МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ

ХАРКІВСЬКА НАЦІОНАЛЬНА АКАДЕМІЯ МІСЬКОГО ГОСПОДАРСТВА

МЕТОДИЧНІ ВКАЗІВКИ

ДО ПРАКТИЧНИХ ЗАНЯТЬ

З КУРСУ

ТЕОРІЯ ІМОВІРНОСТЕЙ І МАТЕМАТИЧНА СТАТИСТИКА

(*для слухачів другої вищої освіти ФПО та ЗН спеціальностей
7.06010107 «Теплогазопостачання і вентиляція»
7.06010101 «Промислове та цивільне будівництво»*)

ХАРКІВ
ХНАМГ
2012

Методичні вказівки до практичних занять з курсу «Теорія імовірностей і математична статистика» (для слухачів другої вищої освіти ФПО та ЗН спеціальностей 7.06010107 «Теплогазопостачання і вентиляція», 7.06010101 «Промислове та цивільне будівництво») / Харк. нац. акад. міськ. госп-ва; уклад.: В. М. Охріменко, Т. Б. Воронкова, О. О. Воронков. – Х.: ХНАМГ, 2012. - 30 с.

Укладачі: доц. В. М. Охріменко,
ст. викл. Т. Б. Воронкова,
ст. викл. О. О. Воронков

Рекомендовано кафедрою «Інформаційних систем і технологій в міському господарстві», протокол № 88 від 11.05.2012 р.

ЗАГАЛЬНІ ПОЛОЖЕННЯ

Курс «Теорія імовірностей і математична статистика» є нормативною дисципліною у навчальному плані перепідготовки за спеціальностями 7.06010107 «Теплогазопостачання і вентиляція» і 7.06010101 «Промислове та цивільне будівництво». Обсяг курсу становить 54 академічних години або 1,5 кредити ECTS, обсяг практичних занять становить 4 аудиторних годин (2 практичних заняття). Відповідно до програми курс розділений на два змістових модуля: «Теорія імовірностей» і «Математична статистика».

Метою вивчення дисципліни «Теорія імовірностей і математична статистика» є формування базових знань в області застосування імовірнісно-статистичного апарата, вивчення закономірностей у масових випадкових явищах, визначення їх імовірнісних характеристик з метою прогнозування.

В результаті вивчення курсу студенти повинні оволодіти основними методами визначення імовірнісних характеристик випадкових величин, статистичного опису результатів спостереження та перевірки статистичних гіпотез для прийняття на їх основі обґрунтованих рішень. Методичні вказівки спрямовані на допомогу студентам оволодіти практичними навичками із застосування імовірнісно-статистичного апарата.

«Теорія імовірностей і математична статистика» вивчає закономірності у випадкових явищах та є основою для побудови кількісних моделей випадкових процесів та явищ. Прикладами таких моделей є моделі планування та керування запасами, теорії ігор, теорії масового обслуговування. Статистичні показники аналізують при оцінці ризику, а також у багатьох сферах управління та прийняття рішень.

Широка сфера застосування теорії імовірностей і математичної статистики зумовлює важливе місце, що займає даний курс у підготовці фахівців вищої кваліфікації.

ПРАКТИЧНЕ ЗАНЯТТЯ 1

ВИЗНАЧЕННЯ ІМОВІРНОСТІ ВИПАДКОВОЇ ПОДІЇ. ЧИСЛОВІ ХАРАКТЕРИСТИКИ ВИПАДКОВОЇ ВЕЛИЧИНИ. НОРМАЛЬНИЙ ЗАКОН РОЗПОДІЛУ

Мета – сформувати вміння визначати імовірності випадкових подій з використанням теорем теорії імовірностей, на підставі універсального закону розподілу випадкової величини визначати імовірності її значень та числові характеристики, користуватися законами розподілу випадкових величин для визначення імовірностей влучення випадкової величини на певний інтервал її значень.

ЗАГАЛЬНІ ВІДОМОСТІ

Випадкова подія – це будь-який факт, що в наслідку досліду може відбутися або не відбутися.

Імовірністю випадкової події називають числову міру ступеню об'єктивної можливості появи цієї події в наслідку досліду.

Імовірність випадкової події можна визначити за класичним методом, якщо наслідки досліду *утворюють повну групу*, *є рівноможливими* та *несумісними*. Імовірність події А визначається як відношення числа можливих наслідків досліду, яки сприяють появі події А, до загального числа можливих наслідків досліду:

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

де n – загальне число можливих наслідків досліду; m – число наслідків досліду, яки сприяють появі події А.

Сумою двох подій А і В називають подію С, яка полягає у появі події А або події В або обох подій разом: С = А + В.

Добутком двох подій А і В називають подію С, що полягає у спільній появі подій А і В: С = А * В.

Протилежними називають дві несумісних події А і $\bar{A}$, якщо вони складають повну групу.

Подію А називають *незалежною* від події В, якщо імовірність події А не зміниться від того, відбулася подія В чи ні. Якщо ж імовірність події А залежить від того, відбулася подія В чи ні, то такі події називають *залежними*.

Імовірність події А, обчислену за умови, що подія В відбулася, називають *умовною імовірністю* події А.

Теорема додавання. Імовірність суми двох несумісних подій А і В дорівнює сумі імовірностей цих подій, тобто

$$P(A + B) = P(A) + P(B).$$

Слідства теореми додавання:

1. Якщо події $A_1, A_2, \ldots, A_i, \ldots, A_n$ утворюють повну групу несумісних подій, сума їх імовірностей дорівнює 1:

$$P(\Sigma A_i) = \Sigma P(A_i) = 1.$$

2. Сума імовірностей двох протилежних подій дорівнює одиниці:

$$P(A) + P(A) = 1,$$

або

$$P(A) = 1 - P(A).$$

Якщо дві події є сумісними, імовірність їх суми дорівнює сумі імовірностей цих подій мінус імовірність їх спільної появи:

$$P(A + B) = P(A) + P(B) - P(A*B).$$

Імовірність суми n сумісних подій

$$P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i} P(A_i) - \sum_{i,j} P(A_i A_j) + \sum_{i,j,k} P(A_i A_j A_k) - \ldots + (-1)^{n-1} P(A_1 A_2 \ldots A_n),$$

де суми поширюються на всі можливі комбінації індексів i, j, k,..., узятих по одному, по два, по три і т.д.

Теорема множення. Імовірність добутку двох подій A і B дорівнює добутку імовірності одного з них на умовну імовірність іншого, обчислену за умови, що перша відбулася:

$$P(A * B) = P(A) * P(B|A).$$

Для імовірності добутку n подій формула має вигляд

$$P(A_1 A_2 \ldots A_n) = P(A_1) P(A_2 | A_1) P(A_3 | A_1 A_2) \ldots P(A_n | A_1 A_2 \ldots A_{n-1})$$

Якщо події A і B незалежні, то умовна імовірність події B дорівнює безумовній імовірності цієї події,

$$P(B|A) = P(B).$$

Імовірність добутку двох незалежних подій дорівнює добутку імовірностей цих подій:

$$P(A * B) = P(A) * P(B).$$

Якщо маємо n незалежних подій:

$$P\left(\prod_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i).$$

Формула повної імовірності. Повна безумовна імовірність події A з урахуванням випадковості умов протікання досліду, дорівнює сумі добутків імовірностей кожної з гіпотез на умовну імовірність події A при кожній з гіпотез.

$$P(A) = \sum_{i=1}^{N} P(H_i A) = \sum_{i=1}^{N} P(H_i) P(A / H_i).$$

Формула Бейеса (теорема гіпотез). Ця формула дозволяє переоцінити імовірності гіпотез після того, як стає відомим наслідок досліду, в результаті якого відбулася подія A. За відомими до проведення досліду (апріорними) імовірностями гіпотез $P(H_i)$ та за результатом досліду (настання події A) обчислюють післядослідні (апостеріорні) імовірності гіпотез $P(H_i|A)$.

$$P(H_i / A) = \frac{P(H_i) * P(A / H_i)}{\sum P(H_i) * P(A / H_i)}.$$

Формула Бернуллі (повторні незалежні випробування). Імовірність того, що в результаті певного числа дослідів подія A з'явиться рівно m разів

$$P_n(m) = C_n^m p^m q^{n-m}.$$

де $P_n(m)$ - імовірність того, що в n випробуваннях подія A з'явиться рівно m разів; C_n^m - число сполучень з n елементів по m; p – імовірність появи події A в одному досліді; q = 1 – p – імовірність не появи події A в одному досліді.

Локальна теорема Лапласа: Якщо імовірність p появи події A в кожному випробуванні постійна і відмінна від нуля та одиниці, імовірність $P_n(m)$ того, що подія A з'явиться в n дослідах рівно m разів, приблизно дорівнює, і тім точніше, чим більше n, значенню функції

$$P_n(m) = \frac{1}{\sqrt{npq}}\varphi(x),$$

де $x = \frac{m - np}{\sqrt{npq}}$, а значення функції $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ визначаються за довідковими таблицями. Функція ((x) парна, тобто φ(-x) = φ(x).

