МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

ТЕОРІЯ ЙМОВІРНОСТЕЙ, ЙМОВІРНІСНІ ПРОЦЕСИ ТА МАТЕМАТИЧНА СТАТИСТИКА

ПРАКТИКУМ
Частина 2

Навчальний посібник

Рекомендовано Методичною радою КПІ ім. Ігоря Сікорського
як навчальний посібник для здобувачів ступеня бакалавра
за спеціальностями 126 «Інформаційні системи та технології»,
121 «Інженерія програмного забезпечення»

Укладачі: О. В. Гавриленко

Електронне мережне навчальне видання

Київ
КПІ ім. Ігоря Сікорського
2023

Рецензент *Павлов О.А.*, д.т.н., професор,
професор кафедри інформатики та програмної інженерії,
факультет інформатики та обчислювальної техніки

Відповідальний редактор *Жураковська О.С.*, к.т.н.,
доцент кафедри інформаційних сиситем та техніки,
факультет інформатики та обчислювальної техніки

Гриф надано Методичною радою КПІ ім. Ігоря Сікорського
(протокол № 7 від 27.04.2023 р.)
за поданням Вченої ради факультету/навчально-наукового інституту
(протокол № 9 від 27.02.2023 р.)

Навчальний посібник призначений для студентів спеціальності 126 «Інформаційні системи та технології» кафедри інформаційних систем та технологій та спеціальності 121 «Інженерія програмного забезпечення» кафедри інформатики та програмної інженерії всіх освітніх програм та всіх форм навчання. У посібнику наведено систематизоване викладення багатьох основних відомостей з теорії ймовірностей, ймовірнісних процесів та математичної статистики з орієнтацією на розв'язання задач з вказаної дисципліни студентами технічних спеціальностей. Метою автора було розвинути навички студентів до розв'язання задач з дисципліни «Ймовірністні моделі та статистичне оцінювання в інформаційно-управляючих системах» для їх застосування в інженерній практиці.

Реєстр. № НП 21/22-283. Обсяг 3,38 авт. арк.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
проспект Перемоги, 37, м. Київ, 03056
https://kpi.ua
Свідоцтво про внесення до Державного реєстру видавців, виготовлювачів
і розповсюджувачів видавничої продукції ДК № 5354 від 25.05.2017 р.

ЗМІСТ

ВСТУП

Метою освоєння дисципліни "Ймовірністні моделі та статистичне оцінювання в інформаційно-управляючих системах" є вивчення закономірностей випадкових явищ, їх властивостей і використання їх для аналізу статистичних даних.

В результаті опанування дисципліни студент

має знати:

основні поняття теорії ймовірностей і математичної статистики та їх властивості;

має вміти:

використовувати імовірнісні моделі при вирішенні завдань, працювати з випадковими величинами,

виконувати розрахунок вибіркових характеристик, оцінювати надійність статистичних даних;

має володіти:

- основною термінологією дисципліни;
- навичками роботи з літературою з дисципліни;
- навичками обґрунтування вибору даних для відповідного статистичного аналізу, з метою досягнення поставленої задачі;
- прийомами розв’язання стандартних задач теорії ймовірностей.

мати досвід:

- роботи з випадковими величинами – даними у вирішенні реальних професійно-орієнтованих завдань на основі володіння глибокими теоретичними знаннями з теорії ймовірностей та математичної статистики, методів аналізу даних з використанням спеціальних програмних систем та пакетів.

КОМПЕТЕНТНОСТІ

Спеціальні (фахові, предметні) компетентності

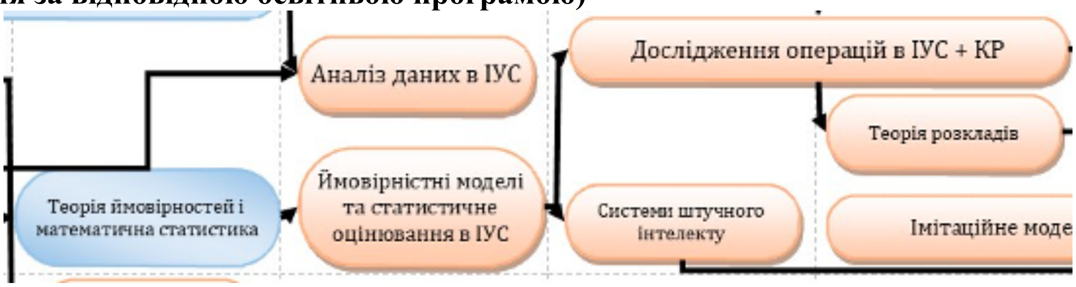
- ФК-5. Здатність оцінювати та враховувати економічні, соціальні, технологічні та екологічні фактори на всіх етапах життєвого циклу інфокомунікаційних систем.
- ФК-11. Здатність до аналізу, синтезу і оптимізації інформаційних систем та технологій з використанням математичних та імітаційних моделей і методів
- ФК-13. Здатність проводити обчислювальні експерименти, порівнювати результати експериментальних даних і отриманих рішень.
- ФК-15. Здатність до алгоритмічного та логічного мислення.
- ФК-18. Здатність до розробки і використання інтелектуальних інформаційних систем, технологій генерації та аналізу знань, алгоритмів штучного інтелекту для вирішення прикладних задач і підтримки прийняття рішень в різних прикладних областях життєдіяльності людини.
- ФК-19. Здатність до застосування методів прийняття управлінських рішень в умовах невизначеності та багатофакторної залежності щодо визначення рішення та ефективності управлінської діяльності.

ПРОГРАМНІ РЕЗУЛЬТАТИ НАВЧАННЯ

- ПРН-1. Знати лінійну та векторну алгебру, диференціальне та інтегральне числення, теорію функцій багатьох змінних, теорію рядів, диференціальні рівняння для функції однієї та багатьох змінних, операційне числення, теорію ймовірностей та математичну статистику в обсязі, необхідному для розробки та використання інформаційних систем, технологій та інфокомунікацій, сервісів та інфраструктури організації.
- ПРН-2. Застосовувати знання фундаментальних і природничих наук, системного аналізу та технологій моделювання, стандартних алгоритмів та дискретного аналізу при розв’язанні задач проєктування і використання інформаційних систем та технологій

- ПРН-24. Вміти розв’язувати складні непередбачувані задачі і проблеми у спеціалізованих сферах професійної діяльності та/або навчання, що передбачають збирання та інтерпретацію та аналіз інформації (даних), вибір методів та інструментальних засобів, застосування інноваційних підходів
- ПРН-27. Розробляти та використовувати математичні моделі для інтерпретації теоретичних та прикладних задач.

Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)



Пререквізити:

- Теорія ймовірностей і математична статистика.

Постреквізити:

- Дослідження операцій в ІУС,
- Системи штучного інтелекту.

ПРАКТИЧНЕ ЗАНЯТТЯ 1

Вибірковий метод

1.1. Генеральна сукупність та вибірка. Варіаційний ряд. Статистичний розподіл вибірки.

Математична статистика — наука про методи збору, систематизації, обробки і використання для наукових і практичних висновків статистичних даних, здобутих у результаті випробувань із випадковими наслідками. Ці дані є можливими значеннями деякої випадкової величини (досліджуваної ознаки) X, наприклад похибки вимірювань фізичної величини, відхилення розмірів виробу від стандарту, процента зайнятості крісел на авіарейсах, прибутку підприємства тощо.

При цьому досліджувані ознаки можуть бути дискретними (набувати окремих ізольованих значень) або неперервними (набувати будь-яких можливих значень з деякого інтервалу).

При дослідженні ознаки X зібрати та обробити всі її можливі значення здебільшого буває неможливо через їх занадто великий об'єм, а також у випадках, коли, наприклад, аналізуючи відповідність стандарту геометричних розмірів виробів, доводиться виконувати масові вимірювання або, аналізуючи міцність виробів, руйнувати їх. Тому з усієї множини N даних про ознаку X, яка називається *генеральною сукупністю* об'ємом n, випадковим чином відбирається досить представницька (репрезентативна) її частина, що містить n даних і називається *вибірковою сукупністю*, або *вибіркою* об'ємом n. При цьому об'єм вибірки n має бути суттєво меншим за об'єм генеральної сукупності N, але дані вибірки мають достатньо повно відбивати властивості ознаки X генеральної сукупності.

Різні можливі значення x_i ознаки X, які потрапили до вибірки, називаються *варіантами*, а система варіант $x_1, x_2, \dots, x_k$, розміщена в зростаючому порядку, називається *варіаційним рядом*.

Якщо варіанти $x_1, x_2, \dots, x_k$ спостерігаються у вибірці відповідно $n_1, n_2, \dots, n_n$ раз, то значення $n_1, n_2, \dots, n_k$ називаються *частотами варіант*, і для них виконується умова

$$\sum_{i=1}^{k} n_i = n,$$

а відношення

$$\omega_1 = \frac{n_1}{n},\ \ \omega_2 = \frac{n_2}{n},\ \dots,\ \ \omega_k = \frac{n_k}{n}$$

називаються відносними частотами варіант, і для них виконується умова

$$\sum_{i=1}^{k} \omega_i = 1.$$

Таблиця, в першому рядку якої наведено впорядковані варіанти x_i (варіаційний ряд), а в другому — відповідні їм частоти n_i (або відносні частоти ω_i), називається дискретним статистичним розподілом вибірки (табл. 1.1).

Таблиця 1.1

x_i	x_1	x_2	...	x_{k-1}	x_k
n_i	n_1	n_2	...	n_{k-1}	n_k

Якщо кількість варіант у вибірці надто велика або досліджувана ознака X є неперервною випадковою величиною, то будується інтервальний статистичний розподіл вибірки, аналогічний дискретному, у першому рядку якого замість дискретного варіаційного ряду записується інтервальний

$$(x_1;\, x_2),\, (x_2;\, x_3),...,(x_m;\, x_{m+1}).$$

Кількість інтервалів m для вибірки обсягом n орієнтовно вважається такою, що дорівнює $\sqrt{n}$ (із округленням до цілого значення), а розмір (крок) h інтервалу обчислюється за формулою

$$h = \frac{x_k - x_1}{m} \tag{1.1}$$

де x_1 і x_k — відповідно найменша і найбільша варіанти у вибірці, m – кількість інтервалів. Частоти n_i^* варіант, які потрапили в кожний частинний інтервал $(x_i;\, x_{i+1})$, записуються в другий рядок таблиці.

Якщо межа двох частинних інтервалів збігається зі значенням варіанти, то частота цієї варіанти або ділиться порівну між цими двома інтервалами, або в усіх таких випадках цілком заноситься у правий (або лівий) від варіанти інтервал.

Правило Стерджеса - емпіричне правило визначення оптимальної кількості інтервалів, на які розбивається діапазон зміни випадкової величини, що спостерігається, при побудові гістограми щільності її розподілу. Названо на ім'я американського статистика Герберта Стерджеса (Herbert Arthur Sturges, 1882-1958).

Кількість інтервалів визначається як:

$$m = 1 + \log_2 n \tag{1.2}$$

де– загальна кількість спостережень величини,– логарифм за основою 2.

Часто зустрічається записаним через десятковий логарифм:

$$m = 1 + 3{,}322 lgn. \tag{1.3}$$

де– загальна кількість спостережень величини,– логарифм за основою 10.

Також зустрічається записаним через натуральний логарифм:

$$m = 1 + 3{,}322 lnn. \tag{1.4}$$

де– загальна кількість спостережень величини,– натуральний логарифм.

Приклад 1.1. У результаті вибіркового аналізу добової виручки авіакомпанії дістали вибірку обсягом n=40 (млн грн):

0,87	0,94	0,99	0,90	0,90	0,87	0,85	0,87
0,90	0,94	0,87	0,87	0,82	0,90	0,94	0,90
0,85	0,85	0,87	0,94	0,81	0,82	0,87	0,97
0,90	0,94	0,85	0,81	0,87	0,85	0,90	0,82
0,99	0,90	0,94	0,82	0,97	0,81	0,85	0,87

Скласти: а) варіаційний ряд; б) дискретний статистичний розподіл вибірки; в) інтервальний розподіл.

Розв'язання. а) Випишемо різні значення варіант, які потрапили у вибірку:

0,87; 0,94; 0,99; 0,90; 0,85; 0,82; 0,81; 0,97.

Розмістивши їх у порядку зростання, дістанемо дискретний варі- аційний ряд:

0,81; 0,82; 0,85; 0,87; 0,90; 0,94; 0,97; 0,99.

б) Обчислимо частоту кожної варіанти із варіаційного ряду і складемо таблицю, відповідну табл. 1.1. Дістанемо дискретний статистичний розподіл вибірки (табл. 1.2).

Таблиця 1.2

x_i	0,81	0,82	0,85	0,87	0,90	0,94	0,97	0,99
n_i	3	4	6	9	8	6	2	2

в) За обсягом вибірки *n=40* визначаємо за формулою (1.2) орієнтовну кількість *m=6* частинних інтервалів в інтервальному статистичному розподілі, а за формулою (1.1) обчислюємо крок інтервалу

$$h = \frac{0,99 - 0,81}{6} = 0,03.$$

Тоді інтервальний варіаційний ряд запишеться у вигляді: (0,81; 0,84), (0,84; 0,87), (0,87; 0,90), (0,90; 0,93), (0,93; 0,96), (0,96; 0,99).

Підсумуємо частоти варіант, які потрапили в кожний із частинних інтервалів, при цьому частоти варіант, які збіглися з межами інтервалів (0,87 і 0,90), поділимо порівну між суміжними інтервалами.

Тоді інтервальний статистичний розподіл вибірки набере вигляду (табл. 1.3):

Таблиця 1.3

Інтервали	(0,81;0,84)	(0,84;0;87)	(0,87;0,90)	(0,90;0,93)	(0,93;0,96)	(0,96;0,99)
n_i^*	7	10	9	4	6	4

1.2. Полігон та гістограма. Емпірична функція розподілу

Для створення наочного уявлення про статистичні розподіли застосовуються полігон та гістограма частот.

Полігоном частот (або відносних частот) називається ламана лінія, відрізки якої сполучають на площині точки з координатами (x_i, n_i (або (x_i, ω_i)). Полігон застосовується також для графічного зображення інтервальних статистичних розподілів вибірки. У цьому випадку за абсциси x_i точок беруться центри частинних інтервалів.

Гістограма застосовується для графічного зображення інтервальних статистичних розподілів. Для побудови гістограми частот (або відносних частот) на осі абсцис відкладаються відрізки, що дорівнюють довжині (кроку) h частинних інтервалів, і на цих відрізках як на основах будуються прямокутники з висотами $\frac{n_i^*}{h}$ (або $\frac{\omega_i^*}{h} = \frac{n_i^*}{nh}$).

Оскільки величини $\frac{n_i^*}{h}(\frac{\omega_i^*}{h})$ є щільностями частот (відносних частот) на відповідних інтервалах, то при достатньо великих обсягах вибірки *n* (і відповідно малих h) гістограма може бути достатньо близьким статистичним аналогом щільності ймовірності досліджуваної ознаки *X* у генеральній сукупності.

Приклад 1.2. Побудувати полігон відносних частот для дискретного статистичного розподілу (табл. 1.2) та гістограму відносних частот для інтервального розподілу (табл. 1.3) із прикладу 1.1.

Розв'язання. Обчислимо відносні частоти $\omega_i^* = \frac{n_i}{n}$ для дискретного розподілу і щільності відносних частот $\frac{\omega_i^*}{h} = \frac{n_i^*}{nh}$ для інтегрального розподілу при *n=40*, *h=0,03*. Дістанемо відповідно розподіли (табл. 1.4 і 1.5).

Таблиця 1.4

x_i	0,81	0,82	0,85	0,87	0,90	0,94	0,97	0,99
n_i	0,075	0,1	0,15	0,225	0,2	0,15	0,05	0,05

Таблиця 1.5

Інтервали	(0,81;0,84)	(0,84;0;87)	(0,87;0,90)	(0,90;0,93)	(0,93;0,96)	(0,96;0,99)
n_i^*	5,83	8,33	7,5	3,33	5	3,33

Полігон та гістограму відносних частот побудовано відповідно на рис. 1.1 і 1.2.

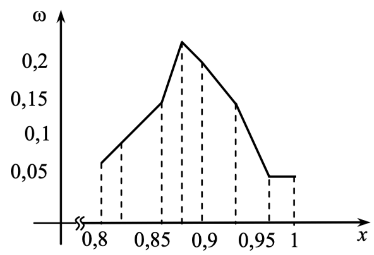


Рис. 1.1

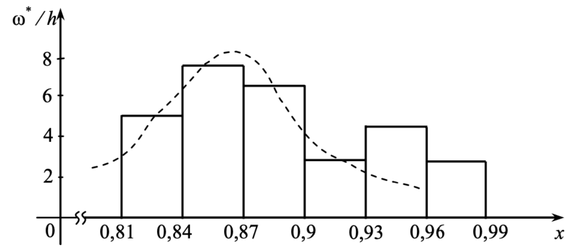


Рис. 1.2

За загальним виглядом гістограми можна скласти певне уявлення про графік щільності розподілу ознаки *X* у генеральній сукупності, наприклад припустити, що він має вигляд, зображений на рис. 1.2 пунктирною лінією (при згладженні деяких сплесків частот, викликаних, можливо, малим обсягом вибірки).

Розподіл ознаки у варіаційному ряду за накопиченими частотами (частинами) зображується за допомогою кумуляти. Кумулята або кумулятивна крива, на відміну від полігону, будується за накопиченими частотами або частотами. У цьому на осі абсцис поміщають значення ознаки, але в осі ординат — накопичені частоти чи частоти.

Кумулята - графічне порівняння двох або більше варіаційних розподілів з рівними чи нерівними інтервалами. Вона будується за кумулятивним розподілом накопичених частот, при цьому використовуються праві кінці інтервалів (рис. 1.3).

Огіва - це різновид кумулятивного розподілу. Вона є дзеркальним відображенням кумуляти. На осі ординат відкладаємо межі інтервалів, по осі абсцис - накопичені частоти (рис. 1.4).

На рис. 1.3 та 1.4 наведено відповідно графіки кумуляти та огіви для наступних даних:

Таблиця 1.6

Вибірка	Частота	Накопичувана частота
0,1	0	0
0,3	5	5
0,5	6	11
0,7	18	29
0,9	16	45
1,1	12	57
1,3	10	67
1,5	1	68
1,7	2	70

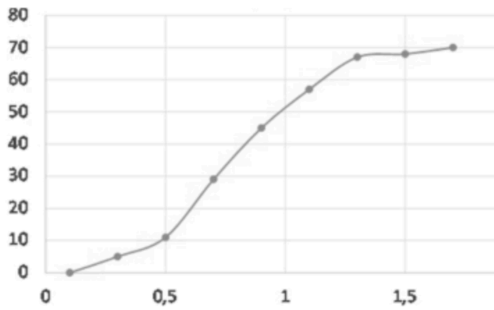


Рис. 1.3

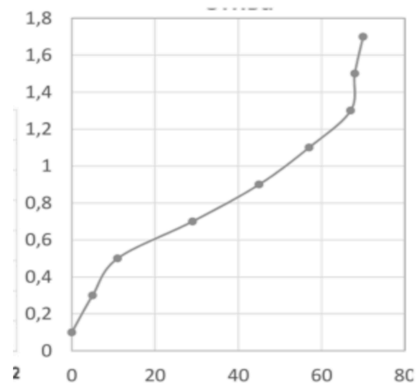


Рис. 1.4

Перейдемо до розгляду емпіричної функції розподілу, яка утворюється за даними вибірки і є наближеною оцінкою реальної або так званої теоретичної функції розподілу *F(x)* ознаки *X* генеральної сукупності.

Емпіричною функцією статистичного розподілу називається функція $F^*(x)$ яка для кожного значення x дорівнює відносній частоті події $X < x$, тобто:

$$F^*(x) = \frac{n_{x_i}}{n},$$

де n_{x_i}— сумарна частота всіх варіант x_i , менших за x.

Для дискретного статистичного розподілу (табл. 1.1) функція $F^*(x)$ набуває таких значень:

$F^*(x) = 0$, при $x \leq x_1$, оскільки в цьому випадку немає варіант, менших за x;

$F^*(x) = \frac{n_1}{n}$, при $x_1 < x \leq x_2$;

$F^*(x) = \frac{n_1+n_2}{n}$, при $x_2 < x \leq x_3$;

....

$F^*(x) = \frac{n_1+n_2+\cdots+n_{k-1}}{n}$, при $x_{k-1} < x \leq x_k$;

$F^*(x) = \frac{n_1+n_2+\cdots+n_k}{n} = 1$, при $x > x_k$, оскільки в цьому випадку додаються всі варіанти.

Графік функції $F^*(x)$ будується аналогічно графіку функції розподілу випадкової величини. Для неперервної ознаки X генеральної сукупності емпірична функція будується за інтервальним статистичним розподілом. Із цією метою обчислюються значення функції $F^*(x)$ у кількох точках, за які найчастіше беруть межі інтервалів. Знайдені точки (x_i, $F^*(x_i)$) наносять на графік і сполучають прямолінійними відрізками.

Якщо x_1 — ліва межа першого інтервалу, а x_{m+1} — права межа останнього інтервалу в інтервальному варіаційному ряду, то $F^*(x)$=0 при $x \leq x_1$ і $F^*(x)$=1 при $x > x_{m+1}$.

Приклад 1.3. Побудувати ем піричну функцію розподілу $F^*(x)$ за дискретним (табл. 1.2) та інтервальним (табл. 1.3) статистичними розподілами вибірки з прикладу 1.1.

Розв'язання. Для обчислення значень емпіричної функції скористаємось дискретним статистичним розподілом із відносними частотами ω_i (табл. 1.4), який знайдено у прикладі 1.2. Значення функції $F^*(x)$обчислюються простим нагромадженням відносних частот:

$$F^*(x)=\begin{cases}0 & \text{при} \quad x\le 0,81;\\ 0+0,075=0,075 & \text{при} \quad 0,81<x\le 0,82;\\ 0,075+0,1=0,175 & \text{при} \quad 0,82<x\le 0,85;\\ 0,175+0,15=0,325 & \text{при} \quad 0,85<x\le 0,87;\\ 0,325+0,225=0,55 & \text{при} \quad 0,87<x\le 0,90;\\ 0,55+0,2=0,75 & \text{при} \quad 0,90<x\le 0,94;\\ 0,75+0,15=0,9 & \text{при} \quad 0,94<x\le 0,97;\\ 0,9+0,05=0,95 & \text{при} \quad 0,97<x\le 0,99;\\ 0,95+0,05=1 & \text{при} \quad x>0,99.\end{cases}$$

Графік функції $F^*(x)$ подано на рис. 1.5 (східчаста розривна лінія).

