Київський національний університет імені Тараса Шевченка

БІОМЕТРІЯ
методичні вказівки для студентів біологічного факультету

Біометрія / Упорядн. Ю.І.Прилуцький, О.В.Оглобля, Ю.П.Скляров. – К.: ВПЦ "Київський університет", 2003. – 46 с.

Рецензент М.В. Макарець, к.ф.-м. наук, доц.

Затверджено вченою радою
біологічного факультету
10 лютого 2003 року

ПЕРЕДМОВА

Це навчальне видання призначене для студентів біологічного факультету Київського національного університету імені Тараса Шевченка, які вивчають загальний курс “Інформатика та математичні методи в біології” упродовж двох семестрів. “Біометрія” є другою частиною цього курсу.

У ньому коротко викладений лекційний матеріал з таких розділів вищої математики, як теорія ймовірності та математична статистика, які на сьогодні найбільш широко застосовуються при аналізі різноманітних біологічних явищ та процесів. Також приведена теорія похибок, знання якої необхідно при обробці експериментальних даних. Видання містить значну кількість типових прикладів і задач з різних розділів біології, які пояснюють теоретичний матеріал, а також достатню кількість завдань як для виконання студентами під час аудиторних занять, так і для самостійної роботи. Основна увага звертається на ясне і чітке розуміння студентами суті тих чи інших математичних понять, методів і формул та вміння їх коректно застосовувати під час розв’язування завдань різного типу складності. Нарешті, у цьому виданні детально розглянуті деякі алгоритми пакету програми “Excel”, які корисні в статистичних обчисленнях.

§1. Основні поняття теорії ймовірності.

Загальним для усіх експериментів є те, що кожен з них може реалізуватися за певних умов скільки завгодно разів. Для нас несуттєва реальна природа цих результатів, важливим є лише те, що їх кількість n скінчена. Кожний такий результат прийнято називати *елементарною подією* ω.

Означення 1. Множина Ω усіх можливих результатів експерименту утворює простір елементарних подій ω: Ω (ω).

Об’єднання деякої множини елементарних подій будемо називати просто подією A.

<u>Приклад 1.</u> У закритій клітці є 100 кролів, з яких 50 – альбіносів. Навмання вибираємо кроля. Простір елементарних подій дорівнює 100. Далі потрібно вибрати саме кроля-альбіноса. У цьому разі подія A об’єднує 50 елементарних подій.

Означення 2. Запереченням події A називається подія $\overline{A}$, яка включає усі елементарні події, що не входять в A.

Тобто, подія $\overline{A}$ відбувається тоді, коли не відбувається подія A (подія $\overline{A}$ протилежна події A).

Події бувають достовірні, неможливі та випадкові.

Означення 3. Достовірної подією називається подія Ω, яка включає усі елементарні події.

Тобто, достовірна подія обов’язково відбувається при будь-якому експерименті.

Означення 4. Неможливою подією називається подія $\varnothing$, яка не містить жодної елементарної події.

Тобто, неможлива подія ніколи не відбувається в експерименті.

Означення 5. Випадковою подією називається подія, яка може відбутися, а може і не відбутися в експерименті.

Ознайомимось з алгеброю випадкових подій.

Означення 6. Об'єднанням (сумою) двох подій A і B називається подія $A \cup B$, яка складається з усіх елементарних подій, що входять хоча б в одну з цих подій: A або B.

Тобто, подія $A \cup B$ відбувається тоді, коли відбувається хоча б одна з випадкових подій A або B.

Означення 7. Добутком (перетином) двох подій A і B називається подія $A \cdot B$, яка складається з елементарних подій, що входять як в A, так і в B.

Тобто, подія $A \cdot B$ відбувається тоді, коли одночасно відбуваються обидві події A і B.

Означення 8. Дві події A і B називаються несумісними, якщо їх добуток є неможливою подією: $A \cdot B = \varnothing$.

Тобто, поява події A виключає появу події B і навпаки.

Якщо розглядати випадкову подію багато разів за однакових умов експерименту, то можна виявити певну закономірність її появи або не появи.

Аксіома. Кожній елементарній події ω простору елементарних подій Ω ставиться у відповідність деяке число $p(\omega)$, яке назвемо ймовірністю елементарної події ω. Величина $p(\omega)$ невід'ємна і

$$\sum_{\omega} p(\omega) = 1 . \qquad (1)$$

Далі будемо припускати, що ймовірність $P(A)$ будь-якої події A дорівнює:

$$P(A) = \underset{\substack{\downarrow \\ \text{усі елементарні події } \omega \text{ входять в } A}}{\sum_{\omega \in A}} p(\omega). \qquad (2)$$

З цієї аксіоми випливають такі властивості ймовірності:

1. $0 \le P(A) \le 1$.

2. $P(\varnothing)=0;\ P(\Omega)=1;\ P(\overline{A})=1-P(A)$.

3. Для несумісних подій, тобто коли $A_i \cdot A_j=\varnothing,\quad i\neq j$, $1\leq i,\ j\leq n$, справедлива ***<u>формула додавання ймовірностей</u>***:

$$P(A_1 \cup A_2 \cup \ldots \cup A_n)=P(A_1)+P(A_2)+\ldots+P(A_n). \tag{3}$$

4. Для будь-яких подій *A* і *B* маємо:

$$P(A \cup B)=P(A)+P(B)-P(A\cdot B). \tag{4}$$

Класичне означення ймовірності: ймовірність події *A* дорівнює відношенню числа елементарних подій *m*, які сприяють появі *A*, до загального числа усіх можливих елементарних подій *n*:

$$P(A)=\frac{m}{n}. \tag{5}$$

Приклад 2. Яка ймовірність того, що навмання взятий з клітки кріль є альбінос?

Оскільки *n=100*, а *m=50*, маємо: $P(A)=\frac{50}{100}=\frac{1}{2}$.

Зустрічаються однак події, результат яких відхиляється від їх ймовірності. Яскравим прикладом такої події є співвідношення статі у потомстві багатьох тварин і людини. Відомо, що стать потомства визначається в момент запліднення, коли в зиготу привносяться або XX-, або XY - хромосоми. Таким чином, ймовірність появи у потомстві чоловічих та жіночих особин одна і та ж: *P*=1/2. Однак, статистика свідчить, що, наприклад, на 1000 новонароджених за певний проміжок часу 52% припадає на появу чоловічої статі, а 48% - жіночої. Зрозуміло, що зі збільшенням кількості новонароджених (числа випробувань) відхилення появи чоловічої статі від жіночої буде зменшуватися, тобто ймовірність їх появи буде прямувати до *P*=1/2. У цьому факті проявляється дія *закону великих чисел*: відношення *m/n* у формулі (5) буде як завгодно близько наближатися до ймовірності *P(A)* події *A,* якщо число випробувань необмежено зростає.

Означення 9. Умовною ймовірністю події *A* за умови, що відбулася подія *B* (при $P(B)>0$), називають величину

$$P\left(A/B\right)=\frac{P\left(A\cdot B\right)}{P\left(B\right)}. \qquad (6)$$

Звідси впливає ***формула добутку подій*** *A* і *B* (ймовірність сумісної появи двох подій *A* і *B*):

$$P\left(A\cdot B\right)=P\left(A/B\right)\cdot P\left(B\right)=P\left(B/A\right)\cdot P\left(A\right). \qquad (7)$$

Означення 10. Події *A* і *B* називаються незалежними, якщо

$$P\left(A\cdot B\right)=P\left(A\right)P\left(B\right),$$

$$\text{або } P\left(A/B\right)=P\left(A\right) \text{ чи } P\left(B/A\right)=P\left(B\right). \qquad (8)$$

Приклад 3. У людському суспільстві 65% людей палять та 40% хворіють на рак легенів. Знайти ймовірність того, що навмання взята людина:

а) не палить, але має рак легенів; б) палить, але не має раку легенів; в) ніколи не палить і не має раку легенів; г) палить і має рак легенів.

Нехай подія *A* – людина палить; *B* – хворіє на рак легенів. Тоді за умовою задачі маємо: $P\left(A\right)=0.65$; $P\left(B\right)=0.4$; $P\left(\overline{A}\right)=0.35$; $P\left(\overline{B}\right)=0.6$.

а) $P\left(\overline{A}\cdot B\right)=0.35\cdot 0.4=0.14$; б) $P\left(A\cdot \overline{B}\right)=0.65\cdot 0.6=0.39$;

в) $P\left(\overline{A}\cdot \overline{B}\right)=0.35\cdot 0.6=0.21$;

г) $P\left(A \cup B\right)=P\left(A\right)+P\left(B\right)-P\left(\overline{A}\cdot \overline{B}\right)=0.65+0.4-0.21=0.74$.

Означення 11. Множина подій $H_1, H_2 \ldots H_n$ називається повною системою подій, якщо усі вони попарно несумісні ($H_i \cdot H_j = \varnothing$ для $i \neq j$) і в сумі складають достовірну подію $\left(H_1 + \ldots + H_n = \Omega\right)$.

Тобто, внаслідок випробування хоча б одна з цих подій H_i з'явиться обов'язково. У цьому випадку справедлива ***формула повної ймовірності*** (подія A може з'явитися лише разом з однією із несумісних між собою H_i подій)

$$P(A) = \sum_{i=1}^{n} P\left(A/H_i\right) P\left(H_i\right) \tag{9}$$

і ***формула Байєса***:

$$P\left(H_i/A\right) = \frac{P\left(A/H_i\right) \cdot P\left(H_i\right)}{\sum_{j=1}^{n} P\left(A/H_j\right) \cdot P\left(H_j\right)}. \tag{10}$$

Приклад 4. Два незалежних ехолота повідомляють про появу косяка риби. 0.2 і 0.1 ймовірності того, що при появі косяка риби спрацює перший та другий ехолот відповідно. Знайти ймовірність того, що при появі косяка риби спрацює лише перший ехолот.

Позначимо такі події: A – спрацює перший ехолот, B – спрацює другий ехолот і C – спрацює хоча б один ехолот. Отже, маємо чотири гіпотези:

$$H_1 = A \cdot B; \quad H_2 = A \cdot \overline{B}; \quad H_3 = \overline{A} \cdot B; \quad H_4 = \overline{A} \cdot \overline{B}.$$

Ймовірності цих гіпотез дорівнюють: $P\left(H_1\right) = 0.2 \cdot 0.1 = 0.02$; $P\left(H_2\right) = 0.2 \cdot 0.9 = 0.18$; $P\left(H_3\right) = 0.8 \cdot 0.1 = 0.08$ і $P\left(H_4\right) = 0.8 \cdot 0.9 = 0.72$.

Оскільки сума $H_1+H_2+H_3+H_4$ є достовірна подія, то $P(H_1)+P(H_2)+P(H_3)+P(H_4)=1$. Умовні ймовірності події С дорівнюють:

$$P(C/H_1)=0;\quad P(C/H_2)=1;\quad P(C/H_3)=1;\quad P(C/H_4)=0.$$

Тоді за формулою Байєса маємо:

$$P(H_2/C)=\frac{P(C/H_2)\cdot P(H_2)}{P(C/H_2)\cdot P(H_2)+P(C/H_3)\cdot P(H_3)}=$$

$$\frac{1\cdot 0.18}{1\cdot 0.18+1\cdot 0.08}=0.7.$$

§2. Основні поняття комбінаторики.

Часто для знаходження чисел m та n, що фігурують у класичному означенні ймовірності події (5), потрібно знати кількість різноманітних комбінацій, які можна одержати з n елементарних наслідків.

Означення 1. Різні групи, складені з будь-яких елементів, які відрізняються елементами або порядком цих елементів, називають комбінаціями цих елементів.

Ознайомимось із різновидами комбінацій.

Означення 2. Комбінації з n елементів, що відрізняються лише порядком елементів, називають переставленням цих елементів.

Кількість переставлень з n елементів знаходять за формулою:

$$P_n = n! \qquad (11)$$

Означення 3. Розміщенням з n елементів по m називають комбінації, які складаються з m елементів, взятих з даних n елементів ($m<n$) і відрізняються як порядком, так і елементами.

Кількість розміщень з n елементів по m знаходять за формулою:

$$A_n^m = \frac{n!}{(n-m)!}. \qquad (12)$$

Означення 4. Сполученням з n елементів по m називають комбінації, що складаються з m елементів, взятих з даних n елементів ($m<n$) і які відрізняються хоча б одним елементом.

Кількість сполучень з n елементів по m знаходять за формулою:

$$C_n^m = \frac{n!}{m!(n-m)!}. \qquad (13)$$

Між кількістю переставлень, розміщень та сполучень існує простий зв'язок:

$$C_n^m = \frac{A_n^m}{P_m}. \qquad (14)$$

Часто доцільно використовувати такі властивості сполучень:

$$C_n^1 = n; \; C_n^n = C_n^0 = 1; \; C_n^k = C_n^{n-k}; \; C_n^0 + C_n^1 + C_n^2 + \ldots + C_n^n = 2^n. \qquad (15)$$

Знаючи визначення сполучення з n елементів по m, формулу додавання ймовірностей для будь-яких подій A і B (4) можна узагальнити на будь-яку їх кількість. Наприклад, ймовірність $P_{n,m}$ того, що здійсниться рівно m подій із $A_1, \ldots, A_n$ дорівнює:

$$P_{n,m} = S_m - C_{m+1}^1 S_{m+1} + C_{m+2}^2 S_{m+2} - \ldots + (-1)^{n-m} C_n^{n-m} S_n.$$

Ймовірність $\tilde{P}_{n,m}$ того, що здійсниться не менше m подій із $A_1, \ldots, A_n$ дорівнює:

$$\tilde{P}_{n,m} = S_m - C_m^1 S_{m+1} + C_{m+1}^2 S_{m+2} - \ldots + (-1)^{n-m} C_{n-1}^{n-m} S_n.$$

Тут $S_m = \sum_{i_1<i_2<...<i_m}^{n} P(A_{i_1} \cdot A_{i_2} \cdot ...A_{i_m})$, $S_0=1$; $i_1, ...i_m=1, ...n;$ $m=1, ...n.$

Приклад 1. У басейні плаває 10 риб, з яких 8 коропів і 2 карасі. Саком навмання виловлюють 6 риб. Знайти ймовірність того, що усі риби будуть коропи.

Кількість усіх можливих елементарних подій $C_{10}^{6} = \frac{10!}{6!4!} = n$, шуканій події будуть сприяти лише 6 риб з 8: $C_8^6 = \frac{8!}{6!2!} = m$. Отже, $P = \frac{C_8^6}{C_{10}^6} = \frac{2}{15}$.

Приклад 2. У клітці знаходяться 4 кролі заражені вірусом B_1 та 5 кролів – вірусом B_2. Навмання беруть з клітки 2 кролі. Знайти ймовірність того, що будуть взяті кролі, заражені різним вірусом.

Усього кролів 9. Тому кількість можливих подій дорівнює $C_9^2 = n$. Шуканій події A буде сприяти така кількість подій $m_1 \cdot m_2 = C_4^1 \cdot C_5^1 = m$ (формула добутку подій). У результаті маємо: $P\left(A\right) = \frac{m}{n} = \frac{5}{54}$.