Інтегральна теорема Лапласа: Якщо імовірність p настання події A в кожному випробуванні постійна і відмінна від нуля та одиниці, приблизно імовірність $P_n(m_1,m_2)$, того, що подія A з'явиться у випробуваннях від m_1 до m_2 разів,

$$P_n(m_1, m_2) = \frac{1}{\sqrt{2\pi}}\int_{x'}^{x''} e^{-x^2/2}dx$$

де $x' = \frac{m_1 - np}{\sqrt{npq}}$; $x'' = \frac{m_2 - np}{\sqrt{npq}}$.

Формула Пуассона. Якщо число незалежних випробувань n велике, але значення добутку np залишається невеликим, імовірність того, що в цих випробуваннях подія A відбудеться m разів, можна визначити за формулою:

$$P_n(m) = \frac{(np)^m}{m!}e^{-np}.$$

Випадковою називають величину, яка в результаті досліду може прийняти те або інше значення. *Дискретною* називають випадкову величину, число значень якої скінченне, або нескінченне, але рахункове (яка може приймати тільки окремі значення). *Безперервною* називають випадкову величину, число значень якої нескінченне навіть на невеликому інтервалі.

Законом розподілу випадкової величини називають будь-яке правило, що дозволяє будь-якому значенню випадкової величини поставити у відповідність його імовірність.

Ряд розподілу - це таблиця, у верхньому рядку якої перелічені всі значення випадкової величини $x_1, x_2, \ldots, x_n$ в порядку зростання, а у нижньому - імовірності появи цих значень $p_1, p_2, \ldots, p_n$:

x_i	x_1	x_2	…	x_n
p_i	p_1	p_2	…	p_n

де $p_i = P\{X=x_i\}$.

Оскільки події $\{X=x_1\}$, $\{X=x_2\}$,…,$\{X=x_n\}$несумісні та утворюють повну групу, сума їх імовірностей дорівнює одиниці $\sum p_i = 1$.

Найбільш загальною формою закону розподілу для всіх випадкових величин (дискретних та безперервних) є функція розподілу.

Функція розподілу випадкової величини X - це імовірність того, що випадкова величина X прийме значення, менше за x:

$$F(x) = P\{X \le x\}.$$

Функція розподілу має наступні властивості.

1. Значення функції розподілу належать відрізкові [0; 1]: $0 \le F(x_2) \le 1$;

2. Функція розподілу - неубутна функція, тобто $F(x_2) \ge F(x_1)$, якщо $x_2 > x_1$.

3. Імовірність того, що випадкова величина X прийме значення, укладене в інтервалі (x_1, x_2), дорівнює приросту функції розподілу на цьому інтервалі:

$$P\{ x_1 \le X \le x_2\} = F(x_2) - F(x_1).$$

4. Якщо всі можливі значення випадкової величини X належать інтервалу $(-\infty, +\infty)$, то при мінус нескінченності функція розподілу дорівнює нулю, а при плюс нескінченності - одиниці, тобто $F(-\infty) = 0$; $F(+\infty) = 1$.

Щільністю розподілу випадкової величини X у точці x називається похідна функції розподілу X у цій точці (передбачається, що F(x) безперервна і диференційована):

$$f(x) = \frac{dF(x)}{dx}.$$

Властивості щільності розподілу імовірностей: 1. Щільність розподілу є невід'ємною, тобто $f(x) \ge 0$ як похідна неубутної функції; 2. Функція розподілу визначається за співвідношенням:

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

3. Інтеграл від щільності розподілу у нескінченних межах дорівнює одиниці:

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

де f(x)dx - елемент імовірності, тобто імовірність влучення випадкової величини X на елементарну ділянку dx.

4. Імовірність влучення безперервної випадкової величини на інтервал (x_1, x_2) дорівнює інтегралу щільності розподілу в межах від x_1 до x_2.

$$P\{x_1 \le X \le x_2\} = \int_{x_1}^{x_2} f(x)dx.$$

Математичним сподіванням випадкової величини X називають суму добутків всіх можливих її значень на імовірності цих значень

$$M[X] = \sum_{i=1}^{n} x_i p_i.$$

Для безперервної випадкової величини математичне сподівання

$$m_x = \int_{-\infty}^{+\infty} xf(x)dx.$$

Математичне сподівання характеризує середнє значення випадкової величини.

Другий початковий момент α_2:

$$\alpha_2 = M[X^2]$$

Під центрованою випадковою величиною розуміють її відхилення від математичного сподівання:

$$\overset{0}{X} = X - m_x .$$

Дисперсія випадкової величини для дискретної X:

$$D_x = \sum_{i=1}^{n} \overset{0}{x}_i^{\,2} p_i = \sum_{i=1}^{n} (x_i - m_x)^2 p_i .$$

Для безперервної випадкової величини:

$$D_x = \int_{-\infty}^{+\infty} \overset{0}{x}^{\,2} f(x)dx = \int_{-\infty}^{+\infty} (x - m_x)^2 f(x)dx .$$

Дисперсія випадкової величини є характеристикою розсіювання цієї величини навколо математичного сподівання.

Середнє квадратичне відхилення X:

$$\sigma_x = \sqrt{D_x}.$$

Експонентний закон розподілу. Функцію розподілу T обчислюють за формулою: $F(t) = P\{T < t\} = 1 - e^{-\lambda t}$.

Щільність розподілу T як похідна функції розподілу F(t) має вигляд:

$$f(t) = d(t)/dt = \lambda e^{-\lambda t} .$$

Математичне сподівання випадкової величини, розподіленої за експонентним законом, зворотно параметру розподілу λ: $m_x = 1/\lambda$.

Дисперсія: $D_t = \dfrac{1}{\lambda^2}$, середнє квадратичне відхилення: $\sigma_x = 1/\lambda$.

Імовірність влучення випадкової величини, що має експонентний розподіл, в інтервал значень (α, β): $P\{\alpha \le t \le \beta\} = e^{-\lambda\alpha} - e^{-\lambda\beta}$.

Нормальний закон розподілу імовірностей визначається двома параметрами m_x і σ_x.

Імовірність влучення випадкової величини X на ділянку значень (α, β) виражається через функцію Лапласа формулою:

$$P\{\alpha < X < \beta\} = \Phi\left(\frac{\beta - m_x}{\sigma}\right) - \Phi\left(\frac{\alpha - m_x}{\sigma}\right).$$

Функція розподілу для випадкової величини, розподіленої нормально визначається за формулою

$$F(x) = \Phi\left(\frac{x - m_x}{\sigma_x}\right) + 0{,}5 .$$

Імовірність влучення нормально розподіленої випадкової величини в ділянку значень, симетричну щодо її математичного сподівання обчислюють за формулою:

$$P\{|x - m_x| < l\} = 2\Phi\left(\frac{l}{\sigma_x}\right).$$

Вказівки до виконання завдання

Після уважного вивчення умови завдання треба сформулювати подію, імовірність якої треба визначити. Потім обґрунтувати, які теореми теорії імовірностей (теорему додавання або теорему множення) або формули (формулу повної імовірності, формулу Бейєса та ін.) треба застосувати для розв’язання задачі. У процесі розв’язання задачі треба чітко висловлювати міркування.

Задача 1.1

Партія виробів містить N виробів, з яких M є дефектними. Навмання з цієї партії вибирають k виробів для контролю. Визначити імовірність того, що серед них буде рівно *l* дефектних виробів.

Розв'язання

Запишемо подію, для якої необхідно визначити імовірність

A={у контрольній партії рівно *l* дефектних виробів}.

Загальна кількість можливих наслідків досліду дорівнює $n = C_N^k$. Кількість наслідків досліду, які сприяють появі події A={у контрольній партії рівно *l* дефектних виробів} визначимо в такий спосіб. Кількість випадків, які сприяють появі *l* дефектних виробів

$$m_д = C_m^l .$$

Кількість випадків, які сприяють появі k-l не дефектних виробів

$$m_г = C_{N-M}^{k-l} .$$

Імовірність події A

$$P(A) = \frac{m_д * m_г}{n} = \frac{C_m^l * C_{N-M}^{k-l}}{C_N^k} .$$

Задача 1.2

Два стрільці роблять по одному пострілу у мішень. Для першого з них імовірність влучення дорівнює 0,8, а для другого - 0,4. Мішень пробита один раз (одне влучення). Знайти імовірність того, що мішень уражена першим стрільцем.

Розв'язання

Є факт, тобто подія A={мішень уражена один раз}, тобто один із стрільців промахнувся. Висуваємо гіпотези: H_1={мішень уражена першим стрільцем}; H_2={мішень уражена другим стрільцем}. Визначимо імовірності гіпотез. Мішень уражена першим стрільцем, якщо він при пострілі потрапив у мішень, а другий стрілець промахнувся, тоді $P(H_1) = 0{,}8*(1-0{,}4) = 0{,}48$. Мішень уражена другим стрільцем, якщо він при пострілі потрапив у мішень, а перший стрілець промахнувся, тоді $P(H_2) = (1-0{,}8) *0{,}4= 0{,}08$. Умовна імовірність події A, за умови, що має місце гіпотеза H_1 дорівнює $P(A|H_1)=1$, і за умови, що має місце гіпотеза H_2 дорівнює $P(A|H_1)=1$, тому що в цих випадках мішень буде напевно уражена один раз. Скористуємося теоремою гіпотез і визначимо імовірність реалізації гіпотези H_1:

$$P(H_i / A) = \frac{P(H_i) * P(A / H_i)}{\sum P(H_i) * P(A / H_i)} = \frac{0{,}48 * 1}{0{,}48 * 1 + 0{,}08 * 1} = 0{,}884 .$$

Задача 1.3

Оптова база постачає товари у 10 магазинів, від кожного з яких може надійти заявка на черговий день з імовірністю 0,4, незалежно від заявок інших магазинів. Знайти найімовірніше число заявок у день та імовірність одержання цього числа заявок.