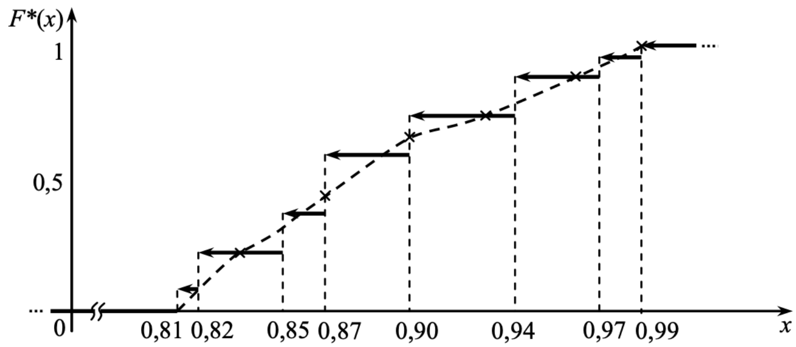


Рис.1.5

Для побудови функції $F^*(x)$за інтервальним статистичним розподілом (табл. 1.3) обчислимо значення функції на межах інтервалів:

x=0,81, $F^*(x)$=0, оскільки варіант, менших за 0,81, у вибірці немає;

x=0,84, $F^*(x)$=7/40=0,175, оскільки варіанти, менші за 0,84, містяться в першому інтервалі, і їх кількість 7;

$$x=0,87, \quad F^*(x)=\frac{1}{40}(7+10)=0,425;$$

$$x=0,90, \quad F^*(x)=\frac{1}{40}(7+10+9)=0,65;$$

$$x=0,93, \quad F^*(x)=\frac{1}{40}(7+10+9+4)=0,75;$$

$$x=0,96, \quad F^*(x)=\frac{1}{40}(7+10+9+4+6)=0,9;$$

$$x=0,99, \quad F^*(x)=\frac{1}{40}(7+10+9+4+6+4)=1.$$

Нанесемо знайдені точки на графік рис. 1.5 і сполучимо їх відрізками прямої. Дістанемо графік функції $F^*(x)$, зображений на рис. 1.5 пунктирною лінією.

ДОМАШНЄ ЗАВДАННЯ 1

Варіанти 1-5

У результаті вибіркового аналізу кількості пасажирів, що беруть квитки лише до проміжного пункту польоту, дістали такі дані:

12	14	19	15	14	18	13	16	17	12
20	17	15	13	17	16	20	14	14	13
17	16	15	19	16	15	18	17	15	14
16	15	15	18	15	15	19	14	16	18
18	15	15	17	15	16	16	14	14	17

Скласти:
1) варіаційний ряд;
2) дискретний статистичний розподіл вибірки;
3) групований варіативний ряд.
Побудувати:
1) полігон частот і відносних частот;
2) кумуляту та огіву вибірки;
3) гістограму групованого варіативного ряду
4) емпіричну функцію розподілу і зобразити її графічно.
Обчислити:
1) Моду та медіану вибірки;
2) Коефіцієнт асиметрії та ексцес вибірки.

Варіанти 6-10

У результаті вибіркового аналізу кількості проданих авіаквитків за добу дістали такі дані, шт.:

59	55	60	46	57	54	57	54	55	46
57	59	57	46	54	55	57	59	54	54
46	55	57	60	55	54	46	55	57	54

Скласти:
1) варіаційний ряд;
2) дискретний статистичний розподіл вибірки;
3) групований варіативний ряд.
Побудувати:
1) полігон частот і відносних частот;
2) кумуляту та огіву вибірки;
3) гістограму групованого варіативного ряду
4) емпіричну функцію розподілу і зобразити її графічно.
Обчислити:
1) Моду та медіану вибірки;
2) Коефіцієнт асиметрії та ексцес вибірки.

Варіанти 11-15

Щогодини реєструвались покази про витрати пального в польоті. Дістали такі дані, т:

6,80	6,50	6,53	7,30	6,95	6,90	6,93	6,72	6,54	6,79
6,80	6,75	6,62	6,65	7,00	6,83	6,70	6,69	6,54	6,97
6,82	6,67	6,91	6,93	6,76	6,71	6,66	6,76	6,79	6,82

Скласти:

1) варіаційний ряд;
2) дискретний статистичний розподіл вибірки;
3) групований варіативний ряд.
Побудувати:
1) полігон частот і відносних частот;
2) кумуляту та огіву вибірки;
3) гістограму групованого варіативного ряду
4) емпіричну функцію розподілу і зобразити її графічно.
Обчислити:
1) Моду та медіану вибірки;
2) Коефіцієнт асиметрії та ексцес вибірки.

Варіанти 16-20

Покази температури в Києві, t:

22	22	25	24	21	24	20	22	26	25	27	25	22
24	20	22	21	24	25	27	26	24	23	27	28	27
21	28	27	28	26	27	29	30	29	28	27	29	30

Скласти:
1) варіаційний ряд;
2) дискретний статистичний розподіл вибірки;
3) групований варіативний ряд.
Побудувати:
1) полігон частот і відносних частот;
2) кумуляту та огіву вибірки;
3) гістограму групованого варіативного ряду
4) емпіричну функцію розподілу і зобразити її графічно.
Обчислити:
1) Моду та медіану вибірки;
2) Коефіцієнт асиметрії та ексцес вибірки.

Варіанти 21-25

Бали, отримані студентом за деякі навчальні дисципліни:

82	82	85	94	91	84	90	92	76	75	77	75	92
94	80	92	91	84	75	77	76	79	83	87	78	81
81	78	77	88	86	77	89	90	89	78	87	79	75
81	78	77	87	86	77	89	90	89	79	87	89	76

Скласти:
1) варіаційний ряд;
2) дискретний статистичний розподіл вибірки;
3) групований варіативний ряд.
Побудувати:
1) полігон частот і відносних частот;
2) кумуляту та огіву вибірки;
3) гістограму групованого варіативного ряду
4) емпіричну функцію розподілу і зобразити її графічно.
Обчислити:
1) Моду та медіану вибірки;
2) Коефіцієнт асиметрії та ексцес вибірки.

Варіанти 26-30

Бали студента, отримані з дисципліни «Математична статистика»:

4 5 4 3 4 5 3 5 4 3 4 1 4 5 4 3 4 5 3 4 5 4 3 4 5 3
3 4 4 3 2 5 1 5 4 3 4 1 4 5 4 3 4 5 3 4 3 5 3 1 5 2
5 3 4 3 4 5 2 5 4 3 4 1 4 5 4 3 4 5 3 4 5 4 3 3 5 1

Скласти:

1) варіаційний ряд;
2) дискретний статистичний розподіл вибірки;
3) групований варіативний ряд.

Побудувати:

1) полігон частот і відносних частот;
2) кумуляту та огіву вибірки;
3) гістограму групованого варіативного ряду
4) емпіричну функцію розподілу і зобразити її графічно.

Обчислити:

1) Моду та медіану вибірки;
2) Коефіцієнт асиметрії та ексцес вибірки.

ПРАКТИЧНЕ ЗАНЯТТЯ 2

Точкові оцінки параметрів розподілу

2.1. Статистичні оцінки параметрів розподілу

Параметри досліджуваної ознаки *X* генеральної сукупності, з якої зроблено вибірку $x_1, x_2, \dots, x_n$ є величинами сталими, але їхні числові значення невідомі. Ці параметри оцінюються параметрами вибірки.

Нехай θ — невідомий параметр (наприклад, математичне сподівання, дисперсія тощо) досліджуваної ознаки *X*.

Вибірковою оцінкою параметра θ називається число θ^*, знайдене за даними вибірки або за статистичним розподілом, залежне від $x_1, x_2, \dots, x_n$, і таке, що наближено дорівнює оцінюваному параметру: $\theta^* \approx \theta$.

Ця оцінка, яка визначається у даній вибірці одним числом, точкою, називається *точковою*.

Оскільки θ^* є випадковою величиною, то вона може бути зміщеною і незміщеною. Коли математичне сподівання цієї оцінки точно дорівнює оцінюваному параметру θ, тобто:

$$M(\theta^*)=\theta,$$

то оцінка θ^*називається *незміщеною*.

Якщо $M(\theta^*) \neq 0$, то статистична оцінка θ^* називається *зміщеною* відносно параметра θ генеральної сукупності.

Різниця

$$\theta^* - \theta = \delta$$

називається *зміщенням* статистичної оцінки θ^*.

Точкова статистична оцінка θ^* називається *спроможною*, якщо, у разі необмеженого збільшення обсягу вибірки, θ^* наближається до оцінюваного параметра θ:

$$\lim_{n\to\infty} P\{|\theta^* - \theta| < \delta\} = 1.$$

Оцінюваний параметр може мати кілька точкових незміщених статистичних оцінок. Точкова статистична оцінка називається *ефективною*, коли при заданому обсязі вибірки вона має мінімальну дисперсію.

Для невідомого математичного сподівання *M(X)* досліджуваної ознаки *X* точковою оцінкою є *вибіркова середня*, яка позначається $\overline{x_{\text{В}}}$ і обчислюється за формулою

$$\overline{x}_{\text{В}} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i. \tag{2.1}$$

Вона задовольняє всі три перелічені умови. Для невідомої дисперсії *D(X)* спроможною оцінкою є *вибіркова дисперсія* $D_{\text{В}}$, обчислювана за формулою

$$D_{\text{В}} = \frac{1}{n}\sum_{i=1}^{k} (x_i - \overline{x}_{\text{В}})^2 n_i. \tag{2.2}$$

Проте ця оцінка є зміщеною. Тому для оцінки невідомої дисперсії застосовують так звану виправлену дисперсію, яку позначають s^2 і обчислюють за формулою

$$s^2 = \frac{n}{n-1} D_{\text{В}} = \frac{1}{n-1}\sum_{i=1}^{k} (x_i - \overline{x}_{\text{В}})^2 n_i\ . \tag{2.3}$$

Оцінкою для середнього квадратичного відхилення правлене вибіркове середнє квадратичне відхилення

$$\sigma_В = \sqrt{s^2} \qquad (2.4)$$

Для неперервної ознаки X, для якої побудовано інтервальний статистичний розподіл, точкові оцінки невідомих математичного сподівання і дисперсії також знаходять за формулами (2.2) і (2.3), в які замість варіант x_i підставляють середини (центри) x_i^* інтервалів.

Приклад 2.1. Знайти точкові оцінки математичного сподівання, дисперсії і середнього квадратичного відхилення добової виручки авіакомпанії за даними вибірки з прикладу 1.1.

Розв'язання. За дискретним статистичним розподілом (табл. 1.2) обчислюємо вибіркову середню $\overline{x_В}$, виправлену вибіркову дисперсію s^2 і виправлене вибіркове середнє квадратичне відхилення $\sigma_В$, використовуючи формули (2.2) — (2.4):

$$\overline{x}_В = \frac{1}{40}(0,81\cdot 3 + 0,82\cdot 4 + 0,85\cdot 6 + 0,87\cdot 9 + 0,9\cdot 8 +$$
$$+0,94\cdot 6 + 0,97\cdot 2 + 0,99\cdot 2) = 0,885;$$

$$s^2 = \frac{1}{39}[(0,81-0,885)^2\cdot 3 + (0,82-0,885)^2\cdot 4 +$$
$$+(0,85-0,885)^2\cdot 6 + (0,87-0,885)^2\cdot 9 +$$
$$+(0,9-0,885)^2\cdot 8 + (0,94-0,885)^2\cdot 6 +$$
$$+(0,97-0,885)^2\cdot 2 + (0,99-0,885)^2\cdot 2] = 0,0025;$$
$$\sigma_В = 0,05.$$

Знайдемо оцінки $\overline{x_В}$, s^2 і $\sigma_В$ за інтервальним статистичним розподілом (табл. 1.3), замінюючи у формулах (2.2) і (2.3) x_i центрами частинних інтервалів x_i^*. Для цього перейдемо спочатку від інтервального статистичного розподілу до дискретного, який набирає такого вигляду:

x_i^*	0,825	0,855	0,885	0,915	0,945	0,975
n^*_i	7	10	9	4	6	4

Тоді

$$\overline{x}_В = \frac{1}{40}(0,825\cdot 7 + 0,855\cdot 10 + 0,885\cdot 9 + 0,915\cdot 4 +$$
$$+0,945\cdot 6 + 0,975\cdot 4) = 0,888;$$

$$s^2 = \frac{1}{39}[(0,825-0,888)^2\cdot 7 + (0,855-0,888)^2\cdot 10 + (0,885-0,888)^2\cdot 9 +$$
$$+(0,915-0,888)^2\cdot 4 + (0,945-0,888)^2\cdot 6 + (0,975-0,888)^2\cdot 4] = 0,0023;$$
$$\sigma_В = 0,048.$$

2.2. Метод моментів

Метод моментів ґрунтується на тому, що невідомі параметри теоретичного розподілу (розподілу генеральної сукупності) визначаються із рівнянь, які добуваються прирівнюванням важливіших числових характеристик (моментів) теоретичного розподілу відповідним числовим характеристикам емпіричного розподілу.

Так, нехай заданий, наприклад, вид теоретичного розподілу, який визначається невідомим параметром. Для знаходження одного параметра необхідне одне рівняння відносно даного параметра. Для цього використовується момент 1-го порядку (математичне сподівання) теоретичного розподілу і відповідна числова характеристика емпіричного розподілу – вибіркове середнє.

Знаходимо математичне сподівання:

$$M(X) = \int_{-\infty}^{\infty} x dF(x, \theta) = M(\theta)$$

і вибіркове середнє:

$$\overline{x}_{\text{В}} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i.$$

Порівнюючи їх, одержуємо рівняння для визначення оцінки невідомого параметра:

$$M(\theta) = \overline{x_{\text{В}}}. \quad (2.5)$$

Для знаходження оцінок двох невідомих параметрів θ_1, θ_2 звичайно беруть математичне сподівання і дисперсію теоретичного розподілу та відповідні їм числові характеристики емпіричного розподілу – вибіркове середнє і вибіркову дисперсію. Одержують два рівняння:

$$M(\theta_1, \theta_2) = \overline{x_{\text{В}}},$$
$$\sigma^2(\theta_1, \theta_2) = s^2. \quad (2.6)$$

В деяких підручниках, зокрема Гмурмана та Кремера, в системі (2.6) дисперсію теоретичного розподілу (генеральну дисперсію) оцінюють за допомогою вибіркової дисперсії $D_{\text{В}}$.

Розв'язуючи цю систему, знаходять відповідні оцінки $\tilde{\theta}_1, \tilde{\theta}_2$. Оцінки методу моментів звичайно є *слушними*, однак за ефективністю вони не є «найкращими». Тим не менш, метод моментів часто використовується на практиці, оскільки приводить до порівняно простих обчислень.

Приклад 2.2. Знаходження методом моментів оцінки невідомого параметра експоненціального розподілу.

Статистична модель. Генеральна сукупність має експоненціальний розподіл із щільністю розподілу $f(x, \lambda) = \lambda e^{-\lambda x}$ $(x \geq 0)$, де – невідомий параметр. Вибіркові значення $x_1, x_2, \ldots, x_n$, узяті із однієї і тієї ж генеральної сукупності. Знайдемо методом моментів оцінку $\tilde{\lambda}$ невідомого параметра розподілу .

Розв'язання. Визначимо математичне сподівання експоненціального розподілу:

$$M(\lambda) = \lambda \int_0^{\infty} x e^{-\lambda x} dx = \frac{1}{\lambda}.$$

Далі по вибірці знаходимо вибіркове середнє емпіричного розподілу :

$$\overline{x}_{\text{В}} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i.$$

Із рівняння

$$M(\lambda) = \overline{x_{\text{В}}} = \frac{1}{\lambda}$$

знаходимо оцінку параметра:

$$\tilde{\lambda} = \frac{1}{\overline{x_{\text{В}}}}.$$

Приклад 2.3. Особливо зручне застосування методу моментів, коли шукані параметри розподілу самі є деякими числовими характеристиками. Знайдемо методом моментів оцінки невідомих параметрів a і нормального розподілу.

Статистична модель. Вибірка $x_1, x_2, \dots, x_n$, одержана із генеральної сукупності з нормальним розподілом із щільністю розподілу:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, (-\infty < x < \infty),$$

де a і σ – невідомі параметри розподілу.

Знайдемо методом моментів оцінки $\tilde{a}$ і $\tilde{\sigma}$ невідомих параметрів a і σ.

Розв'язання. Визначимо математичне сподівання і дисперсію розподілу:

$$a = \int_{-\infty}^{\infty} xf(x)dx, \sigma^2 = \int_{-\infty}^{\infty} (x-a)^2 f(x)dx.$$

За даними вибірки знаходимо вибіркові числові характеристики $\overline{x_\text{В}}$, s і прирівнюємо їх до відповідних числових характеристик теоретичного розподілу:

$$a = \overline{x_\text{В}}, \sigma = s.$$

У результаті одержуємо шукані оцінки параметрів: $\tilde{a}$ і $\tilde{\sigma}$.

2.3. Метод максимальної правдоподібності

Одним із найбільш універсальних методів одержання оцінок параметрів розподілів генеральної сукупності є метод *максимальної правдоподібності.*

Основу метода складає *функція правдоподібності*, яка виражає щільність ймовірності (ймовірність) сумісної появи результатів вибірки $x_1, x_2, \dots, x_n$:

$$L(x_1, x_2, \dots, x_n, \theta) = f(x_1, \theta) f(x_2, \theta) \dots f(x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta) \quad (2.7)$$

або

$$L(x_1, x_2, \dots, x_n, \theta) = p(x_1, \theta) p(x_2, \theta) \dots p(x_n, \theta) = \prod_{i=1}^{n} p(x_i, \theta). \quad (2.7')$$

Згідно з методом максимальної правдоподібності за оцінку невідомого параметра приймається таке його значення , яке максимізує функцію $L(\theta)$. Величина , при якій функція правдоподібності досягає максимального значення, називається *оцінкою максимальної правдоподібності*.

Природність такого підходу до визначення статистичних оцінок випливає із смислу функції правдоподібності, яка при кожному фіксованому значенні є *мірою правдоподібності* одержання вибірки. Оцінка є такою, що вибірка, яка одержана у результаті спостережень, є найбільш вірогідною. Знаходження оцінки спрощується, якщо максимізувати не саму функцію $L(\theta)$, а $\ln L(\theta)$, оскільки максимум обох функцій досягається при одному і тому ж значенні .

Для одержання оцінки максимальної правдоподібності треба розв'язати рівняння:

$$\frac{d\ln L(\theta)}{d\theta} = 0.$$

Якщо потрібно оцінити не один, а декілька параметрів $\theta_1, \theta_2, \dots, \theta_m$, то оцінка максимальної правдоподібності цих параметрів знаходиться із системи рівнянь:

$$\frac{\partial L(\theta_1, \theta_2, \dots, \theta_m)}{\partial \theta_i} \; (i=1, \dots, m).$$

Перевага методу максимальної правдоподібності полягає у тому, що для широкого класу розподілів він приводить до оцінок, які є слушними, асимптотично ефективними, мають асимптотично нормальний розподіл і, якщо для параметра існує ефективна оцінка, то рівняння правдоподібності має єдиний розв'язок, який співпадає з нею. Однак оцінка максимальної правдоподібності може виявитись зсуненою.

Приклад 2.4. Методом максимальної правдоподібності оцінимо параметр розподілу Пуассона.

Статистична модель. Генеральна сукупність має розподіл Пуассона.

$$P_n(X = x_i) = \frac{\lambda^{x_i}}{x_i!} e^{-\lambda},$$

де *n* – кількість випробувань у кожній серії, x_i – кількість появ події у i-й серії *(i=1,…,n)*. Знайдемо оцінку невідомого параметра по вибірці .

Розв'язання. Складемо функцію правдоподібності.

$$L(x_1, x_2, \ldots, x_n, \lambda) = p(x_1, \lambda) p(x_2, \lambda) \ldots p(x_n, \lambda)$$

Визначимо логарифм цієї функції

$$lnL(\lambda) = \sum x_i ln\lambda - n\lambda - \ln(x_1! \ldots x_n!) =$$
$$= \frac{\lambda^{x_1}}{x_1!} e^{-\lambda} \cdot \ldots \cdot \frac{\lambda^{x_n}}{x_n!} e^{-\lambda} = \frac{\lambda^{\sum x_i}}{x_1! \ldots x_n!} e^{-n\lambda}.$$

Прирівнюючи похідну цієї функції по до нуля, одержуємо рівняння правдоподібності

$$\frac{dlnL(\lambda)}{d\lambda} = \frac{\sum x_i}{\lambda} - n = 0.$$

Розв'язуючи це рівняння відносно , знаходимо:

$$\tilde{\lambda} = \frac{\sum x_i}{n} = \bar{x}_B.$$

Оскільки

$$\frac{d^2 lnL(\lambda)}{d\lambda^2} = -\frac{\sum x_i}{\lambda^2} < 0,$$

то $\tilde{\lambda} = \bar{x}_B$ є точкою максимуму функції $lnL(\lambda)$.

З вищенаведеного одержуємо, що є оцінка максимальної правдоподібності параметра розподілу Пуассона.

Приклад 2.5. Методом максимальної правдоподібності знайдемо оцінку параметрів і нормального розподілу.

Статистична модель. Вибірка $x_1, x_2, \ldots, x_n$ одержана із генеральної сукупності, розподіленої за нормальним законом з відповідними параметрами. Знайдемо оцінки параметрів і методом максимальної правдоподібності.