§3. Випадкові величини та їх розподіл.

Випадковою величиною називають таку величину, яка внаслідок експерименту може прийняти лише одне заздалегідь невідоме числове значення.

Випадкові величини бувають дискретними та неперервними.

Означення 1. Дискретною випадковою величиною називають величину, яка може приймати відокремлені, ізольовані одне від одного числові значення (їх можна пронумерувати) з відповідними ймовірностями.

Означення 2. Неперервною випадковою величиною називають величину, яка може приймати будь-яке числове значення з деякого скінченого або нескінченого інтервалу. Кількість можливих значень такої величини є нескінчена.

Для повної характеристики випадкової величини потрібно вказати не лише усі її можливі значення, але й закон, за яким знаходять ймовірності кожного значення.

Означення 3. Функцією розподілу випадкової величини ξ називається функція $F(x)$, яка дорівнює: $F(x) = P(\xi \le x)$.

Тобто, функція розподілу визначає ймовірність того, що випадкова величина ξ прийме значення, менше або рівне x.

Властивості функції розподілу:

1. $0 \le F(x) \le 1$;
2. $F(x)$ - зростаюча функція, тобто $F(x_2) > F(x_1)$, якщо $x_2 > x_1$.

Позначимо значення, які приймає дискретна випадкова величина ξ через $x_1, x_2, \ldots$, а ймовірності, з якими ці значення приймаються, через $p_1, p_2, \ldots$. Тоді $\sum_i p_i = 1$.

Розподіл дискретної випадкової величини ξ буде повністю описаний, якщо вказати, що для будь-якого номера i ймовірність того, що ξ приймає значення x_i, дорівнює $p_i = P(\xi = x_i)$.

Функція розподілу $F(x)$ у цьому випадку дорівнює:

$$F(x) = P(\xi \le x) = \sum_{x_i \le x} p_i . \qquad (16)$$

Таким чином, $F(x)$ є ступенева функція, яка має в кожній точці x_i стрибок p_i (Рис.1).

Прикладами дискретних випадкових величин є:

1) біноміально розподілена випадкова величина.

Означення 4. Ймовірність того, що в n незалежних випробуваннях, у кожному з яких ймовірність появи події дорівнює p, ця подія наступить рівно k разів, визначається за ***формулою Бернуллі***

$$p_k = P(\xi = k) = C_n^k p^k q^{n-k}, \tag{17}$$

де $q = 1 - p$.

Тобто, випадкова величина ξ приймає n+1 значення k =0, 1, 2, …, n з відповідними ймовірностями (17). Їх сукупність називається ***біноміальним розподілом***.

Ймовірність того, що подія наступить не менше k_1 і не більше k_2 разів, дорівнює:

$$p_{k_1;k_2} = \sum_{k=k_1}^{k_2} p_k . \tag{18}$$

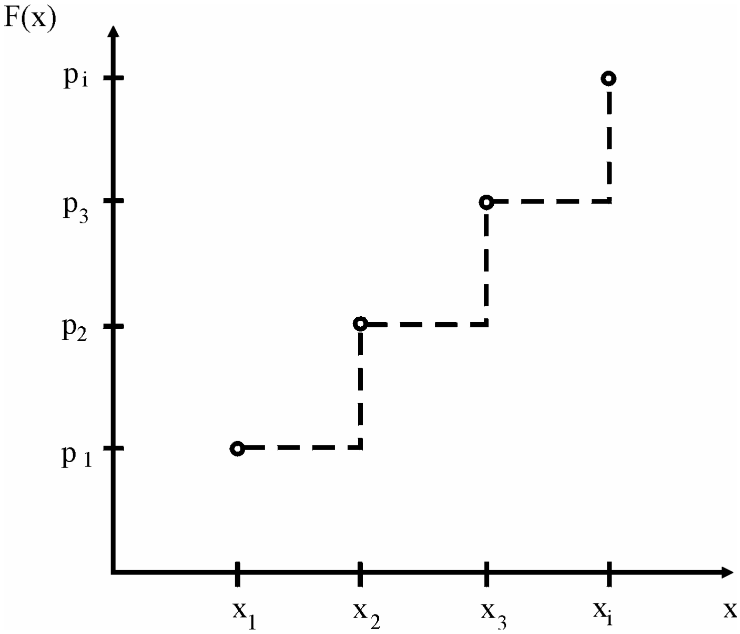


Рис. 1. Функція розподілу (16) для дискретної випадкової величини.

Приклад 1. У лабораторії проводять експерименти з дослідження вірусів. 2-м кролям ввели вірус B_1, 3-м – вірус B_2 і 5-ти – вірус B_3. Ймовірність зараження вірусом B_1 дорівнює 0.1, вірусом B_2 – 0.3 і вірусом B_3 – 0.5. Навмання взятий кріль виявився зараженим. Який вірус найімовірніше йому вводили?

Усього маємо n=10 кролів. Тому згідно формули Бернуллі, навмання взятий кріль виявиться зараженим вірусом B_1 з ймовірністю $P_{10}\left(k=2\right)=C_{10}^{2}(0.1)^{2}(0.9)^{8}=0.19$; вірусом B_2 - $P_{10}\left(k=3\right)=C_{10}^{3}(0.3)^{3}(0.7)^{7}=0.27$; вірусом B_3 -

$P_{10}\left(k=5\right)=C_{10}^{5}(0.5)^{5}(0.5)^{5}=0.25$. Отже, навмання взятий кріль виявився зараженим вірусом B_2.

2) випадкова величина, розподілена за ***<u>законом Пуассона</u>***:

$$p_k=P\left(\xi=k\right)=\frac{\lambda^k e^{-\lambda}}{k!}, \tag{19}$$

де $\lambda>0$ – параметр розподілу Пуассона.

Тобто, випадкова величина ξ приймає злічену множину значень k = 0, 1, 2, … з відповідними ймовірностями (19).

Зауважимо, що у випадку, коли число n незалежних випробувань досить велике, $n>>1$, а $p<<1$, біноміальний розподіл (17) апроксимує розподіл Пуассона, параметр якого $\lambda=np$. На практиці формула (19) є добрим наближенням для (17), якщо $n\geq 100$, $0\leq np\leq 10$.

<u>Приклад 2.</u> У лабораторії проводять експерименти з реєстрації β-частинок. Ймовірність зареєструвати β-частинку дорівнює 10^{-3}. Яка найменша кількість β-частинок повинна вилетіти з джерела для того, щоб із ймовірністю не менше 0.99 зареєструвати більше трьох β-частинок?

Нехай k – шукана кількість β-частинок. Подія A – лічильник зареєстрував більше трьох частинок. Тоді $P(A)=1-P\left(\overline{A}\right)$ і, згідно формули (19), $P\left(\overline{A}\right)=P(0)+P(1)+P(2)+P(3)=$ $e^{-\lambda}\left(1+\lambda+\frac{\lambda^2}{2}+\frac{\lambda^3}{6}\right)\leq 0.01$. Звідси знаходимо, $\lambda\approx 11$ і $k=\lambda/p\approx 11\cdot 10^3$.

Для неперервної випадкової величини ξ функція розподілу $F\left(x\right)$ неперервна. Крім того, існує функція густини $f\left(x\right)$ така, що:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt. \qquad (20)$$

Тобто, $f(x) = F'(x)$.

Розглянемо основні властивості функції густини неперервної випадкової величини ξ:

1. $f(x) \ge 0$;
2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$, оскільки подія $-\infty < x < \infty$ - достовірна;
3. $P(x_1 < \xi \le x_2) = \int_{x_1}^{x_2} f(t)\,dt$.

Найпростішим прикладом неперервної випадкової величини ξ є рівномірно розподілена випадкова величина ξ на інтервалі $(a;b)$ з функцією густини (Рис. 2):

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b \\ 0, & x \notin (a;b). \end{cases} \qquad (21)$$

<u>Приклад 3.</u> Випадкова величина має функцію густини $f(x) = \dfrac{\alpha}{1+x^2}$. Визначити параметр α та функцію розподілу неперервної випадкової величини.

Параметр α знайдемо, використовуючи властивість (2) функції густини: $1 = \int_{-\infty}^{\infty} \dfrac{\alpha}{1+x^2}dx = \alpha arctgx\Big|_{-\infty}^{\infty} = \alpha\pi$. Отже, α=1/π. Функцію розподілу знайдемо за формулою (20):

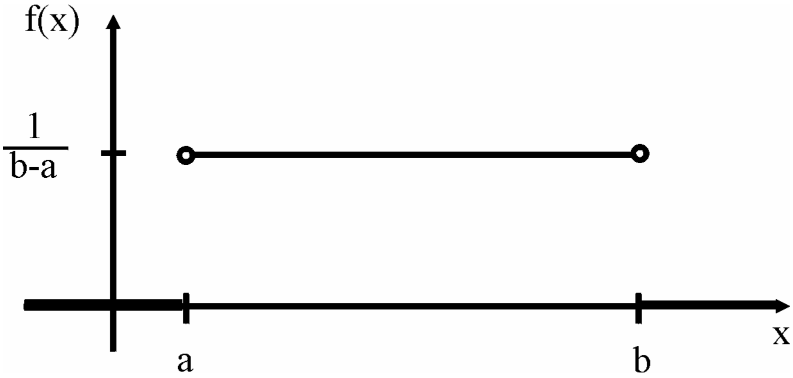


Рис. 2. Функція густини (21) для рівномірно розподіленої випадкової величини.

$$F\left(x\right)=\int\limits_{-\infty}^{x} f\left(t\right)dt=\frac{1}{\pi}\int\limits_{-\infty}^{x}\frac{1}{1+t^2}dt=\frac{1}{\pi}arctgt\Big|_{-\infty}^{x}=\frac{1}{2}+\frac{1}{\pi}arctgx\,.$$

Приклад 4. Випадкова величина задана функцією розподілу: $F\left(x\right)=x^2-2x+1$. Визначити область значень цієї випадкової величини та ймовірність того, що $x \geq 0.5$.

Згідно властивості (1) функції розподілу маємо: $0\le x^2-2x+1\le 1$. Отже, $x\in(0;2)$. Тепер знайдемо ймовірність $P\left(x\ge 0.5\right)$. Подія $x<0.5$ буде протилежною, тому $P\left(x\ge 0.5\right)=1-P\left(x<0.5\right)=1-F(0.5)=1-0.25=0.75$.

§4. Числові характеристики випадкових величин.

На практиці не завжди вдається одержати закон розподілу випадкової величини, або цей закон є надто складним для практичних розрахунків. Тому з’явилася потреба характеризувати випадкову величину за допомогою таких числових характеристик: математичне сподівання, дисперсія та середнє квадратичне відхилення.

Означення 1. Математичним сподіванням $M\left(\xi\right)$ випадкової величини ξ називається число, яке дорівнює сумі добутків усіх можливих значень x на відповідні їм ймовірності, якщо ξ - дискретна випадкова величина

$$M\left(\xi\right) = \sum_{i} x_i p_i \tag{22}$$

і

$$M\left(\xi\right) = \int_{-\infty}^{\infty} xf\left(x\right)dx\,, \tag{23}$$

якщо ξ - неперервна випадкова величина.

Математичне сподівання дискретної випадкової величини ξ характеризує її середнє значення $\overline{\xi}$ (середнє арифметичне)

$$\overline{\xi} = \xi_c = M(\xi) = \frac{\sum_{i=1}^{n} \xi_i}{n} \tag{24}$$

із врахуванням ймовірностей його можливих значень. На практиці під математичним сподіванням розуміють центр розподілу випадкової величини.

Розглянемо основні властивості математичного сподівання:

1) $M\left(C\right) = C, \quad C = const$;
2) $M\left(C\xi\right) = CM\left(\xi\right)$;
3) $M\left(\xi + \eta\right) = M\left(\xi\right) + M\left(\eta\right)$;
4) $M\left(\xi \cdot \eta\right) = M\left(\xi\right) \cdot M\left(\eta\right)$, якщо ξ та η - незалежні випадкові величини;

5) для деякої функції $\varphi(x)$:

$M(\varphi(\xi)) = \sum_i \varphi(x_i) p_i$, якщо ξ - дискретна випадкова величина і

$M(\varphi(\xi)) = \int_{-\infty}^{\infty} \varphi(x) f(x) dx$, якщо ξ - неперервна випадкова величина.

Приклад 1. Незалежні випадкові величини ξ та η розподілені так:

ξ	2	4
p	0.3	0.1

η	5	7	8
p	0.4	0.3	0.2

Знайти математичне сподівання випадкової величини $\xi \cdot \eta$.

Спочатку знайдемо математичні сподівання для кожної з цих величин. За формулою (22) маємо: $M(\xi) = \sum_{i=1}^{2} x_i p_i$ =0.6+0.4=1 і $M(\eta) = \sum_{i=1}^{3} x_i p_i$ =2+2.1+1.6=5.7. Нарешті, згідно властивості (4) одержимо: $M(\xi \cdot \eta) = M(\xi) \cdot M(\eta)$=5.7.

Означення 2. Дисперсією $D(\xi)$ дискретної випадкової величини ξ називається число, яке дорівнює математичному сподіванню квадрата відхилення випадкової величини від її математичного сподівання:

$$D(\xi) = M\left(\left[\xi - M(\xi)\right]^2\right). \qquad (25)$$

Дисперсію $D(\xi)$ неперервної випадкової величини визначають так:

$$D(\xi) = \int_{-\infty}^{\infty} (x - M(x))^2 f(x)\,dx = \int_{-\infty}^{\infty} x^2 f(x)\,dx - [M(\xi)]^2 . \quad (26)$$

Таким чином, дисперсія випадкової величини характеризує розсіювання можливих значень випадкової величини відносно центру розподілу. Дисперсія вимірюється у квадратних одиницях розмірності випадкової величини.

Означення 3. Середньоквадратичне відхилення $\sigma(\xi)$ випадкової величини ξ дорівнює квадратному кореню з дисперсії:

$$\sigma(\xi) = \sqrt{D(\xi)} . \quad (27)$$

Отже, середньоквадратичне відхилення має розмірність випадкової величини.

Розглянемо основні властивості дисперсії:

1) $D(\xi) = M(\xi^2) - [M(\xi)]^2 \geq 0$;
2) $D(C) = 0, \quad C = const$;
3) $D(C\xi) = C^2 D(\xi)$;
4) $D(\xi \pm \eta) = D(\xi) + D(\eta)$, якщо ξ та η - незалежні випадкові величини.

Приклад 2. Знайти дисперсію випадкової величини, яка розподілена так:

ξ	5	-7	-3
p	0.4	0.3	0.2

Будемо шукати $D(\xi)$ за формулою, яка фігурує у властивості (1): $D(\xi)=M\left(\xi^2\right)-\left[M(\xi)\right]^2$. Математичне сподівання $M(\xi)=\sum_{i=1}^{3} x_i p_i$ =2-2.1-0.6=-0.7. Тоді $\left[M(\xi)\right]^2$=0.49. $M\left(\xi^2\right)=\sum_{i=1}^{3} x_i^2 p_i$ =10+14.7+1.8=26.5. У результаті маємо: $D(\xi)$=26.5-0.49=26.01.

