізняються певними ознаками.

Найважливішою особливістю статистичної сукупності є однорідність, одноякісність її елементів.

Статистичний критерій – це, обчислений на основі фактичних спостережень оціночний показник, відповідно до якого приймають або відхиляють нульову гіпотезу.

Стохастичний зв'язок – зв'язок, при якому кожному значенню ознаки відповідає певна множина значень ознаки „у" , які утворюють так званий умо-вний розподіл. Якщо умовні розподіли замінюють одним параметром – середнім значенням ys , то такий зв'язок називають кореляційним.

Суцільне спостереження – одержання інформації про всі одиниці досліджуваної сукупності. Також: спостереження, при якому закономірності й характеристики визначають дослідженням всіх одиниць генеральної сукупності. Забезпечує найбільш повну інформацію про загальну кількість одиниць сукупності і дає вірогідні узагальнені статистичні характеристики явищ, що аналізуються.

Теоретична крива розподілу – крива, що виражає загальну закономірність даного типу розподілу в чистому вигляді, що виключає вплив випадкових факторів.

Точність статистичного спостереження – ступінь відповідності величин якогось показника, отриманого за матеріалами статистичного спостереження, до дійсної його величини.

F-розподіл – це спільний закон розподілу двох взаємопов'язаних вибіркових дисперсій для випадкових величин x і y, кожна з яких розподілена нормально.

Факторна ознака – статистична ознака, яка впливає на іншу ознаку і зумовлює її зміну. Факторне групування – групування, в якому групувальною ознакою єфакторний показник, що впливає на зміну результативної ознаки. Якщофакторна ознака істотна, а кількість одиниць у групі досить велика, то іншіумови в середньому по групі вирівнюватимуться, і зміна результативного показника визначатиметься зміною факторного.

Факторний аналіз – перехід від вихідної інформації до узагальнених факторів, які є результатом їх первісної агрегації і лінійної комбінації.

Функціональний зв'язок – зв'язок, при якому кожному значенню факторної ознаки (аргументу), що характеризує певне явище, в усіх випадках відповідає одне або кілька значень результативної ознаки (функції).

Центральна тенденція – це властивість значень досліджуваної ознакигрупуватися навколо центра розподілу частот, статистичною характеристикоюякого є середня величина.

Часткові коефіцієнти кореляції – показники, які характеризують тісноту зв'язку результативної ознаки з однією факторною ознакою при умові, що інші факторні ознаки перебувають на постійному рівні. Парний коефіцієнт кореляції між результативною і факторною ознаками, як правило, не дорівнює відповідному частковому коефіцієнту.

Частота – кількість одиниць спостереження, що мають однакове значенняознаки. Іноді замість частот використовують частості.

РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Базова

1. Руденко В. М. Математична статистика. Навч. посіб. - К.: Центр учбової літератури, 2012. - 304 с.
2. Гласс Дж. и Стенли Дж., Статистические методы в педагогике и психологии. / Пер. с англ. М., 1976
3. Наследов А.Д. Математические методы психологического исследования. Анализ и интерпретация данных. – СПб.: Речь, 2004.
4. Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО «Речь», 2003.
5. Суходольский Г.В. Основы математической статистики для психологов. Л., 1972
6. Тарасов С.Г. Основы применения математических методов в психологии. СПб, 1998

Додаткова

1. Харман Г. Современный факторный анализ. – М.: Статистика,1972.
2. Иберла К. Факторный анализ. Статистика. 1980. 398 с.
3. Гмурман В.Е. Теория вероятностей и математическая статистика. – М.: Высшая школа, 2000.
4. Ермолаев О.Ю. Математическая статистика для психологов. – М.: Московский психолого-социальный институт: Флинта, 2003.