Розв'язання. Записуємо функцію правдоподібності:

$$\ln \mathrm{L}(\mathrm{a}, \sigma) = \frac{n}{2}(ln2n + ln\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - a)^2.$$

Диференціюючи $\ln \mathrm{L}(\mathrm{a}, \sigma)$ по a і σ^2 одержуємо систему рівнянь:

$$\sum_{i=1}^{n} (x_i - a) \quad = 0,$$
$$-\frac{n}{2} + \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - a)^2 = 0.$$

.

Звідки знаходимо оцінки:

$$\tilde{a} = \overline{x_{\mathrm{B}}}, \ , \tilde{\sigma} = s.$$

Ці оцінки співпадають з оцінками методу моментів. Вони слушні, причому $\tilde{a}$ є незміщеною, а $\tilde{\sigma}$ – зміщеною і, як було сказано раніше, – ефективною (асимптотично ефективною).

2.4. Метод найменших квадратів

Сутність методу найменших квадратів полягає у тому, що оцінки невідомих параметрів розподілу $\theta = (\theta_1, \theta_2, \ldots, \theta_m)$ визначаються із умови мінімізації суми квадратів відхилень вибіркових даних x_i *(i=1,…,n)* від оцінки, що визначається. Наприклад, знайдемо оцінку за методом найменших квадратів для генеральної середньої. Згідно з цим методом оцінку знаходимо з умови

$$s(\tilde{\theta}) = \sum_{i=1}^{n}(x_i - \theta)^2 \rightarrow min. \tag{2.8}$$

Використовуючи необхідну умову екстремуму функції, прирівнюємо до нуля похідну

$$\frac{ds}{d\theta} = -2\sum_{i=1}^{n}(x_i - \theta) = 0 \text{ або } \sum_{i=1}^{n} x_i - n\theta = 0.$$

Звідки

$$\tilde{\theta} = \overline{x_{\text{В}}}, \tag{2.9}$$

Отже, оцінка генеральної середньої є вибіркова середня $\overline{x_{\text{В}}}$,.

Метод найменших квадратів має широке застосування у практиці статистичних досліджень, оскільки, по-перше, не потребує знання закону розподілу вибіркових даних; по-друге, для нього досить добре розроблений математичний апарат чисельної реалізації.

Метод найменших квадратів застосовується у моделях кореляційного і регресійного аналізу.

2.5. Метод мінімуму χ^2

Статистична модель. Нехай $x_1, x_2, \dots, x_n$ – вибірка незалежних спостережень над випадковою величиною *X*, розподіл якої належить класу розподілів $F(x, \theta)$, який залежить від невідомого параметра $\theta = (\theta_1, \theta_2, \dots, \theta_m)$. Вибірка одержана при деяких конкретних значеннях цього параметра.

Припустимо, множина значень *X* розбита на m інтервалів $\Delta_1, \dots, \Delta_m$, які не перетинаються. Позначимо через n_iчисло спостережень у вибірці, які потрапили у інтервал Δ_i. Якщо множина значень *X* скінченна, тобто величина приймає лише скінченне число значень, то можна вважати, що Δ_i - одноточкова множина. Таким чином, проведено групування результатів спостережень, у результаті чого одержано інтервальний варіаційний ряд.

Позначимо через $p_i(\theta) = F(\Delta_i, \theta), i = 1, \dots, m$ ймовірність попадання значень випадкової величини *X* у *i*-й інтервал. Складемо величину

$$\chi^2 = \sum_{i=1}^{n} \frac{(n_i - np_i(\theta))^2}{np_i(\theta)} \tag{2.10}$$

яка служить оцінкою параметра на основі даної вибірки. Оскільки ймовірності $p_i(\theta)$ є функціями вибіркових значень, то і величина χ^2 є функцією вибірки.

Оцінка $\tilde{\theta}(x_1, x_2, \dots, x_n)$ називається *оцінкою за методом мінімуму* χ^2, якщо вона одержана мінімізацією величини χ^2. Якщо θ - *m*-вимірний параметр, то для знаходження оцінки за методом мінімуму одержуємо систему рівнянь

$$\sum_{i=1}^{n}\left(\frac{n_i - np_i(\theta)}{p_i(\theta)} + \frac{(n_i - np_i(\theta))^2}{2np_i(\theta)}\right)\frac{\partial p_i(\theta)}{\partial \theta_k},\ k = 1, \dots, m \tag{2.11}$$

За своїми асимптотичними властивостями оцінки, які добуваються за методом χ^2, дуже близькі до оцінок максимальної правдоподібності. Наприклад, при деяких умовах з імовірністю 1 є лише один слушний корінь відповідних рівнянь і він дійсно дає абсолютний мінімум величини. При достатньо великому об’ємі вибірки другим членом в (2.11) можна знехтувати. Дійсно, згідно теореми Бернуллі, при відносна частота $\frac{n_i}{n}$ по ймовірності збігається до ймовірності , тому другий член в (2.11) прямує до 0. Отже, систему рівнянь (2.11) можна замінити близькою до неї системою рівнянь

$$\sum_{i=1}^{n} \frac{n_i - np_i(\theta)}{p_i(\theta)} \cdot \frac{\partial p_i(\theta)}{\partial \theta_k}, \qquad k = 1, \dots, m.$$

Ця система еквівалентна системі

$$\sum_{i=1}^{n} \frac{n_i}{p_i(\theta)} \cdot \frac{\partial p_i(\theta)}{\partial \theta_k}, \; k = 1, \dots, m. \qquad (2.12)$$

Дійсно

$$\sum_{i=1}^{n} \frac{n_i - np_i(\theta)}{p_i(\theta)} \cdot \frac{\partial p_i(\theta)}{\partial \theta_k} = \sum_{i=1}^{n} \frac{n_i}{p_i(\theta)} \cdot \frac{\partial p_i(\theta)}{\partial \theta_k} - n \sum_{i=1}^{n} \frac{\partial p_i(\theta)}{\partial \theta_k} = \sum_{i=1}^{n} \frac{n_i}{p_i(\theta)} \cdot \frac{\partial p_i(\theta)}{\partial \theta_k} - n \frac{\partial}{\partial \theta_k} \left(\sum_{i=1}^{n} p_i(\theta)\right) = \sum_{i=1}^{n} \frac{n_i}{p_i(\theta)} \cdot \frac{\partial p_i(\theta)}{\partial \theta_k} - n \frac{\partial}{\partial \theta_k}(1) = \sum_{i=1}^{n} \frac{n_i}{p_i(\theta)} \cdot \frac{\partial p_i(\theta)}{\partial \theta_k}.$$

Розбиття гіпотетичного простору X на інтервалів , що не перетинаються, породжує дискретну випадкову змінну, функція правдоподібності якої є

$$L = \prod_{i=1}^{m} p_i^{m_i}. \qquad (2.13)$$

Отже, система рівнянь (2.11) еквівалентна системі рівнянь

$$\frac{\partial L}{\partial \theta_k} = 0, k = 1, \dots, m. \qquad (2.14)$$

ДОМАШНЄ ЗАВДАННЯ 2

Варіанти 1-5

Задано вибірку:

12	14	19	15	14	18	13	16	17	12
20	17	15	13	17	16	20	14	14	13
17	16	15	19	16	15	18	17	15	14
16	15	15	18	15	15	19	14	16	18
18	15	15	17	15	16	16	14	14	17

Обчислити
1) вибіркове середнє;
2) вибіркову дисперсію;
3) виправлену дисперсію.

Оцінити параметри розподілу генеральної сукупності, з якої зроблено дану вибірку, вважаючи, що вона має нормальний розподіл. Використовувати
1) метод моментів;
2) метод максимальної правдоподібності.

Записати функцію щільності для розподілу генеральної сукупності, враховуючи попередні оцінки. Намалювати графік.

Варіанти 6-10

Задано вибірку:

59	55	60	46	57	54	57	54	55	46
57	59	57	46	54	55	57	59	54	54
46	55	57	60	55	54	46	55	57	54

Обчислити
1) вибіркове середнє;
2) вибіркову дисперсію;
3) виправлену дисперсію.

Оцінити параметри розподілу генеральної сукупності, з якої зроблено дану вибірку, вважаючи, що вона має нормальний розподіл. Використовувати
1) метод моментів;

2) метод максимальної правдоподібності.
Записати функцію щільності для розподілу генеральної сукупності, враховуючи попередні оцінки. Намалювати графік.

Варіанти 11-15

Задано вибірку:

6,80	6,50	6,53	7,30	6,95	6,90	6,93	6,72	6,54	6,79
6,80	6,75	6,62	6,65	7,00	6,83	6,70	6,69	6,54	6,97
6,82	6,67	6,91	6,93	6,76	6,71	6,66	6,76	6,79	6,82

Обчислити
1) вибіркове середнє;
2) вибіркову дисперсію;
3) виправлену дисперсію.
Оцінити параметри розподілу генеральної сукупності, з якої зроблено дану вибірку, вважаючи, що вона має нормальний розподіл. Використовувати
1) метод моментів;
2) метод максимальної правдоподібності.
Записати функцію щільності для розподілу генеральної сукупності, враховуючи попередні оцінки. Намалювати графік.

Варіанти 16-20

Задано вибірку:

82	82	85	94	91	84	90	92	76	75	77	75	92
94	80	92	91	84	75	77	76	79	83	87	78	81
81	78	77	88	86	77	89	90	89	78	87	79	75
81	78	77	87	86	77	89	90	89	79	87	89	76

Обчислити
1) вибіркове середнє;
2) вибіркову дисперсію;
3) виправлену дисперсію.
Оцінити параметри розподілу генеральної сукупності, з якої зроблено дану вибірку, вважаючи, що вона має нормальний розподіл. Використовувати
1) метод моментів;
2) метод максимальної правдоподібності.
Записати функцію щільності для розподілу генеральної сукупності, враховуючи попередні оцінки. Намалювати графік.

Варіанти 21-25

Задано вибірку:

22	22	25	24	21	24	20	22	26	25	27	25	22	23
24	20	22	21	24	25	27	26	24	23	27	28	27	24
21	28	27	28	26	27	29	30	29	28	27	29	30	27

Обчислити
1) вибіркове середнє;
2) вибіркову дисперсію;
3) виправлену дисперсію.
Оцінити параметри розподілу генеральної сукупності, з якої зроблено дану вибірку, вважаючи, що вона має нормальний розподіл. Використовувати
1) метод моментів;

2) метод максимальної правдоподібності.
Записати функцію щільності для розподілу генеральної сукупності, враховуючи попередні оцінки. Намалювати графік.

Варіанти 26-30

Задано вибірку:

4 5 4 3 4 5 3 5 4 3 4 1 4 5 4 3 4 5 3 4 5 4 3 4 5 3
3 4 4 3 2 5 1 5 4 3 5 2 4 5 4 3 4 2 3 4 3 5 3 1 5 1
5 3 4 3 4 5 2 5 4 3 4 1 4 5 4 3 4 5 3 4 5 4 3 3 5 2

Обчислити
1) вибіркове середнє;
2) вибіркову дисперсію;
3) виправлену дисперсію.

Оцінити параметри розподілу генеральної сукупності, з якої зроблено дану вибірку, вважаючи, що вона має нормальний розподіл. Використовувати
1) метод моментів;
2) метод максимальної правдоподібності.

Записати функцію щільності для розподілу генеральної сукупності, враховуючи попередні оцінки. Намалювати графік.

ПРАКТИЧНЕ ЗАНЯТТЯ 3

Інтервальні оцінки параметрів розподілу

3.1. Інтервальні статистичні оцінки. Довірча ймовірність та довірчий інтервал

Точкові оцінки параметрів розподілу не дають можливості зробити висновки про їхню точність та надійність, оскільки є по суті випадковими величинами. Щоб мати уявлення про точність та надійність оцінки, у математичній статистиці користуються довірчими ймовірностями та довірчими інтервалами.

Інтервал ($-\theta^*;\theta^*$) називається *довірчим* для оцінюваного параметра θ з довірчою ймовірністю (надійністю) γ, якщо він покриває цей параметр θ з імовірністю, не меншою за γ, тобто

$$P\{\theta_1^* < \theta < \theta_2^*\} \geq \gamma.$$

Межі $-\theta^*$ і θ^* довірчого інтервалу називаються довірчими межами для оцінюваного параметра θ. Статистична оцінка, що визначається двома числами — кінцями інтервалів, називається інтервальною. Надійність γ вибирається, як правило, достатньо великою, наприклад, 0,9, 0,95 або 0,99.

Як приклад наведемо схему побудови довірчого інтервалу для невідомого параметра а нормально розподіленої ознаки X. При відомому параметрі σ довірчим для а є інтервал

$$\left(\overline{x}_{\text{В}} - t\frac{\sigma}{\sqrt{n}};\quad \overline{x}_{\text{В}} + t\frac{\sigma}{\sqrt{n}}\right), \tag{3.1}$$

де $\overline{x_{\text{В}}}$ — вибіркова середня; n — об’єм вибірки; t — розв’язок рівняння $2\Phi(t)=\gamma$ при заданій надійності γ. Для функції Лапласа $\Phi(x)$ складено таблицю значень (дод. 2 Гмурман), за допомогою якої за заданим значенням γ/2 функції знаходять значення аргументу t .

Приклад 3.1. Маємо такі дані про розміри виручки (у тис. грн) на 30 випадково вибраних рейсах:

42 24 49 67 45 27 39 21 58 40
28 78 44 66 20 62 70 81 7 68
94 76 63 88 65 14 46 20 72 91

Побудувати інтервальний статистичний розподіл із довжиною кроку h=20 (тис. грн). Припускаючи, що виручка розподілена за нормальним законом, знайти з надійністю γ=0,95 довірчий інтервал для невідомого математичного сподівання, якщо середнє квадратичне відхилення σ=25 (тис. грн).

Розв’язання. Інтервальний статистичний розподіл буде таким:

Інтервали	[7;27)	[27;47)	[47;67)	[67;87)	[87;107)
n_i	6	8	6	7	3

Для визначення $\overline{x_{\text{В}}}$ необхідно побудувати дискретний статистичний розподіл, що має такий вигляд:

x_i^*	17	37	57	77	97
n_i	6	8	6	7	3

де x_i^* — середини відповідних інтервалів.

Тоді

$$\overline{x}_{\text{В}}=\frac{\sum_i x_i^* n_i}{n}=\frac{17\cdot 6+37\cdot 8+57\cdot 6+77\cdot 7+97\cdot 3}{30}\approx 52,3.$$

Для побудови довірчого інтервалу із заданою надійністю γ=0,95 необхідно знайти *t*: *Ф(t)=γ/2* =0,475. Звідси *t*=1,96 (дод. 2 Гмурмана).

Знаходимо межі інтервалу:

$$\overline{x}_{\text{В}}-\frac{t\sigma}{\sqrt{n}}=52,3-\frac{1,96\cdot 25}{\sqrt{30}}=52,3-\frac{1,96\cdot 25}{5,5}=43,4$$

$$\overline{x}_{\text{В}}+\frac{t\sigma}{\sqrt{n}}=52,3+\frac{1,96\cdot 25}{\sqrt{30}}=52,3+\frac{1,96\cdot 25}{5,5}=61,2.$$

Отже, з ймовірністю, не меншою за 0,95, середнє значення виручки на даному рейсі міститься в межах від 43,4 до 61,2 тис. грн.

Приклад 3.2. Визначити обсяг вибірки *n* нормально розподіленої випадкової величини, за якого похибка *ε*=0,01 при обчисленні математичного сподівання гарантується з ймовірністю 0,999, якщо середнє квадратичне відхилення σ=5.

Розв'язання. Оскільки $\varepsilon=t\sigma/\sqrt{n}$, то звідси $n=t^2\sigma^2/\varepsilon^2$. Значення *t* знаходимо з рівняння *Ф(t)*=γ/2=^{0,999}/2=0,4995. За таблицею функції Лапласа (дод. 2 Гмурмана) знаходимо *t*=3,4, отже, *n*=2890000.

Оскільки при розв'язуванні задач математичної статистики ми, як правило, маємо в розпорядженні лише вибірку, тобто параметр σ невідомий, то для знаходження наближених меж довірчого інтервалу замість σ використовується його вибіркова точкова оцінка σ_B. При великих обсягах вибірки така заміна приводить до цілком прийнятних похибок.

При малих вибірках для оцінювання математичного сподівання якщо невідоме значення середнього квадратичного відхилення, застосовується випадкова величина t , що має розподіл Стьюдента з $k=n-1$ ступенями свободи. При цьому для кожного $\gamma:(0<\gamma<1)$ існує єдине значення t_γ , яке визначається за таблицею t-розподілу (розподілу Стьюдента) з $n-1$ ступенями свободи (дод. 4).

Обчисливши за даним статистичним розподілом $\overline{x_{\text{В}}}$, s^2 і визначивши за таблицею розподілу Стьюдента значення t_γ, будуємо довірчий інтервал (дод. 3 Гмурмана)

$$\left(\overline{x}_{\text{В}}-\frac{t_\gamma\sqrt{s^2}}{\sqrt{n}};\ \overline{x}_{\text{В}}+\frac{t_\gamma\sqrt{s^2}}{\sqrt{n}}\right) \tag{3.2}$$

Приклад 3.3. Побудувати довірчий інтервал з надійністю γ=0,99 для математичного сподівання максимальної швидкості, м/с, літака, припускаючи, що вона має нормальний розподіл, за даними 20 вимірювань:

504,5 485,2 512,0 497,1 502,4 488,3 491,9 489,3 502,7 509,2
514,3 502,3 497,8 511,3 515,7 506,3 499,2 497,4 485,2 507,8

Розв'язання. Для побудови довірчого інтервалу знайдемо середнє вибіркове і виправлену дисперсію.

$$\overline{x}_{\text{в}} = \frac{1}{20}(504,5 + 485,2 + 512,0 + 497,1 + 502,4 + 488,3 + 491,9 +$$
$$+489,3 + 502,7 + 509,2 + 514,3 + 502,3 + 497,8 + 511,3 + 515,7 +$$
$$+506,3 + 499,2 + 497,4 + 485,2 + 507,8) \approx 501;$$
$$s^2 = \frac{1}{19}(3,5^2 + 15,8^2 + 11^2 + 3,9^2 + 1,4^2 + 12,7^2 + 9,1^2 + 11,7^2 +$$
$$+1,7^2 + 8,2^2 + 13,3^2 + 1,3^2 + 3,2^2 + 10,3^2 + 14,7^2 + 5,3^2 + 1,8^2 +$$
$$+3,6^2 + 15,8^2 + 6,8^2) = \frac{1}{19}(12,25 + 249,64 + 121 + 15,21 + 1,96 +$$
$$+161,29 + 82,81 + 136,89 + 2,89 + 67,24 + 176,89 + 1,69 + 10,24 +$$
$$+106,09 + 216,09 + 28,09 + 3,24 + 12,96 + 249,64 + 46,24) \approx 89,6.$$

За таблицею розподілу Стьюдента (дод. 4) за даною надійністю γ=0,99 і кількістю ступенів свободи $k=n-1$=20–1=19 знаходимо значення t_γ=2,861. Обчислимо межі довірчого інтервалу:

$$\overline{x}_{\text{в}} - \frac{t_\gamma\sqrt{s^2}}{\sqrt{n}} = 501 - \frac{2,861\cdot\sqrt{89,6}}{\sqrt{20}} = 501 - \frac{2,861\cdot 9,47}{4,47} = 495,$$

$$\overline{x}_{\text{в}} + \frac{t_\gamma\sqrt{s^2}}{\sqrt{n}} = 501 + \frac{2,861\cdot\sqrt{89,6}}{\sqrt{20}} = 501 + \frac{2,861\cdot 9,47}{4,47} = 507.$$

Отже, з ймовірністю, не меншою за 0,99, можна стверджувати, що математичне сподівання максимальної швидкості даного літака міститься в інтервалі (495; 507).

При невідомому математичному сподіванні довірчий інтервал із надійністю γ для дисперсії $D(X)$ нормального розподілу має вигляд

$$\left(\frac{(n-1)s^2}{\chi_1^2(\gamma)};\ \frac{(n-1)s^2}{\chi_2^2(\gamma)}\right), \tag{3.3}$$

а для середнього квадратичного відхилення дістаємо:

$$\left(\frac{\sqrt{(n-1)s^2}}{\chi_1(\gamma)};\ \frac{\sqrt{(n-1)s^2}}{\chi_2(\gamma)}\right). \tag{3.4}$$

Додатні числа $\chi_1(\gamma)$ та $\chi_2(\gamma)$ визначаються з рівностей

$$F(\chi_1^2(\gamma)) = \frac{1-\gamma}{2};\ F(\chi_2^2(\gamma)) = \frac{1+\gamma}{2}, \tag{3.5}$$

де $F(x)$ — функція розподілу випадкової ведичини χ^2 (хі-квадрат) з ($n-1$) ступенями свободи (дод. 5 Гмурмана).

Приклад 3.4. Під час вибіркового аналізу максимального завантаження літака при певній кількості пального, дістали такі дані, т:

78 80 85 88 89 90 88 93

88 85 93 89 80 90 85 89

Припускаючи, що максимальне завантаження літака має нормальний розподіл, знайти довірчий інтервал для дисперсії та середнього квадратичного відхилення з надійністю γ=0,95.

Розв'язання. Для побудови довірчих інтервалів необхідно знайти s^2. Для цього обчислимо значення $\overline{x_B}$.

$$\overline{x}_B = \frac{1}{16}(78 + 80 \cdot 2 + 85 \cdot 3 + 88 \cdot 3 + 89 \cdot 3 + 90 \cdot 2 + 92 + 93) = 86{,}8;$$

$$s^2 = \frac{1}{15}(8{,}8^2 + 6{,}8^2 \cdot 2 + 1{,}8^2 \cdot 3 + 1{,}2^2 \cdot 3 + 2{,}2^2 \cdot 3 +$$
$$+ 3{,}2^2 \cdot 2 + 5{,}2^2 + 6{,}2^2) = 19.$$

За формулами (3.5) дістаємо:

$$F(\chi_1^2(\gamma)) = \frac{1-0{,}95}{2} = 0{,}025; \quad F(\chi_2^2(\gamma)) = \frac{1+0{,}95}{2} = 0{,}975.$$

За таблицею значень величини χ^2 (дод. 5 Гмурмана) з $n-1$=15 ступенями свободи знаходимо χ_2^2=6, χ_1^2=25 і за формулами (3.3) та (3.4) дістаємо довірчі інтервали відповідно для дисперсії та середнього квадратичного відхилення (11,4; 47,5) і (3,4; 6,9).