<u>Приклад 3.</u> Обчислити $M(\xi)$, $D(\xi)$, $\sigma(\xi)$, якщо випадкова величина ξ: а) рівномірно розподілена на інтервалі $(a;b)$; б) розподілена за законом Пуассона і в) розподілена за біноміальним законом.

а) $M(\xi)=\int_{-\infty}^{\infty} xf(x)dx=\int_a^b \frac{x}{b-a}dx=\frac{b+a}{2}$;

$$D(\xi)=\int_{-\infty}^{\infty}\left[x-M(x)\right]^2 f(x)dx=\int_a^b \frac{\left(x-\frac{b+a}{2}\right)^2}{b-a}dx=\frac{(b-a)^2}{12};$$

$\sigma(\xi)=\frac{b-a}{2\sqrt{3}}$;

б) $M\left(\xi\right)=\sum\limits_{k=0}^{\infty}k\frac{\lambda^k}{k!}e^{-\lambda}=e^{-\lambda}\sum\limits_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}=\lambda e^{-\lambda}\sum\limits_{k=0}^{\infty}\frac{\lambda^k}{k!}=\lambda;$

$$D\left(\xi\right)=\sum_{k=0}^{\infty}k^2\frac{\lambda^k}{k!}e^{-\lambda}-\lambda^2=e^{-\lambda}\sum_{k=0}^{\infty}k(k-1)\frac{\lambda^k}{k!}+e^{-\lambda}\sum_{k=0}^{\infty}k\frac{\lambda^k}{k!}-\lambda^2=$$

$$\lambda^2e^{-\lambda}\frac{d^2}{d\lambda^2}\left(\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\right)+\lambda e^{-\lambda}\frac{d}{d\lambda}\left(\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\right)-\lambda^2=$$

$$e^{-\lambda}\left(\lambda^2e^{\lambda}+\lambda e^{\lambda}\right)-\lambda^2=\lambda;$$

$$\sigma\left(\xi\right)=\sqrt{\lambda}.$$

в) Нехай ξ_i – значення величини ξ в і-тому випробуванні, причому ξ_i приймає два значення: 1 з ймовірністю p та 0 з ймовірністю $q=1-p$. Тоді

$$M\left(\xi_i\right)=1\cdot p+0\cdot q=p;$$

$$D\left(\xi_i\right)=1\cdot p+0\cdot q-p^2=p(1-p)=pq;$$

$$\sigma\left(\xi_i\right)=\sqrt{pq}.$$

Враховуючи, що усіх незалежних випробувань n і $\xi=\xi_1+\ldots+\xi_n$ маємо: $M\left(\xi\right)=\sum\limits_{i=1}^{n}M\left(\xi_i\right)=np;$ $D\left(\xi\right)=\sum\limits_{i=1}^{n}D\left(\xi_i\right)=npq$ і $\sigma\left(\xi\right)=\sqrt{npq}$.

<u>Приклад 4.</u> Знайти числові характеристики випадкової величини, яка задана функцією розподілу:

$$F(x)=\begin{cases}0, & x \le 0;\\ \dfrac{x^2}{10}, & 0 < x < 1;\\ 2, & x > 1.\end{cases}$$

Спочатку знайдемо функцію густини: $f(x)=\begin{cases}\dfrac{x}{5}, & 0 \le x \le 1;\\ 0, & x \notin [0;1].\end{cases}$

У результаті маємо:

$M(\xi)=\int\limits_0^1 x\cdot\frac{x}{5}dx=\frac{1}{15}$; $D(\xi)=\int\limits_0^1 x^2\cdot\frac{x}{5}dx-\frac{1}{225}=\frac{41}{900}$;

$\sigma(\xi)=\sqrt{\frac{41}{900}}\approx 0.21$.

§5. Закони розподілу неперервної випадкової величини.

Означення 1. Випадкова величина ξ називається нормально розподіленою випадковою величиною з параметрами m та σ^2 $\left(\xi \sim N\left(m;\sigma^2\right)\right)$, якщо її функція густини має такий вигляд:

$$f(x)=\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{(x-m)^2}{2\sigma^2}}. \tag{28}$$

Графік функції (28) називають нормальною кривою або ***кривою Гаусса*** (Рис. 3). Якщо m=0 і σ=1, то формула (28) називається функцією Лапласа.

Функцію нормального закону розподілу визначають так (інтегральна функція Лапласа, яка табульована (див. Таблицю 1)):

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(y-m)^2}{2\sigma^2}} dy = \left| \frac{y-m}{\sigma} = z \right| = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-m}{\sigma}} e^{-\frac{z^2}{2}} dz \,. \tag{29}$$

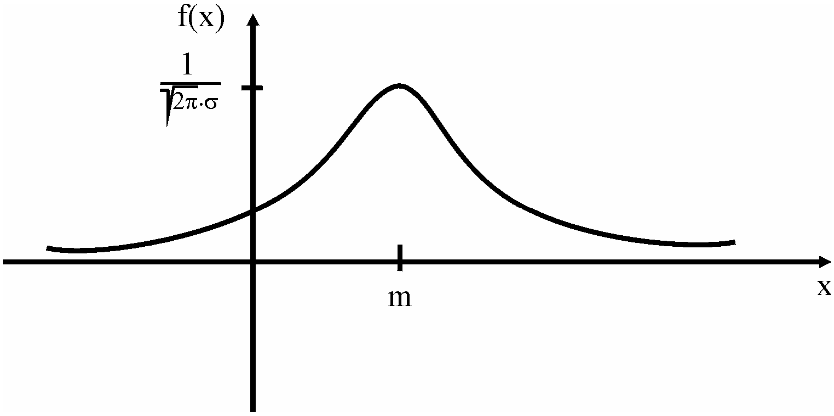


Рис. 3. Функція густини (28) для нормально розподіленої випадкової величини.

Ймовірність влучення в інтервал $(a;b)$ нормально розподіленої випадкової величини знаходять за формулою:

$$P(a < \xi < b) = F\left(\frac{b-m}{\sigma}\right) - F\left(\frac{a-m}{\sigma}\right). \tag{30}$$

Приклад 1. Використовуючи значення інтегралу Пуассона $\int\limits_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz = \sqrt{2\pi}$, знайти числові характеристики нормально розподіленої випадкової величини.

$$M(\xi) = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{\infty} x e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \left| \frac{x-m}{\sigma} = z \right| =$$

$$\frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} (\sigma z + m) e^{-\frac{z^2}{2}} dz = m;$$

$$D(\xi) = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{\infty} (x-m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \left| \frac{x-m}{\sigma} = z \right| =$$

$$\frac{\sigma^2}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} z^2 e^{-\frac{z^2}{2}} dz = -\frac{2\sigma^2}{\sqrt{2\pi}} \frac{\partial}{\partial \alpha} \int\limits_{-\infty}^{\infty} e^{-\frac{\alpha z^2}{2}} dz \Bigg|_{\alpha=1} = \sigma^2 \text{ і } \sigma(\xi) = \sigma.$$

Приклад 2. Зріст досліджуваних тварин розподілено за нормальним законом, причому $M(\xi) = m = 75$ мм і $\sigma(\xi) = \sigma = 6$ мм. Визначити ймовірність того, що хоча б одна з трьох навмання взятих тварин буде мати зріст від 70 до 80 мм.

Позначимо події: A – із 3-х навмання взятих тварин зріст хоча б однієї належить проміжку (70;80); $\overline{A}$ - зріст усіх 3-х тварин не належить проміжку (70;80). Зріст досліджуваних тварин розподілено за нормальним законом, тому за формулою (30) маємо:

$$P\left(70<\xi<80\right)=F\left(\frac{80-75}{6}\right)-F\left(\frac{70-75}{6}\right)=2F\left(\frac{5}{6}\right)\approx 0.59$$

(див. Таблицю 1). Тут ми врахували, що інтегральна функція Лапласа є непарна, тобто $F\left(x\right)=-F\left(x\right)$. Ймовірність того, що зріст однієї тварини не належить проміжку (70;80) буде: $p=1-P\left(70<\xi<80\right)=0.41$. Застосовуючи формулу добутку ймовірностей незалежних подій, знаходимо ймовірність події $\overline{A}$:

Таблиця 1. Значення інтегральної функції Лапласа:

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

x	$F(x)$	x	$F(x)$	x	$F(x)$	x	$F(x)$
0.0	0.000	0.55	0.2088	1.3	0.4032	2.5	0.4938
0.01	0.0040	0.6	0.2257	1.4	0.4192	2.6	0.4953
0.05	0.0199	0.65	0.2422	1.5	0.4332	2.7	0.4965
0.1	0.0398	0.7	0.2580	1.6	0.4452	2.8	0.4974
0.15	0.0596	0.75	0.2734	1.7	0.4554	2.9	0.4981
0.2	0.0798	0.8	0.2881	1.8	0.4641	3.0	0.4986
0.25	0.0987	0.85	0.3023	1.9	0.4713	3.2	0.4993
0.3	0.1179	0.9	0.3159	2.0	0.4772	3.4	0.4997
0.35	0.1368	0.95	0.3289	2.1	0.4821	3.6	0.49984
0.4	0.1554	1.00	0.3413	2.2	0.4861	3.8	0.49992
0.45	0.1736	1.1	0.3643	2.3	0.4893	4.0	0.49996
0.5	0.1915	1.2	0.3849	2.4	0.4918	4.5	0.49997
						5.0	0.49998

$P(\overline{A}) = (0.41)^3 \approx 0.07$. Отже, ймовірність шуканої події A дорівнює: $P(A) = 1 - P(\overline{A}) = 0.93$.

<u>Приклад 3.</u> При обстеженні групи підлітків виявилося, що їх середній зріст характеризується такими показниками: $\overline{x}$=164.8 см і σ_x=5.8 см. У групі виявився підліток, зріст якого дорівнює x_i=172.4 см. Визначити відхилення зросту цього підлітка від середньої величини цієї ознаки у досліджуваній групі.

Шукане відхилення, яке ще називають *нормованим відхиленням*, визначається за формулою:

$$t = \frac{x_i - \overline{x}}{\sigma_x} = +\,1.31.$$

Зауваження. Отримуючи значення нормованих відхилень для різних ознак, можна порівняти місця, які займають індивіди по кожній з цих ознак у їх розподілах. Наприклад, нормоване відхилення у досліджуваного підлітка по ширині плеч дорівнює –0.41. Тоді можна стверджувати, що у нього довжина тіла відхиляється від середньої у бік великих значень цієї ознаки, а ширина плеч – у бік малих, тобто має місце відносно вузькоплечий тип тіла.

Правило трьох сигм. Якщо закон розподілу випадкової величини ξ невідомий, але $\left|\xi - m\right| < 3\sigma$, тоді можна припустити, що ξ розподілена нормально.

Припустимо, що кожна з n незалежних випадкових величин $\xi_1, \xi_2, \ldots, \xi_n$ розподілена нормально з параметрами 0 і 1 $\left(\xi_i \sim N\left(0;1\right)\right)$. Розглянемо випадкову величину:

$$\chi^2 = \sum_{i=1}^{n} \xi_i^2 . \qquad (31)$$

Означення 2. Розподіл випадкової величини χ^2 називається хі-квадрат розподілом з n ступенями вільності.

Числові характеристики χ^2 – розподілу дорівнюють: $M\left(\chi^2\right) = n$, $D\left(\chi^2\right) = 2n$ і $\sigma\left(\chi^2\right) = \sqrt{2n}$. На Рис. 4. зображені графіки деяких функцій густини випадкової величини, яка має χ^2 - розподіл.

Нехай кожна з $n+1$ незалежних випадкових величин ξ_0, ξ_1, ..., ξ_n розподілена нормально з параметрами 0 і σ^2 $\left(\xi_i \sim N\left(0;\sigma^2\right)\right)$. Розглянемо випадкову величину:

$$t = \frac{\xi_0}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\xi_i^2}} \; . \qquad (32)$$

Означення 3. Розподіл випадкової величини t називається розподілом Стьюдента або t – розподілом з n ступенями вільності.

Числові характеристики t – розподілу дорівнюють:

$M\left(t\right) = 0$, $D\left(t\right) = \frac{n}{n-2}$ і $\sigma\left(t\right) = \sqrt{\frac{n}{n-2}}$. Як приклад, на Рис. 5 зображений графік функції густини випадкової величини, яка розподілена за Стьюдентом.

Нехай кожна з $m+n$ незалежних випадкових величин $\xi_1, \ldots, \xi_m$; $\eta_1, \ldots, \eta_n$ розподілена нормально з параметрами 0 і σ^2 $\left(\xi_i, \eta_j \sim N\left(0;\sigma^2\right)\right)$. Розглянемо випадкову величину:

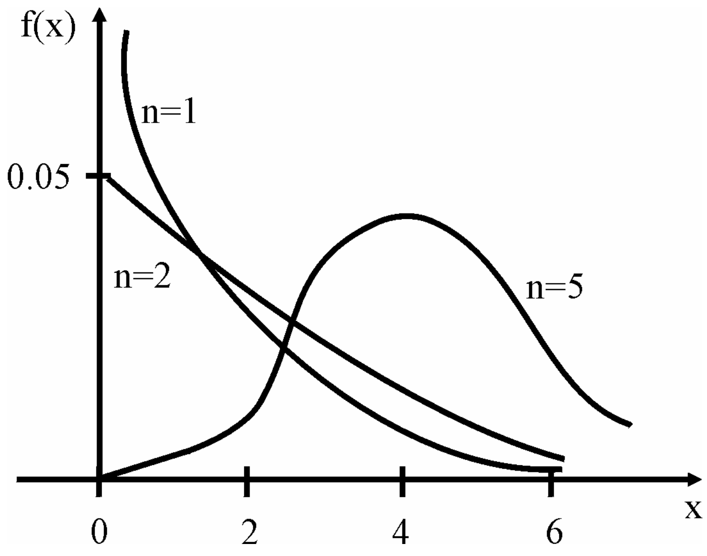


Рис. 4. Графіки функції густини випадкової величини, яка має χ^2 - розподіл.

$$F = \frac{\frac{1}{m}\sum\limits_{i=1}^{m}\xi_i^2}{\frac{1}{n}\sum\limits_{i=1}^{n}\eta_i^2}. \qquad (33)$$

Означення 4. Розподіл випадкової величини z називається розподілом Фішера або F – розподілом зі ступенями вільності m і n.

На Рис. 6 зображені графіки функцій густини випадкової величини, яка розподілена за Фішером.