Розв'язання

У цьому випадку n = 10, p=0,4. Добуток np дорівнює цілому числу np = 10*0,4 = 4. Імовірність того, що від магазинів буде отримано 4 заявки, можна визначити за формулою Бернуллі:

$$P_n(m) = C_n^m p^m q^{n-m},$$

де m=4, q=1-0,4=0,6. Тоді імовірність того, що від магазинів буде отримано 4 заявки:

$$P_{10}(4) = C_{10}^4 0{,}4^4 0{,}6^{10-4} = 0{,}251$$

Задача 1.4

Визначимо імовірність того, що деяка подія А в 150 дослідах з'явиться рівно 12 разів. Імовірність появи події А в одному досліді дорівнює 0,1.

Розв'язання

У цьому випадку n = 150, p=0,1. Тоді q=1-0,1=0,9, m=12. Для визначення імовірності, скористаємося локальною теоремою Лапласа:

$$P_n(m) = \frac{1}{\sqrt{npq}}\varphi(x).$$

Обчислимо $x = \frac{m - np}{\sqrt{npq}} = \frac{12 - 150 * 0{,}1}{\sqrt{150 * 0{,}1 * 0{,}9}} = -0{,}82$, значення функції визначимо за довідковою таблицею φ(-0,82) =0,2939, тоді:

$$P_{150}(12) = \frac{1}{\sqrt{150 * 0{,}1 * 0{,}9}} * 0{,}2939 = 0{,}08.$$

Визначимо імовірність того, що подія А в 150 дослідах з'явиться не менше за 10 і не більше за 20 разів. Для цього скористаємося інтегральною теоремою Лапласа.

Знайдемо $x' = \frac{m_1 - np}{\sqrt{npq}} = \frac{10 - 150 * 0{,}1}{\sqrt{150 * 0{,}1 * 0{,}9}} = -1{,}36$ та $x'' = \frac{m_2 - np}{\sqrt{npq}} = \frac{14 - 150 * 0{,}1}{\sqrt{150 * 0{,}1 * 0{,}9}} = -0{,}27$.

$$P_n(m_1,m_2) = \Phi(x'')-\Phi(x') = 0{,}4131 – 0{,}1064 = 0{,}307.$$

Задача 1.5

Імовірність виготовлення нестандартної деталі дорівнює 0,004. Знайти ймовірність того, що серед 1000 деталей буде 5 нестандартних.

Розв'язання

Формалізуємо задачу: n = 1000, p = 0,004, a = np = 1000 * 0,004 = 4. Для знаходження ймовірності події $P_{1000}(5)$ уживемо формулу Пуассона:

$$P_{1000}(5) = \frac{(4)^5}{5!}e^{-4} = 0{,}1563.$$

Задача 1.6

Тричі кидають монету. Випадкова величина X - число появ герба. Побудувати ряд розподілу випадкової величини X. Визначити функцію розподілу випадкової величини X та побудувати її графік.

Розв'язання

Побудуємо ряд розподілу X. Очевидно, що число появ герба при трира-зовому киданні монети може приймати чотири значення 0, 1, 2, 3. Для визначення імовірностей цих значень скористаємося формулою Бернуллі. Число дослідів n=3, імовірність появи герба в одному досліді p=0,5, імовірність непояви герба в одному досліді q=1-p=1-0,5=0,5. Отже, ряд розподілу випадкової величини X має вигляд:

x_i	0	1	2	3
p_i	1/8	3/8	3/8	1/8

Виконаємо перевірку: $\sum p_i$ = 1/8+3/8+3/8+1/8=1.

За визначенням функція розподілу випадкової величини X - це імовірність того, що випадкова величина X прийме значення, менше x: $F(x) = P\{X \leq x\}$.

Обчислимо значення функції розподілу:

F(0) = P{X (0} = 0;
F(1) = P{X (1} = 1/8;
F(2) = P{X (2} = 1/8 + 3/8 = 4/8;
F(3) = P{X (3} = 4/8 + 3/8 = 7/8;
при X(3 F(x) = 1.

Побудуємо графік F(x) (рис. 1.1)

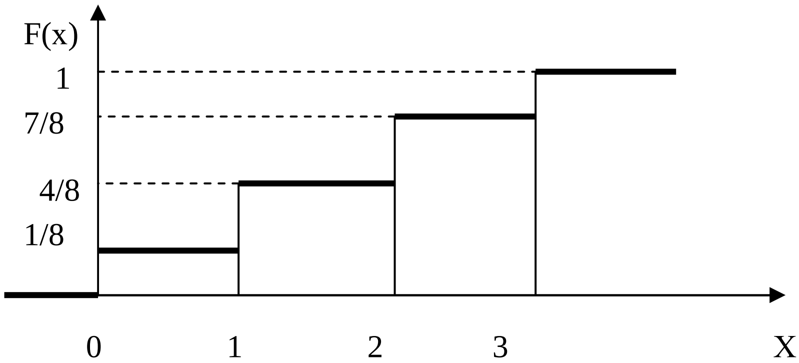

Рис. 1.1 - Графік функції розподілу

Задача 1.7

Визначимо числові характеристики дискретної випадкової величини для умов прикладу задачі 2.1. Маємо ряд розподілу:

x_i	0	1	2	3
p_i	1/8	3/8	3/8	1/8

Розв'язання

Визначимо математичне сподівання випадкової величини X:

$$m_x = \Sigma x_i * p_i = 0*1/8 + 1*3/8 + 2*3/8 + 3*1/8 = 1{,}5.$$

Дисперсію визначимо за двома способами:

за формулою другого центрального моменту:

$$D_x = \Sigma(x_i-m_x)^2 * p_i = (0-1{,}5)^2*1/8 + (1-1{,}5)^2*3/8 + (2-1{,}5)^2*3/8 + (3-1{,}5)^2*1/8 = 0{,}75.$$

та за формулою, що містить другий початковий момент α_2:

$$D_x = \alpha_2 - m_x^2$$

$$\alpha_2 = \Sigma x_i^2 * p_i = 0^2*1/8 + 1^2*3/8 + 2^2*3/8 + 3^2*1/8 = 3.$$

$$D_x = \alpha_2 - m_x^2 = 3-1,5^2 = 0,75.$$

Визначимо середнє квадратичне відхилення випадкової величини X:

$$\sigma_x = \sqrt{D_x} = \sqrt{0,75} = 0,855.$$

Задача 1.8

Є випадкова величина X з експонентним законом розподілу. Параметр розподілу λ=0,4. Визначити числові характеристики та функцію розподілу випадкової величини X, а також імовірність того, що вона прийме значення в інтервалі (6, 10).

Розв'язання

Числові характеристики випадкової величини X визначимо за формулами:

$$m_x = 1/\lambda=1/0,4 =2,5;\ D_x = 1/\lambda^2=1/(0,4)^2=6,25;\ \sigma_x = m_x = 2,5.$$

Щільність розподілу :

$$f(x) = \begin{cases} 0, & \text{при} \quad x < 0 \\ 0,4e^{-0,4x}, & \text{при} \quad x \geq 0 \end{cases}$$

Функція розподілу:

$$F(x) = \begin{cases} 0, & \text{при} \quad x < 0 \\ 1 - e^{-0,4x}, & \text{при} \quad x \geq 0 \end{cases}$$

Визначимо імовірність того, що випадкова величина X прийме значення в інтервалі (6, 10):

$$P\{6 \leq X \leq 10\} = e^{-\lambda\alpha} - e^{-\lambda\beta} = e^{-0,4*6} - e^{-0,4*10} = 0,0907 - 0,0183 = 0,0724.$$

Задача 1.9

Відомі імовірнісні характеристики нормально розподіленої випадкової величини X: m =17; σ=0,6. Знайти імовірність події $P(\alpha<X<\beta)$; імовірність того, що $P(|x-m| < \delta)$, якщо α =16,8; β = 17,2; δ=0,3.

Розв'язання

Обчислимо імовірність, що X належить інтервалу (16,8; 17,2).

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - m}{\sigma}\right) - \Phi\left(\frac{\alpha - m}{\sigma}\right) = \Phi\left(\frac{17,2-17}{0,6}\right) - \Phi\left(\frac{16,8-17}{0,6}\right) = \Phi\left(\frac{1}{3}\right) - \Phi\left(-\frac{1}{3}\right) = 2*\Phi\left(\frac{1}{3}\right) = 0,26.$$

Визначимо імовірність того, що X відхилиться від свого середнього значення m менше чим на δ:

$$P(|x-17| < 0,3) = 2*\Phi\left(\frac{0,3}{0,6}\right) = 0,38.$$

Практичне заняття 2

ВАРІАЦІЙНИЙ РЯД. ВЛАСТИВОСТІ ВИБІРКОВИХ ЧИСЛОВИХ ХАРАКТЕРИСТИК. ДОВІРЧИЙ ІНТЕРВАЛ І ДОВІРЧА ІМОВІРНІСТЬ. ЕЛЕМЕНТИ ТЕОРІЇ КОРЕЛЯЦІЇ.