Також інтервальною оцінкою з надійністю γ середнього квадратичного відхилення σ нормально розподіленої кількісної ознаки X по виправленому середньому квадратичному відхиленню s є довірчій інтервал

$$s(1-q) < \sigma < s(1+q) \text{ при } q < 1,$$
$$0 < \sigma < s(1+q) \text{ при } q > 1, \tag{3.6}$$

де q знаходиться за таблицею в додатку 4 Гмурмана за заданими n та γ.

Приклад 3.5. Під час вибіркового аналізу максимального завантаження літака при певній кількості пального, дістали такі дані, т:

78 80 85 88 89 90 88 93
88 85 93 89 80 90 85 89

Припускаючи, що максимальне завантаження літака має нормальний розподіл, знайти довірчий інтервал для середнього квадратичного відхилення з надійністю γ=0,95.

Розв'язання. У прикладі 3.4. наведено обчислення s. З додатку 4 отримуємо q=0,44. Згідно (3.6) довірчій інтервал для середнього квадратичного відхилення (2,9; 6,3).

Інтервальною оцінкою з надійністю γ невідомої ймовірності p біноміального розподілу по відносній частоті w є довірчій інтервал

$$p_1 < p < p_2, \tag{3.7}$$

де

$$p_1 = \frac{n}{t^2+n}\left(w + \frac{t^2}{2n} - \sqrt{\frac{w(1-w)}{n} + \left(\frac{t}{2n}\right)^2}\right),$$

$$p_2 = \frac{n}{t^2+n}\left(w + \frac{t^2}{2n} + \sqrt{\frac{w(1-w)}{n} + \left(\frac{t}{2n}\right)^2}\right),$$

де n – загальна кількість випробувань, m – число появ події, w – відносна частота, t — розв'язок рівняння $2\Phi(t)=\gamma$ при заданій надійності γ.

Зауваження: при великих значеннях n (порядку сотні)

$$p_1 = w - t\sqrt{\frac{w(1-w)}{n}},$$

$$p_2 = w + t\sqrt{\frac{w(1-w)}{n}}. \qquad (3.8)$$

Приклад 3.6 Проводиться 60 незалежних випробувань. В них подія A з'являється 15 разів. Отримати інтервальну оцінку з надійністю γ=0,95 невідомої ймовірності p.

Розв'язання. Згідно умови задачі n=60, m=15, w=0,25.

2Φ(t)=0,96, тобто t=1,96 (див. дод.2 Гмурмана).

Згідно (3.7) довірчій інтервал для ймовірності p – (0,16;0,37).

ДОМАШНЄ ЗАВДАННЯ 3

Варіанти 1-5

Задано вибірку:

12	14	19	15	14	18	13	16	17	12
20	17	15	13	17	16	20	14	14	13
17	16	15	19	16	15	18	17	15	14
16	15	15	18	15	15	19	14	16	18
18	15	15	17	15	16	16	14	14	17

1) Побудувати інтервальний статистичний розподіл із оптимальною довжиною кроку.

2) Припускаючи, що дані розподілені за нормальним законом, знайти з надійністю γ=0,95 довірчий інтервал для невідомого математичного сподівання, якщо середнє квадратичне відхилення σ=3;

3) Побудувати довірчий інтервал з надійністю γ=0,99 для математичного сподівання, припускаючи, що дані мають нормальний розподіл;

4) Припускаючи, що дані мають нормальний розподіл, знайти довірчий інтервал для дисперсії та середнього квадратичного відхилення з надійністю γ=0,95.

Варіанти 6-10

Задано вибірку:

59	55	60	46	57	54	57	54	55	46
57	59	57	46	54	55	57	59	54	54
46	55	57	60	55	54	46	55	57	54

1) Побудувати інтервальний статистичний розподіл із оптимальною довжиною кроку.

2) Припускаючи, що дані розподілені за нормальним законом, знайти з надійністю γ=0,95 довірчий інтервал для невідомого математичного сподівання, якщо середнє квадратичне відхилення σ=5;

3) Побудувати довірчий інтервал з надійністю γ=0,99 для математичного сподівання, припускаючи, що дані мають нормальний розподіл;

4) Припускаючи, що дані мають нормальний розподіл, знайти довірчий інтервал для дисперсії та середнього квадратичного відхилення з надійністю γ=0,95.

Варіанти 11-15

Задано вибірку:

6,80	6,50	6,53	7,30	6,95	6,90	6,93	6,72	6,54	6,79
6,80	6,75	6,62	6,65	7,00	6,83	6,70	6,69	6,54	6,97
6,82	6,67	6,91	6,93	6,76	6,71	6,66	6,76	6,79	6,82

1) Побудувати інтервальний статистичний розподіл із оптимальною довжиною кроку.

2) Припускаючи, що дані розподілені за нормальним законом, знайти з надійністю γ=0,95 довірчий інтервал для невідомого математичного сподівання, якщо середнє квадратичне відхилення σ=0,2;
3) Побудувати довірчий інтервал з надійністю γ=0,99 для математичного сподівання, припускаючи, що дані мають нормальний розподіл;
4) Припускаючи, що дані мають нормальний розподіл, знайти довірчий інтервал для дисперсії та середнього квадратичного відхилення з надійністю γ=0,95.

Варіанти 16-20

Задано вибірку:

12	15	14	14	19	18	17	15	17	19
13	17	19	16	15	18	20	20	12	13
13	17	12	16	15	18	20	14	12	13
12	17	19	16	15	18	20	19	12	13

1) Побудувати інтервальний статистичний розподіл із оптимальною довжиною кроку.
2) Припускаючи, що дані розподілені за нормальним законом, знайти з надійністю γ=0,95 довірчий інтервал для невідомого математичного сподівання, якщо середнє квадратичне відхилення σ=2;
3) Побудувати довірчий інтервал з надійністю γ=0,99 для математичного сподівання, припускаючи, що дані мають нормальний розподіл;
4) Припускаючи, що дані мають нормальний розподіл, знайти довірчий інтервал для дисперсії та середнього квадратичного відхилення з надійністю γ=0,95.

Варіанти 21-25

Задано вибірку:

82	82	85	94	91	84	90	92	76	75	77	75	92
94	80	92	91	84	75	77	76	79	83	87	78	81
81	78	77	88	86	77	89	90	89	78	87	79	75
81	78	77	87	86	77	89	90	89	79	87	89	76

1) Побудувати інтервальний статистичний розподіл із оптимальною довжиною кроку.
2) Припускаючи, що дані розподілені за нормальним законом, знайти з надійністю γ=0,95 довірчий інтервал для невідомого математичного сподівання, якщо середнє квадратичне відхилення σ=3;
3) Побудувати довірчий інтервал з надійністю γ=0,99 для математичного сподівання, припускаючи, що дані мають нормальний розподіл;
4) Припускаючи, що дані мають нормальний розподіл, знайти довірчий інтервал для дисперсії та середнього квадратичного відхилення з надійністю γ=0,95.

Варіанти 26-30

Задано вибірку:

22	22	25	24	21	24	20	22	26	25	27	25	22	23
24	20	22	21	24	25	27	26	24	23	27	28	27	24
21	28	27	28	26	27	29	30	29	28	27	29	30	27

1) Побудувати інтервальний статистичний розподіл із оптимальною довжиною кроку.
2) Припускаючи, що дані розподілені за нормальним законом, знайти з надійністю γ=0,95 довірчий інтервал для невідомого математичного сподівання, якщо середнє квадратичне відхилення σ=1;
3) Побудувати довірчий інтервал з надійністю γ=0,99 для математичного сподівання, припускаючи, що дані мають нормальний розподіл;

4) Припускаючи, що дані мають нормальний розподіл, знайти довірчий інтервал для дисперсії та середнього квадратичного відхилення з надійністю $\gamma=0,95$.

ПРАКТИЧНЕ ЗАНЯТТЯ 4

Перевірочна контрольна робота 1

1). У результаті вибіркового аналізу кількості проданих авіаквитків за добу дістали такі дані, шт.:

59 55 60 46 57 54 57 54 55 46
57 59 57 46 54 55 57 59 54 54
46 55 57 60 55 54 46 55 57 54

1) Скласти дискретний статистичний розподіл;
2) побудувати полігон частот і відносних частот;
3) побудувати кумуляту та огіву для відносних накопичуваних частот.

2). Дістали такі дані, т:

6,80 6,50 6,80 6,75 6,82 6,67 6,53 7,30 6,95 6,62
6,65 7,00 6,91 6,93 6,76 6,90 6,93 6,83 6,70 6,71
6,66 6,72 6,54 6,79 6,69 6,54 6,97 6,76 6,79 6,82

Скласти інтервальний розподіл вибірки і побудувати за ним емпіричну функцію розподілу $F^*(x)$ та гістограму.

3). Вимірювання відхилень розрахункового часу польоту літака від фактичного дали такі результати, хв:

x_i	-2,5	-2	-1,5	-1	-0,5	0	0,5	1	1,5	2	2,5
n_i	5	8	12	21	20	40	30	20	20	14	10

Знайти точкові незміщені оцінки середнього значення і середнього квадратичного відхилення відхилень розрахункового часу польоту літака від фактичного.

4). Знайти довірчі інтервали та точність відповідних інтервальних оцінок для невідомого математичного сподівання нормально розподіленої випадкової величини X, якщо $\sigma^2=D(X)=25$, обсяг вибірки $n=36$ і $\overline{x_{\text{В}}}=7$. Довірчі інтервали та відповідні їм точності побудувати для надійності $\gamma=0,95$; $\gamma=0,99$; $\gamma=0,999$.

ПРАКТИЧНЕ ЗАНЯТТЯ 5

Перевірка статистичних гіпотез

5.1. Основні поняття

Оскільки висновки про параметри і закони розподілу ознаки X генеральної сукупності робляться на основі обмеженого вибіркового матеріалу, вони не можуть бути категоричними і мають сприйматись лише як деякі припущення. Такі припущення про закони розподілу або їхні параметри називаються статистичними гіпотезами. Наприклад, статистичною гіпотезою є припущення: ознака X генеральної сукупності розподілена за нормальним законом.

Задача статистичної перевірки гіпотези полягає в побудові за вибірковими даними статистичного критерію (критерію узгодження), який дає змогу прийняти або відхилити висунуту гіпотезу.

Статистичною називають гіпотезу щодо типу невідомого розподілу або щодо параметрів відомого розподілу.

Висунуту гіпотезу H_0 називається *основною* або *нульовою*.

Конкуруючою (альтернативною) називають гипотезу H_1, яка суперечить нульовій.

Розрізняють гіпотези, які містять одне або більше припущень.

Простою називають гіпотезу, що містить одне припущення.

Складною називають гіпотезу, що містить скінчену або нескінчену кількість простих гіпотез.

Число α також називається ймовірністю *помилки 1-го ро*ду, яка полягає у відкиданні вірної нульової гіпотези.

Число β називається ймовірністю *помилки 2-го роду*, яка полягає у прийнятті невірної нульової гіпотези.

Статистичним критерієм називається випадкова величина K який призначений для перевірки гіпотез.

Спостережуваним (емпіричне) значенням називають те значення $K_{сп}$, яке обчислено по вибіркам.

Критичною областю називають сукупність значень критерію , при яких нульову гіпотезу відкидають.

Областю прийняття гіпотези називають сукупність значень критерію , при яких нульову гіпотезу приймають.

Критичною точкою називають число $K_{кр}$, яке відрізняють критичну область від області прийняття гіпотез.

Правосторонньою називається критична область, що визначається нерівність $K>K_{кр}$, де $K_{кр}$ – додатнє число.

Лівосторонньою називається критична область, що визначається нерівність $K<K_{кр}$, де $K_{кр}$ – від’ємне число.

Двосторонньою називається критична область, що визначається нерівність $|К|>K_{кр}$,

Для визначення критичної області задається рівень значущості та шукають критичні точки з наступних умов:

1. Для правосторонньої області

$$P(K>K_{кр})=\alpha \ (K_{кр}>0) \tag{5.1}$$

2. Для лівосторонньої області

$$P(K<K_{кр})=\alpha \; (K_{кр}<0) \quad (5.2)$$

3. Для двосторонньої симетричної області

$$P(K>K_{кр})=\alpha/2,\; P(K<-K_{кр})=\alpha/2 \; (K_{кр}>0) \quad (5.3)$$

Потужністю критерію називають ймовірність попадання в критичну область, за умовою, що справедлива конкуруюча гіпотеза.

З усього вищесказаного можна підсумувати, що розрізняють два типи статистичних критеріїв:

1). В яких в умові задано випадкову величину з відомим законом розподілу. Нульова гіпотеза полягає у рівності певних числових характеристик генеральної сукупності (математичне сподівання, дисперсія, коефіцієнт кореляції, тощо) або їхня рівність наперед заданому гіпотетичному значенню.

2). В яких в умові дано вибірку, взяту з певної генеральної сукупності. Нульова гіпотеза полягає у припущенні щодо типу розподілу всієї генеральної сукупності.

Етапи перевірки статистичної гіпотези:

1. Формулювання нульової та альтернативної гіпотез.
2. Розрахунок необхідних статистичних характеристик.
3. Вибір рівня значущості α.
4. Вибір та обчислення розрахункового значення критерію для перевірки гіпотез. Критичне значення визначається за відповідною таблицею.
5. Порівняння розрахункового і табличного значень критерію, на основі чого робиться висновок про прийняття чи відхилення гіпотези.

5.2. Одновибірковий t-критерій (критерій Стьюдента).

Застосовується для перевірки нульової гіпотези про рівність математичного очікування деякому відомому значенню.

Розглянемо даний критерій на прикладі конкретної прикладної задачі.

<u>Завдання:</u> Визначити, чи відповідає рівень фізичного розвитку школярів певного віку загальноприйнятій нормі.

Вихідні дані:

1.Результати вимірювань рівня фізичного розвитку школярів (вибірка обсягом n).

2. μ – норма фізичного розвитку школярів.

Порядок виконання обчислень:

1. Сформулювати нульову гіпотезу про рівень фізичного розвитку школярів.

2. Визначити вибіркові числові характеристики, а саме:

а) середнє арифметичне значення:

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

б) оцінку значення дисперсії генеральної сукупності:

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

в) стандартну похибку середнього арифметичного:

$$S_{\bar{x}}=\frac{\sigma}{\sqrt{n}}$$

3. Обчислити розрахункове значення статистичного t-критерію Стьюдента за формулою:

$$t_p = \frac{\bar{x} - \mu}{S_{\bar{x}}} \tag{5.4}$$

4. Обрати рівень значущості, наприклад, $\alpha = 0{,}05$, що відповідає ймовірності

$$\beta = (1-\alpha) \cdot 100\% = 95\%. \tag{5.5}$$

5. Визначити значення t-критерію Стьюдента $k(\alpha, \nu)$ згідно таблиці при $\nu = n-1$, де ν – число ступенів свободи, n – обсяг вибірки (див. дод. 6 Гмурмана).

6. Порівняти табличні значення t – критерію з розрахунковими і зробити висновок про прийняття чи відхилення гіпотези:

а) якщо $t_P < t_\alpha$, то з ймовірністю $\beta = (1-\alpha) \cdot 100\% = 95\%$. можна стверджувати, що нульова гіпотеза відхиляється і рівень фізичного розвитку школярів не відповідає нормі.

б) якщо $t_P > t_\alpha$, то з ймовірністю $\beta = (1-\alpha) \cdot 100\% = 95\%$. можна стверджувати, що нульова гіпотеза відхиляється і рівень фізичного розвитку школярів не відповідає нормі.

5.3. Двовибірковий t-критерій для незалежних вибірок.

Нехай є дві незалежні вибірки об'ємаминормально розподілених випадкових величин. Необхідно перевірити за вибірковими даними нульову гіпотезу рівності математичних очікувань цих випадкових величин

Розглянемо даний критерій на прикладі конкретної прикладної задачі.

Завдання: порівняти рівень фізичного розвитку студентів 2-х спеціалізацій.

Вихідні дані:

1. Вибіркові параметри x, які характеризують рівень фізичного стану студентів 1-ї спеціалізації.

2. Вибіркові параметри y, які характеризують рівень фізичного стану студентів 2-ї спеціалізації.

Порядок виконання обчислень:

1. Сформулювати нульову гіпотезу про рівність фізичного стану студентів 2-х спеціалізацій.

2. Визначити вибіркові числові характеристики:

а) середні арифметичні значення кожної вибірки:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,\ \bar{y} = \frac{1}{m}\sum_{i=1}^{m} y_i,$$

б) оцінки дисперсії для кожної вибірки:

$$D_X = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2,\ D_Y = \frac{1}{m}\sum_{i=1}^{m}(y_i - \bar{y})^2.$$

3. Визначити стандартну похибку різниці середніх арифметичних незв'язаних сукупностей:

$$S_{\bar{x}-\bar{y}} = \sqrt{\frac{D_X}{n} + \frac{D_Y}{m}}. \tag{5.6}$$

4. Обчислити розрахункове значення статистичного t-критерію Стьюдента за формулою:

$$t_p = \frac{|\bar{x} - \bar{y}|}{S_{\bar{x}-\bar{y}}}. \tag{5.7}$$

5. Обрати рівень значущості α, що відповідає $\beta = (1-\alpha) \cdot 100\% = 95\%$, і визначити табличне значення t-критерію Стьюдента (див. дод. 6 Гмурмана). Число ступенів свободи ν розраховується за формулою:

$$\nu = (n_X - 1) + (n_Y - 1) \tag{5.8}$$

6. Порівняти розрахункове значення t-критерію Стьюдента з табличним і зробити висновок про прийняття чи відхилення нульової гіпотези:

а) якщо розрахункове значення t-критерію менше за табличне, то фізичний стан цих 2-х груп є однаковий і нульова гіпотеза приймається;

б) якщо розрахункове значення t-критерію більше або рівне табличному, то фізичний стан цих двох груп не є однаковий і нульова гіпотеза відхиляється.

5.4. Двовибірковий t-критерій для залежних вибірок

Використовується для обчислення емпіричного значення-критерію в ситуації перевірки гіпотези про відмінності між двома залежними вибірками (наприклад, двома пробами одного і того ж тесту з тимчасовим інтервалом).

Розглянемо даний критерій на прикладі конкретної прикладної задачі.

Завдання: визначити, чи відбуваються зміни спортивних показників студентів за час тренувань, що проходять за даним методом.

Вихідні дані:

Результати спортивних показників студентів: x_1 – значення показника першого виміру (початок експерименту); x_2 – значення показника другого виміру (кінець експерименту).

Порядок виконання обчислень:

1. На основі припущення про нормальний розподіл різниці

$$d_i = x_{1i} - x_{2i}$$

сформулювати нульову гіпотезу про рівність двох середніх арифметичних спортивних показників.

2. Визначити вибіркові характеристики:

а) різницю між результатами першого і другого вимірів

$$d_i = x_{1i} - x_{2i}$$

б) середню арифметичну різницю

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$$

в) дисперсію різниці

$$D_d = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$$

г) стандартну похибку середньої різниці

$$S_{\bar{d}} = \sqrt{\frac{D_d}{n}}$$

3. Обчислити розрахункове значення статистичного t-критерію Стьюдента за формулою:

$$t_P = \frac{|\bar{d}|}{S_{\bar{d}}} \tag{5.9}$$

4. Обрати рівень значущості, наприклад, α=0,05, що відповідає ймовірності β=95% і визначити табличне значення t-критерію Стьюдента t_a, де $\nu=n-1$ – число ступенів свободи (див. дод. 6 Гмурмана).

5. Порівняти розрахункове значення t-критерію з табличним і зробити висновок про суттєві зміни спортивних показників студентів:

а) якщо $t_P<t_\alpha$ - суттєвих змін за час тренувань не сталося.

б) якщо $t_P>t_\alpha$ - спортивний рівень або покращився, або погіршився.

5.5. F-критерій (критерій Фішера).

Опис критерію: Задані дві вибірки $X=(x_1, x_2, \dots, x_n)$, $x_i \in \mathbb{R}$; $Y=(y_1, y_2, \dots, xy_n)$, $y_i \in \mathbb{R}$.

Позначимо через σ_1^2 і σ_2^2 дисперсії вибірок *X* і *Y*, s_1^2 і s_2^2 — вибіркові оцінки дисперсій σ_1^2 і σ_2^2:

$$s_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2,\ s_2^2 = \frac{1}{m-1}\sum_{i=1}^{m}(y_i - \bar{y})^2.$$

де $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{y} = \frac{1}{m}\sum_{i=1}^{m} y_i$— вибіркові средніх вибірок *X* і *Y*.

Додаткове припущення:

1) вибірки *X* і *Y* є нормальними. Критерій Фішера чутливий до порушення припущення про нормальність.
2) $s_1^2 > s_2^2$ Звичайно в чисельнику ставиться більша із двох порівнювальних дисперсій.

Нульова гіпотеза: H_0: $\sigma_1^2 = \sigma_2^2$

Статистика критерію Фішера:

$$F_{\text{сп}} = s_1^2 / s_2^2 \tag{5.10}$$

має росподіл Фішера з $n-1$ і $m-1$ степенями свободи.

Одже критичною областью критерия є правий хвіст розподілу Фішера, що. відповідає альтернативній гіпотезі H_1.

Критерій (при рівні значущості α):

- проти альтернативи H_1: $\sigma_1^2 \neq \sigma_2^2$

якщо $F_{\text{сп}} < F_{\alpha/2}(n-1,m-1)$ або $F_{\text{сп}} > F_{1-\alpha/2}(n-1,m-1)$, то нульова гіпотеза H_0 відкидається на користь альтернативи H_1.

- проти альтернативи H_1: $\sigma_1^2 > \sigma_2^2$

якщо $F_{\text{сп}} > F_{1-\alpha}(n-1,m-1)$, то нульова гіпотеза H_0 відкидається на користь альтернативи H_1; де $F_\alpha(n-1,m-1)$ є α-квантиль розподілу Фішера з $n-1$ і $m-1$ степенями свободи (див. дод. 7 Гмурмана).