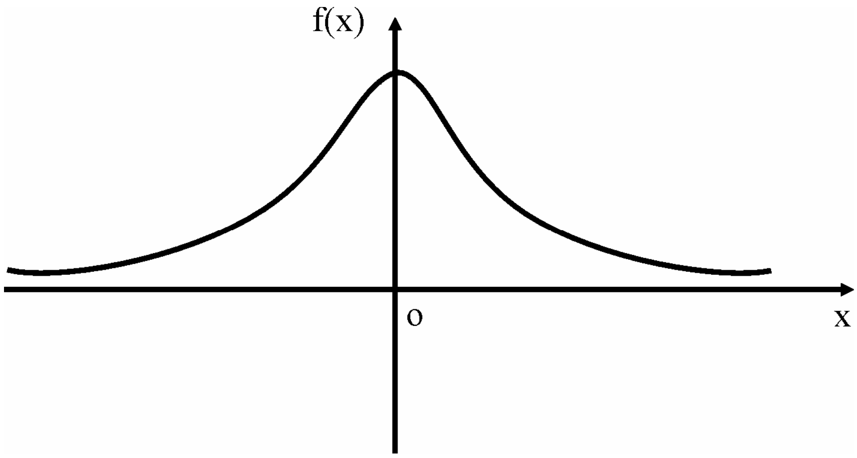


Рис. 5. Графік функції густини випадкової величини, яка розподілена за Стьюдентом.

§6. Двовимірні випадкові величини.

Значення n – вимірної випадкової величини є набір з n чисел. Розглянемо властивості багатовимірних випадкових величин на прикладі двовимірної випадкової величини $\left(\xi_1, \xi_2\right)$.

Означення 1. Функцією розподілу двовимірної випадкової величини $\left(\xi_1, \xi_2\right)$ називають функцію:

$$F\left(x, y\right) = P\left(\xi_1 \le x; \xi_2 \le y\right). \quad (34)$$

Розглянемо основні властивості функції розподілу двовимірної випадкової величини:

1) $0 \le F\left(x, y\right) \le 1$;

2) $F\left(x, y\right)$ – не спадна функція за кожним аргументом x та y;

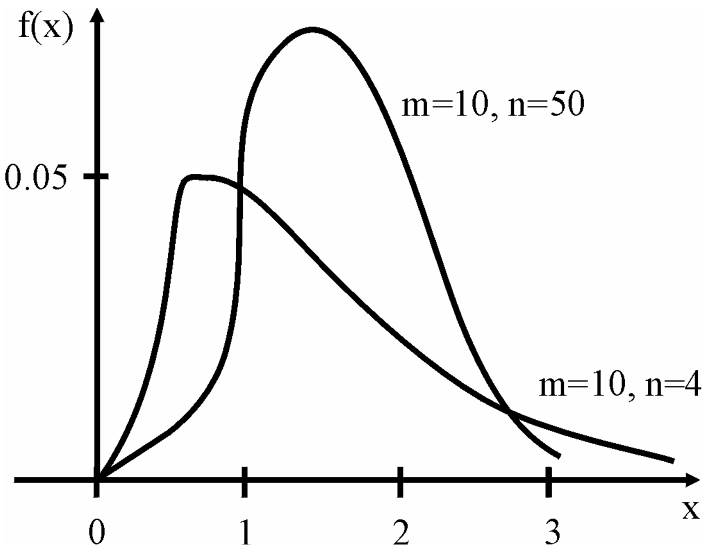


Рис. 6. Графіки функції густини випадкової величини, яка розподілена за Фішером.

3) $\lim\limits_{x\to-\infty} F(x,y)=0;\ \lim\limits_{y\to-\infty} F(x,y)=0;$

$\lim\limits_{\substack{x\to-\infty\\ y\to-\infty}} F(x,y)=0$ і $\lim\limits_{\substack{x\to+\infty\\ y\to+\infty}} F(x,y)=1$;

4) $\lim\limits_{y\to+\infty} F(x,y)=F_1(x)$ і $\lim\limits_{x\to+\infty} F(x,y)=F_2(y)$, де $F_1(x)$ і $F_2(x)$ – функції розподілу одновимірних випадкових величин ξ_1 і ξ_2 відповідно.

Функція розподілу дискретної двовимірної випадкової величини (ξ_1,ξ_2) має такий вигляд

$$F(x,y)=P(\xi_1 \le x; \xi_2 \le y)=\sum_{\substack{x_i \le x,\\ y_j \le y}} p_{ij}, \qquad (35)$$

де $p_{ij}=P(\xi_1=x_i;\ \xi_2=y_j)$ – ймовірність того, що для будь-яких i та j величина (ξ_1,ξ_2) приймає значення (x_i, y_j). Причому, $\sum_{i,j} p_{ij}=1$.

Приклад 1. Знайти закони розподілу компонент двовимірної випадкової величини (ξ_1,ξ_2), закон розподілу якої заданий таблицею:

ξ_2	ξ_1		
	x_1	x_2	x_3
y_1	0.06	0.18	0.16
y_2	0.1	0.3	0.2

Ймовірності відповідних значень ξ_1 і ξ_2 знаходимо так: $p(x_1)$=0.06+0.1=0.16; $p(x_2)$=0.18+0.3=0.48; $p(x_3)$=0.16+0.2=0.36; $p(y_1)$=0.06+0.18+0.16=0.4; $p(y_2)$=0.1+0.3+0.2=0.6. Причому, $\sum_{i=1}^{3} p(x_i)=1$ і $\sum_{i=1}^{2} p(y_i)=1$. Отже, закони розподілу для одновимірних величин ξ_1 і ξ_2 будуть мати такий вигляд:

ξ_1	x_1	x_2	x_3
p	0.16	0.48	0.36

ξ_2	y_1	y_2
p	0.4	0.6

Функція густини $f(x, y)$ для неперервної двовимірної випадкової величини (ξ_1, ξ_2) задовольняє рівності:

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x, y) dx dy . \qquad (36)$$

Розглянемо основні властивості функції густини неперервної двовимірної випадкової величини (ξ_1, ξ_2):

1) $f(x, y) \geq 0$;

2) $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$;

3) $P(a < \xi_1 \leq b;\ c < \xi_2 \leq d) = \int_{a}^{b} \int_{c}^{d} f(x, y) dx dy$.

Математичне сподівання двовимірної випадкової величини (ξ_1, ξ_2) характеризує координати її центру розподілу. Ці координати у випадку неперервних величин знаходять за формулами:

$$m(\xi_1) = m_1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dx dy ;$$

$$m(\xi_2) = m_2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x, y) dx dy . \qquad (37)$$

Дисперсії $D(\xi_1)$ і $D(\xi_2)$ характеризують розсіювання випадкової величини (ξ_1, ξ_2) вздовж координатних осей Ох та Оу відповідно. Їх знаходять за такими формулами:

$$D(\xi_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^2 f(x, y) dx dy - (m_1)^2 ;$$

$$D(\xi_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^2 f(x,y)dxdy - \left(m_2\right)^2. \qquad (38)$$

Характеристикою зв'язку між двома випадковими величинами ξ_1 і ξ_2 є коефіцієнт кореляції

$$r\left(\xi_1,\xi_2\right) = \frac{\operatorname{cov}\left(\xi_1,\xi_2\right)}{\sigma_1\sigma_2}, \qquad (39)$$

де

$$\operatorname{cov}\left(\xi_1,\xi_2\right) = M\left((\xi_1 - m_1)(\xi_2 - m_2)\right) =$$
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-m_1)(y-m_2) f(x,y)dxdy \qquad (40)$$

коефіцієнт коваріації, а $\sigma_1 = \sigma\left(\xi_1\right) = \sqrt{D\left(\xi_1\right)}$ і $\sigma_2 = \sigma\left(\xi_2\right) = \sqrt{D\left(\xi_2\right)}$ - середньоквадратичні відхилення.

Якщо випадкові величини ξ_1 і ξ_2 дискретні, то в формулах (37-38), (40) інтегрування замінюють на сумування по усім можливим значенням випадкових величин.

Означення 2. Якщо коефіцієнт кореляції дорівнює нулеві, $r\left(\xi_1,\xi_2\right)=0$, то випадкові величини ξ_1 і ξ_2 називаються некорельованими.

Розглянемо основні властивості коефіцієнта кореляції:

1) $\left|r\left(\xi_1,\xi_2\right)\right| \le 1$;

2) якщо ξ_1 і ξ_2 - незалежні випадкові величини, то $r\left(\xi_1,\xi_2\right)=0$;

3) якщо $\xi_2 = a\xi_1 + b$, де a і b – деякі сталі, то $\left|r\left(\xi_1,\xi_2\right)\right| = 1$.

Функція густини для двовимірного нормального розподілу випадкової величини $\left(\xi_1, \xi_2\right)$ має такий вигляд:

$$f\left(x, y\right) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} e^{-\frac{1}{2\left(1-r^2\right)}\left[\frac{\left(x-m_1\right)^2}{\sigma_1^2} + \frac{\left(y-m_2\right)^2}{\sigma_2^2} - 2r\frac{\left(x-m_1\right)\left(y-m_2\right)}{\sigma_1\sigma_2}\right]}. \quad (41)$$

З формули (41) випливає: якщо складові нормально розподіленої величини $\left(\xi_1, \xi_2\right)$ некорельовані, то вони є незалежними випадковими величинами. Дійсно, за умови $r\left(\xi_1, \xi_2\right) = 0$ маємо:

$$f\left(x, y\right) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{\left(x-m_1\right)}{2\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{\left(y-m_2\right)}{2\sigma_2^2}} = f_1\left(x\right) \cdot f_2\left(y\right).$$

Зауважимо, що у загальному випадку із незалежності двох величин випливає їх некорельованість, але із некорельованості ще не випливає незалежність цих величин.

§7. Випадкова вибірка.

Будемо розглядати експеримент, пов'язаний з реалізацією одновимірної величини ξ, яка має функцію розподілу $F\left(x\right)$. Виконавши n незалежних повторень експерименту, отримаємо послідовність n значень випадкової величини ξ, які позначимо x_1, x_2, ..., x_n. Кожне з цих значень будемо називати вибірковим значенням, а усю множину вибіркових значень x_1, x_2, ..., x_n – випадковою вибіркою з розподілу $F\left(x\right)$. Крім того, генеральною називають сукупність об'єктів, з яких зроблено вибірку. Наприклад, якщо з 100 кролів для дослідження взято 50, тоді об'єм генеральної сукупності N=100, а об'єм вибірки n=50.

Розглянемо основні числові характеристики вибіркової сукупності.

Означення 1. Вибірковою середньою називають:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i . \qquad (42)$$

Означення 2. Вибірковою дисперсією називають:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} n_i x_i^2 - n\cdot\bar{x}^2\right). \qquad (43)$$

Означення 3. Вибірковим середньоквадратичним відхиленням називають:

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} . \qquad (44)$$

Означення 4. Нехай $x_1, x_2, \ldots, x_n$ – випадкова вибірка з розподілу $F_1\left(x\right)$, а $y_1, y_2, \ldots, y_n$ – випадкова вибірка з розподілу $F_2\left(x\right)$. Вибірковим коефіцієнтом кореляції називають:

$$r = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}} . \qquad (45)$$

Приклад 1. Для випадкової вибірки з Таблиці 2 знайти вибіркові характеристики.

Враховуючи, що $n = 10$, маємо:

$\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 17.1;$ $\quad S^2 = \frac{1}{9}\sum_{i=1}^{10}\left(x_i - \bar{x}\right)^2 = 5.4;$

$S = \sqrt{\frac{1}{9}\sum_{i=1}^{10}\left(x_i - \bar{x}\right)^2} = 2.3.$

Таблиця 2. Вміст крохмалю у картоплі (дані Хальда).

n	*x*,%
1	21.7
2	13.7
3	18.3
4	17.5
5	18.5
6	15.6
7	17.7
8	16.6
9	14.0
10	17.2

Таблиця 3. Співвідношення між процентним вмістом сухої речовини *x* у свіжому шпинаті і процентним вмістом аскорбінової кислоти *y*, яка збереглася після висушування шпинату при 90^o C (дані Петерсена).

n	*x*, %	*y*, %
1	10.0	70.9
2	8.9	74.0
3	8.9	58.6
4	9.2	80.6
5	7.8	69.4
6	10.1	76.0

Приклад 2. Знайти вибірковий коефіцієнт кореляції для двох вибірок з Таблиці 3.

Спочатку визначимо вибіркові середні для двох вибірок *x* та *y*:

$\overline{x} = \frac{1}{6}\sum_{i=1}^{6} x_i = 9.2$; $\overline{y} = \frac{1}{6}\sum_{i=1}^{6} y_i = 71.6$. Тоді

$$r = \frac{\sum_{i=1}^{6}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sqrt{\sum_{i=1}^{6}\left(x_i - \overline{x}\right)^2 \sum_{i=1}^{6}\left(y_i - \overline{y}\right)^2}} = 0.3.$$

На закінчення, розглянемо ще один простий приклад для ілюстрації відмінності між задачами теорії ймовірності і математичної статистики.

Нехай існує N кролів, серед яких доля альбіносів складає 0.1. Припустимо, що із загальної кількості N кролів навмання вибрано 50 кролів. Позначимо через ξ число альбіносів у цій групі кролів. Це число може дорівнювати 0,1,2,3,…,50. Якщо виконати подібний вибір багато разів, то найчастіше число альбіносів у вибірках буде дорівнювати $\xi = 50 \cdot 0.1 = 5$. Перед теорією ймовірності стоїть питання: з якою ймовірністю випадкова величина ξ буде приймати свої можливі значення 0,1,2,…?

І, навпаки, припустимо, що ми маємо вибірку об'єму 50, у якій виявилось ξ альбіносів. Що можна сказати про долю альбіносів серед N кролів? Це і є питання математичної статистики. У теорії ймовірності ми, коректно підібравши математичну модель задачі, досліджуємо поведінку системи. У математичній статистиці ми за даними спостережень за системою, знаючи математичну модель, оцінюємо параметри цієї моделі, порівнюючи її з реальністю.

§8. Довірчий інтервал.

Виходячи з вибіркових значень $x_1, x_2, \ldots, x_n$, розглянемо можливість побудови деякого інтервалу, який містить істинне значення параметра θ з наперед заданою довірчою ймовірністю.

Означення 1. Довірчою ймовірністю оцінки параметра θ за θ^* називають ймовірність $P\left(\left|\theta - \theta^*\right| < \delta\right)$, з якою виконується нерівність $\left|\theta - \theta^*\right| < \delta$.

Найчастіше довірчу ймовірність вибирають рівною P=0.95 або 0.99 або 0.999. З означення випливає, що інтервал ($\theta_н$, $\theta_в$) , де $\theta_н=\theta^*-\delta$ і $\theta_в=\theta^*+\delta$, містить невідомий параметр θ генеральної сукупності. Цей інтервал називають довірчим. Число δ характеризує точність оцінки визначення параметра θ.

Як приклад, приведемо довірчі інтервали ($\theta_н$, $\theta_в$) для випадкової вибірки $x_1, x_2,\ldots,x_n$ з нормальною розподілу $N(m;\sigma^2)$.

1. Довірчий інтервал для оцінки математичного сподівання m.

а) Припустимо, що значення дисперсії σ^2 відоме. Згідно із властивістю нормально розподіленої випадкової величини ξ (29-30) маємо: $P\left(m-\delta<\xi<m+\delta\right)=2F\left(\frac{\delta}{\sigma}\right)$=$2F(t)$. Але $\overline{\xi}$ випадкова величина, яка нормально розподілена $\overline{\xi}\sim N(m;\frac{\sigma^2}{n})$.