Мета - сформувати вміння побудувати закон розподілу випадкової величини на підставі статистичних даних та визначати статистичні оцінки параметрів розподілу, оцінювати їх властивості та визначати можливі похибки та їх імовірності, а також щонайменшу кількість дослідів для їх запобігання, на підставі статистичних даних будувати рівняння регресії та визначати його параметри за методом найменших квадратів.

ЗАГАЛЬНІ ВІДОМОСТІ

Варіаційним рядом називають таблицю, в одному рядку якої розташовуються варіанти $x_1, x_2, \ldots, x_n$ у зростаючому або убутному порядку, а в другому — відповідні їм частоти $m_1, m_2, \ldots, m_n$...

x_i	x_1	x_2	…	x_n
m_i	m_1	m_2	…	m_n

де $\sum_{i=1}^{n} m_i = n$.

Дискретною називають варіацію, при якій окремі значення ознаки (варіанти) відрізняються одна від одної на певну скінчену величину.

Безперервною називають варіацію, при якій значення ознаки можуть відрізнятися одне від одного на як завгодно малу величину.

Замість абсолютних частот m_i зазвичай використовують відносні – p^*_i. Для одержання відносних частот необхідно відповідну частоту розділити на суму всіх частот: $p_1^* = \dfrac{m_1}{\sum_{i=1}^{n} m_i}$, $p_2^* = \dfrac{m_2}{\sum_{i=1}^{n} m_i}$, …, $p_n^* = \dfrac{m_n}{\sum_{i=1}^{n} m_i}$. Сума всіх відносних частот дорівнює одиниці: $\sum_{i=1}^{n} p_i^* = 1$.

Полігон розподілу будують у прямокутній системі координат. Величину ознаки відкладають на осі абсцис, частоти або відносні частоти - на осі ординат.

Кумулятивна крива (кумулята) утворюється при зображенні варіаційного ряду з накопиченими відносними частотами у прямокутній системі координат. Накопичена частота певної варіанти утворюється підсумовуванням всіх частот варіант, які передують даній, із частотою цієї варіанти. При $\Delta x_i \to 0$ кумулята прагне до безперервної кривої, яка є статистичною функцію розподілу досліджуваної ознаки.

Гістограму розподілу будують аналогічно полігону у прямокутній системі координат. При побудові гістограми на осі абсцис вибирають відрізки, що відповідають інтервалам, на яких будують прямокутники із площею, пропорційною відносним частотам інтервалів p_i^*. З умови побудови гістограми випливає, що вся її площа дорівнює 1. При $\Delta x_i \to 0$ гістограма прагне до безперервної залежності, що є статистичною щільністю розподілу випадкової величини.

Однією з задач математичної статистики є отримання оцінок числових характеристик досліджуваної ознаки. Іншою задачею є визначення закону розподілу досліджуваної ознаки за статистичним даними. При цьому виникає задача із згладжування статистичних рядів за допомогою аналітичних виразів.

Щоб знайти статистичну функцію розподілу F*(x), підраховують число варіант, в яких ознака X прийняла значення менші за x, тобто X<x. Якщо число таких варіант m(x), а обсяг сукупності дорівнює n, то $F^*(x) = \frac{m(x)}{n}$.

Якщо результати спостереження зведені у групований статистичний ряд, на його підставі будують гістограму, з урахуванням того, що частота появ ознаки на i-му інтервалі $p^* = \frac{m_i}{n}$, де m_i – кількість появ x на i-му інтервалі.

Для оформлення статистичного ряду у вигляді гістограми на осі абсцис відкладають інтервали ряду, а потім на кожному інтервалі будують прямокутник, площа якого дорівнює p^*_i.

Задача вирівнювання статистичних рядів полягає у знаходженні теоретичної кривої розподілу. Якщо клас функцій, що описують розподіл, відомий, то задача зводиться до раціонального вибору параметрів розподілу. Одним з методів розв’язання цієї задачі є метод моментів, відповідно до якого числові параметри розподілу вибирають так, щоб найважливіші числові характеристики дорівнювали їх статистичним оцінкам. Зокрема, математичне сподівання та дисперсія приймаються рівними їх статистичним оцінкам - вибірковій середній та вибірковій дисперсії: $m_x = \tilde{X}$, $D_x = \sigma^2_{виб}$. Інший метод визначення статистичних оцінок параметрів розподілу є метод максимальної правдоподібності.

Числові характеристики варіаційного ряду. Однією з найважливіших характеристик варіаційного ряду є його *середнє значення*, яку визначають як відношення суми добутків варіантів на відповідні частоти до суми всіх частот:

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i m_i}{\sum_{i=1}^{k} m_i}, \qquad \overline{x} = \sum_{i=1}^{k} x_i p_i^*, \qquad \text{або} \qquad \overline{x} = \frac{\sum_{i=1}^{n} x_i}{n},$$

де m_i – частоти варіаційного ряду; p_i* - відносні частоти; k – число груп.

Дисперсія варіаційного ряду – це середня арифметична квадрата відхилення значень ознак ряду від їх середньої арифметичної. Дисперсія обчислюється за формулами:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{n} \qquad \sigma^2 = \frac{\sum_{i=1}^{k}(x_i - \overline{X})^2 m_i}{\sum_{i=1}^{k} m_i}$$

Стандартне відхилення варіаційного ряду визначають як арифметичне значення квадратного кореня з дисперсії.

$$\sigma = \sqrt{\sigma^2} \ .$$

Оцінка параметра a^* є *спроможною*, якщо при $n\to\infty$ вона збігається за імовірністю до оцінюваного параметра a: $\lim\limits_{n\to\infty} a^* = a$.

Оцінка параметра a^* є *незміщеною* (тобто не містить систематичної помилки), якщо її математичне сподівання дорівнює оцінюваному параметру a: $M[a^*] = a$.

Оцінка параметра a^* є *ефективною*, якщо при заданому обсязі вибірки вона має найменшу дисперсію. Ступінь ефективності оцінюють відношенням дисперсій: $F = \frac{\sigma_1^2}{\sigma_2^2}$, тобто якщо F>1, то σ^2_2 ефективніша, і навпаки.

Математичне сподівання вибіркової середньої не залежить від числа дослідів n і дорівнює генеральній середній: $M[\tilde{X}] = \bar{X}$.

Дисперсія вибіркової середньої: $\sigma^2[\tilde{X}] = \frac{\sigma^2_{ген}}{n}$.

Середнє квадратичне відхилення вибіркової середньої: $\sigma[\tilde{X}] = \frac{\sigma_{ген}}{\sqrt{n}}$.

Отже, вибіркова середня є незміщеною оцінкою генеральної середньої.

Незміщену оцінку дисперсії позначають S^2 та визначають за формулою:

$$S^2 = \frac{n}{n-1}\sigma^2_{выб} = \frac{\sum\left(x_i - \tilde{X}\right)^2}{n-1}.$$

Оцінку кореляційного моменту визначають за формулою:

$$K_{хувув} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1}.$$

а оцінку коефіцієнта кореляції - за формулою

$$r_{хувув} = \frac{K_{хувув}}{\sigma_{выб_x}\sigma_{выб_y}}.$$

Приведені оцінки є спроможними та незміщеними.

Довірчий інтервал і довірча імовірність. Якщо точкову оцінку параметра визначено на підставі вибірки малого обсягу, вона може значно відрізнятися від оцінюваного параметра. Для визначення помилки від заміни генерального параметра його оцінкою використовують поняття довірчого інтервалу та довірчої імовірності.

При визначенні дійсного значення шуканого параметра m_x величину погрішності характеризує довірчий інтервал L, а ступінь впевненості, що похибка не перевищить L, характеризує довірча імовірність β.

Якщо для певного параметру розподілу, наприклад, математичного сподівання m_x отримано спроможну та незміщену оцінку a^*, потрібно знати, до яких помилок може призвести заміна параметра m_x його точковою оцінкою, і з яким ступенем впевненості можна очікувати, що ці помилки не вийдуть за певні межі. Призначають досить велику імовірність β (0,95; 0,99) таку, що подію $A=\{|a^* - m_x| < l\}$, яка характеризується цією імовірністю, можна вважати практично вірогідною, потім знаходять таке значення l, для якого справедлива рівність

$$P(A) = P\{(a - l) < m_x < (a + l)\} = \beta.$$

Тобто з імовірністю β невідоме значення параметра m_x буде перебувати в інтервалі L=[a^*- l, a^* + l]. Більші за абсолютним значенням помилки, зустрічатимуться з імовірністю α= 1 - β. Границі інтервалу називають довірчими границями: $a_1 = a^*–l$; $a_2 = a^*+l$.

Утруднення полягає в тому, що закон розподілу a^* залежить від закону розподілу досліджуваної ознаки X і, отже, від його невідомих параметрів (зокрема і від самого параметра a). Щоб обійти це утруднення, застосовують наступний грубо наближений прийом: замінюють у виразі для l невідомі параметри їх точковими значеннями. При 20-30 дослідах цей прийом зазвичай дає задовільні за точністю результати.