Приклад 5.1. У двох третіх класах проводилося тестування розумового розвитку за тестом десяти учнів. Отримані значення величин середніх достовірно не розрізнялися, проте психолога цікавить питання - чи є відмінності в ступені однорідності показників розумового розвитку між класами.

Розв'язання. Для критерію Фішера необхідно порівняти дисперсії тестових оцінок в класах.

Результати тестування представлені в табл. 5.1.

Таблиця 5.1

№	ПІБ	3-А клас	3-Б клас
1	Андрєєв Л.	90	41
2	Валєєва І.	29	49
3	Енокян А.	39	56
4	Іванов П.	79	64
5	Сидорова М	88	72
6	Петров А.	53	65
7	Попова Л.	34	63
8	Шестаков І.	40	87
9	Еховий М.	75	77
10	Якушева А.	79	62
	СУМА	606	636
	СЕРЕДНЄ	60,6	63,6

Розрахувавши дисперсії для змінних X і Y, отримуємо:

$$\sigma_X^2=572{,}83;\ \sigma_Y^2=174{,}04.$$

За формулою для розрахунку F критерію знаходимо:

$$F_{сп}=572{,}83/174{,}04=3{,}29.$$

По таблиці критичних значень для критерію Фішера при ступенях свободи в обох випадках рівних $k_1=k_2=10\text{-}1=9$ знаходимо $F_{кр}=3{,}18$ (<3,29), отже, в термінах статистичних гіпотез можна стверджувати, що H_0 (гіпотеза про подібність) може бути відкинута на рівні 5%, а приймається в цьому випадку гіпотеза H_1. Дослідник може стверджувати, що за ступенем однорідності такого показника, як розумовий розвиток, є відмінність між вибірками з двох класів.

5.6. Перевірка гіпотези про вибірковий коефіцієнт кореляції.

Нехай із генеральної сукупності двовимірної випадкової величини (X,Y), розподіленої за нормальним законом, здійснено вибірку об'єму n і за даними вибірки $(x_i;\ y_i)$ обчислено вибірковий коефіцієнт кореляції $r_В$ який виявився відмінним від нуля.

Оскільки вибірка випадкова, то не можна зробити висновок, що коефіцієнт кореляції r генеральної сукупності також відмінний від нуля. Тому виникає необхідність при заданому рівні значущості α перевірити нульову гіпотезу H_0: $r=0$ при альтернативній гіпотезі H_1: $r\neq 0$.

Якщо нульова гіпотеза буде відхилена, то це означає, що вибірковий коефіцієнт кореляції суттєво відрізняється від нуля, тобто значущий, і випадкові величини X і Y – корельовані. Якщо нульова гіпотеза буде прийнята, то вибірковий коефіцієнт кореляції мало відрізняється від нуля і випадкові величини X і Y – некорельовані.

За критерій перевірки нульової гіпотези беруть випадкову величину:

$$T_{сп}=\sqrt{\frac{r_В^2(n-2)}{1-r_В^2}}, \qquad (5.11)$$

яка при справедливості нульової гіпотези має розподіл Стьюдента з $k=n-2$ ступенями свободи. Оскільки альтернативна гіпотеза має вигляд $r\neq 0$, то критична область – двостороння.

Для перевірки нульової гіпотези при заданому рівні значущості α потрібно за формулою (5.11) обчислити спостережуване значення критерію $T_{сп}$, а за даними таблиці критичних точок розподілу Стьюдента для двосторонньої області при заданому рівні значущості α і числі ступенів свободи k відшукати критичну точку $t_{кр}(\alpha;k)$ (див. дод. 6 Гмурмана).

Якщо $|T_{сп}|<t_{кр}$, то нульову гіпотезу приймають, а у випадку $|T_{сп}|\geq t_{кр}$ – відхиляють.

Приклад 5.2. Оцінити значущість кореляційного зв'язку між успішністю виконання тестових завдань з фізики (X) і математики (Y) учнями загальноосвітньої школи (рис. 5.1).

	A	B	C
1	i	$\bar{X}$	$\bar{Y}$
2	1	22	18
3	2	25	22
4	3	18	20
5	4	24	25
6	5	14	15
7	6	20	18
8	7	16	20
9	8	15	13
10	9	24	18
11	10	20	15
12	11	11	10
13	12	11	13
14	r_{xy} =	0,77	
15	t_r =	3,87	
16	α =	0,01	
17	t_{st} =	3,17	

Рис 5.1

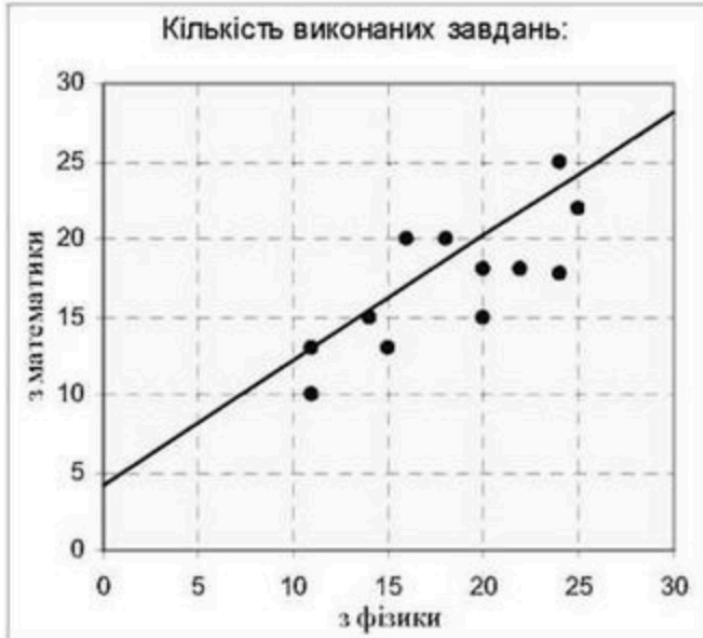


Рис 5.2

Розрахунки та інтерпретація результаті наведені на рис.5.1., рис 5.2.

5.7. Порівняння виправленої вибіркової дисперсії з гіпотетичною генеральною дисперсією нормальної сукупності

Нехай генеральна сукупність розподілена нормально, причому генеральна дисперсія хоч і невідома, але є підстави припускати, що вона дорівнює гіпотетичному (передбачуваному) значенню σ_0^2. Насправді σ_0^2 встановлюється виходячи з попереднього досвіду чи теоретично.

Нехай із генеральної сукупності вилучено вибірку об'єму *n* і з ній знайдено виправлена вибіркова дисперсія s^2 з *k=n—1* ступенями свободи. Потрібно по виправленої дисперсії при заданому рівні значущості перевірити нульову гіпотезу, яка полягає в тому, що генеральна дисперсія аналізованої сукупності дорівнює гіпотетичному значенню σ_0^2. Враховуючи, що s^2 є незміщеною оцінкою генеральної дисперсії, нульову гіпотезу можна записати так:

$$H_0 : M(s^2) = \sigma_0^2.$$

Отже, потрібно перевірити, що математичне очікування виправленої дисперсії дорівнює гіпотетичному значенню генеральної дисперсії. Інакше кажучи, потрібно встановити, значимо чи незначно розрізняються виправлена вибіркова і гіпотетична генеральна дисперсії.

На практиці аналізована гіпотеза перевіряється, якщо потрібно перевірити точність приладів, інструментів, верстатів, методів дослідження та стійкість технологічних процесів. Наприклад, якщо відома допустима характеристика розсіювання контрольованого розміру деталей, що виготовляються верстатом-автоматом, дорівнює σ_0^2 а знайдена за вибіркою виявиться значуще більше σ_0^2, то верстат вимагає підналагодження.

Як критерій перевірки нульової гіпотези приймемо випадкову величину $(n-1)s^2/\sigma_0^2$. Ця величина випадкова, тому що в різних дослідах s^2 приймає різні, наперед невідомі значення. Оскільки можна довести, що вона має розподіл χ^2 з *k=n—1* ступенями свободи, позначимо її через $\chi^2_{\text{сп}}$.

Отже, критерій перевірки нульової гіпотези

$$\chi^2_{сп}=(n-1)s^2/\sigma_0^2 \tag{5.12}$$

Критична область будується залежно від виду конкуруючої гіпотези.

Перший випадок. Нульова гіпотеза H_0: $\sigma^2=\sigma_0^2$. Конкуруюча гіпотеза H_1: $\sigma^2>\sigma_0^2$

У цьому випадку будують правосторонню критичну область, виходячи з вимоги, щоб ймовірність попадання критерію в цю область у припущенні справедливості нульової гіпотези дорівнювала прийнятому рівню значимості:

$$P[\chi^2_{сп}>\chi^2_{кр}(\alpha,k)]=\alpha.$$

Критичну точку $\chi^2_{кр}(\alpha,k)$ знаходять по таблиці критичних точок розподілу χ^2, і тоді правостороння критична область визначається нерівністю $\chi^2_{сп}>\chi^2_{кр}$, а сфера прийняття нульової гіпотези - нерівністю $\chi^2_{сп}<\chi^2_{кр}$.

Сформулюємо правило перевірки нульової гіпотези.

Правило 1. Для того щоб при заданому рівні значущості α перевірити нульову гіпотезу H_0: $\sigma^2=\sigma_0^2$ про рівність невідомої генеральної дисперсії нормальної сукупності гіпотетичного значення при конкуруючій гіпотезі H_1: $\sigma^2>\sigma_0^2$ за формулою (5.12) обчислюємо $\chi^2_{сп}$ і за таблицею критичних точок розподілу χ^2 за заданим рівнем значущості α і числом ступенів свободи $k=n-1$ знайти критичну точку $\chi^2_{кр}(\alpha,k)$ (див. дод. 5 Гмурмана).

Якщо $\chi^2_{сп}<\chi^2_{кр}$ — немає підстав відкинути нульову гіпотезу.

Якщо $\chi^2_{сп}>\chi^2_{кр}$ -нульову гіпотезу відкидають.

Приклад 5.3. З нормальної генеральної сукупності вилучено вибірку обсягу n=13 і по ній знайдено виправлену вибіркову дисперсію s^2=14,6. Потрібно при рівні значущості 0,01 перевірити нульову гіпотезу H_0: $\sigma^2=\sigma_0^2=12$, прийнявши в якості конкуруючої гіпотези H_1: $\sigma^2>12$.

Розв'язання. Знайдемо спостерігане значення критерію:

$$\chi^2_{сп}=(n-1)s^2/\sigma_0^2=((13-1)\cdot 14{,}6)/12=14{,}6.$$

За умовою, конкуруюча гіпотеза має вигляд H_1: $\sigma^2>12$, тому критична область правостороння.

За рівнем значущості 0,01 і числом ступенів свободи $k=n-1=13-1=12$ знаходимо критичну точку $\chi^2_{кр}(0{,}01;\ 12)=26{,}2$.

Оскільки $\chi^2_{сп}<\chi^2_{кр}$ — немає підстав відкинути нульову гіпотезу.

Іншими словами, різниця між виправленою дисперсією (14,6) та гіпотетичною генеральною дисперсією (12) – незначна.

Другий випадок. Нульова гіпотеза H_0: $\sigma^2=\sigma_0^2$. Конкуруюча гіпотеза H_1: $\sigma^2\neq\sigma_0^2$.

У цьому випадку будують двосторонню критичну область, виходячи з вимоги, щоб ймовірність попадання критерію в цю область у припущенні справедливості нульової гіпотези дорівнювала прийнятому рівню значимості α.

Критичні точки - ліву і праву межі критичної області - знаходять, вимагаючи, щоб ймовірність попадання критерію в кожній з двох інтервалів критичної області дорівнювала α/2:

$$P[\chi^2_{сп}<\chi^2_{лів.кр}(\alpha/2,k)]=\alpha/2,$$

$$P[\chi^2_{сп}>\chi^2_{лів.кр}(\alpha/2,k)]=\alpha/2.$$

У таблиці критичних точок розподілу χ^2 вказані лише «праві» критичні точки, тому виникає утруднення у відшуканні «лівої» критичної точки. Це ускладнення легко подолати, якщо взяти до уваги, що події $\chi^2_{сп}<\chi^2_{лів.кр}(\alpha/2,k)$ і $\chi^2_{сп}>\chi^2_{лів.кр}(\alpha/2,k)$ протилежні і, отже, сума їх ймовірностей дорівнює одиниці:

$$P[\chi^2_{сп}<\chi^2_{лів.кр}(\alpha/2,k)]+P[\chi^2_{сп}>\chi^2_{лів.кр}(\alpha/2,k)]=1.$$

Звідси

$$P[\chi^2_{сп}<\chi^2_{лів.кр}(\alpha/2,k)]=1-P[\chi^2_{сп}>\chi^2_{лів.кр}(\alpha/2,k)].$$

Ми бачимо, що ліву критичну точку можна шукати як праву (і значить, її можна знайти за таблицею), виходячи з вимоги, щоб ймовірність попадання критерію в інтервал, розташований правіше цієї точки, дорівнювала 1—(α/2).

Правило 2. Для того, щоб при заданому рівні значимості а перевірити нульову гіпотезу про рівність невідомої генеральної дисперсії σ^2 нормальної сукупності гіпотетичного значення σ_0^2 при конкуруючій гіпотезі H_1: $\sigma^2\neq\sigma_0^2$ треба обчислити спостерігане значення за формулою (5.12) і по таблиці знайти ліву критичну точку $\chi^2_{кр}(1-\alpha/2;\ k)$ і праву критичну точку $\chi^2_{кр}(\alpha/2;\ k)$.

Якщо $\chi^2_{лів.кр}<\chi^2_{сп}<\chi^2_{пр.кр}$ — немає підстав відкинути нульову гіпотезу.

Якщо $\chi^2_{сп}<\chi^2_{лів.кр}$ або $\chi^2_{сп}>\chi^2_{пр.кр}$ — нульову гіпотезу відкидають.

Правило 3. При конкуруючій гіпотезі H_1: $\sigma^2<\sigma^2$знаходять критичну точку $\chi^2_{кр}(1-\alpha;\ k)$.

Якщо $\chi^2_{сп}>\chi^2_{кр}(1-\alpha;\ k)$ - немає підстав відкинути нульову гіпотезу.

Якщо $\chi^2_{сп}<\chi^2_{кр}(1-\alpha;\ k)$-нульову гіпотезу відкидають.

ДОМАШНЄ ЗАВДАННЯ 5

Варіанти 1-5

Задано вибірку X_1:

12	14	19	15	14	18	13	16	17	12
20	17	15	13	17	16	20	14	14	13
17	16	15	19	16	15	18	17	15	14
16	15	15	18	15	15	19	14	16	18
18	15	15	17	15	16	16	14	14	17

та вибірку X_2:

12	15	14	14	19	18	17	15	17	19
13	17	19	16	15	18	20	20	12	13
13	17	12	16	15	18	20	14	12	13
12	17	19	16	15	18	20	19	12	13

1) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність дисперсій нормальних генеральний сукупностей для всіх можливих конкуруючих гіпотез;

2) При рівні значущості $\alpha = 0{,}05$ для вибірки X_1 перевірити гіпотезу про рівність виправленої вибіркової дисперсії гіпотетичному значенню $\sigma^2 = \sigma_0^2 = 0{,}18$ для всіх можливих конкуруючих гіпотез;

3) При рівні значущості $\alpha = 0{,}05$ для вибірки X_2 перевірити гіпотезу про рівність математичного сподівання нормальної генеральної сукупності гіпотетичному значенню $M(X)$=16 для всіх можливих конкуруючих гіпотез

4) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність математичних сподівань нормальних генеральних сукупностей для всіх можливих конкуруючих гіпотез;

5) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність нулю коефіцієнту кореляції для всіх можливих конкуруючих гіпотез.

Варіанти 6-10

Задано вибірку X_1:

59	55	60	46	57	54	57	54	55	46
57	59	57	46	54	55	57	59	54	54
46	55	57	60	55	54	46	55	57	54
59	47	46	55	60	47	54	57	59	60

та вибірку X_2:

55	55	60	46	54	57	59	60	57	46
55	46	60	46	54	57	60	60	57	59
47	47	55	46	54	60	55	54	57	59
47	57	59	46	55	54	59	59	54	60

1) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність дисперсій нормальних генеральний сукупностей для всіх можливих конкуруючих гіпотез;

2) При рівні значущості $\alpha = 0{,}05$ для вибірки X_2 перевірити гіпотезу про рівність математичного сподівання нормальної генеральної сукупності гіпотетичному значенню: $M(X)$=54 для всіх можливих конкуруючих гіпотез;

3) При рівні значущості $\alpha = 0{,}05$ для вибірки X_1 перевірити гіпотезу про рівність виправленої вибіркової дисперсії гіпотетичному значенню $\sigma^2 = \sigma_0^2 = 0{,}15$ для всіх можливих конкуруючих гіпотез;

4) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність математичних сподівань нормальних генеральних сукупностей для всіх можливих конкуруючих гіпотез;

5) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність нулю коефіцієнту кореляції для всіх можливих конкуруючих гіпотез.

Варіанти 11-15

Задано вибірку X_1:

6,80	6,50	6,53	7,30	6,95	6,90	6,93	6,72	6,54	6,79
6,80	6,75	6,62	6,65	7,00	6,83	6,70	6,69	6,54	6,97
6,82	6,67	6,91	6,93	6,76	6,71	6,66	6,76	6,79	6,82
6,71	6,91	6,90	7,00	6,53	6,72	6,66	6,80	6,82	6,81

та вибірку X_2:

6,80	6,54	6,90	7,00	6,53	6,93	6,79	6,53	6,57	6,82
6,80	6,91	6,90	7,00	6,76	6,93	6,79	6,53	6,57	6,81
6,71	6,90	6,90	7,00	6,53	6,93	6,79	6,53	6,72	6,97
6,80	6,54	6,90	6,93	6,93	6,93	6,79	6,53	6,57	6,54

1) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність дисперсій нормальних генеральний сукупностей для всіх можливих конкуруючих гіпотез;

2) При рівні значущості $\alpha = 0{,}05$ для вибірки X_1 перевірити гіпотезу про рівність виправленої вибіркової дисперсії гіпотетичному значенню $\sigma^2 = \sigma_0^2 = 0{,}2$ для всіх можливих конкуруючих гіпотез

3) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність математичних сподівань нормальних генеральних сукупностей для всіх можливих конкуруючих гіпотез;

4) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність нулю коефіцієнту кореляції для всіх можливих конкуруючих гіпотез;

5) При рівні значущості $\alpha = 0{,}05$ для вибірки X_2 перевірити гіпотезу про рівність математичного сподівання нормальної генеральної сукупності гіпотетичному значенню: $M(X)$=6,60 для всіх можливих конкуруючих гіпотез.

Варіанти 16-20

Задано вибірку X_1:

55	55	60	46	54	57	59	60	57	46
55	46	60	46	54	57	60	60	57	59
47	47	55	46	54	60	55	54	57	59
47	57	59	46	55	54	59	59	54	60

та вибірку X_2:

6,80	6,54	6,90	7,00	6,53	6,93	6,79	6,53	6,57	6,82
6,80	6,91	6,90	7,00	6,76	6,93	6,79	6,53	6,57	6,81
6,71	6,90	6,90	7,00	6,53	6,93	6,79	6,53	6,72	6,97
6,80	6,54	6,90	6,93	6,93	6,93	6,79	6,53	6,57	6,54

1) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність дисперсій нормальних генеральний сукупностей для всіх можливих конкуруючих гіпотез;

2) При рівні значущості $\alpha = 0{,}05$ для вибірки X_1 перевірити гіпотезу про рівність виправленої вибіркової дисперсії гіпотетичному значенню $\sigma^2 = \sigma_0^2 = 0{,}1$ для всіх можливих конкуруючих гіпотез;

3) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність математичних сподівань нормальних генеральних сукупностей для всіх можливих конкуруючих гіпотез;

4) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність нулю коефіцієнту кореляції для всіх можливих конкуруючих гіпотез;

5) При рівні значущості $\alpha = 0{,}05$ для вибірки X_2 перевірити гіпотезу про рівність математичного сподівання нормальної генеральної сукупності гіпотетичному значенню: $M(X)$=6,65 для всіх можливих конкуруючих гіпотез.

Варіанти 21-25

Задано вибірку X_1:

75	85	84	81	84	80	82	76	75	77
80	82	81	84	85	77	76	84	83	84
78	77	88	86	87	79	80	79	78	87
78	77	88	86	84	79	80	79	78	81

та вибірку X_2:

12	14	19	15	14	18	13	16	17	12
20	17	15	13	17	16	20	14	14	13
17	16	15	19	16	15	18	17	15	14
16	15	15	18	15	15	19	14	16	18
18	15	15	17	15	16	16	14	14	17

1) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність дисперсій нормальних генеральний сукупностей для всіх можливих конкуруючих гіпотез;

2) При рівні значущості $\alpha = 0{,}05$ для вибірки X_2 перевірити гіпотезу про рівність математичного сподівання нормальної генеральної сукупності гіпотетичному значенню $M(X)=15$ для всіх можливих конкуруючих гіпотез;

3) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність математичних сподівань нормальних генеральних сукупностей для всіх можливих конкуруючих гіпотез;

4) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність нулю коефіцієнту кореляції для всіх можливих конкуруючих гіпотез;

5) При рівні значущості $\alpha = 0{,}05$ для вибірки X_1 перевірити гіпотезу про рівність виправленої вибіркової дисперсії гіпотетичному значенню $\sigma^2 = \sigma_0^2 = 0{,}12$ для всіх можливих конкуруючих гіпотез.