Тому при заміні ξ на $\overline{\xi}$, а σ на $\frac{\sigma}{\sqrt{n}}$, отримуємо точність оцінки: $\delta=\frac{t\sigma}{\sqrt{n}}$. Число t визначають з рівності $F(t)=P/2$ з використанням Таблиці 1 значень інтегральної функції Лапласа.

Отже, нижня і верхня границі довірчого інтервалу для математичного сподівання m, яке відповідає довірчій ймовірності P, дорівнюють:

$$\theta_н=\overline{x}-\frac{t\sigma}{\sqrt{n}} \text{ і } \theta_в=\overline{x}+\frac{t\sigma}{\sqrt{n}}. \qquad (46)$$

З формули (46) випливає, що при зростанні об'єму вибірки n величина $\frac{t\sigma}{\sqrt{n}}$ зменшується, а це означає, що точність оцінки δ істинного значення θ збільшується.

б) Припустимо, що дисперсія σ^2 невідома. У цьому випадку нижня і верхня границі довірчого інтервалу для математичного сподівання m, яке відповідає заданій довірчій ймовірності P, дорівнюють

$$\theta_н = \bar{x} - \frac{t_{P,n-1}S}{\sqrt{n}} \quad \text{і} \quad \theta_в = \bar{x} + \frac{t_{P,n-1}S}{\sqrt{n}}, \tag{47}$$

де $t_{P,n-1}$ – коефіцієнт розподілу Стьюдента з n-1 ступенями вільності; S – вибіркове середньоквадратичне відхилення.

Зазначимо, що значення коефіцієнтів $t_{P,n}$ табульовані (див. Таблицю 4).

<u>Приклад 1.</u> Вважаючи, що випадкова вибірка з Таблиці 1 є вибірка з t – розподілу, визначити довірчий інтервал істинної величини з довірчою ймовірністю P=0.96.

Враховуючи, що $\bar{x} = 17.1$, $S = 2.3$ (див. Приклад 1, наведений у попередньому параграфі), а також $t_{0.96,\ 9}$=2.3 (див. Таблицю 3), маємо:

$\theta_н = \bar{x} - \frac{t_{0.96,\ 9}S}{\sqrt{10}}$ =15.4; $\theta_в = \bar{x} + \frac{t_{0.96,\ 9}S}{\sqrt{10}} = 18.8$. Отже, шукана величина знаходиться в такому довірчому інтервалі: 15.4≤ x≤ 18.8.

<u>Приклад 2.</u> Знайти мінімальний об'єм n вибірки, щоб з довірчою ймовірністю P=0.95, точністю δ =0.1 випадкова величина $\bar{\xi}$ була нормально розподілена $\bar{\xi}$ ~$N(m; \frac{0.25}{n})$.

Таблиця 4. Коефіцієнти Стьюдента $t_{P,n}$.

n	P						
	0.7	0.8	0.9	0.96	0.98	0.99	0.999
2	2.0	3.1	6.3	12.7	31.8	63.7	636.6
3	1.3	1.9	2.9	4.3	7.0	9.9	31.6
4	1.3	1.6	2.4	3.2	4.5	5.8	12.9
5	1.2	1.5	2.1	2.8	3.7	4.6	8.7
6	1.2	1.5	2.0	2.6	3.4	4.0	6.9
7	1.1	1.4	1.9	2.4	3.1	3.7	6.0
8	1.1	1.4	1.9	2.4	3.0	3.5	5.4
9	1.1	1.4	1.9	2.3	2.9	3.4	5.0
10	1.1	1.4	1.8	2.3	2.8	3.3	4.8
11	1.1	1.4	1.8	2.2	2.8	3.2	4.6
12	1.1	1.4	1.8	2.2	2.7	3.2	4.5
13	1.1	1.4	1.8	2.2	2.7	3.1	4.3
14	1.1	1.4	1.8	2.2	2.7	3.0	4.2
15	1.1	1.3	1.8	2.1	2.6	3.0	4.1
16	1.1	1.3	1.8	2.1	2.6	2.9	4.0
17	1.1	1.3	1.7	2.1	2.6	2.9	4.0
18	1.1	1.3	1.7	2.1	2.6	2.9	4.0
19	1.1	1.3	1.7	2.1	2.6	2.9	3.9
20	1.1	1.3	1.7	2.1	2.5	2.9	3.9

Для P=0.95 маємо: $F(t)$=0.475. Звідси, з використанням Таблиці 1 значень інтегральної функції Лапласа, знаходимо: t=1.96. Враховуючи, що σ^2=0.25, нарешті отримуємо: $n = \dfrac{t^2 \cdot \sigma^2}{\delta^2} \approx 96$.

§9. Багатовимірні дані. Аналіз статистичних зв'язків.

При аналізі багатовимірних даних важливу роль відіграє поняття типу змінних, які описують об'єкт. Виділяють два типи змінних: кількісні і номінальні. Значення кількісних змінних отримують шляхом вимірювання або підрахунку (ріст, вага, час, температура, число організмів і т. д.). Для кількісної змінної має зміст обчислити середнє із декількох її значень. Номінальні змінні можуть приймати два і більше значень, які не є числами і не впорядковані між собою. Значення номінальної змінної відповідає віднесенню об’єкта до деякого класу або категорії (стать, вид, колір, місце проживання і т. д.).

Формою представлення багатовимірних даних є матриця, яка містить N рядків і M стовпчиків:

$$x = \begin{pmatrix} x_{11} & \ldots & x_{1M} \\ \vdots & & \vdots \\ x_{N1} & \cdots & x_{NM} \end{pmatrix}. \tag{49}$$

Тут елемент x_{ij} цієї матриці визначає значення j-тої змінної X_j для i-того спостереження.

Якщо змінні вимірюються у різних одиницях або діапазони їх зміни значно відрізняються, то у цьому випадку потрібно змінити шкали змінних таким чином, щоб усі вони змінювались приблизно в однакових межах. Наприклад, застосовуючи таке перетворення

$$z_{ij} = \frac{x_{ij} - A_j}{S_j}; \quad A_j = \sum_{i=1}^{N} \frac{x_{ij}}{N};$$

$$S_j = \sqrt{\sum_{i=1}^{N} \frac{\left(x_{ij} - A_j\right)^2}{N-1}}\ , \tag{50}$$

де A_j - середнє значення j-тої змінної, а S_j – середнє квадратичне відхилення цієї змінної, отримуємо, що середнє значення Z_j буде дорівнювати нулю, основна маса значень буде лежати в інтервалі

(–2; 2) і досить рідко будуть зустрічатися значення поза інтервалом (–3; 3).

У якості міри зв'язку між кількісними змінними використовують коефіцієнт кореляції між ними

$$R_{jl} = \frac{\sum_{i=1}^{N}\left(x_{ij} - A_j\right)\left(x_{il} - A_l\right)}{\left(N-1\right)S_j S_l}, \qquad (51)$$

де A_j і A_l – середні значення змінних X_j та X_l, відповідно; S_j і S_l – середні квадратичні відхилення цих змінних.

Метод аналізу зв'язку припускає розбиття вихідної сукупності змінних на дві множини M_1 і M_2. Нехай, наприклад, перша множина змінних $X_1,\ldots,X_{M-1}$ ($M_1=M-1$) відіграє роль пояснюючих змінних, а друга множина змінних Y ($M_2=1$) - роль пояснювальної змінної.

На практиці можуть реалізуватися такі випадки:

1. Якщо $X_1,\ldots,X_{M-1}$ та Y – кількісні змінні, то має місце регресійний аналіз.
2. Якщо $X_1,\ldots,X_{M-1}$ – кількісні змінні, а Y – номінальна змінна, то задача розв'язується методом дискримінантного аналізу.
3. Якщо $X_1,\ldots,X_{M-1}$ – номінальні змінні, а Y – кількісна змінна, то має місце дисперсійний аналіз.
4. Якщо $X_1,\ldots,X_{M-1}$ та Y – номінальні змінні, то така задача носить назву розпізнання образів.

1. Розглянемо задачу регресійного аналізу, яка полягає в знаходженні зв'язку між змінними $X_1,\ldots,X_{M-1}$ та змінною Y шляхом побудови відповідної регресійної функції (наприклад, лінійної комбінації з $X_1,\ldots,X_{M-1}$, яка б була найбільш близької до Y).

Нехай Y – випадкова величина, математичне сподівання якої лінійно залежить від однієї змінної X

$$M(Y)=\theta_0 + \theta_1 X, \qquad (52)$$

де θ_0 і θ_1 – деякі невідомі сталі.

Запишемо нашу модель у такому вигляді

$$y_i=\theta_0 + \theta_1 x_i+ e_i,\ i=1,...,N, \qquad (53)$$

де e_i – відхилення від шуканої моделі (помилки), причому $M(e_i)=0$; $D(e_i)=\sigma^2$; $\mathrm{cov}(e_i, e_j)=0$, де $i, j=1,...,N$, $i\neq j$.

Розглянемо суму квадратів відхилень:

$$R^2 = \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} \left(y_i - \theta_0 - \theta_1 x_i \right)^2 . \qquad (54)$$

Будемо підбирати оцінки $\tilde{\theta}_0$ і $\tilde{\theta}_1$ параметрів θ_0 і θ_1 так, щоб мінімізувати величину R^2. Диференціюючи R^2 по θ_0 і θ_1 і прирівнюючи похідні до нуля, отримуємо алгебраїчну систему з двох рівнянь. Розв'язуючи цю систему, маємо такі оцінки шуканих параметрів:

$$\tilde{\theta}_1 = \frac{\sum_{i=1}^{N} \left(x_i - \overline{x} \right)\left(y_i - \overline{y} \right)}{\sum_{i=1}^{N} \left(x_i - \overline{x} \right)^2} \text{ і } \tilde{\theta}_0 = \overline{y} - \tilde{\theta}_1 \overline{x} . \qquad (55)$$

Такий метод оцінювання називається методом найменших квадратів. Підставивши $\tilde{\theta}_0$ і $\tilde{\theta}_1$ в модель (52), отримаємо лінію регресії: $y = \tilde{\theta}_0 + \tilde{\theta}_1 x$. Величини $\Delta_i = y_i - \tilde{\theta}_0 - \tilde{\theta}_1 x_i$ - це відхилення результатів експерименту від лінії регресії. Використовуючи їх, можна визначити оцінку дисперсії помилки спостережень:

$$S^2 = \frac{\sum_{i=1}^{N} \left(y_i - \tilde{\theta}_0 - \tilde{\theta}_1 x_i \right)^2}{N-2} . \qquad (56)$$

Далі переходимо до регресійного аналізу, тобто до статистичного дослідження отриманої регресії.

Припустимо, що помилки e_i мають нормальний розподіл $N(0;\sigma^2)$ і розглянемо перевірку такої початкової статистичної гіпотези H_0: $\theta_1 = a$. Для цього потрібно, по-перше, задати рівень значущості α (в біологічних задачах α=0.05 або 0.01) і, по-друге, розрахувати, наприклад, статистичний t-критерій:

$$t = \frac{\hat{\theta}_1 - a}{\sqrt{S^2 \Big/ \sum\limits_{i=1}^{N} \left(x_i - \bar{x}\right)^2}}\,. \qquad (57)$$

Якщо розрахований t-критерій приймає значення з області прийняття гіпотези H_0 $(-t_{P,n-2}\,; t_{P,n-2})$, то вважається, що гіпотеза H_0 узгоджується з експериментальними даними.

2. Розглянемо тепер випадок, коли X – номінальна, а Y – кількісна змінна. Це ситуація однофакторного дисперсійного аналізу.

Нехай один фактор X варіюється в експерименті на p рівнях. На i-тому рівні реєструється n_i спостережень залежної змінної Y. Результат j-того спостереження на i-тому рівні позначимо y_{ij} $\left(i = 1,..., p;\ j = 1,..., n_i\right)$. Результати усіх спостережень зручно представити у вигляді моделі: $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, де μ – загальне середнє Y; α_i – ефект, обумовлений i-тим рівнем фактора X; ε_{ij} – помилка j–того спостереження на i-тому рівні. У цьому випадку μ і α_i – невідомі параметри моделі, які можна оцінити за допомогою методу найменших квадратів. При цьому потрібно врахувати додаткове обмеження $\sum\limits_{i=1}^{p} \alpha_i = 0$. Оцінки цих параметрів мають такий вигляд

$$\hat{\mu} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} y_{ij} \Big/ N\,; \quad \hat{\alpha}_i = \left[\sum_{j=1}^{n_i} y_{ij} \Big/ n_i\right] - \hat{\mu}, \qquad (58)$$

де N – загальне число спостережень.

Для прийнятої моделі з перевірки гіпотези H_0: $\alpha_1 = \ldots = \alpha_p = 0$ застосуємо метод дисперсійного аналізу. Такого типу гіпотези перевіряються, наприклад, при порівнянні ефективності дії декількох ліків, при порівнянні середньої ваги тварин певного виду, які мешкають на різних територіях т. д.

Припускаємо, що помилки спостережень ε_{ij} – це незалежні випадкові величини, які мають нормальний розподіл $N(0;\sigma^2)$ (дисперсія σ^2 однакова для усіх рівней зміни фактора X). У дисперсійному аналізу прийняті наступні зручні позначення: точка замість індекса означає усереднення по цьому індексу. Наприклад,

$$y_{1\cdot} = \sum_{j=1}^{n_i} y_{1j} \Big/ n_1 \; ; \; y_{\cdot\cdot} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} y_{ij} \Big/ N \; . \qquad (59)$$

Запишемо тотожність $y_{ij} - y_{\cdot\cdot} = (y_{i\cdot} - y_{\cdot\cdot}) + (y_{ij} - y_{i\cdot})$. Підносячи обидві частини до квадрату і сумуючи по i та j, отримуємо

$$S = \sum_{i=1}^{p} \sum_{j=1}^{n_i} \left(y_{ij} - y_{\cdot\cdot} \right)^2 = \sum_{i=1}^{p} n_i \left(y_{i\cdot} - y_{\cdot\cdot} \right)^2 +$$

$$\sum_{i=1}^{p} \sum_{j=1}^{n_i} \left(y_{ij} - y_{i\cdot} \right)^2 = S_M + S_B \; , \qquad (60)$$

де S_M – випадкова величина, яка характеризує ступінь розсіювання спостережень між рівнями фактора X, а S_B – всередині рівнів. Отже, ми отримали розклад загального розсіювання спостережень на два доданки, кожен з яких відповідає за своє джерело помилок.

Далі необхідно побудувати критерій перевірки гіпотези H_0 при заданому рівні значущості α. Для цього розглянемо, наприклад, статистику:

$$F = \frac{S_M \big/ (p-1)}{S_B \big/ (N-p)} \; . \qquad (61)$$

Так вибраний критерій називається F-критерієм. Якщо гіпотеза H_0 вірна, то величина F близька до 1. Тому у якості області прийняття H_0 потрібно взяти область значень

$\left(-\infty; t_{p-1,N-p}\right)$, де $t_{p-1,N-p}$ – коефіцієнт розподілу Фішера з p–1 і N–p ступенями вільності.