Кореляційний аналіз заснований на використанні рівняння регресії.

Регресією Y на X називають умовне математичне сподівання випадкової величини Y за умови, що X прийняла значення x_i. Лінію, що з'єднує точки $\overline{y}_i$, називають *лінією регресії*.

Для апроксимації лінії регресії аналітичним виразом використовують *рівняння регресії* $y_x = \varphi(\overline{x})$. Вибір виду залежності $\overline{y}_x = \varphi(x)$ можна здійснити з теоретичних міркувань або графічно за допомогою *поля кореляції*. На практиці найчастіше використовують лінійне рівняння регресії: Y = ρ_{yx} X + b. Коефіцієнт при перемінній X ρ_{yx} називають *коефіцієнтом регресії*.

Для визначення параметрів залежності, що згладжує $\overline{y}_x = \varphi(x)$, і зокрема, значень параметрів ρ_{yx} та b рівняння регресії, застосовують *метод найменших квадратів* (МНК). Цей метод дозволяє при відомому класі апроксимуючої залежності $\overline{y}_x = \varphi(x)$ так вибрати значення її параметрів, щоб ця залежність щонайкраще відображала дані спостережень. При використанні МНК вимога найкращого узгодження апроксимуючої кривої $\overline{y}_x = \varphi(x)$ із дослідними даними зводиться до того, щоб сума квадратів відхилень цієї кривої від експериментальних точок оберталася у мінімум: $\sum_{i=1}^{n}(y_i - y_{ip})^2 \to \min$, де y_i – значення результативної ознаки Y, отримані в результаті спостережень; y_{ip} - розрахункові значення результативної ознаки Y, отримані на підставі аналітичного вираження кривої, яка згладжує $\overline{y}_x = \varphi(x)$.

Отримані для лінійного рівняння регресії вирази для коефіцієнта регресії ρ_{yx} та вільного члена b мають вигляд:

$$\rho_{yx} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}; \qquad b = \frac{\sum x_i^2 * \sum y_i - \sum x_i * \sum x_i y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}.$$

Вибірковий коефіцієнт кореляції слугує для оцінки тісноти лінійної кореляційної залежності, він визначається за формулою $r_B = \rho * \frac{\sigma_x}{\sigma_y}$.

Коефіцієнт кореляції $r_в$ дозволяє оцінити величину лінійного зв'язку між двома випадковими величинами X та Y. Вибірковий коефіцієнт кореляції приймає значення від -1 до +1 і характеризує тісноту лінійного зв'язку між ознаками у вибірці. Якщо $r_в = 0$, то лінійний зв'язок відсутній, чим ближче значення $|r_в|$ до одиниці, тим тісніше зв'язок, при $|r_в| = 1$ він стає функціональним.

Для оцінки тісноти нелінійного кореляційного зв'язку застосовують *вибіркове кореляційне відношення* η. Вибірковим кореляційним відношенням Y до X називають відношення міжгрупового середнього квадратичного відхилення до загального середнього квадратичного відхилення результативної ознаки Y: $\eta_{yx} = \frac{\sigma_{\overline{y_x}}}{\sigma_y}$, де $\sigma_{\overline{y_x}}$ -міжгрупове середнє квадратичне відхилення. Його визначають за формулою $\sigma_{\overline{y_x}} = \sqrt{D_{межгр}} = \sqrt{\frac{\sum N_j(\overline{y_j} - \overline{y})^2}{n}}$, де $\overline{y_j}$ - умовна середня значень Y j-ї групи; N_j – обсяг j-ї групи.

Міжгрупова дисперсія – це дисперсія групових середніх відносно загальної середньої.

Внутрішньогрупова дисперсія є середнім арифметичним групових дисперсій: $D_{вн.гр} = \frac{\sum N_j S_y}{n}$.

Загальна дисперсія результативної ознаки Y є сумою внутрішньогрупової та міжгрупової дисперсій: $D_y = D_{вн.гр} + D_{міжгр}$.

При функціональному зв'язку між X та Y кореляційне відношення дорівнює одиниці: $\eta_{yx} = \frac{\sigma_{\overline{y_x}}}{\sigma_y} = 1$. Якщо кореляційне відношення дорівнює нулю, $\eta=0$, то між X та Y зв'язок відсутній. Значення кореляційного відношення лежать у межах від 0 до 1: $0 \le \eta \le 1$; значення кореляційного відношення перевершують або дорівнюють вибірковому коефіцієнту кореляції: $\eta \ge |r_в|$; якщо кореляційне відношення дорівнює вибірковому коефіцієнту кореляції, $\eta = |r_в|$, то між X та Y є лінійна кореляційна залежність.

Задача 2.1

З метою дослідження точності прибору зробили 500 вимірювань помилки. Результати вимірювань звели у групований варіаційний ряд:

групи	-4 ...-3	-3 ...-2	-2 ...-1	-1 ...0	0 ...1	1 ...2	2 ...3	3 ...4
m_i	6	25	72	133	120	88	46	10

Побудувати гістограму розподілу та визначити параметри розподілу.

Розв'язання

Визначимо частоти для кожного розряду групованого варіаційного ряду, користуючись формулою

$$p_i^* = \frac{m_i}{n},$$

де m_i – число значень помилки X, що потрапили в i-у групу; n – число зроблених вимірів, n = 500.

групи	-4 …-3	-3 …-2	-2 …-1	-1 …0	0 …1	1 …2	2 …3	3 …4
m_i	6	25	72	133	120	88	46	10
p_i*	0,012	0,05	0,144	0,266	0,24	0,176	0,092	0,02

Для побудови гістограми визначимо значення щільності частот для кожної групи значень за формулою

$$f_i^* = \frac{p_i^*}{l},$$

де l - довжина групи, l= 1.

групи	-4 …-3	-3 …-2	-2 …-1	-1 …0	0 …1	1 …2	2 …3	3 …4
m_i	6	25	72	133	120	88	46	10
p_i*	0,012	0,05	0,144	0,266	0,24	0,176	0,092	0,02
f_i*	0,012	0,05	0,144	0,266	0,24	0,176	0,092	0,02

Побудуємо графік гістограми (рис. 2.1).

З вигляду гістограми можна припустити, що її можна згладити за допомогою нормального закону (припущення підтверджується тим, що досліджувана випадкова величина є помилкою вимірювання, а отже розподілена нормально), щільність розподілу якого визначають за виразом

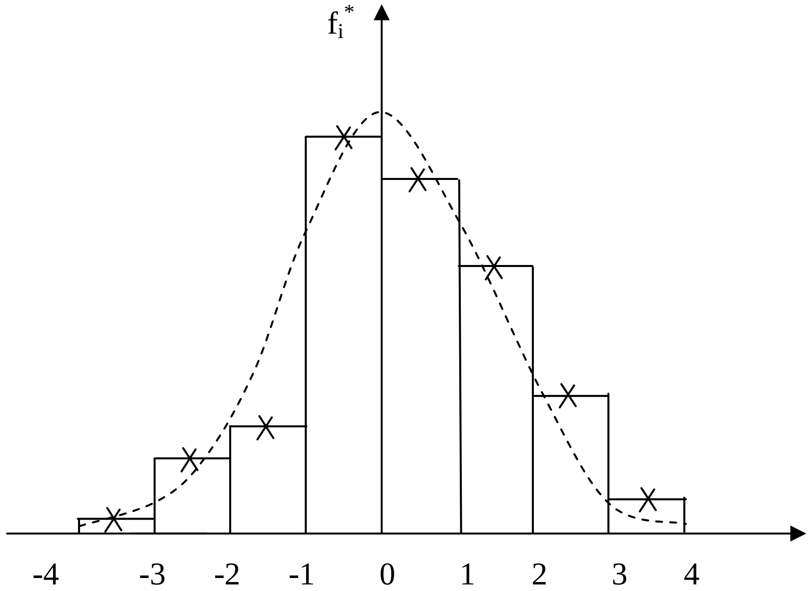

Рис. 2.1 - Графік гістограми

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}.$$

Параметри m та σ, що входять у вираз щільності f(x), виберемо так, щоб щонайкраще погодити аналітичний вираз зі статистичним розподілом. Визначимо вибіркове середнє $\tilde{X}$ та вибіркове середнє квадратичне відхилення $\sigma_{выб}$ за даними групованого варіаційного ряду. Як значення x_i виберемо середину i-ї групи і цьому значенню поставимо у відповідність як імовірність його частоту p_i*, отримаємо

$$\tilde{X} = \sum_{i=1}^{8} x_i p_i^* = -3{,}5*0{,}012 - 2{,}5*0{,}05 - 1{,}5*0{,}144 - 0{,}5*0{,}266 + 0{,}5*0{,}24 + 1{,}5*0{,}176 + 2{,}5*0{,}092 + 3{,}5*0{,}02 = 0{,}168;$$

$$\tilde{\alpha}_2 = \sum_{i=1}^{8} x_i^2 p_i^* = (-3{,}5)^2*0{,}012 + (-2{,}5)^2*0{,}05 + (-1{,}5)^2*0{,}144 + (-0{,}5)^2*0{,}266 + \\ + (0{,}5)^2*0{,}24 + (1{,}5)^2*0{,}176 + (2{,}5)^2*0{,}092 + (3{,}5)^2*0{,}02 = 2{,}126;$$

$$\sigma^2_{выб} = \tilde{\alpha}_2 - \tilde{X}^2 = 2{,}126 - (0{,}168)^2 = 2{,}098;$$

$$\sigma_{выб} = \sqrt{\sigma^2_{выб}} = \sqrt{2{,}098} = 1{,}448.$$

Дістали розподіл $f^*(x) = \dfrac{1}{1{,}448\sqrt{2\pi}} \exp\left\{-\dfrac{(x-0{,}168)^2}{2*2{,}126}\right\}$, користуючись яким, підрахуємо значення f*(x) на межах груп:

$f_i(-4) = 0{,}0045$; $f_i(1) = 0{,}2343$;
$f_i(-3) = 0{,}0256$; $f_i(2) = 0{,}1244$;
$f_i(-2) = 0{,}0895$; $f_i(3) = 0{,}0435$;
$f_i(-1) = 0{,}1986$; $f_i(4) = 0{,}0087$;
$f_i(0) = 0{,}274$.