Варіанти 26-30

Задано вибірку X_1:

75	85	84	81	84	80	82	76	75	79
80	82	81	84	85	76	76	84	83	84
78	77	88	86	87	79	80	79	78	87
78	75	88	86	84	79	80	79	78	82

та вибірку X_2:

75	85	84	81	83	80	82	76	75	71
80	82	81	84	85	77	76	84	83	84
78	77	88	86	87	79	80	79	78	87
78	77	81	87	84	79	80	79	78	81

1) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність дисперсій нормальних генеральний сукупностей для всіх можливих конкуруючих гіпотез;

2) При рівні значущості $\alpha = 0{,}05$ для вибірки X_2 перевірити гіпотезу про рівність математичного сподівання нормальної генеральної сукупності гіпотетичному значенню: $M(X)$=2

3) При рівні значущості $\alpha = 0{,}05$ для вибірки X_1 перевірити гіпотезу про рівність виправленої вибіркової дисперсії гіпотетичному значенню $\sigma^2 = \sigma_0^2 = 0{,}1$ для всіх можливих конкуруючих гіпотез;

4) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність математичних сподівань нормальних генеральних сукупностей для всіх можливих конкуруючих гіпотез;

5) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність нулю коефіцієнту кореляції для всіх можливих конкуруючих гіпотез.

ПРАКТИЧНЕ ЗАНЯТТЯ 6

Перевірка статистичних гіпотез (продовження)

6.1. Критерій χ^2-Пірсона

Критерієм згоди називають критерій перевірки гіпотези про вигляд невідомого закону розподілу.

Є ряд критеріїв згоди: χ^2-Пірсона, Колмогорова, Смірнова та інші.

Нехай за вибіркою об’єму n отримано інтервальний статистичний ряд (табл. 6.1) із частковими інтервалами однакової довжини

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

Таблиця 6.1

$(x_i,\ x_{i+1})$	$(x_1,\ x_2)$	$(x_2,\ x_3)$	…	$(x_\ell,\ x_{\ell+1})$
n_i	n_1	n_2	…	n_ℓ

Потрібно при рівні значущості α перевірити нульову гіпотезу: генеральна сукупність розподілена за нормальним законом. Критерієм перевірки цієї гіпотези беруть випадкову величину

$$\chi^2 = \sum_{i=1}^{\ell} \frac{(n_i - n_i')^2}{n_i'}, \tag{6.1}$$

де l – число часткових інтервалів в статистичному ряді, n_i $(i=1,\ldots,l)$ – їх частоти (*емпіричні частоти*), і n' – *теоретичні частоти*, обчислені за припущенням, що генеральна сукупність розподілена нормально (число появ в n випробуваннях значень нормально розподіленої випадкової величини X із i-го часткового інтервалу). При $n \to +\infty$ закон розподілу цієї випадкової величини прямує до закону розподілу χ^2 з k - ступенями вільності. Якщо передбачається нормальний розподіл, то число ступенів вільності $k=l-3$, де l – число часткових інтервалів. Область прийняття нульової гіпотези визначається нерівністю:

$$\chi^2_{\text{сп}} < \chi^2_{\text{кр}}(\alpha; k). \tag{6.2}$$

Щоб знайти теоретичні частоти нормального розподілу, необхідно:

1) знайти середини x_i^* $(i=\overline{1,l})$ часткових інтервалів даного статистичного ряду (усереднені варіанти);

2) обчислити вибіркову середню $\overline{x^*}$ та вибіркове середнє квадратичне відхилення σ^* за даним розподілом вибірки, беручи в якості варіант середини часткових інтервалів:

$$\overline{x^*} = \frac{1}{n}\sum_{i=1}^{l} n_i\, x_i^*,\ \sigma^* = \sqrt{\overline{(x^*)^2} - \overline{x^*}^2} \tag{6.3}$$

де $\overline{(x^*)^2} = \frac{1}{n}\sum_{i=1}^{l} n_i\,(x_i^*)^2$.

3) нормалізувати випадкову величину X, тобто перейти до величини $Z = \frac{X-\overline{x^*}}{\sigma^*}$ і обчислити кінці інтервалів (z_i, z_{i+1}):

$$z_i = \frac{x_i-\overline{x^*}}{\sigma^*},\ z_{i+1} = \frac{x_{i+1}-\overline{x^*}}{\sigma^*} \tag{6.4}$$

покладаючи при цьому $z_i=-\infty$, $z_{i+1}=+\infty$;

4) обчислити теоретичні ймовірності потрапляння Z в інтервал (z_i, z_{i+1}):

$$P_i = \Phi(z_{i+1}) - \Phi(z_i) \tag{6.5}$$

5) знайти теоретичні частоти:

$$n'_i = nP_i \qquad (6.6)$$

За формулою (6.1) обчислюємо спостережуване значення критерію $\chi^2_{сп}$. За даним рівнем значущості α і числом ступенів свободи k із таблиці критичних точок розподілу χ^2 знаходимо критичну точку $\chi^2_{кр}(\alpha;k)$. Якщо $\chi^2_{сп}<\chi^2_{кр}(\alpha;k)$, то немає підстав відкинути нульову гіпотезу. Якщо $\chi^2_{сп}>\chi^2_{кр}(\alpha;k)$, то нульову гіпотезу відкидаємо.

Слід зазначити, що критерій Пірсона використовується також для перевірки гіпотез про інші типи розподілів генеральної сукупності (біноміальний, експоненційний тощо). Для цього в формулі (6.6) P_i шукається в залежності від типу розподілу.

Приклад 6.1. За даним інтервальним розподілом вибірки об'єму n=100 при рівні значущості α=0,05 за критерієм згоди Пірсона перевірити гіпотезу про нормальний розподіл генеральної сукупності.

(x_i, x_{i+1})	(3,8)	(8,13)	(13,18)	(18,23)	(23,28)	(28,33)	(33,38)
n_i	6	8	15	40	16	8	7

Розв'язання. Побудуємо статистичний розподіл, варіантами якого є середини даних інтервалів:

x_i^*	5,5	10,5	15,5	20,5	25,5	30,5	35,5
n_i	6	8	15	40	16	8	7

За формулою (6.3) обчислимо вибіркову середню та вибіркове середнє квадратичне відхилення:

$$\overline{x^*} \approx 20{,}7;\ \sigma^* \approx 7{,}28.$$

Використовуючи формули (6.4)–(6.6), обчислимо теоретичні частоти і n'. Для цього складемо розрахункову таблицю 6.2:

Таблиця 6.2

i	x_i	x_{i+1}	z_i	z_{i+1}	$\Phi(z_i)$	$\Phi(z_{i+1})$	P_i	n_i'
1	3	8	$-\infty$	–1,74	–0,5000	–0,4591	0,0409	4,09
2	8	13	–1,74	–1,06	–0,4591	–0,3554	0,1037	10,37
3	13	18	–1,06	–0,37	–0,3554	–0,1443	0,2111	21,11
4	18	23	–0,37	0,32	–0,1443	0,1255	0,2698	26,98
5	23	28	0,32	1,00	0,1255	0,3413	0,2158	21,58
6	28	33	1,00	1,69	0,3413	0,4545	0,1132	11,32
7	33	38	1,69	$+\infty$	0,4545	0,5000	0,0455	4,55
						$\sum$	1	100

Обчислимо спостережуване значення критерію $\chi^2_{сп}$, для чого складемо розрахункову таблицю 6.3:

Таблиця 6.3

i	n_i	n_i'	$n_i - n_i'$	$(n_i - n_i')^2$	$(n_i - n_i')^2 / n_i'$
1	6	4	2	4	1,0000
2	8	10	–2	4	0,4000
3	15	21	–6	36	1,7143
4	40	27	13	169	6,2593
5	16	22	–6	36	1,6364
6	8	11	–3	9	0,8181
7	7	5	2	2	0,4000
$\sum$	100	100			12,2281

Отже, $\chi^2_{\text{сп}}\approx 12{,}2281$. За даними таблиці критичних точок розподілу χ^2 при рівні значущості $\alpha=0{,}05$ та числі ступенів вільності $k=4$ знаходимо критичну точку $\chi^2_{\text{кр}}(0{,}05;4)=9{,}5$. Оскільки $\chi^2_{\text{сп}}>\chi^2_{\text{кр}}$, то гіпотезу про нормальний розподіл генеральної сукупності X відкидаємо. Це означає, що емпіричні та теоретичні частоти відрізняються істотно, тобто дані спостережень не узгоджуються з гіпотезою про нормальний розподіл генеральної сукупності.

6.2. Критерій узгодження Колмогорова.

Критерій Колмогорова (також відомий, як критерій узгодження *Колмогорова–Смірнова*) є одним з основних і найбільш широко використовуваних непараметричних методів у силу своєї достатньої чутливості до розходжень у досліджуваних розподілах.

Критерій Колмогорова–Смірнова призначений для зіставлення:

- емпіричного розподілу з теоретичним розподілом;
- одного емпіричного розподілу з іншим.

Критерій дозволяє знайти точку, у якій сума накопичених розбіжностей між двома розподілами є найбільшою, і оцінити ймовірність цієї розбіжності.

Обмеження для коректного застосування критерію:

1) Вибірка досить велика (більше 50 спостережень).
2) Класи інтервалів повинні бути впорядковані по зростанню або убуванню деякої ознаки. Вони обов'язково повинні відображати її направлену зміну.

Вважаємо, що експериментальний розподіл описується емпіричною функцією розподілу $F_n(x)$, а теоретичний розподіл, який висунуто відповідно до певної гіпотези для опису ЕД, характеризується функцією розподілу $F(x)$. За міру розбіжності між теоретичним і статистичним розподілами взяте максимальне значення модуля різниці між емпіричною функцією розподілу $F_n(x)$, і відповідною теоретичною функцією розподілу *F(x)*

$$D_n = |F_n(x) - F(x)|. \tag{6.7}$$

Аргументами зручності в застосуванні цього критерію є простота обчислення величини D_n, а також те, що її розподіл має досить простий вид. А.М. Колмогоров довів, що, яка б не була функція розподілу $F(x)$ неперервної випадкової величини X, при обмеженні збільшення числа незалежних спостережень n ймовірність нерівності $D_n\sqrt{n} \geq \lambda$ прямує до границі

$$P(\lambda) = 1 - \sum_{k=-\infty}^{\infty}(-1)^k e^{-2k^2\lambda^2}, \tag{6.8}$$

що характеризує ймовірність того, що, якщо випадкова величина X дійсно розподілена за законом $F(x)$, через чисто випадкові причини максимальна розбіжність між $F_n(x)$ і $F(x)$ буде не меншою, чим фактично спостережувана.

Фрагмент таблиці для даного критерію наведено в підручнику Кремера та [4] з інформаційних ресурсів.

<u>*Схема застосування критерію Колмогорова*</u>:

- побудова емпіричної функції розподілу $F_n(x)$;
- побудова теоретичної функції розподілу $F(x)$;
- знаходження максимуму D_n модуля різниці між ними за формулою (6.7) і визначення величини $\lambda = D_n\sqrt{n}$;
- знаходження ймовірності $P(\lambda)$:
- якщо ця ймовірність дуже мала (менше рівня значущості), гіпотезу про розподіл випадкової величини за запропонованим законом варто відхилити;
- у випадку досить великих значень $P(\lambda)$ (більших рівня значущості) гіпотеза про запропонований закон розподілу випадкової величини *X* визнається не суперечною експерименту. Розбіжності між $F_n(x)$ і $F(x)$ вважаються несуттєвими і можуть бути пояснені випадковими факторами.

Критерій Колмогорова простіше критерію χ^2 - Пірсона, тому його охоче застосовують на практиці. Однак, цей критерій можна застосовувати тільки у випадку, коли гіпотетичний розподіл $F(x)$ разом з усіма параметрами, що в нього входять. Такий випадок порівняно нечасто зустрічається на практиці. Звичайно з теоретичних міркувань відомо тільки загальний вид функції $F(x)$. При застосуванні критерію Пірсона ця обставина враховується відповідним зменшенням числа степенів вільності розподілу χ^2. Критерій Колмогорова такого узгодження не передбачає.

Приклад 6.2. На основі вибірки

Інтервали	[0;5)	[5;10)	[10;15)	[15;20)	[20;25)	[25;30)	[30;35)
n_I	30	16	7	3	2	1	1

побудовано експоненціальний закон розподілу часу обслуговування 60 заявок на АТС, що визначається функцією розподілу $F(x) = \begin{cases} 1 - e^{-0,14x}, x \geq 0, \\ 0, x < 0 \end{cases}$ і щільністю ймовірності $f(x) = \begin{cases} 0,14e^{-0,14x}, x \geq 0, \\ 0, x < 0 \end{cases}$. Перевірити узгодженість теоретичних і емпіричних частот за критерієм Колмогорова.

<u>***Розв'язання:***</u> по заданому статистичному розподілу побудуємо емпіричну функцію розподілу $F_n(x) = \frac{n_X}{n}$, де n_X – кількість значень ознаки, які не перевищують величину X.

Наприклад,

$F_n(x) = \frac{30}{60} = 0,5$, при $0 < x \leq 5$,

$F_n(x) = \frac{46}{60} = 0,77$, при $5 < x \leq 10$.

Значення теоретичної функції розподілу для кожного з інтервалів обчислюємо за формулою: $F(x) = \begin{cases} 1 - e^{-0,14x}, x \geq 0, \\ 0, x < 0 \end{cases}$ для правого кінця проміжку.

Наприклад,

$F(x) = 1 - e^{-0,14 \cdot 5} = 0,503$, при $0 < x \leq 5$,

$F(x) = 1 - e^{-0,14 \cdot 10} = 0,753$, при $5 < x \leq 10$.

Обчислення проведемо в табл.6.4.

Таблиця 6.4

Розрахунок величини D_n

Інтервали	n_i	$F_n(x)$	$F(x)$	$\lvert F_n(x) - F(x)\rvert$
< 0	0	0	0	0
[0;5)	30	0,5	0,503	0,003
[5;10)	16	0,77	0,753	0,017
[10;15)	7	0,88	0,878	0,002
[15;20)	3	0,93	0,939	0,009
[20;25)	2	0,97	0,969	0,001
[25;30)	1	0,98	0,985	0,005
[30;35)	1	1	0,993	0,007

Найбільше значення в останньому стовпці $D_n = 0{,}017$ визначає величину $\lambda = 0{,}017\sqrt{60} = 0{,}132$. По таблицях значень функції (6.8) маємо P(0,132)=1 – це за критерієм Колмогорова свідчить про те, гіпотезу про розподіл часу обслуговування заявок на АТС за експоненціальним законом варто визнати не суперечною експерименту.

ДОМАШНЄ ЗАВДАННЯ 6

Варіанти 1-5

Задано вибірку *X*:

12	14	19	15	14	18	13	16	17	12
20	17	15	13	17	16	20	14	14	13
17	16	15	19	16	15	18	17	15	14
16	15	15	18	15	15	19	14	16	18
18	15	15	17	15	16	16	14	14	17

1) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про нормальність розподілу генеральної сукупності;

2) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про біноміальність розподілу генеральної сукупності;

3) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про експоненційність розподілу генеральної сукупності;

4) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність математичних сподівань нормальних генеральних сукупностей за допомогою критерію Колмогорова;

5) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про нормальність розподілу генеральної сукупності за допомогою методу спрямлюючих діаграм (Гмурман, стор. 259).

Варіанти 6-10

Задано вибірку *X*:

59	55	60	46	57	54	57	54	55	46
57	59	57	46	54	55	57	59	54	54
46	55	57	60	55	54	46	55	57	54
59	47	46	55	60	47	54	57	59	60

1) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про нормальність розподілу генеральної сукупності;

2) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про біноміальність розподілу генеральної сукупності;

3) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про експоненційність розподілу генеральної сукупності;
4) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність математичних сподівань нормальних генеральних сукупностей за допомогою критерію Колмогорова;
5) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про нормальність розподілу генеральної сукупності за допомогою методу спрямлюючих діаграм (Гмурман, стор. 259).

Варіанти 11-15

Задано вибірку *X*:

6,80	6,50	6,53	7,30	6,95	6,90	6,93	6,72	6,54	6,79
6,80	6,75	6,62	6,65	7,00	6,83	6,70	6,69	6,54	6,97
6,82	6,67	6,91	6,93	6,76	6,71	6,66	6,76	6,79	6,82
6,71	6,91	6,90	7,00	6,53	6,72	6,66	6,80	6,82	6,81

1) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про нормальність розподілу генеральної сукупності;
2) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про біноміальність розподілу генеральної сукупності;
3) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про експоненційність розподілу генеральної сукупності;
4) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність математичних сподівань нормальних генеральних сукупностей за допомогою критерію Колмогорова;
5) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про нормальність розподілу генеральної сукупності за допомогою методу спрямлюючих діаграм (Гмурман, стор. 259).

Варіанти 16-20

Задано вибірку *X*:

12	15	14	14	19	18	17	15	17	19
13	17	19	16	15	18	20	20	12	13
13	17	12	16	15	18	20	14	12	13
12	17	19	16	15	18	20	19	12	13

1) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про нормальність розподілу генеральної сукупності;
2) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про біноміальність розподілу генеральної сукупності;
3) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про експоненційність розподілу генеральної сукупності;
4) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність математичних сподівань нормальних генеральних сукупностей за допомогою критерію Колмогорова;
5) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про нормальність розподілу генеральної сукупності за допомогою методу спрямлюючих діаграм (Гмурман, стор. 259).

Варіанти 21-25

Задано вибірку *X*:

75	85	84	81	84	80	82	76	75	77
80	82	81	84	85	77	76	84	83	84
78	77	88	86	87	79	80	79	78	87
78	77	88	86	84	79	80	79	78	81

1) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про нормальність розподілу генеральної сукупності;
2) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про біноміальність розподілу генеральної сукупності;
3) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про експоненційність розподілу генеральної сукупності;
4) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність математичних сподівань нормальних генеральних сукупностей за допомогою критерію Колмогорова;
5) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про нормальність розподілу генеральної сукупності за допомогою методу спрямлюючих діаграм (Гмурман, стор. 259)

Варіанти 26-30

Задано вибірку *X*:

55	55	60	46	54	57	59	60	57	46
55	46	60	46	54	57	60	60	57	59
47	47	55	46	54	60	55	54	57	59
47	57	59	46	55	54	59	59	54	60

1) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про нормальність розподілу генеральної сукупності;
2) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про біноміальність розподілу генеральної сукупності;
3) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про експоненційність розподілу генеральної сукупності;
4) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність математичних сподівань нормальних генеральних сукупностей за допомогою критерію Колмогорова;
5) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про нормальність розподілу генеральної сукупності за допомогою методу спрямлюючих діаграм (Гмурман, стор. 259)

ПРАКТИЧНЕ ЗАНЯТТЯ 7

Кореляційний та регресійний аналіз. Оцінки лінійної, поліноміальної та багатофакторної регресії за методом найменших квадратів

При одночасній появі двох і більше величин у результаті проведення експерименту дослідник має підстави для встановлення певної залежності між ними.

Строгої функціональної залежності між змінними, у буквальному розумінні цього слова, у реальному світі не існує, бо вони перебувають під впливом випадкових факторів, наслідки яких передбачити практично неможливо. Тому між змінними існує особлива форма зв'язку, яку називають стохастичною і яка трансформується, не змінюючи своєї суті, у статистичну залежність.

Показником, що вимірює стохастичний зв'язок між змінними, є *коефіцієнт кореляції*, який свідчить з певною мірою ймовірності, наскільки зв'язок між змінними близький до строгої лінійної залежності.

Для двовимірного статистичного розподілу вибірки введемо поняття статистичної залежності між ознаками *X* та *Y*.

*Статичною залежніст*ю ознаки *X* від ознаки *Y* називають таку залежність, за якої при зміні значень ознаки $Y = y_i$ змінюється умовний статистичний розподіл ознаки *X*.

Статистичною залежністю ознаки Y від ознаки *X* називають таку залежність, за якої зі зміною значень ознаки $X = x_i$ змінюється умовний статистичний розподіл ознаки *Y*.

У разі зміни умовних статистичних розподілів змінюватимуться й умовні числові характеристики.

Звідси випливає визначення кореляційної залежності між ознаками *X* і *Y*.

Кореляційною залежністю ознаки *Y* від *X*, або *регресією Y* на *X*, називається функціональна залежність умовного середнього y_{x_j} від аргументу *x*:

$$\overline{y_x} = \varphi(x). \tag{7.1}$$

Аналогічно *кореляційною залежністю* ознаки *X* від *Y* називається функціональна залежність умовного середнього x_{y_i} від аргументу у:

$$\overline{x_y} = \psi(y). \tag{7.2}$$

Графіки функцій $\varphi(x)$ та $\psi(y)$ *називають кривими регресії*.

Між ознаками *X* та *Y* може існувати статистична залежність і за відсутності кореляційної. Але коли існує кореляційна залежність між ознаками *X* та *Y*, то обов'язково між ними існуватиме і статистична залежність.

Лінійна регресія

Для функції $\varphi(x)$ припускаються відомими властивості гладкості або її вигляд з точністю до деяких параметрів. Наприклад,

$$\overline{y_x} = \varphi(x, \beta_0, \beta_1) \text{ або } \overline{y_x} = \varphi\left(x, \beta_0, \beta_{1,}, \dots \beta_{m,}\right)$$

У частинному випадку, якщо

$$\varphi(x, \beta_0, \beta_1) = \beta_0 + \beta_1 x,$$

то говорять про *лінійну регресію Y* на *X*.

Пряма

$$y = \beta_0 + \beta_1 x$$

називається *прямою лінійної регресії* з невідомими параметрами β_0 і β_1

Для вибіркових даних X:$\{x_1, x_2, \dots, x_n\}$ і Y:$\{y_1, y_2, \dots, y_n\}$ природно визначити залежність рівностями

$$y_i = \beta_0 + \beta_1 x_1 + \delta_i, (i = \overline{1; n}),$$

де δ_i — незалежні випадкові величини, що характеризують відхилення у від гіпотетичної теоретичної регресії з $M(\delta_i)=0$ і однаковими дисперсіями $D(\delta_i)=\sigma^2$. При цьому елементи послідовності $\delta_1, \delta_2, \dots, \delta_n$ є некорельованими ($K_{ij}=0$).