Приклад 1. Побудувати лінію регресії для Таблиці 3. Перевірити гіпотезу H_0: θ_1=0 з рівнем значущості 0.04 (довірчою ймовірністю 0.96).

Згідно формули (55) маємо: $\tilde{\theta}_1 = \dfrac{\sum\limits_{i=1}^{6}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum\limits_{i=1}^{6}\left(x_i - \bar{x}\right)^2}$ =2.7 і $\tilde{\theta}_0 = \bar{y} - \tilde{\theta}_1\bar{x}$ =46.8. Отже, шукана лінія регресії має такий вигляд: $y = \tilde{\theta}_0 + \tilde{\theta}_1 x$ =46.8+2.7x. Гіпотезу H_0 перевіримо, використовуючи статистику (57)

$$t = \frac{\tilde{\theta}_1}{\sqrt{S^2 \Big/ \sum\limits_{i=1}^{6}\left(x_i - \bar{x}\right)^2}} = 0.6,$$

де $S^2 = \dfrac{\sum\limits_{i=1}^{6}\left(y_i - \tilde{\theta}_0 - \tilde{\theta}_1 x_i\right)^2}{4}$ =63.5. Оскільки розрахований t-критерій приймає значення з області прийняття гіпотези H_0 (–3.2;3.2) ($C_{0.96,\ 4}$=3.2, див. Таблицю 4), то вважається, що гіпотеза H_0 узгоджується з експериментальними даними.

Задачі для самостійної роботи

1. Скількома способами можна розмістити 5 студентів за столом, біля якого стоять 5 стільців?
2. Товариство з 7 чоловік сідає за круглий стіл. Знайти ймовірність того, що певні дві особи будуть сидіти поряд.

3. Студенти 2-го курсу вивчають 10 дисциплін. На один день можна планувати 3 дисципліни. Скількома способами можна скласти розклад занять на один день?
4. У воді плаває 4 коропи і 5 карасів. Навмання виловлюють 2 риби. Знайти імовірність того, що ці риби будуть різні.
5. 25 екзаменаційних білетів містить по 2 питання, які не повторюються. Студент може відповісти лише на 45 питань. Яка ймовірність того, що взятий студентом екзаменаційний білет містить відомі йому питання?
6. У клітці є 20 кролів, з яких 5 альбіносів. Навмання взяли 3 кролі. Яка ймовірність того, що серед взятих кролів: а) усі не альбіноси; б) усі альбіноси; в) один альбінос і два не альбіноси?
7. На 2-му курсі навчається 175 студентів. Яка ймовірність того, що хоча б у двох студентів збігаються дні народження?
8. Є 30 екзаменаційних білетів, серед яких є 5 "щасливих". Кому вигідніше тягнути білет – першому чи другому студенту?
9. За статистикою 68% чоловіків, які досягли 60-ліття, досягають також і 70-ліття. Яка ймовірність того, що 60-річний чоловік не досягне свого 70-річчя?
10. Серед студентів 2-го курсу є 45% тих, що палять, 40% тих, що інколи вживають алкоголь, 18% тих, що палять та інколи вживають алкоголь, Знайти ймовірність того, що будь-який студент цього курсу:
    а) не палить, але іноді вживає алкоголь;
    б) палить, але ніколи не вживає алкоголь;
    в) не палить та ніколи не вживає алкоголь;
    г) або лише палить або лише вживає алкоголь.
11. Для сіяння пшениці заготовлено насіння 95% першого сорту, 3% другого сорту і 2% третього сорту. Ймовірність того, що із насіння виросте колосся, яке містить 50 насінин, становить для першого сорту 0.5, для другого сорту – 0.2 і для третього сорту – 0.1. Знайти ймовірність того, що навмання взяте колосся буде мати не менше 50 насінин.

12. В обчислювальному центрі є 6 ПК IBM AT та 4 ПК IBM XT. Ймовірність виходу з ладу під час роботи для ПК IBM AT дорівнює 0.05, а для ПК IBM XT – 0.2. Студент виконує лабораторну роботу на обраному навмання ПК. Знайти ймовірність того, що під час роботи ПК не вийде з ладу.
13. У лікарню поступають 50% хворих на грип, 30% хворих на ангіну та 20% хворих на запалення легенів. Ймовірність повного одужання від грипу дорівнює – 0.7, від ангіни – 0.8 та запалення легенів – 0.9. Виписано хворого, який повністю одужав. Знайти ймовірність того, що він був хворий на грип.
14. У рибалки є три улюблені місця, куди він приходить з однаковою ймовірністю. Ймовірність кльову на першому місці дорівнює $\frac{1}{3}$, на другому - $\frac{1}{2}$ і на третьому - $\frac{1}{4}$. Рибалка закинув вудку у навмання вибраному місці і риба клюнула. Знайти ймовірність того, що рибалка закинув вудку в першому місці.
15. Серед студентів 2-го курсу є 50% тих, що одержують на екзаменах оцінку “добре”, 20% тих, що одержують “відмінно” і 10% тих, що одержують “добре” та “відмінно”. Знайти ймовірність того, що будь-який студент цього курсу:
    а) не одержує ”відмінно”, а одержує “добре”;
    б) одержує “відмінно” і не одержує “добре”;
    в) не одержує “добре” та не одержує “відмінно”;
    г) одержує або лише “добре” або лише “відмінно”.
16. Група піддослідних тварин складається з 20 кролів, 6 щурів та 4 мишей. Ймовірність заразитися вірусом B_1 для кроля дорівнює 0.9, для щура – 0.8 і для миші – 0.75. Знайти ймовірність того, що навмання взята тварина буде заражена цим вірусом.
17. У клітці є тварини двох підвидів – першого та другого, причому тварин другого підвиду в 1.5 рази більше, ніж тварин

першого підвиду. Знайти ймовірність того, що серед трьох навмання взятих тварин хоча б одна буде першого підвиду.

18. Ймовірність появи події A хоча б один раз в 5-ти незалежних випробуваннях дорівнює 0.9. Яка ймовірність появи події A в одному випробуванні, якщо при кожному випробуванні ця ймовірність однакова?
19. Побудувати графік функції розподілу біноміально розподіленої випадкової величини для $n = 1$.
20. Нехай ξ – деяка випадкова величина. Що можна сказати про є залежність випадкових величин ξ і ξ^2?
21. Нехай $\xi_1, \ldots \xi_n$ – незалежні випадкові величини, які однаково нормально розподіленні: $\xi_i \sim N(m, \sigma^2)$. Обчислити $M\left(\overline{\xi}\right)$, $D\left(\overline{\xi}\right)$ і $\sigma\left(\overline{\xi}\right)$.
22. Дискретна випадкова величина X може приймати лише два значення x_1 та x_2 $(x_1 < x_2)$. Відомі ймовірність p_1=0.2 появи можливого значення x_1, математичне сподівання $M(X)$=3.8 та дисперсія $D(X)$=0.16. Знайти закон розподілу цієї випадкової величини.
23. Випадкова величина x задана функцією розподілу: $F(x)=x^2 - 4x +4$. Визначити область значень випадкової величини x та ймовірність того, що $x \geq 2{,}3$.
24. Знайти дисперсію випадкової величини x, яка задана законом:

x	-5	0	4	5
p	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$

25. Знайти $M(x)$, $D(x)$ і $\sigma(x)$ випадкової величини x, яка задана функцією розподілу:

$$F(x)=\begin{cases}0, & x\le 0;\\ \dfrac{x^2}{25}, & 0\le x\le 5;\\ 1, & x\ge 5.\end{cases}$$

26. Випадкову величину X задано функцією розподілу:

$$F(x)=\begin{cases}0, & x\le 0;\\ x^3, & 0<x\le 1;\\ 1, & x>1.\end{cases}$$

Знайти густину розподілу $f(x)$, математичне сподівання та дисперсію цієї випадкової величини. Побудувати графіки $F(X)$ та $f(x)$.

27. Знайти числові характеристики випадкової величини, розподіленої за показниковим законом:

$$f(x)=\begin{cases}4e^{-4x}, & x\ge 0;\\ 0, & x<0.\end{cases}$$

28. Вибіркова сукупність задана таблицею:

x_i	1	2	3	4
n_i	15	10	20	5

Тут n_i – частота появи значення x_i. Знайти вибіркові характеристики.

29. Знайти вибіркові характеристики ряду розподілу кальція у сироватці крові мавп:

x_i, мг	9.0	9.8	10.6	11.4	12.2	13.0	13.8	14.6
n_i	2	6	15	23	25	17	7	5

30. Випадкова величина розподілена за нормальним законом з параметром σ=2. Зроблена вибірка об’єму n=25. З довірчою

ймовірністю P=0.96 знайти довірчий інтервал невідомого параметра цього розподілу.

31. Випадкова величина розподілена за нормальним законом з математичним сподіванням m=8 та середньоквадратичним відхиленням σ=1. Знайти ймовірність влучення цієї величини в інтервал (2;13).
32. Проростання пшеничного насіння становить 85%. Знайти ймовірність того, що з 2000 посіяних насінин проросте від 1880 до 1920.
33. Випадкова величина X розподілена рівномірно на інтервалі (3;8). Знайти її густину розподілу, математичне сподівання, дисперсію та ймовірність того, що $X \in (1;4)$.
34. Ймовірність присутності студента на лекції дорівнює 0.8. Знайти ймовірність того, що із 100 студентів на лекції будуть присутні не менше 75 та не більше 90.
35. Застосовуючи метод найменших квадратів, скласти рівняння параболи $y = ax^2 + bx + c$, яка проходить найближче до експериментальних точок:

x_i	-4	-3	-2	-1	0	1	2	3	4
y_i	6	3	1	0.3	-0.1	-0.2	0	0.2	1

36. За дослідними даними методом найменших квадратів знайти параметри лінійної залежності $y = ax + b$:

x_i	1	2	3	4
y_i	1.1	0.48	-0.1	-0.8

37. Побудувати лінію регресії для такого розподілу:

x_i	1	3	5	7	9
y_i	0.48	1.26	2.85	3.96	5.15

Використовуючи t-критерій, перевірити гіпотезу H_0: θ_1=-0.3 з довірчою ймовірністю 0.96.

§10. Статистична обробка результатів вимірювань (теорія похибок).

При проведенні експериментальних досліджень значення фізичної величини можна отримувати через прямі та непрямі вимірювання.

Прямі вимірювання – це вимірювання, коли шукане значення фізичної величини отримують безпосередньо в результаті проведеного експерименту. Наприклад, вимірювання довжини предмета лінійкою, вимірювання сили струму в електричному колі за допомогою амперметра тощо.

Непрямі вимірювання – це вимірювання, коли шукане значення фізичної величини знаходять на основі відомої залежності між цією величиною та величинами, які отримують за допомогою прямих вимірювань. Наприклад, знаходження площі тіла (S) за результатами вимірювань його довжини (a) і ширини (b): $S=ab$; знаходження опору резистора (R) за результатами вимірювань напруги (U) і сили струму (I) в електричному колі (закон Ома): $R=U/I$.

Досвід показує, що довільні вимірювання фізичної величини завжди супроводжуються відхиленнями від істинного значення величини, що вимірюється. Отже, *похибка вимірювання* фізичної величини – це відхилення результату вимірювання від істинного значення вимірюваної фізичної величини.

За способом вираження розрізняють похибки абсолютні та відносні.

Абсолютною похибкою називають похибку, яка виражена в одиницях вимірюваної величини

$$\Delta X = X - X_{вим}, \qquad (62)$$

де X – істинне значення вимірюваної фізичної величини, $X_{вим}$ – значення, яке одержане в результаті проведеного вимірювання.

Відносною похибкою називають похибку, яка дорівнює відношенню абсолютної похибки до істинного значення вимірюваної величини:

$$\delta = \frac{\Delta X}{X}. \qquad (63)$$

Точність вимірювань визначає якість вимірювання, що відображає близькість результатів до істинного значення X. Висока точність означає малі похибки. Кількісно точність вимірювання визначається величиною, яка обернена до відносної похибки, тобто $\frac{1}{\delta}$.

Похибка вимірювання обраного методу складається із систематичної та випадкової складових.

Систематична складова похибки вимірювання – це така її частина, яка при повторних вимірах однієї і тієї ж величини, що виконуються при незмінних умовах, залишається постійною або закономірно змінюється. Систематична складова прямого вимірювання зумовлена, головним чином, похибками засобів вимірювання (*інструментальними похибками*), а непрямого вимірювання – як інструментальними похибками, так і недоліками методів вимірювання (*похибки методу*) та *похибками обчислень*. Наприклад, вимірювання розмірів тіла проводяться за допомогою міліметрової лінійки. Похибка вимірювання визначається процедурою відліку, тобто абсолютна похибка дорівнює ціні поділки лінійки: ΔX=1 мм.

Випадкова складова похибки вимірювання – це така частина похибки вимірювання, яка змінюється випадковим чином при проведенні повторних вимірювань однієї й тієї ж величини. Випадкова складова зумовлена такими змінами умов експерименту, які не контролюються.

Для оцінки випадкової складової похибки вимірювання використовують результати теорії випадкових величин, а саме такі числові характеристики як математичне сподівання $M(X) = x_c$, дисперсію $D(X)$, середньоквадратичну похибку (стандартну похибку) методу $S = \sqrt{D(X)}$, середньоквадратична похибка результату $S_{x_c} = \frac{\sigma}{\sqrt{n}}$ та довірчий інтервал вимірювання $\Delta X = t_{P,n} \cdot S_{x_c} = \frac{t_{P,n} \cdot S}{\sqrt{n}}$, де $t_{P,n}$ – коефіцієнт Стьюдента (P – значення довірчої ймовірності в серії з n вимірів, яке береться з Таблиці 4). Тоді істинне (шукане) значення фізичної величини буде таким:

$$X = x_c \pm \Delta X . \quad (64)$$

При проведенні прямих вимірювань можуть виникнути три ситуації:

1. систематична складова набагато більша за випадкову: $\Delta X_{сист} >> \Delta X_{вип}$. У цьому випадку довірчий інтервал вимірювання визначається більшою з них, тобто систематичною складовою. У результаті маємо: $X = x_c \pm \Delta X_{сист}$;
2. випадкова складова набагато більша за систематичну: $\Delta X_{вип} >> \Delta X_{сист}$. У цьому випадку довірчий інтервал вимірювання визначається лише випадковою складовою. Результат записується у вигляді: $X = x_c \pm \Delta X_{вип}$;
3. систематична і випадкова складові одного порядку: $\Delta X_{сист} \sim \Delta X_{вип}$. У цьому випадку довірчий інтервал вимірювання визначається за формулою:

$$\Delta X = \sqrt{\Delta X^2_{сист} + \Delta X^2_{вип}} . \quad (65)$$

Зауважимо, що випадкова складова зі збільшенням кількості вимірів зменшується, оскільки $\Delta X \sim \frac{1}{\sqrt{n}}$. Чи є необхідним таке зменшення випадкової складової? Виявляється, що ні. Досить, якщо вона буде одного порядку з систематичною, яка для даної серії вимірів визначає мінімальну похибку, що її дає метод вимірювання.