Відкладемо на графіку отримані точки та проведемо плавну криву (рис. 2.1).

Задача 2.2

Для визначення точності вимірювального прибору було зроблено п'ять незалежних вимірювань, результати яких зведені у таблицю:

Номер вимірювання	1	2	3	4	5
x_i	2781	2836	2807	2763	2858

Визначити незміщену оцінку дисперсії помилок вимірювального прибору, якщо дійсне значення вимірюваної величини:
а) відомо і дорівнює 2800; б) невідомо.

Розв'язання

а) якщо значення вимірюваної величини відомо, то генеральна середня $\overline{X} = 2800$, незміщену оцінку дисперсії у цьому випадку можна визначити за формулою

$$\sigma^2_{выб} = \frac{\Sigma(x_i - \overline{x})^2}{n} = \frac{(2781-2800)^2 + (2836-2800)^2 + (2807-2800)^2 + (2763-2800)^2 + (2858-2800)^2}{5} = 1287{,}8$$

б) якщо значення вимірюваної величини невідомо, слід визначити вибіркову середню, а незміщену оцінку дисперсії обчислити за формулою

$$S^2 = \frac{n}{n-1}\sigma^2_{выб} = \frac{\Sigma(x_i - \tilde{x})^2}{n-1} = \frac{(2781-2809)^2 + (2836-2809)^2 + (2807-2809)^2 + (2763-2809)^2 + (2858-2809)^2}{5-1} = 1508{,}5$$

Задача 2.3

Зроблені вимірювання випадкової величини Y при різних значеннях випадкової величини X. Визначити вибірковий коефіцієнт кореляції цих величин.

x_i	-8	-10	22	2
y_i	-10	-2	4	-1

Розв'язання

Коефіцієнт кореляції визначимо за формулою

$$r_{xy} = \frac{K_{xy}}{\sigma_x \sigma_y}.$$

Для його обчислення необхідно знайти вибірковий кореляційний момент

$$K_{хувув} = \frac{\sum_i (x_i - \overline{X})(y_i - \overline{Y})}{n-1}.$$

Для оцінки середніх значень X та Y визначимо вибіркові середні

$$\tilde{X} = \frac{-8+10+22+2}{4} = 6{,}5 \qquad \tilde{Y} = \frac{-10-2+4-1}{4} = -2{,}25.$$

Визначимо вибіркові дисперсії X та Y

$$S_x^2 = \frac{(-8-6{,}5)^2 + (10-6{,}5)^2 + (22-6{,}5)^2 + (2-6{,}5)^2}{4-1} = 161;$$

$$S_y^2 = \frac{(-10+2{,}25)^2 + (-2+2{,}25)^2 + (4+2{,}25)^2 + (-1+2{,}25)^2}{4-1} = 33{,}6.$$

Визначимо вибіркові середні квадратичні відхилення:

$$\sigma_{выб_x} = \sqrt{S_x^2} = \sqrt{161} = 12{,}7 \qquad \sigma_{выб_y} = \sqrt{S_y^2} = \sqrt{33{,}6} = 5{,}8.$$

Тепер розрахуємо кореляційний момент

$$K_{хувув} = \frac{(-8-6{,}5)(-10+2{,}25)+(10-6{,}5)(-2+2{,}25)+(22-6{,}5)(4+2{,}25)+(2-6{,}5)(-1+2{,}25)}{4-1} = 68{,}2$$

та коефіцієнт кореляції

$$r_{xy} = \frac{K_{хувув}}{\sigma_{выб_x}\sigma_{выб_y}} = \frac{68{,}2}{12{,}7 * 5{,}8} = 0{,}926.$$

Задача 2.4

Знайдемо довірчий інтервал для вибіркової середньої досліджуваної ознаки X. Нехай зроблено n незалежних дослідів та визначені спроможні і незміщені оцінки параметрів цієї ознаки $\tilde{X}$ та $\sigma^2_{виб}$.

Розв'язання

Нехай $\tilde{X}$ = 10, $\sigma^2_{виб}$ = 4, n = 40. Задамося значенням довірчої імовірності β = 0,95. Тоді можна записати

$$P\{(\tilde{X} - l) < \overline{X} < (\tilde{X} + l)\} = 0{,}95.$$

Скористуємося тим, що випадкова величина $\tilde{X}$ є функцією n незалежних випадкових величин x_i. Тоді відповідно до центральної граничної теореми щільність розподілу випадкової величини $\tilde{X}$ практично буде підпорядковуватися нормальному закону розподілу з параметрами

$$M[\tilde{X}] = \overline{X} = 10, \; D[\tilde{X}] = \frac{\sigma^2_{ген}}{n} = 4/40 = 0{,}1.$$

Для нормального закону розподілу імовірність влучення випадкової величини на інтервал значень можна виразити за допомогою інтеграла імовірностей:

$$P\{(m_x^* - l) \le m_x \le (m_x^* + l)\} = \left[\Phi\left(\frac{m_x^* + l - m_x}{\sqrt{\ }\sigma_x}\right) - \Phi\left(\frac{m_x^* - l - m_x}{\sqrt{\ }\sigma_x}\right)\right] = \left[\Phi\left(\frac{l}{\sqrt{\ }\sigma_x}\right) - \Phi\left(\frac{-l}{\sqrt{\ }\sigma_x}\right)\right] = 2\Phi\left(\frac{l}{\sqrt{\ }\sigma_x}\right) = \beta$$

Підставимо значення:

$$\Phi\left(\frac{l}{\sqrt{0{,}32}}\right) = 0{,}475; \qquad \frac{l}{0{,}453} = 1{,}4,$$

звідки l = 1,4 * 0,453 = 0,634.

Отже, з імовірністю 0,95 інтервал (9,366; 10,634) накриє генеральну середню досліджуваної ознаки X.

Задача 2.5

Нехай у результаті досліду отримані наступні експериментальні дані:

x_i	1	2	3
y_i	1	3	4

Потрібно визначити параметри лінійної та квадратичної залежностей для X та Y.

Розв'язання

Нанесемо на координатну площину точки з координатами (x_i,y_i) (рис. 2.2).

Із графіка видно, що точки не лежать на одній прямій, і що із зростанням X У має тенденцію до зростання.

а) нехай шукана залежність - лінійна: $y = a_0 + a_1x$,

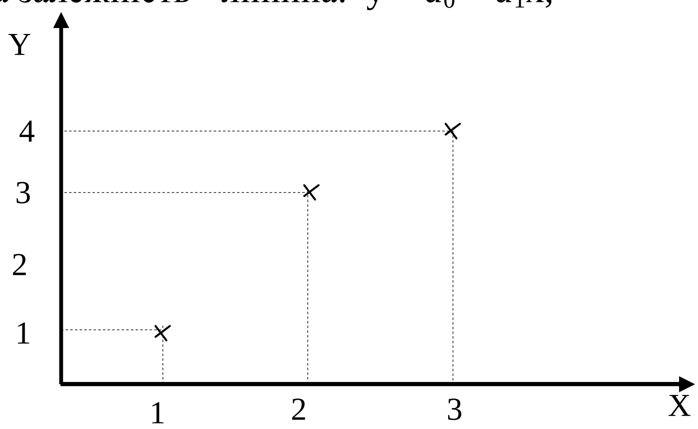

Рис. 2.2 - Побудова поля кореляції

запишемо її як функцію параметрів: $\sum_{i=1}^{n}(y_i - a_0 - a_1x_i)^2 \to \min$ та візьмемо часткові похідні за параметрами a_1 та a_0 і дорівняємо їх до нуля:

$$\begin{cases} 2*\Sigma(y_i - a_0 - a_1x_i)*(-1) = 0; \\ 2*\Sigma(y_i - a_0 - a_1x_i)*(-x_i) = 0; \end{cases}$$

виконаємо перетворення:

$$\begin{cases} \Sigma a_0 + \Sigma a_1x_i = \Sigma y_i \\ \Sigma a_0\, x_i + \Sigma a_1x_i^2 = \Sigma y_i\, x_i \end{cases}$$

$$\begin{cases} na_0 + a_1\Sigma x_i = \Sigma y_i \\ a_0\Sigma\, x_i + a_1\Sigma x_i^2 = \Sigma y_i\, x_i \end{cases}$$

Підставимо значення: n=3; Σx_i =6; Σx_i^2 =14; Σy_i = 8; $\Sigma y_i\, x_i$ = 19 та визначимо параметри.

$$\begin{cases} 3a_0 + 6\, a_1 = 8 \\ 6a_0 + 14a_1 = 19. \end{cases}$$

$$\begin{cases} a_0 = -1/3 \\ a_1 = 1,5. \end{cases}$$

Шукана залежність має вигляд:

$$y = -1/3 + 1,5\ x.$$

Отримана лінійна залежність є найімовірнішою з лінійних залежностей.