Для нелінійного випадку аналогічно припускається, що

$$y_i = \varphi(x_i) + \delta_i, \quad M(\delta_i) = 0, \quad D(\delta_i) = \sigma^2.$$

Щоб визначити параметри функції $\varphi(x)=\varphi(x, \beta_0, \beta_{1,} \dots \beta_m)$, застосовують метод найменших квадратів. Для цього складають суму квадратів відхилень

$$Q = \sum_{i=1}^{n} \delta_i^2, \qquad Q\left(\beta_0, \beta_{1,} \dots \beta_m\right) = \sum_{i=1}^{n} \left(y_i - \varphi\left(x, \beta_0, \beta_{1,} \dots \beta_m\right)\right)^2$$

і вимагають, щоб вона для вибіркових даних була мінімальною. Статистичні оцінки $\overline{\beta_0}$, $\overline{\beta_1}$, ..., $\overline{\beta_m}$ параметрів β_0, β_1,..., β_m знаходять як розв'язки системи рівнянь:

$$\frac{\partial Q}{\partial \beta_0} = 0, \dots, \frac{\partial Q}{\partial \beta_N} = 0.$$

Метод найменших квадратів для параметрів β_0 , β_1 лінійної регресії дає такі оцінки:

$$\bar{\beta}_1 = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}, \quad \bar{\beta}_0 = \bar{y}_B - \bar{\beta}_1 \bar{x}_B. \tag{7.3}$$

Коефіцієнт, $\overline{\beta_1}$ можна знайти також за формулою

$$\bar{\beta}_1 = \frac{\bar{K}_{XY}}{s_X^2}. \tag{7.4}$$

Якщо елементи вибірки (x_i, y_i), $(i = \overline{1; n})$ перетворити за формулами

$$x_i^0 = \frac{x_i - \bar{x}_B}{\sigma_{BX}}, \quad y_i^0 = \frac{y_i - \bar{y}_B}{\sigma_{BY}},$$

то пряма регресії буде визначатися співвідношенням $y^0=r_B x^0$.

Аналогічно оцінки параметрів β_0', β_1' лінійної регресії X на Y обчислюється за формулами

$$\bar{\beta}_1' = \frac{\bar{K}_{XY}}{s_Y^2}; \quad \bar{\beta}_0' = \bar{x}_B - \bar{\beta}_1' \bar{y}_B \tag{7.5}$$

Співвідношення

$$\sqrt{\bar{\beta}_1 \bar{\beta}_1'} = |r_B| \tag{7.6}$$

Приклад 7.1. Дано вибірку

x_i 3,72 3,09 3,47 3,25 3,34 3,11 3,51 3,34 3,68 3,40

y_i 1,49 2,73 3,32 3,32 3,69 3,67 3,30 2,55 3,11 3,60

Обчислити коефіцієнт кореляції, визначити і побудувати прямі регресії Y на X та X на Y.

Розв'язання. Знаходимо $\overline{x_{\text{В}}}=3{,}39$; $\overline{y_{\text{В}}}=3{,}08$. Складаємо табл. 7.1.

Таблиця 7.1

i	x_i	$x_i-\overline{x}_B$	y_i	$y_i-\overline{y}_B$	$(x_i-\overline{x}_B)\,(y_i-\overline{y}_B)$	$(x_i-\overline{x}_B)^2$	$(y_i-\overline{y}_B)^2$
1	3,72	0,33	1,49	– 1,59	– 0,525	0,109	2,528
2	3,09	– 0,3	2,73	– 0,35	0,105	0,09	0,123
3	3,47	0,08	3,32	0,24	0,019	0,006	0,058
4	3,25	– 0,14	3,32	0,24	– 0,034	0,020	0,058
5	3,34	– 0,05	3,69	0,61	– 0,03	0,003	0,372
6	3,11	– 0,28	3,67	0,59	– 0,165	0,078	0,348
7	3,51	0,12	3,30	0,22	0,026	0,014	0,048
8	3,34	– 0,05	2,55	– 0,53	0,027	0,003	0,281
9	3,68	0,29	3,11	0,03	0,009	0,084	0,001
10	3,40	0,01	3,60	0,52	0,005	0,000	0,270
$\sum$	33,91		30,78		– 0,56	0,41	4,09

Знаходимо s_X^2=0,05;σ_{AX}=0,22; s_Y^2 =0,45; σ_{BY}=0,67; $\overline{K_{XY}}$=–0,06; r_B=–0,41.

Прямі регресії Y на X та X на Y мають відповідно вигляд

$$y=\beta_0+\beta_1 x \text{ та } x=\beta'_0+\beta' y.$$

Оцінки для параметрів β_0 і β_1 знаходимо за формулами (7.3) та (7.4):

$$\overline{\beta_1}=-0{,}06/0{,}05=-1{,}2;\ \overline{\beta_0}=3{,}08-(-1{,}2)\cdot 3{,}39=7{,}15$$

Оцінки параметрів β'_0 та $\beta_1{}'$ знаходимо за формулами (7.5):

$$\overline{\beta_1{}'}=-0{,}06/0{,}45=-0{,}13;\ \overline{\beta_0{}'}=3{,}39-(-0{,}13)\cdot 3{,}08=3{,}8.$$

Таким чином, прямі регресії мають такі рівняння: y=7,15–1,2x та x=3,8–0,13y. Для перевірки застосуємо співвідношення (7.6):

$$\sqrt{\overline{\beta}_1\overline{\beta}'_1}=\sqrt{(-1{,}2)\cdot(-0{,}13)}=0{,}39.$$

З урахуванням помилок округлення можна вважати, що

$$\sqrt{\overline{\beta}_1\overline{\beta}'_1}=|r_{\text{В}}|.$$

Прямі регресії подано на рис. 7.1.

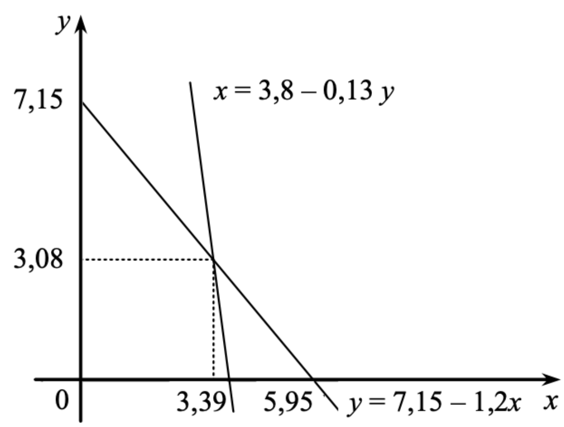


Рис. 7.1. Прямі регресії

Поліноміальна регресія

Як зазначалося вище, лінійна регресія є частинним випадком поліноміальної регресії.

У статистиці, *поліноміальна регресія* є однією з форм регресійного аналізу, в якому залежність між незалежною змінною x і залежною змінною y моделюється як поліном від x степеня n.

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_m x^m + \varepsilon, \quad (7.7)$$

де β_0-β_m – коефіцієнти, а ε – помилка.

Модель поліноміальної регресії можна записати як систему лінійних рівнянь:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^m \\ 1 & x_2 & x_2^2 & \dots & x_2^m \\ 1 & x_3 & x_3^2 & \dots & x_3^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^m \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad (7.8)$$

що при використанні чисто матричної нотації записується як

$$\vec{y} = \mathbf{X}\vec{\beta} + \vec{\varepsilon}.$$

Вектор розрахункових коефіцієнтів поліноміальної регресії (за допомогою оцінки вичайних найменших квадратів) становить

$$\widehat{\vec{\beta}} = (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\,\mathbf{X}^\mathsf{T}\vec{y}. \quad (7.9)$$

Квадратична регресія

У випадку, коли n=2, маємо квадратичну регресію. Її рівняння

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 \quad (7.10)$$

та для оцінок $\overline{\beta_0}$, $\overline{\beta_1}$, $\overline{\beta_2}$ параметрів β_0, β_1, β_2 дістаємо таку систему лінійних рівнянь:

$$\begin{gathered} \bar{\beta}_2\sum_{i=1}^{n} x_i^4 + \bar{\beta}_1\sum_{i=1}^{n} x_i^3 + \bar{\beta}_0\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i^2 y_i, \\ \bar{\beta}_2\sum_{i=1}^{n} x_i^3 + \bar{\beta}_1\sum_{i=1}^{n} x_i^2 + \bar{\beta}_0\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i, \\ \bar{\beta}_2\sum_{i=1}^{n} x_i^2 + \bar{\beta}_1\sum_{i=1}^{n} x_i + \bar{\beta}_0 n = \sum_{i=1}^{n} y_i. \end{gathered} \quad (7.11)$$

Для спрощення обчислень вихідні дані часто попередньо перетворюють.

Приклад 7.2. Заміри температури повітря на протязі доби в другій половині вересня в середній смузі кожні 4 години дали наступні результати:

t, час	0	4	8	12	16	20	24
T, °C	10	6	9	13	17	14	11

Вважаючи, що залежність між цими змінними має вигляд $T=\beta_0 + \beta_1 t + \beta_2 t^2$, знайти оцінки параметрів β_0, β_1, β_2 за методом найменших квадратів.

Розв'язання. Попередньо перетворимо вихідні дані за формулами $x = \frac{t-12}{4}$, $y = T - 11$ і знайдемо оцінки параметрів моделі $y = \beta_0 + \beta_1 x + \beta_2 x^2$. Для підрахунку коефіцієнтів системи (7.11) складемо табл. 7.2.

Таблиця 7.2

t_i	x_i	T_i	y_i	x_iy_i	x_i^2	$x_i^2y_i$	x_i^3	x_i^4
0	– 3	10	– 1	3	9	– 9	– 27	81
4	– 2	6	– 5	10	4	– 20	– 8	16
8	– 1	9	– 2	2	1	– 2	– 1	1
12	0	13	2	0	0	0	0	0
16	1	17	6	6	1	6	1	1
20	2	14	3	6	4	12	8	16
24	3	11	0	0	9	0	27	81
∑	0		3	27	28	– 13	0	196

Система для обчислення оцінок параметрів β_0, β_1, β_2 має вигляд:

$$28\overline{\beta_0} + 196\overline{\beta_2} = -13,$$
$$28\overline{\beta_1} = 27,$$
$$7\overline{\beta_0} + 28\overline{\beta_2} = 3.$$

Розв'язавши цю систему, одержуємо $\overline{\beta_0}=1{,}63$; $\overline{\beta_1}=0{,}96$; $\overline{\beta_2}=-0{,}3$.
Тобто $y = 1{,}63 + 0{,}96_1 x - 0{,}3x^2$.
Повертаючись до вихідних змінних, одержуємо

$$T - 11 = 1{,}63 + 0{,}96\left(\frac{t-12}{4}\right) - 0{,}3\left(\frac{t-12}{4}\right)^2.$$

Або

$$T = 6{,}87 + 0{,}72t - 0{,}02t^2.$$

Багатофакторна регресія

В залежності від кількості вхідних параметрів системи (факторних даних) розрізняють парну або однофакторну залежність, коли є один вхідний аргумент (як розглядалося вище) і, якщо ж аргументів більше, ніж один, то залежність називається *множинною* або *багатофакторною*. Основна мета множинної регресії – побудувати модель з великою кількістю факторів, визначивши при цьому вплив кожного з них окремо, а також сукупну дію на результативний показник. В загальному рівняння багатофакторної регресії має вигляд:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon. \tag{7.12}$$

Рівняння (7.12) отримано з рівняння поліноміальної регресії (7.7) якщо замінити:

$$x_1 = x, x_2 = x^2, \ldots, x_n = x^n.$$

Коефіцієнти визначаються з формули (7.9).

Оцінки якості регресійної моделі

- *коефіцієнт множинної кореляції R* – показує тісноту зв'язку вихідної змінної від вхідних

$$R = \pm\sqrt{1 - \frac{\sum_{i=1}^{n}(Y_{\text{факт}} - Y_{\text{розр}})^2}{\sum_{i=1}^{n}(Y_{\text{факт}} - Y_{\text{сер}})^2}}, \tag{7.13}$$

де $Y_{розр}$ – теоретичні значення залежної змінної на підставі побудованої регресійної моделі; $Y_{сер}$ – загальна середня фактичних даних залежної змінної; $Y_{факт}$ – фактичні індивідуальні значення залежної змінної;

- *коефіцієнт детермінації* R^2 – показує наскільки вихідна змінна, яка визначена на основі побудованої моделі відповідає реальним даним. Якщо R^2 близький до 0, то це означає низьку значимість моделі, відсутні лінійна залежність між параметрами

$$R^2 = \frac{\sum_{i=1}^{n}(Y_{\text{факт}} - Y_{\text{розр}})^2}{\sum_{i=1}^{n}(Y_{\text{факт}} - Y_{\text{сер}})^2}. \tag{7.14}$$

- *стандартна помилка* - показує середню величину відхилення вихідної змінної y^*, яка обрахована на основі моделі, від наявних статистичних значень вихідної змінної y. Обчислюється як відносна похибка

$$\delta = \frac{|Y_{\text{факт}} - Y_{\text{розр}}|}{Y_{\text{факт}}} 100\%.$$

- *F-тест та t-тест* проводять для оцінки значущості моделі. Для цього використовують порівняння так званих фактичних та табличних (теоретичних) значень F-критерію (критерій Фішера) та t-критерію (критерій Стьюдента).

- *t-тест (або t-критерій)* досліджує лінійну залежність між кожною окремою вхідною змінною (X_i) і вихідною змінною (Y), *F-тест (або F-критерій)* досліджує лінійну залежність для всієї моделі, тобто між набором вхідних змінних ($X_1, X_2, \ldots, X_n$) та вихідною змінною (Y) рис.7.2.

Рис.7.2. Пояснення критерію Фішера та Стьюдента.

Опис F-тесту та t-тесту наведено в практичному занятті 5.

ДОМАШНЄ ЗАВДАННЯ 7

Варіанти 1-5

1. Під час обробки даних випробувань літака було зафіксовано такі показники швидкості V відриву літака при зльоті за даної зустрічної швидкості вітру v, м/с:

v	9,8	10	11	11,4	13	14	14,7
V	63,8	64,2	64,7	65,3	66	66,7	68

Знайти статистичні оцінки математичного сподівання, дисперсії, кореляційного момента та коефіцієнта кореляції.

2. За даними спостережень двох випадкових величин *X* і *Y* потрібно:

а) знайти статистичні оцінки математичного сподівання, дисперсії, кореляційного момента та коефіцієнта кореляції випадкових величин, що входять у систему;

б) перевірити гіпотезу про залежність системи випадкових величин (X, Y) на рівні значущості $\alpha=0,05$;

в) знайти статистичні оцінки для параметрів β_0, β_1 лінійної регресії $y = \beta_0 + \beta_1 x$.

г) обчислити коефіцієнт R для регресії з попереднього пункту.

$X = x_i$	0,531	0,524	0,541	0,550	0,559	0,620	0,632	0,672
$Y = y_i$	0,620	0,580	0,640	0,650	0,670	0,680	0,695	0,699

Варіанти 6-10

1. Результати аналізу залежності кількості набраних балів абітурієнтами X від кількості абітурієнтів Y на вступних іспитах подано вибіркою:

X	45	43	42	41	40	39
Y	25	38	65	95	120	140

Знайти статистичні оцінки математичного сподівання, дисперсії, кореляційного момента та коефіцієнта кореляції.

2. За даними спостережень двох випадкових величин X і Y потрібно:

а) знайти статистичні оцінки математичного сподівання, дисперсії, кореляційного момента та коефіцієнта кореляції випадкових величин, що входять у систему;

б) перевірити гіпотезу про залежність системи випадкових величин (X, Y) на рівні значущості α=0,05;

в) знайти статистичні оцінки для параметрів β_0, β_1 лінійної регресії $y = \beta_0 + \beta_1 x$.

г) обчислити коефіцієнт R^2 для регресії з попереднього пункту.

$X = x_i$	44	43	42	41	40	39	38	37
$Y = y_i$	10	25	68	136	152	162	170	180

Варіанти 11-15

1. Залежність річної заробітної плати Y від загального виробітку X подано у вигляді двовимірного статистичного розподілу:

$Y = y_i$	$X = x_j$					
	1,5	2,5	3,5	4,5	5,5	n_{y_i}
0,82	1	3	–	–	–	4
0,86	–	3	2	1	–	6
0,9	–	2	5	9	3	19
0,94	–	–	–	6	4	10
0,98	–	–	–	–	2	2
n_{x_j}	1	8	7	16	9	41

Знайти статистичні оцінки математичного сподівання, дисперсії, кореляційного момента та коефіцієнта кореляції.

2. За даними спостережень двох випадкових величин X і Y потрібно:

а) знайти статистичні оцінки математичного сподівання, дисперсії, кореляційного момента та коефіцієнта кореляції випадкових величин, що входять у систему;

б) перевірити гіпотезу про залежність системи випадкових величин (X, Y) на рівні значущості α=0,05;

в) знайти статистичні оцінки для параметрів β_0, β_1 лінійної регресії $y = \beta_0 + \beta_1 x$.

г) обчислити похибку δ для регресії з попереднього пункту.

$X = x_i$	– 20,2	– 20,5	– 21,4	– 21,8	– 22,0	– 22,5	– 22,8	– 22,8
$Y = y_i$	– 10,2	– 11,5	– 12,4	– 12,8	– 13	– 13,5	– 14,2	– 14,6

Варіанти 16-20

1. Залежність висоти прольоту дальнього привода H, м, під час заходу літака на посадку від мінімальної швидкості літака V, м/с, наведено у вигляді двовимірного статистичного розподілу:

$H = h_i$	$V = v_j$					
	53	53,5	54	54,5	55	n_{h_i}
322	–	1	3	–	1	
324	2	2	–	–	–	
326	–	–	2	3	–	
328	3	1	2	–	–	
330	2	–	1	–	–	
n_{v_j}						

Знайти статистичні оцінки математичного сподівання, дисперсії, кореляційного момента та коефіцієнта кореляції.

2. За даними спостережень двох випадкових величин X і Y потрібно:

а) знайти статистичні оцінки математичного сподівання, дисперсії, кореляційного момента та коефіцієнта кореляції випадкових величин, що входять у систему;

б) перевірити гіпотезу про залежність системи випадкових величин (X, Y) на рівні значущості α=0,05;

в) знайти статистичні оцінки для параметрів β_0, β_1 лінійної регресії $y = \beta_0 + \beta_1 x$.

г) обчислити коефіцієнт R^2 для регресії з попереднього пункту.

$X = x_i$	30	35	31	38	41	48	50	55
$Y = y_i$	45	25	48	52	54	51	59	60

Варіанти 21-25

1. Під час обробки даних випробувань літака було зафіксовано такі показники швидкості V відриву літака при зльоті за даної зустрічної швидкості вітру ν, м/с:

v	9,8	10	11	11,4	13	14	14,7
V	63,8	64,2	64,7	65,3	66	66,7	68

Перевірити гіпотезу про незалежність цих випадкових величин на рівні значущості α=0,001.

2. За даними спостережень двох випадкових величин X і Y потрібно:

а) знайти статистичні оцінки математичного сподівання, дисперсії, кореляційного момента та коефіцієнта кореляції випадкових величин, що входять у систему;

б) перевірити гіпотезу про залежність системи випадкових величин (X, Y) на рівні значущості $\alpha=0,05$;

в) знайти статистичні оцінки для параметрів β_0, β_1 лінійної регресії $y = \beta_0 + \beta_1 x$.

$X = x_i$	0,689	0,692	0,694	0,698	0,690	0,710	0,720	0,725
$Y = y_i$	0,715	0,725	0,781	0,790	0,795	0,800	0,810	0,850

г) провести F-тест для регресії з попереднього пункту.

Варіанти 26-30

1. Результати аналізу залежності кількості набраних балів абітурієнтами X від кількості абітурієнтів Y на вступних іспитах подано вибіркою:

X	45	43	42	41	40	39
Y	25	38	65	95	120	140

Перевірити гіпотезу про незалежність цих випадкових величин на рівні значущості $\alpha=0,001$.

2. За даними спостережень двох випадкових величин X і Y потрібно:

а) знайти статистичні оцінки математичного сподівання, дисперсії, кореляційного момента та коефіцієнта кореляції випадкових величин, що входять у систему;

б) перевірити гіпотезу про залежність системи випадкових величин (X, Y) на рівні значущості $\alpha=0,05$;

в) знайти статистичні оцінки для параметрів β_0, β_1 лінійної регресії $y = \beta_0 + \beta_1 x$.

г) провести t-тест для регресії з попереднього пункту.

$X = x_i$	24	26	28	30	32	34	36	38
$Y = y_i$	30	32	34	36	38	40	42	44

ПРАКТИЧНЕ ЗАНЯТТЯ 8

Випадкові процеси, Марківські процеси, потоки події. Теорія масового обслуговування.

8.1. Основні поняття.

Випадкова функція (випадковий процес) $X(t)$ є функція, яка в результаті експерименту може прийняти той або іншій заздалегідь невідомий конкретний вигляд. Цей конкретний вигляд, який приймає випадкова функція в результаті експерименту, називається *реалізацією випадкової функції* (випадкового процесу).

Переріз випадкової функції є випадкова величина при фіксованому t.

Одновимірним законом розподілу випадкової величини $X(t)$ називається закон розподілу $f(x, t)$ перерізу $x(t)$ випадкової функції.

Випадкова функція $X(t)$ називається *нормальною*, якщо закон розподілу системи любої кількості числа n її перерізів є n-вимірний нормальний закон.

Математичним сподіванням випадкової функції $X(t)$ називається невипадкова функція $m_x(t)$, яка при кожному t є математичне сподівання відповідного перерізу випадкової функції: $m_x(t)=M(X(t))$.

Дисперсія випадкової функції $X(t)$ є невипадкова величин $D_x(t)$, яка при любому значенні змінної t дорівнює дисперсії відповідного перерізу випадкової функції $X(t)$, тобто $D_x(t)=D(X(t))$.

Середнє квадратичне відхилення $\sigma_x(t)$ випадкової функції $X(t)$ дорівнює:

$$\sigma_x(t)=\sqrt{D_x(t)}.$$

Математичне сподівання випадкової функції характеризує середню траєкторію всіх можливих її реалізацій, а середнє квадратичне відхилення – розкид реалізацій відносно середньої траєкторії.

Кореляційною функцією випадкового процесу $X(t)$ називається невипадкова функція $K_x(t_1,t_2)M\left(X(t_1)-m_x(t_1)\right)\left(X(t_2)-m_x(t_2)\right)$, яка характеризує ступень тісноти лінійної залежності між перерізами, а також розкид цих перерізів відносно математичного сподівання.