Як приклад статистичної обробки результатів вимірювань, розглянемо наступне завдання.

Завдання. Визначити константу швидкості гідролітичного розщеплення сахарози.

1. За допомогою напівтіньового поляриметра визначити кут повороту площини поляризації α_t. Провести 20 вимірів через кожні 3 хв.
2. Визначити кут обертання $\alpha_э$, який відповідає закінченню реакції гідролізу.
3. Використовуючи метод найменших квадратів, побудувати графік залежності $y(t)=lg(\alpha_t-\alpha_э)$. Екстраполювавши дану лінію до точки $t=0$, отримаємо відрізок на осі ординат, який дорівнює $lg(\alpha_0-\alpha_э)$, де α_0 – кут обертання площини поляризації на початку реакції.
4. Розрахувати константу швидкості реакції k при даній температурі для кожного моменту часу t за формулою:
$$k = \frac{2.303}{t} lg \frac{\alpha_0 - \alpha_\infty}{\alpha_t - \alpha_\infty}.$$
5. Результати проведених експериментів та розрахунків занести у таблицю. Знайти похибки вимірювань фізичних величин.

1. Результати експерименту 1 та розрахунків 3 занесемо у таблицю, врахувавши попередньо результат експерименту 2: $\alpha_э = -15^0$:

n	t, хв	α_t, град	α_t-$\alpha_э$, град	$y(t)=lg(\alpha_t-\alpha_э)$
1	1	19.3	34.3	1.54
2	4	16.2	31.2	1.49
3	7	13.2	28.2	1.45
4	10	11.1	26.1	1.42
5	13	8.2	23.2	1.37
6	16	5.6	20.6	1.31
7	19	3.4	18.4	1.26
8	22	1.6	16.6	1.22
9	25	0	15	1.18
10	28	-1.5	13.5	1.13
11	31	-2.8	12.2	1.09
12	34	-3.9	11.1	1.05
13	37	-4.7	10.3	1.01
14	40	-5.5	9.5	0.98
15	43	-6.2	8.8	0.94
16	46	-7.0	8.0	0.90
17	49	-8.3	6.7	0.83
18	52	-9.1	5.9	0.77
19	55	-11.4	3.6	0.56
20	58	-12.6	2.4	0.38

2. Будемо шукати функціональну залежність між $lg(\alpha_t-\alpha_э)$ і t у вигляді лінійної функції: $y(t)=lg(\alpha_t-\alpha_э)=at+b$.

Згідно методу найменших квадратів, маємо лінійну систему двох рівнянь з двома невідомими a і b:

$$\begin{cases} \sum\limits_{i=1}^{20} y_i t_i = 543.97 = a\sum\limits_{i=1}^{20} t_i^2 + b\sum\limits_{i=1}^{20} t_i = 23390\,a + 590b, \\ \\ \sum\limits_{i=1}^{20} y_i = 21.88 = 590a + 20b. \end{cases}$$

Розв'язуючи цю систему, знаходимо: $a= -0.017$ і $b=1.596$. Отже, $y(t)= -0.017t+1.596$ – рівняння шуканої апроксимуючої прямої, яка приведена на Рис. 7.

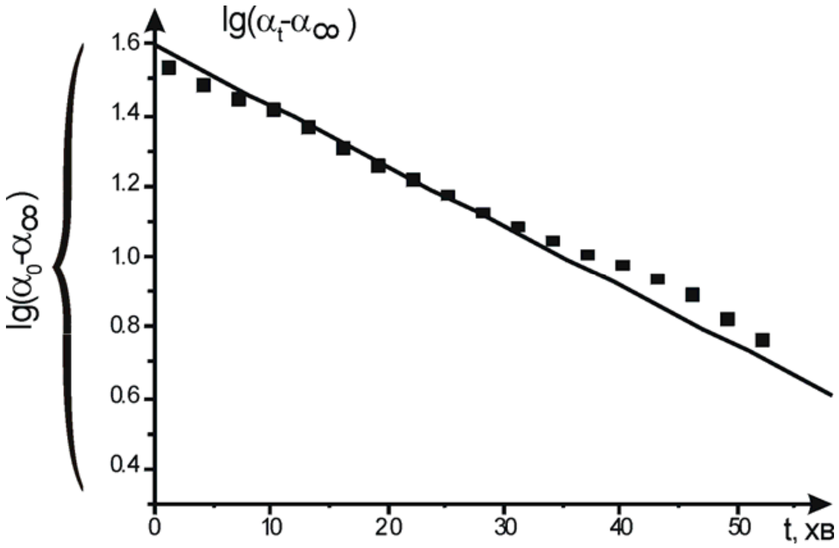


Рис. 7. Побудована залежність $y(t)=lg(\alpha_t-\alpha_э)$ в пакеті програми "Origin".

Нарешті, значення величини $lg(\alpha_0-\alpha_э)$ буде дорівнювати: $lg(\alpha_0-\alpha_э)= \lim\limits_{t\to 0} y(t)=b=$1.596. Звідси знаходимо, що $\alpha_0-\alpha_э$=39.4^0 і α_0=24.4^0.

****Зауваження.*** Усі однотипні розрахунки, у тому числі і числових характеристик вибіркових даних, зручно робити в пакеті програми “Excel” (див. наступний параграф).

3. Результати розрахунку константи швидкості реакції *k* при кімнатній температурі для кожного моменту часу *t* занесемо у таблицю:

Таблиця 5.

n	*t*, хв	*k*, $хв^{-1}$
1	1	0.139
2	4	0.058
3	7	0.021
4	10	0.041
5	13	0.041
6	16	0.041
7	19	0.040
8	22	0.039
9	25	0.039
10	28	0.038
11	31	0.038
12	34	0.037
13	37	0.036
14	40	0.036
15	43	0.035
16	46	0.035
17	49	0.036
18	52	0.037
19	55	0.044
20	58	0.048

Визначимо $k_c = \bar{k} = \frac{1}{20}\sum_{i=1}^{20} k_i = 0.041$ хв$^{-1}$ – середнє значення шуканої величини, потім $S = \sqrt{\frac{1}{19}\sum_{i=1}^{20}\left(k_c - k_i\right)^2}$ =0.023 хв$^{-1}$ – середньоквадратичну похибку методу, далі $S_{k_c} = \frac{S}{\sqrt{20}}$ =0.005 хв$^{-1}$ - середньоквадратичну похибку результату і $\Delta k_{вип} = t_{0.96,\,20} \cdot S_{k_c}$ =0.011 хв$^{-1}$ - довірчий інтервал вимірювання, який визначає випадкову складову похибки вимірювання фізичної величини ($t_{0.96,\,20}$=2.1, див. Таблицю 4). Оскільки випадкова складова перевищує систематичну складову $\Delta k_{сист} = \sqrt{\sum_{i=1}^{20}\left(\frac{\partial k}{\partial t_i}\right)^2 \cdot \Delta t_i^2} = \sqrt{\sum_{i=1}^{20}\left(\frac{k_i}{t_i}\right)^2} \cdot \Delta t$ =0.002 хв$^{-1}$ (тут $\Delta t = \Delta t_i = \Delta t_{сист}$ ≈0.017 хв - систематична складова похибки прямого вимірювання часу t), то у цьому випадку довірчий інтервал вимірювання (абсолютна похибка) визначається лише випадковою складовою. Отже, шукана величина становить: $k = k_c \pm \Delta k_{вип}$ = (0.041 ± 0.011) хв$^{-1}$.

Завдання 1. При обчисленні абсолютної похибки для константи швидкості реакції k врахувати систематичну складову похибки прямого вимірювання кута повороту площини поляризації α_t: $\Delta\alpha_{t\ сист}=1^0$.

Завдання 2. Вимірювальним приладом, який не має систематичних похибок, були зроблені незалежні вимірювання деякої величини. Знайти її істинне значення за результатами вимірювань: 8, 9, 11, 12.

§11. Комп'ютерне розв'язування задач біометрії з використанням програми "Excel".

В програмному забезпеченні "Excel" передбачена можливість вирішення багатьох важливих задач теорії ймовірності та математичної статистики. При цьому забезпечується висока точність обчислень, можливість роботи з великими об'ємами статистичних даних. Умовно відзначимо два рівні використання "Excel" у задачах біометрії:

1. використання вмонтованих в "Excel" спеціальних функцій по статистиці;
2. використання вмонтованого в "Excel" пакету "Статистичний аналіз".

В "Excel" існує 78 функцій для проведення статистичних розрахунків. Щоб познайомитися з набором цих функцій, необхідно у вікні "Excel" вибрати в меню опцію "Вставка", а в меню, що з'явилося, вибрати опцію "Функція". Далі у вікні меню, яке з'явилося, вибрати опцію "Статистичні" (див. Рис. 8). За допомогою клавіш прокрутки можна вибрати будь-яку з наведених функцій. Наприклад, на Рис. 8 вибрана функція СРЗНАЧ, яка обчислює вибіркове середнє. Опис функції можна отримати при натискуванні клавіші зі знаком "?" (Допомога).

У склад "Excel" входить пакет аналізу даних, який призначений для вирішення складних статистичних задач. Щоб ознайомитися з цим пакетом, необхідно у вікні "Excel" вибрати в меню опцію "Сервіс", а в меню, що з'явилося, вибрати опцію "Аналіз даних". За допомогою клавіші прокрутки можна вибрати будь-яку з наведених функцій аналізу. Наприклад, при виборі опції "Вибірка" (див. Рис. 9) на екрані з'явиться меню "Вибірка". За допомогою цієї функції аналізу здійснюється побудова простої випадкової вибірки заданого об'єму по генеральній сукупності.

Приклад 1. Розглянемо знаходження числових характеристик вибіркових даних для константи швидкості гідролітичного розщеплення сахарози k, наведених у Таблиці 5, за допомогою статистичних функцій.

Представимо дані для константи швидкості реакції у вигляді таблиці (Рис. 10). У відповідних комірках вводимо назви: “Вибіркове середнє” (C6-D6); “Вибіркова дисперсія” (C8-D8); “Вибіркове середньоквадратичне відхилення” (C10-D10; C11-D11).

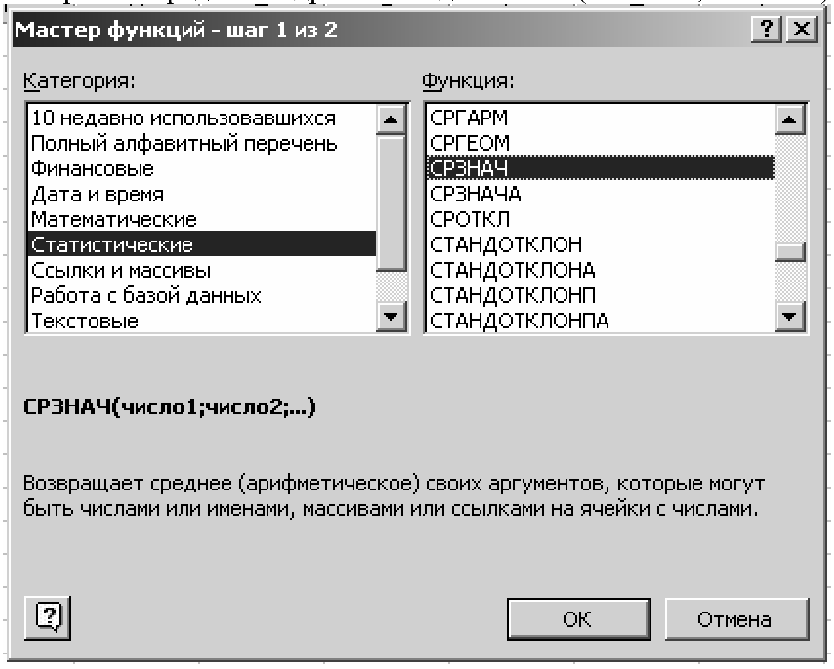


Рис. 8.

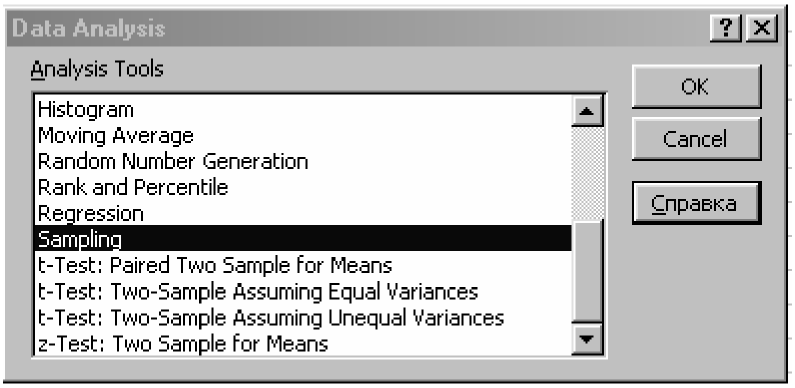
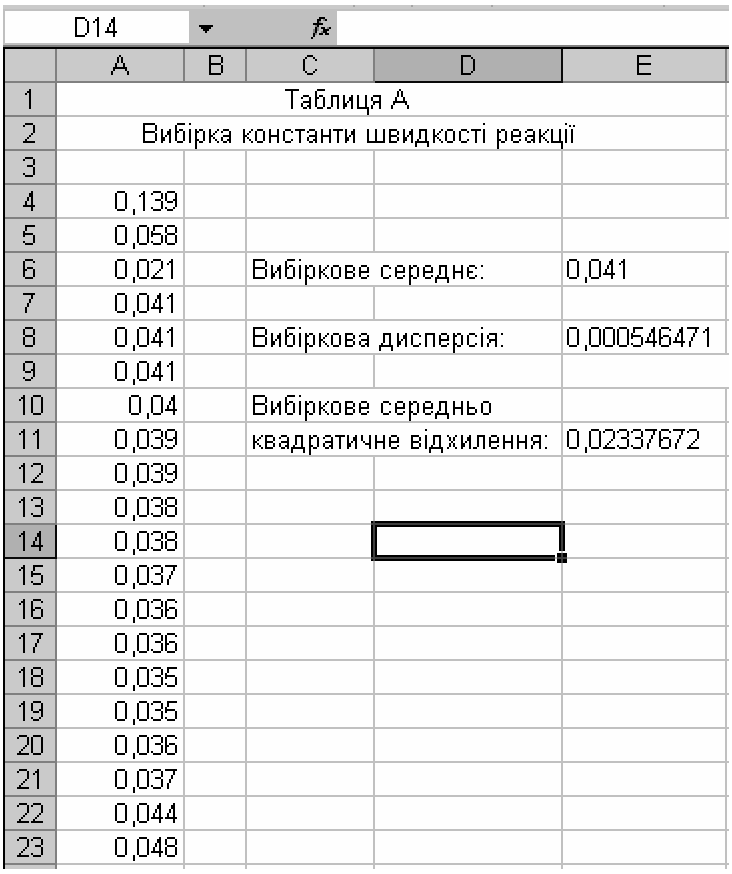

Data Analysis
Analysis Tools
Histogram
Moving Average
Random Number Generation
Rank and Percentile
Regression
Sampling
t-Test: Paired Two Sample for Means
t-Test: Two-Sample Assuming Equal Variances
t-Test: Two-Sample Assuming Unequal Variances
z-Test: Two Sample for Means
OK
Cancel
Справка

Рис. 9.