Протабулюємо її та побудуємо графік (рис. 2.3).

x_i	1	2	3
y_i	1	3	4
$y_i^т$	1,17	2,67	4,17

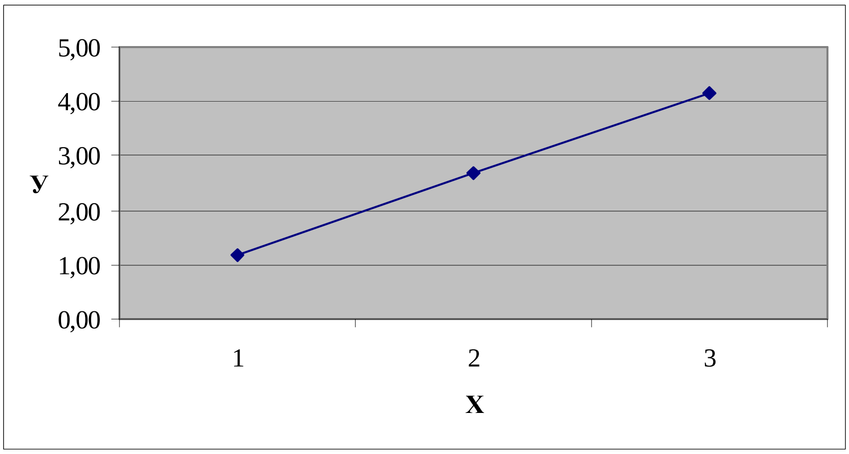

Рис. 2.3 - Лінійна залежність y = -1/3 + 1,5 x.

б) нехай шукана залежність – квадратична: $Y = a_0 + a_1x + a_2x^2$, запишемо її: $\sum_{i=1}^{n}(y_i - a_0 - a_1x_i - a_2x_i^2)^2 \to \min$, візьмемо часткові похідні та дорівняємо їх до нуля:

$$\begin{cases} 2*\Sigma(y_i - a_0 - a_1x_i - a_2x_i^2)*(-1) = 0 \\ 2*\Sigma(y_i - a_0 - a_1x_i - a_2x_i^2)*(-x_i) = 0 \\ 2*\Sigma(y_i - a_0 - a_1x_i - a_2x_i^2)*(-x_i^2) = 0 \end{cases}$$

виконаємо перетворення:

$$\begin{cases} \Sigma y_i - \Sigma a_0 - \Sigma a_1x_i - \Sigma a_2x_i^2 = 0 \\ \Sigma y_i\, x_i - \Sigma\, a_0\, x_i - \Sigma a_1x_i^2 - \Sigma a_2x_i^3 = 0 \\ \Sigma y_i\, x_i^2 - \Sigma a_0\, x_i^2 - \Sigma a_1x_i^3 - \Sigma a_2x_i^4 = 0 \end{cases}$$

$$\begin{cases} na_0 + a_1\Sigma x_i + a_2\Sigma x_i^2 = \Sigma y_i \\ a_0\Sigma x_i + a_1\Sigma x_i^2 + a_2\Sigma x_i^3 = \Sigma y_i\, x_i \\ a_0\, \Sigma x_i^2 + a_1\Sigma x_i^3 + a_2\Sigma x_i^4 = \Sigma y_i\, x_i^2 \end{cases}$$

Підставимо значення: $n=3$; $\Sigma x_i =6$; $\Sigma x_i^2 =14$; $\Sigma y_i = 8$; $\Sigma y_i\, x_i = 19$;$\Sigma x_i^3 =36$; $\Sigma x_i^4 =98$; $\Sigma y_i\, x_i^2 = 49$ і визначимо параметри

$$\begin{cases} 3a_0 + 6\, a_1 + 14a_2 = 8 \\ 6a_0 + 14a_1 + 36a_2 = 19 \\ 14a_0 + 36a_1 + 98a_2 = 49. \end{cases}$$

Звідки : $a_2=-0,091$; $a_1=1,273$; $a_0=0,0455$.

Отже, найімовірніша квадратична залежність матиме вигляд:

$$y = 0{,}0455 + 1{,}273x - 0{,}091x^2.$$

Протабулюємо її, збільшивши для наочності кількість точок:

x_i	0	1	2	3	4	5
y_i^T	0,05	1,23	2,23	3,05	3,68	4,14

Отримаємо графік, показаний на рис. 2.4.

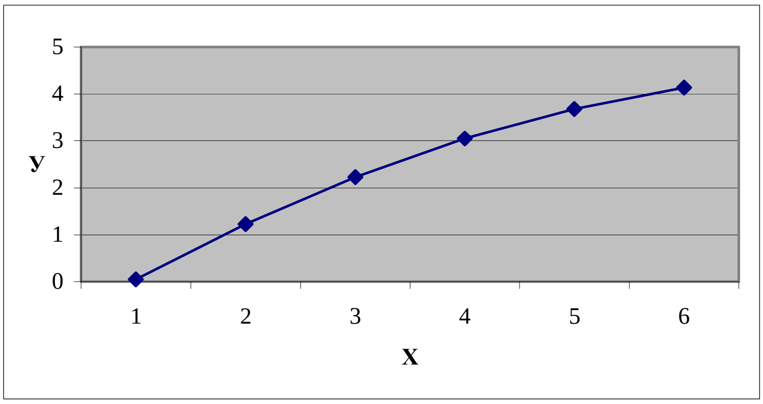

Рис. 2.4 - Квадратична залежність $y = 0{,}0455 + 1{,}273x - 0{,}091x^2$

Задача 2.6

Визначити параметри лінійної регресії для системи випадкових величин X і Y, результати вимірювання яких приведені у задачі 2.5.

№ досліду	**x_i**	**y_i**
1	**2**	**3**
1	4	0,041
2	8	0,05
3	10	0,081
4	14	0,104
5	16	0,12
6	20	0,139
7	19	0,154
8	23	0,18
9	26	0,208
10	30	0,241
11	31	0,25
12	36	0,269
13	37	0,301

Розв'язання

Побудуємо поле кореляції (рис. 2.5). Очевидно, що статистичні дані добре згладжуються лінійною залежністю.

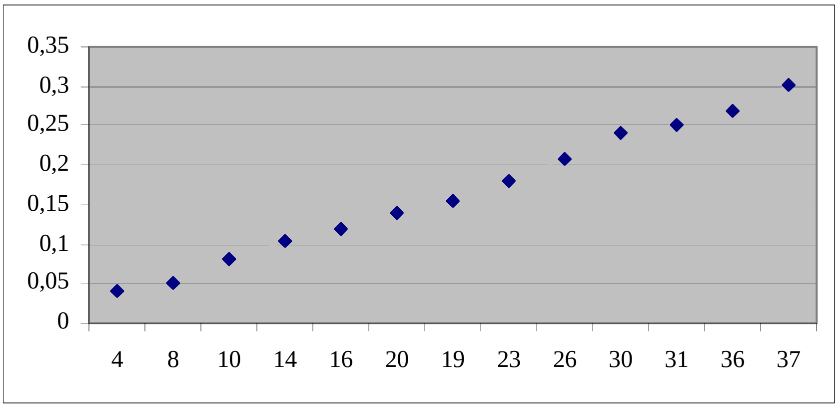

Рис. 2.5 - Побудова поля кореляції

Визначимо параметри лінійної залежності між X та Y, проміжні обчислення зробимо у таблиці

№ досліду	x_i	y_i	x_i^2	$x_i y_i$
1	4	0,041	16	0,164
2	8	0,05	64	0,4
3	10	0,081	100	0,81
4	14	0,104	196	1,456
5	16	0,12	256	1,92
6	20	0,139	400	2,78
7	19	0,154	361	2,926
8	23	0,18	529	4,14
9	26	0,208	676	5,408
10	30	0,241	900	7,23
11	31	0,25	961	7,75
12	36	0,269	1296	9,684
13	37	0,301	1369	11,137
Сума	Σx_i=274	Σy_i=2,138	Σx_i^2= 7124	$\Sigma x_i{*}y_i$= 55,805

Отримаємо

$$\rho_{yx} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} = \frac{13*55{,}08 - 274*2{,}138}{13*7124 - 274^2} = 0{,}00796;$$

$$b = \frac{\sum x_i^2 * \sum y_i - \sum x_i * \sum x_i y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} = \frac{7124*2{,}138 - 274*55{,}8}{13*7124 - 274^2} = -0{,}0034.$$

Отже, шукана залежність матиме вигляд:

$$y = 0{,}00796x - 0{,}0034.$$

СПИСОК ДЖЕРЕЛ

1. Гмурман В. Э. Теория вероятностей и математическая статистика. -М.: Высш. школа, 1977. - 498 с.

2. Гмурман В.Э. Руководство к решению задач по теории вероятностей и математической статистике. - М.: Высш. школа, 1975. - 330с.

3. Ачкасов А.Є., Плакіда В.Т., Воронков О.О., Воронкова Т.Б. Теорія імовірностей і математична статистика: Навчальний посібник.- Харків, ХНАМГ, 2008.- 249 с.

4. Вентцель Е. С. Теория вероятностей. - М.: Высшая школа, 1999.

5. Теория статистики с основами теории вероятностей: Учебное пособие для вузов/ Под ред. И.И. Елисеевой.- М.: ЮНИТИ-ДАНА, 2001.- 446 с.

6. http://teorver.ru

7. http://www.artspb.com

8. http://www.matburo.ru

9. http://stud-project.ru

10. http://www.statsoft.ru

11. http://www.alife.narod.ru

12. http://neuro.net.ua

ДОДАТКИ

Додаток 1

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

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	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	3653	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	0008	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

Додаток 2

Таблиця значень функції $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{t^2}{2}} dt$

X	Φ(x)	X	Φ(x)	X	Φ(x)	X	Φ(x)
0,00	0,0000	0,36	0,1406	0,72	0,2642	1,08	0,3599
0,01	0,0040	0,37	0,1443	0,73	0,2673	1,09	0,3621
0,02	0,0080	0,38	0,1480	0,74	0,2703	1,10	0,3643
0,03	0,0120	0,39	0,1517	0,75	0,2734	1,11	0,3665
0,04	0,0160	0,40	0,1554	0,76	0,2764	1,12	0,3686
0,05	0,0199	0,41	0,1591	0,77	0,2794	1,13	0,3708
0,06	0,0239	0,42	0,1628	0,78	0,2823	1,14	0,3729
0,07	0,0279	0,43	0,1664	0,79	0,2852	1,15	0,3749
0,08	0,0319	0,44	0,1700	0,80	0,2818	1,16	0,3770
0,09	0,0359	0,45	0,1736	0,81	0,2910	1,17	0,3790
0,10	0,0398	0.46	0,1772	0,82	0,2939	1,18	0,3810
0,11	0,0438	0,47	0,1808	0,83	0,2967	1,19	0,3830
0,12	0,0478	0,48	0,1844	0,84	0,2995	1,20	0,3849
0,13	0,0517	0,49	0,1879	0,85	0,3023	1.21	0,3869
0,14	0,0557	0,50	0,1915	0,86	0,3051	1,22	0,3883
0,15	0,0596	0,51	0,1950	0,87	0,3078	1,23	0,3907
0,16	0,0636	0,52	0,1985	0,88	0,3106	1,24	0,3925
0,17	0,0675	0,53	0,2019	0,89	0,3133	1,25	0,3944
0,18	0,0714	0,54	0,2054	0,90	0,3159	1,26	0,3926
0,19	0,0753	0,55	0,2088	0,91	0,3186	1,27	0,3980
0,20	0,0793	0,56	0,2123	0,92	0,3212	1,28	0,3997
0,21	0,0832	0,57	0,2157	0,93	0,3238	1,29	0,4015
0,22	0,0871	0,58	0,2190	0,94	0,3264	1,30	0,4032
0,23	0,0910	0,59	0,2224	0,95	0,3289	1,31	0,4049
0,24	0,0948	0,60	0,2257	0,96	0,3315	1,32	0,4066
0,25	0,0987	0,61	0,2291	0,97	0,3340	1,33	0,4082
0,26	0,1026	0,62	0,2324	0,98	0,3365	1,34	0,4099
0.27	0,1064	0,63	0,2357	0,99	0,3389	1,35	0,4115
0,28	0,1103	0,64	0,2389	1,00	0,3413	1,36	0,4131
0,29	0,1141	0,65	0,2422	1,01	0,3438	1,37	0,4147
0,30	0,1179	0,66	0,2454	1,02	0,3461	1,38	0,4162
0,31	0,1217	0,67	0,2486	1,03	0,3485	1,39	0,4177
0,32	0,1255	0,68	0,2517	1,04	0,3508	1,40	0,4192
0,33	0,1293	0,69	0,2549	1,05	0,3531	1,41	0,4207
0,34	0,1331	0,70	0,2580	1,06	0,3554	1,42	0,4222
0,35	0,1368	0,71	0,2611	1,07	0,3577	1,43	0,4236

Продовження додатка 2

x	*Ф(x)*	*X*	*Ф(x)*	*X*	*Ф(x)*	*X*	*Ф(x)*
1,44	0,4251	1,73	0,4582	2,04	0,4793	2,62	0,4956
1,45	0,4265	1,74	0,4591	2,06	0,4803	2,64	0,4959
1,46	0,4279	1,75	0,4599	2,08	0,4812	2,66	0,4961
1,47	0,4292	1,76	0,4608	2,10	0,4821	2,68	0,4963
1,48	0,4306	1,77	0,4616	2,12	0,4830	2,70	0,4965
1,49	0,4319	1,78	0,4625	2,14	0,4838	2,72	0,4967
1,50	0,4332	1,79	0,4633	2,16	0,4846	2,74	0,4969
1,51	0,4345	1,80	0,4641	2,18	0,4854	2,76	0,4971
1,52	0,4357	1,81	0,4649	2,20	0,4861	2,78	0,4973
1,53	0,4370	1,82	0,4656	2,22	0,4868	2,80	0,4974
1,54	0,4382	1,83	0,4664	2,24	0,4875	2,82	0,4976
1,55	0,4394	1,84	0,4671	2,26	0,4881	2,84	0,4977
1,56	0,4406	1,85	0,4678	2,28	0,4887	2,86	0,4979
1,57	0,4418	1,86	0,4686	2,30	0,4893	2,88	0,4980
1,58	0,4429	1,87	0,4693	2,32	0,4898	2,90	0,4981
1,59	0,4441	1,88	0,4699	2,34	0,4904	2,92	0,4982
1,60	0,4452	1,89	0,4706	2,36	0,4909	2,94	0,4984
1,61	0,4463	1,90	0,4713	2,38	0,4913	2,96	0,4985
1,62	0,4474	1,91	0,4719	2,40	0,4918	2,98	0,4986
1,63	0,4484	1,92	0,4726	2,42	0,4922	3,00	0,49865
1,64	0,4495	1,93	0,4732	2,44	0,4927	3,20	0,49931
1,65	0,4505	1,94	0,4738	2,46	0,4931	3,40	0,49966
1,66	0,4515	1,95	0,4744	2,48	0,4934	3,60	0,499841
1,67	0,4525	1,96	0,4750	2,50	0,4938	3,80	0,499928
1,68	0,4535	1,97	0,4756	2,52	0,4941	4,00	0,499968
1,69	0,4545	1,98	0,4761	2,54	0,4945	4,50	0,499997
1,70	0,4554	1,99	0,4767	2,56	0,4948	5,00	0,499997
1,71	0,4564	2,00	0,4772	2,58	0,4951		
1,72	0,4573	2,02	0,4783	2,60	0,4953		

ЗМІСТ

Загальні положення....... 3
Практичне заняття 1. Визначення імовірності випадкової події. Числові характеристики випадкової величини. Нормальний закон розподілу 4
Загальні відомості 4
Вказівки до виконання завдання 8
Задача 1.1 9
Задача 1.2 9
Задача 1.3 9
Задача 1.4 10
Задача 1.5 10
Задача 1.6 10
Задача 1.7 11
Задача 1.8 12
Задача 1.9 12
Практичне заняття 2. Варіаційний ряд. Властивості вибіркових числових характеристик. Довірчий інтервал і довірча імовірність. Елементи теорії кореляції 13
Загальні відомості 13
Задача 2.1 17
Задача 2.2 19
Задача 2.3 19
Задача 2.4 20
Задача 2.5 21
Задача 2.6 23
Список джерел 25
Додатки 26

Навчальне видання

МЕТОДИЧНІ ВКАЗІВКИ
ДО ПРАКТИЧНИХ ЗАНЯТЬ
З КУРСУ

ТЕОРІЯ ІМОВІРНОСТЕЙ І МАТЕМАТИЧНА СТАТИСТИКА

(*для слухачів другої вищої освіти ФПО та ЗН спеціальностей 7.06010107 «Теплогазопостачання і вентиляція» 7.06010101 «Промислове та цивільне будівництво»*)

Укладачі: доц. **ОХРІМЕНКО** Вячеслав Миколайович,
ст. викл. **ВОРОНКОВА** Тетяна Борисівна,
ст. викл. **ВОРОНКОВ** Олексій Олександрович

Відповідальний за випуск *А. І. Кузнєцов*

За авторською редакцією

Комп'ютерне верстання *К. А. Алексанян*

План 2012, поз. 610 М

Підп. до друку 02.10.2012	Формат 60×84/16
Друк на різографі.	Ум. друк. арк. 1,7
Зам. №	Тираж 50 пр.

Видавець і виготовлювач:
Харківська національна академія міського господарства,
вул. Революції, 12, Харків, 61002
Електронна адреса: rectorat@ksame.kharkov.ua
Свідоцтво суб'єкта видавничої справи:
ДК № 4064 від 12.05.2011 р.