	A	B	C	D	E
1	Таблиця А				
2	Вибірка константи швидкості реакції				
3					
4	0,139				
5	0,058				
6	0,021		Вибіркове середнє:		0,041
7	0,041				
8	0,041		Вибіркова дисперсія:		0,000546471
9	0,041				
10	0,04		Вибіркове середньо		
11	0,039		квадратичне відхилення:		0,02337672
12	0,039				
13	0,038				
14	0,038				
15	0,037				
16	0,036				
17	0,036				
18	0,035				
19	0,035				
20	0,036				
21	0,037				
22	0,044				
23	0,048				

Рис. 10

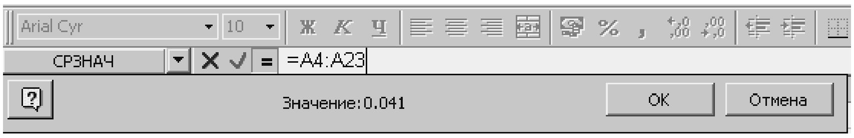


Рис. 11.

Обчислення вибіркового середнього $\bar{k} = \frac{1}{n}\sum_{i=1}^{n} k_i$ проводимо за допомогою функції СРЗНАЧ (число 1; число 2;…), де “число 1”, “число 2”, … - це від 1 до 30 числових аргументів, які відповідають вибірці з генеральної сукупності. Далі активізуємо комірку Е6 і вводимо функцію (див. Рис. 11): =(А4:А23). Після натискання клавіші ”Enter” результат обчислення автоматично заноситься в комірку Е6 (див. Рис. 10). Аналогічно проводимо розрахунки вибіркової дисперсії $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(k_i - \bar{k}\right)^2$ та вибіркового середньоквадратичного відхилення $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(k_i - \bar{k}\right)^2}$, використовуючи функції ДИСП(число 1; число 2;…) та СТАНДОТКЛ (число 1; число 2;…), відповідно. Результати обчислень автоматично заносяться у попередньо активізовані комірки Е8 та Е11 (див. Рис. 10).

<u>Приклад 2</u>. Вага деяких тварин, які проживають на різних територіях, підпорядкована нормальному закону із середнім значенням 5 кг і стандартним відхиленням 0.1. Знайти ймовірність того, що навмання взята тварина:

а) важить менше 4.8 кг;

б) важить більше 5.1 кг;

в) вага тварини лежить в межах від 4.8 до 5.1 кг.

У випадку нормального розподілу випадкової величини використовуємо функцію НОРМРАСП (х, середнє, стандартне_відхл, інтегральна), де х – це значення, для якого будується розподіл; “середнє” – це математичне сподівання *m* (середнє арифметичне розподілу); “стандартне_відхл” – це середньоквадратичне відхилення σ (стандартне відхилення розподілу); “інтегральна” – це логічне значення, яке визначає

	A	B	C	D
1		Задача а)	Задача б)	Задача в)
2	Вага тварини	менше 4.8 кг	більше 5.1 кг	в межах від 4.8 до 5.1 кг
3	Ймовірність			

Рис. 12.

форму функції: $f\left(x\right)=\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{(x-m)^2}{2\sigma^2}}$ (густина розподілу), якщо "інтегральна" = ЛОЖЬ і $F\left(x\right)=\frac{1}{\sqrt{2\pi\sigma}}\int\limits_{-\infty}^{x}e^{-\frac{(y-m)^2}{2\sigma^2}}dy$ (функція розподілу), якщо "інтегральна" =ИСТИНА. За умовою задачі "середнє"=$m=5$ і "стандартне_відхл"=σ=0.1. Будуємо таблицю (Рис. 12). Далі розглядаємо такі випадки:

а) для підрахунку ймовірності $P(X<4.8)$ події, що навмання взята тварина важить менше 4.8 кг, використаємо функцію НОРМРАСП за умови, що "інтегральна" =ІСТИНА. Для цього активізуємо комірку В3. Набираємо команди: "Вставка", "Функція", "Статистичні" і НОРМРАСП. У вікні меню функції НОРМРАСП, що з'явилося, набираємо параметри задачі (Рис. 13 і 14). Результат $P(X<4.8)=0.02275$ можна прочитати в меню функції. Після натиснення клавіші "ОК" він також автоматично заноситься у комірку В3 (див. Рис. 13).

б) Для підрахунку ймовірності $P(X>5.1)$ події, що навмання взята тварина важить більше 5.1 кг, використаємо співвідношення: $P(X>5.1)=1 - P(X<5.1)$. Для цього активізуємо комірку С3. Далі набираємо функцію (див. Рис. 15): =1 - НОРМРАСП(5.1;5;0.1;

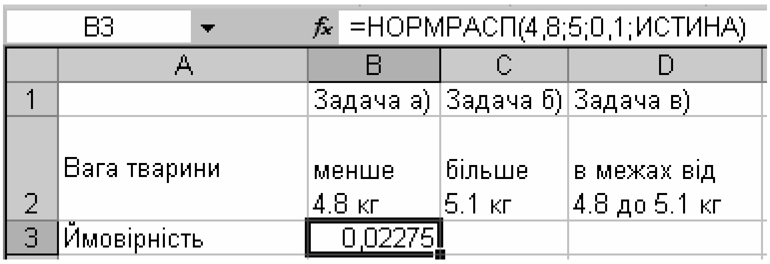


Рис. 13.

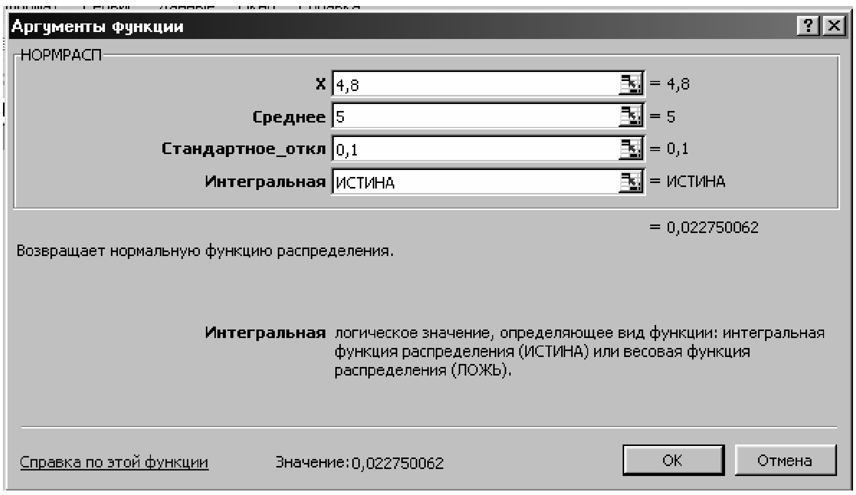


Рис. 14.

ИСТИНА). Після натиснення клавіші “Enter”, результат *P(X>5.1)=0.158655* автоматично заноситься у комірку C3 (Рис. 15).

в) Для підрахунку ймовірності $P(4.8<X<5.1)$ події, що вага навмання взятої тварини лежить в межах від 4.8 до 5.1 кг, використаємо співвідношення: $P(4.8<X<5.1) = P(X<5.1) - P(X<4.8)$. Для цього активізуємо комірку D3. Потім набираємо функцію

C3 =1-НОРМРАСП(5,1;5;0,1;ИСТИНА)

	A	B	C	D
1		Задача а)	Задача б)	Задача в)
2	Вага тварини	менше 4.8 кг	більше 5.1 кг	в межах від 4.8 до 5.1 кг
3	Ймовірність	0,02275	0,158655	

Рис. 15.

D3 =НОРМРАСП(5,1;5;0,1;ИСТИНА)-НОРМРАСП(4,8;5;0,1;ИСТИНА)

	A	B	C	D	E	F	G
1		Задача а)	Задача б)	Задача в)			
2	Вага тварини	менше 4.8 кг	більше 5.1 кг	в межах від 4.8 до 5.1 кг			
3	Ймовірність	0,02275	0,158655	0,818594678			

Рис. 16.

(див. Рис. 16): =НОРМРАСП(5.1;5;0.1;ИСТИНА) – НОРМРАСП (4.8;5;0.1;ИСТИНА). Після натиснення клавіші “Enter”, результат $P(4.8<X<5.1) = 0.8186$ автоматично заноситься у комірку D3 (Рис. 16).

Приклад 3. Є вибірка із 100 тварин. Випадкова величина – їх зріст, розподілена за нормальним законом. Середнє значення величини росту тварин складає 45.5 см, стандартне відхилення генеральної сукупності дорівнює 3.24 см. Визначити довірчий інтервал для середнього значення m росту тварин з заданою ймовірністю $P=0.95$.

Точність оцінки δ, яка пов’язана з визначенням довірчого інтервалу шуканої величини ($x=m \pm \delta$), знаходиться за допомогою функції ДОВЕРИТ (альфа; станд_відх; розмір), де “альфа” $=1 - P$ – це рівень значущості; “станд_відх” – це стандартне відхилення генеральної сукупності; “розмір” – це об’єм вибірки. За умовою задачі маємо: “альфа” =0.05; “станд_відх”=3.24 і “розмір”=100. Будуємо таблицю “Точність оцінки” і активізуємо комірку А2. Набираємо команди: “Вставка”, “Функція”, “Статистичні” і ДОВЕРИТ. У вікні меню функції ДОВЕРИТ, що з’явилося, набираємо числові значення її аргументів (Рис. 17). Результат δ=*0.635* можна прочитати у вікні меню цієї функції (див. Рис. 17). Після натиснення клавіші ”ОК” він також автоматично заноситься у комірку А2 (див. Рис. 18).

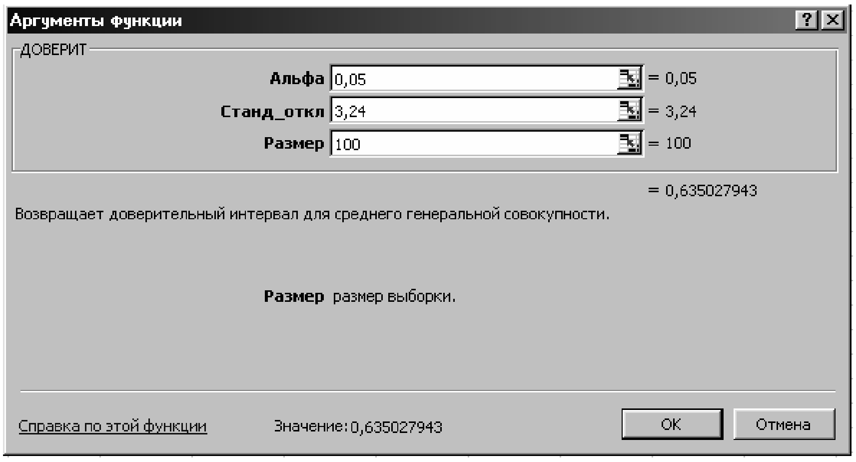


Рис. 17.

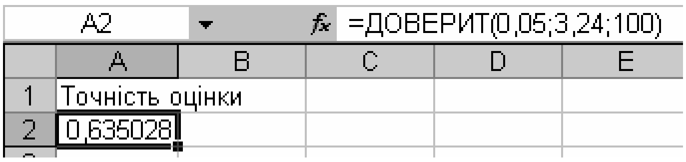


Рис. 18.

Завдання 1. Обчислити вибіркове середнє, вибіркову дисперсію та вибіркове середньоквадратичне відхилення для такої вибірки (відсотки виконання учбової програми до зимової сесії у вибірці з n=200 студентів):

Виконання програми, %	Кількість студентів, n
<10	2
10-20	3
20-30	10
30-40	15
40-50	25
50-60	55
70-80	90
80-90	150
90-100	50

Завдання 2. За допомогою функцій ФАКТР(число), ЧИСЛОКОМБ(число; число_вибраних) та ПЕРЕСТ(число; число_вибраних), де “число” – це ціле число, що задає кількість об’єктів n; “число_вибраних” – це ціле число, що задає кількість об’єктів у кожній комбінації або перестановці m, розрахувати відповідно 10!, C_{10}^{3} та A_{10}^{3}.

Завдання 3. Використовуючи функцію БИНОМРАСПР (число_успіхів; число_випробувань; ймовірність_успіху; інтегральна), де “число_успіхів” – це кількість успішних випробувань m; “число_випробувань” – це число незалежних випробувань n; “ймовірність_успіху” – це ймовірність успіху кожного випробування p; “інтегральна” – це логічне значення, що визначає форму функції: $P_n\left(m\right) = C_n^m p^m (1-p)^{n-m}$ (ймовірність події, яка полягає в тому, що, наступить рівно m успіхів в n

випробуваннях), якщо “інтегральна”=ЛОЖЬ і $P_n(k \le m) = \sum_{k_1=0}^{m} P_n(k_1)$ (ймовірність події, яка полягає в тому, що наступить не більше m успіхів в n випробуваннях), якщо “інтегральна”=ИСТИНА, розглянути таку задачу: Після багаторічних спостережень виявилося, що ймовірність зараження піддослідних тварин вірусом гепатиту складає 0.4. У припущенні, що в певний день випадковим чином відібрано 25 тварин, визначити ймовірність таких подій:

а) вірусом гепатиту будуть заражені рівно 10 тварин;

б) не більше 10 тварин;

в) не менше 10 тварин;

г) від 10 до 15 тварин.

Рекомендована література

1. В.Ю. Урбах. Статистический анализ в биологических и медицинских исследованиях.-М.: Высшая школа, 1975.
2. В.Е. Гмурман. Теория вероятностей и математическая статистика.-М.: Высшая школа, 1980.
3. Г.Ф. Лакин. Биометрия.-М.: Высшая школа, 1990.
4. А. Гончаров. Microsoft Excel 97 в примерах.-С.-Пб.: Питер, 1997.
5.
6. О.І. Конділенко, М.І. Міщенко. Похибки вимірювань фізичних величин: Методичні рекомендації до лабораторного практикуму з курсу загальної фізики.-Житомир: ЖІТІ, 2000.-46 с.
7. Ю.І. Прилуцький, О.В. Оглобля, К.І. Богуцька. Основи вищої математики. – К.: ВПЦ “Київський університет”, 2002. – 46 с.

ЗМІСТ

Навчальне видання

БІОМЕТРІЯ
методичні вказівки для студентів біологічного факультету

Упорядники: ПРИЛУЦЬКИЙ Юрій Іванович
ОГЛОБЛЯ Олександр Володимирович
СКЛЯРОВ Юрій Петрович

Редактор