Нормованою кореляційною функцією випадкового процесу $X(t)$ називається функція $r_x(t_1,t_2)$, яка обчислюється за формулою:

$$r_x(t_1,t_2)=\frac{k_x(t_1,t_2)}{\sigma_x(t_1)\sigma_x(t_2)}.$$

Приклад 8.1. Випадкова функція $X(t)$ є ступінчата знакозмінна функція (рис. 8.1), яка через одиничні інтервали приймає значення: +1, –1. Положення ступінчатої функції відносно початку відліку випадково; випадкова величина Т, що характеризує зсув першої точки зміни знаку відносно початку координат є випадковою величиною, розподіленою рівномірно в інтервалі (0,1). Знайти математичне сподівання $m_x(t)$, дисперсію $D_x(t)$ і кореляційну функцію $r_x(t_1,t_2)$.

Розв'язування. Розглянемо переріз випадкової функції $X(t)$. Він може з однаковою ймовірністю попасти як на ділянку, де випадкова функція дорівнює 1, так і на ділянку, де вона рівна –1. Тому ряд розподілу для любого перерізу має вигляд:

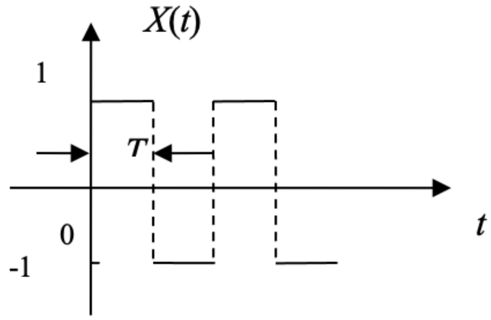


Рис.8.1

Ряд розподілу:

–1	1
1/2	1/2

Числові характеристики:

$$m_x(t) = (-1)\cdot 1/2 + 1\cdot 1/2 = 0;$$

$$D_x(t) = (-1)^2 \cdot \frac{1}{2} + 1^2 \cdot \frac{1}{2} = 1\cdot$$

$$K_x(t, t+\tau) = M\big(X(t)\cdot X(t+\tau)\big).$$

Оскільки, добуток $X(t)\cdot X(t+\tau)$ може приймати тільки два значення +1 або –1, то $X(t)\cdot X(t+\tau) = 1\cdot p_1 + (-1)(1-p_1) = 2p_1 - 1$, де p_1 – ймовірність того, що точки t і $t+\tau$ попадуть на ділянки, в яких $X(t)$ та $X(t+\tau)$ мають однаковий знак. Через рівномірності розподілу зсуву T можна перенести початок відлину до лівого кінця тої ділянки, на якої знаходиться точка t_i вважати, що точка t рівномірно розподілена в інтервалі (0,1). Тоді ймовірність p_1 є ймовірність того, $t+\tau$ попадає в деякий з інтервалів, ($2n,2n+1$); $n = 0, \pm1, \pm2, \ldots$.

Обчислимо цю ймовірність для різних значень τ.

При $0 \le \tau \le 1$ точка $t+\tau$ може попасти в інтервал (0,1)**,** або в інтервал (1,2). Тому

$$p_1 = p(t+\tau < 1) = p(t < 1-\tau) = 1-\tau.$$

При $1 \le \tau \le 2$ точка $(t+\tau)$ може попасти в інтервал (1,2), або в інтервал (2,3). Тому

$$p_1 = p(t+\tau > 2) = p(t > 2-\tau) = 1-(2-\tau) = \tau - 1.$$

Якщо продовжити ці міркування, то

$$P_1 = \begin{cases} -1-(\tau-2n), & 2n < \tau < 2n+1,\ x < 0 \\ (\tau-2n)-1, & 2n+1 < \tau < 2n+2 \end{cases},$$

$$K_x(t, t+\tau) = k_x(\tau) = \begin{cases} 4n+1-2\tau, & 2n < \tau < 2n+1,\ x<0 \\ 2\tau-(4n+3), & 2n+1 < \tau < 2n+2 \end{cases}.$$

Графік кореляційної функції $k_2(\tau)$ представлено на рис. 8.2.

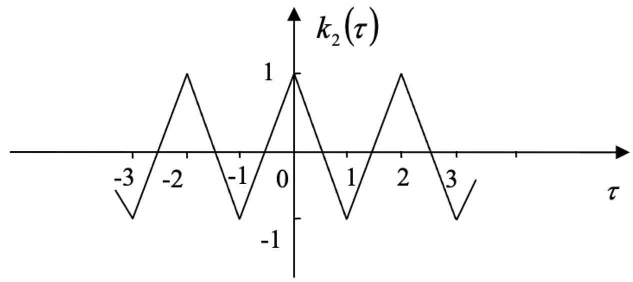


Рис. 8.2. Графік кореляційної функції

Випадковий процес є процес переходу системи зі стану в стан під впливом випадкових факторів.

Система X називається *системою з дискретним станами,* якщо вона має злічену (в частковому випадку – скінчену) множину можливих станів $X_1, X_2, \dots, X_n$ і перехід з одного стану в інший відбувається стрибком. Можливі стани системи X зображуються, за допомогою графіків станів, на якому стани системи зображуються прямокутниками, а можливі переходи системи зі стану в стан – стрілками, які з'єднують відповідні прямокутники. Для того, щоб описати випадковий процес – в системі з дискретними станами, використовують ймовірності станів $p_1(t), p_2(t), \dots, p_n(t)$, де $p_k(t)$ (k=1,…n) – ймовірність того, що в момент t система знаходиться в стані X_k причому $\sum_{k=1}^{n} p_k(t) = 1$.

Випадковий процес, який відбувається в системі X, називається *процесом з дискретним часом*, якщо перехід системи зі стану в стан можливий тільки у визначені моменти часу $t_1, t_2, \dots, t_n$

Якщо переходи можливі в будь-який момент часу, то процес називається *процесом з неперервним часом*.

Потік подій – це послідовність подій, які відбуваються один за одним у випадкові моменти часу.

Щільністю (інтенсивністю) потоку подій називається середнє число подій в одиницю часу.

Потік подій називається *потоком без післядії*, якщо ймовірність появи на будь-якому проміжку часу того або іншого числа подій не залежить від того, яке число подій попало на інші проміжки. які не перетинаються з даним проміжком. Наприклад, потік пасажирів, які входять у метро.

Потік подій називається *ординарним*, якщо ймовірність появи на елементарному проміжку часу Δt двох або більш подій мала в порівнянні з ймовірністю появи однієї події, тобто події з'являються на проміжку Δt поодинці, а не групами. Наприклад, потік потягів, які під'їжджають до станції, ординарне, а потік вагонів – ні.

Потік подій називається *регулярним*, якщо події слідують одна за одній через рівні визначені проміжки часу. Наприклад, потік виробів на конвеєрі.

Потік подій називається *стаціонарним*, якщо його ймовірності характеристики не залежать від часу. Наприклад, потік автомобілів на міській вулиці в час пік, протягом доби цей потік нестаціонарний.

Ординарний потік подій без післядії називається *пуассонівським потоком*. Якщо пуассоновський потік стаціонарне, то він називається *стаціонарним пуассонівським або найпростішим*.

Випадковий процес з дискретними станами називається *марківським*, якщо його ймовірнісні характеристики в майбутньому залежать тільки від того, в якому стані цей процес знаходиться в теперішній момент часу і не залежить від того, яким чином цей процес відбувався в минулому (майбутнє залежить від минулого тільки через теперішнє).

Якщо процес марківським, то всі потоки подій, які переводять систему зі стану до стану, є пуассонівськими.

8.2. Система рівнянь Колмогорова

Якщо процес є марківським, то для ймовірностей станів $p_1(t), p_2(t), \dots, p_n(t)$ можна скласти систему диференціальних рівнянь (рівняння Колмогорова). Для цього використовується розмічений граф станів системи, на якому проти кожної стрілки, що з'єднує стани, проставлена інтенсивність потоку подій. Інтенсивність потоку подій, яка переводить систему зі стану X_i до стану X_j позначається λ_{ij}.

Правило складання рівнянь Колмогорова:

В лівій частині кожного з рівнянь стоїть похідна ймовірності i-го стану $\frac{dp_i(t)}{dt}$.

В правій частині – стільки членів, скільки стрілок зв'язано безпосередньо з даним станом; якщо стрілка веде до цього стану, член має знак «+», якщо веде з цього стану, то член має знак «–». Кожен доданок дорівнює добутку інтенсивності потоку подій, який переводить систему за даній стрілки, на ймовірність того стану, з якого стрілка виходить.

Приклад 8.2. Записати рівняння Колмогорова для графа на рис. 8.3.

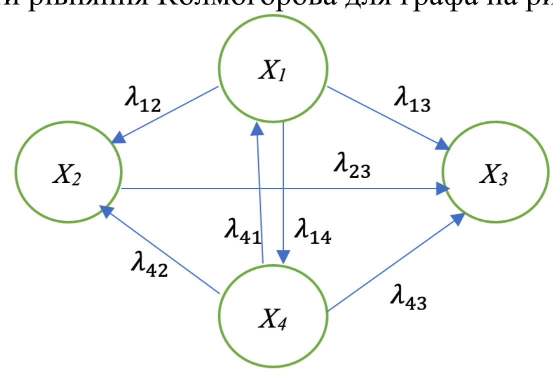


Рис. 8.3

Розв'язання.

$$
\begin{cases}
\dfrac{dp_1(t)}{dt} = -(\lambda_{12} + \lambda_{13} + \lambda_{14})p_1(t) + \lambda_{41}p_4(t) \\
\dfrac{dp_2(t)}{dt} = \lambda_{12}p_1(t) + \lambda_{42}p_4(t) - \lambda_{23}p_2(t) \\
\dfrac{dp_3(t)}{dt} = \lambda_{13}p_1(t) + \lambda_{23}p_2(t) + \lambda_{43}p_4(t) \\
\dfrac{dp_4(t)}{dt} = \lambda_{14}p_1(t) - (\lambda_{41} + \lambda_{42} + \lambda_{43})p_4(t) \\
p_1(t) + p_2(t) + p_3(t) + p_4(t) = 1
\end{cases}
$$

Число рівнянь в системі може бути зменшено на одиницю за рахунок останнього рівняння.

Початкові умови для інтегрування системи відображають стан системи в початковий момент.

Якщо система при $t = 0$ була в стані X_i, то $p_i(0)$=1, $p_j(0)$=0, $i \neq j$

Граничним режимом для системи X називається випадковий процес, який утворюється в системі при $t \to \infty$.

Оскільки граничні ймовірності станів є сталі, то в лівої частині маємо нулі, а система диференціальних рівнянь перетворюється на систему алгебраїчних рівнянь.

Для системи X, розмічений граф станів якої представлено на рис. 8.3, система алгебраїчних рівнянь, яка визначає граничний режим, буде такою:

$$\begin{cases} 0 = -(\lambda_{12} + \lambda_{13} + \lambda_{14})p_1(t) + \lambda_{41}p_4(t) \\ 0 = \lambda_{12}p_1(t) + \lambda_{42}p_4(t) - \lambda_{23}p_2(t) \\ 0 = \lambda_{13}p_1(t) + \lambda_{23}p_2(t) + \lambda_{43}p_4(t) \\ 0 = \lambda_{14}p_1(t) - (\lambda_{41} + \lambda_{42} + \lambda_{43})p_4(t) \\ p_1(t) + p_2(t) + p_3(t) + p_4(t) = 1 \end{cases}$$

Регулярним потоком подій називається потік, в якому події слідують одна за одною через строго визначені проміжки часу.

Процес «загибелі та розмноження» має вигляд на рис. 8.4:

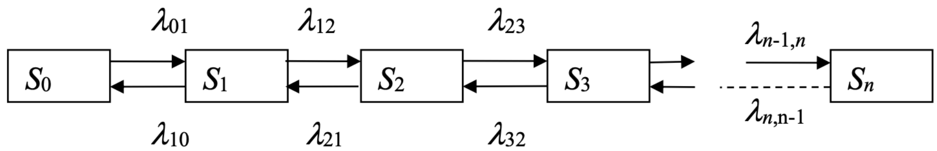


Рис. 8.4

Записуючи рівняння для граничних ймовірностей. Маємо систему алгебраїчних рівнянь:

$$\begin{cases} \lambda_{01}p_0(t) = \lambda_{10}p_1(t) \\ \lambda_{12}p_1(t) = \lambda_{21}p_2(t) \\ \lambda_{23}p_2(t) = \lambda_{32}p_3(t) \\ \dots \\ \lambda_{n-1n}p_{n-1}(t) = \lambda_{nn-1}p_n(t) \\ p_0(t) + p_1(t) + p_2(t) + p_3(t) + \dots + p_n(t) = 1 \end{cases}$$

Тоді

$$p_0(t) = \left(1 + \frac{\lambda_{01}}{\lambda_{10}} + \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} + \frac{\lambda_{23}\lambda_{12}\lambda_{01}}{\lambda_{32}\lambda_{21}\lambda_{10}} + \dots + \frac{\lambda_{n-1n} \dots \lambda_{23}\lambda_{12}\lambda_{01}}{\lambda_{nn-1} \dots \lambda_{32}\lambda_{21}\lambda_{10}}\right)^{-1}$$

$$p_1(t) = \frac{\lambda_{01}}{\lambda_{10}}p_0(t), p_2(t) = \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}}p_0(t), p_3(t) = \frac{\lambda_{23}\lambda_{12}\lambda_{01}}{\lambda_{32}\lambda_{21}\lambda_{10}}p_0(t), \dots,$$

$$p_n(t) = \frac{\lambda_{n-1n} \dots \lambda_{23}\lambda_{12}\lambda_{01}}{\lambda_{nn-1} \dots \lambda_{32}\lambda_{21}\lambda_{10}}p_0(t)$$

Приклад 8.3. Процес «загибелі та розмноження» представлено графом (рис. 8.5). Знайти граничні ймовірності станів.

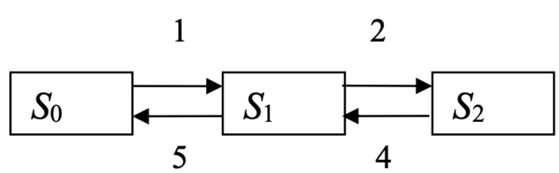


Рис. 8.5

Розв’язання.

$$p_0(t) = \left(1 + \frac{\lambda_{01}}{\lambda_{10}} + \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}}\right)^{-1} = \left(1 + \frac{1}{5} + \frac{1}{10}\right)^{-1} = \frac{10}{13},$$

$$p_1(t) = \frac{\lambda_{01}}{\lambda_{10}} p_0(t) = \frac{1}{5} p_0(t) = \frac{2}{13},$$

$$p_2(t) = \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} p_0(t) = \frac{1}{10} p_0(t) = \frac{1}{13}.$$

Циклічний процес. Граф станів циклічного процесу має вигляд (рис. 8.6).

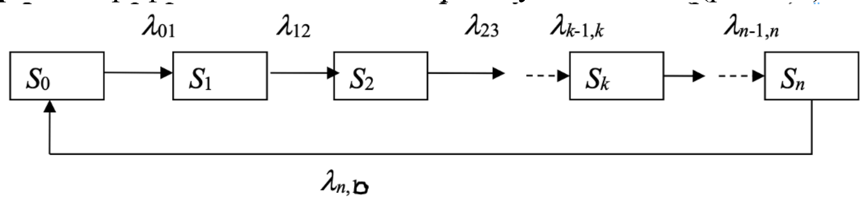


Рис. 8.6

Система алгебраїчних рівнянь для цього графу має вигляд:

$$\begin{cases} \lambda_{n0} p_n(t) = \lambda_{01} p_0(t) \\ \lambda_{12} p_1(t) = \lambda_{01} p_0(t) \\ \lambda_{23} p_2(t) = \lambda_{12} p_1(t) \\ \dots \\ \lambda_{n-1n} p_{n-1}(t) = \lambda_{n0} p_n(t) \\ p_0(t) + p_1(t) + p_2(t) + p_3(t) + \dots + p_n(t) = 1 \end{cases}$$

Отримаємо:

$$p_0(t) = \left(1 + \lambda_{01}\left(\frac{1}{\lambda_{12}} + \frac{1}{\lambda_{23}} + \frac{1}{\lambda_{34}} + \dots + \frac{1}{\lambda_{n0}}\right)\right)^{-1}$$

$$p_1(t) = \frac{\lambda_{01}}{\lambda_{12}} p_0(t), p_2(t) = \frac{\lambda_{01}}{\lambda_{23}} p_0(t), p_3(t) = \frac{\lambda_{01}}{\lambda_{34}} p_0(t) \dots,$$

$$p_n(t) = \frac{\lambda_{01}}{\lambda_{n0}} p_0(t)$$

Приклад 8.3. Циклічний процес представлено графом (рис. 8.7). Знайти граничні ймовірності станів.

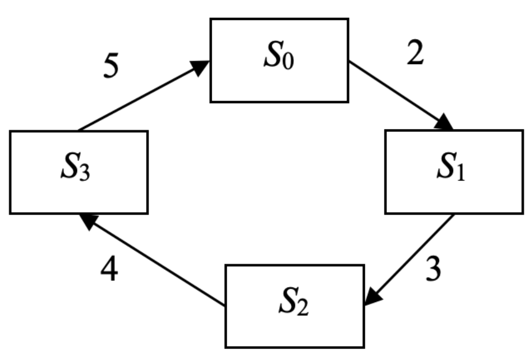


Рис. 8.7

Розв’язання.

$$p_0(t) = \left(1 + \lambda_{01}\left(\frac{1}{\lambda_{12}} + \frac{1}{\lambda_{23}} + \frac{1}{\lambda_{30}}\right)\right)^{-1} = \left(1 + 2\left(\frac{1}{3} + \frac{1}{4} + \frac{1}{5}\right)\right)^{-1} = \frac{30}{77}$$

$$p_1(t) = \frac{2}{3}p_0(t) = \frac{20}{77}, p_2(t) = \frac{1}{2}p_0(t) = \frac{15}{77}, p_3(t) = \frac{2}{5}p_0(t) = \frac{12}{77}.$$

8.3. Матриця переходів

Припустимо, що проводиться серія експериментів з можливими результатами $X_1, X_2, \dots, X_n$, які називаються *станами*.

$p_i^{(0)}$ – ймовірність того, що ми починаємо в стані X_i.

p_{ij}- ймовірність того, що в результаті експеримента система перейшла зі стану X_i в стан X_j.

Якщо $p_i^{(1)}$ – ймовірність того, що результатом експерименту буде стан X_i, то

$$p_i^{(1)} = p_1^{(0)}p_{1i} + p_2^{(0)}p_{2i} + p_3^{(0)}p_{3i} + \dots + p_n^{(0)}p_{ni}.$$

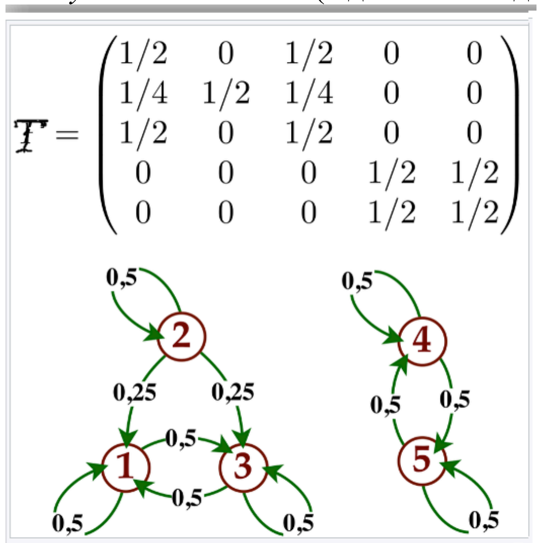
Очевидно, що

$$p_{j1} + p_{j2} + p_{j3} + \dots + p_{jn} = 1.$$

Матриця вигляду

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1n} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2n} \\ p_{31} & p_{32} & p_{33} & \cdots & p_{3n} \\ p_{41} & p_{42} & p_{43} & \cdots & p_{4n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & p_{n3} & \cdots & p_{nn} \end{bmatrix}.$$

називається *матрицею переходу*. Позначимо її T (в деяких випадках вона позначається P).



Матриця ймовірностей переходу і граф переходів однорідного ланцюга Маркова з п'ятьма станами

Рис. 8.8. Приклад матриці переходу

Нехай

$$p^{(0)} = (p_1^{(0)}, p_2^{(0)}, p_3^{(0)}, \dots, p_n^{(0)})$$

та

$$p^{(1)} = (p_1^{(1)}, p_2^{(1)}, p_3^{(1)}, \dots, p_n^{(1)}),$$

тоді

$$(p_1^{(1)}, p_2^{(1)}, p_3^{(1)} \dots, p_n^{(1)}) = (p_1^{(0)}, p_2^{(0)}, p_3^{(0)} \dots, p_n^{(0)}) \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1n} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2n} \\ p_{31} & p_{32} & p_{33} & \cdots & p_{3n} \\ p_{41} & p_{42} & p_{43} & \cdots & p_{4n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & p_{n3} & \cdots & p_{nn} \end{bmatrix}. \quad (8.1)$$

Приклад 8.4. Прогноз погоди.

Погода класифікується як ясна, помірно похмура та похмура.

1. Якщо погода ясна, то з ймовірністю 0,5 вона буде ясною, з ймовірністю 0,4 – помірно похмурою, з ймовірністю 0,1 – похмурою.
2. Якщо погода помірно похмура, то з ймовірністю 0,3 вона буде ясною, з ймовірністю 0,5 – помірно похмурою, з ймовірністю 0,2 – похмурою.
3. Якщо погода похмура то з ймовірністю 0,2 вона буде ясною, з ймовірністю 0,4 – помірно похмурою, з ймовірністю 0,4 – похмурою.

Питання 1. Якщо ймовірність ясної погоди в неділю дорівнює 0,6, а ймовірність помірно похмурої – 0,4, то яка ймовірність, що погода в понеділок буде ясною?
Питання 2. Яка ймовірність що у вівторок погода буде помірно похмурою.

Розв'язання.

Згідно умови

$$p^{(0)} = (0.6, 0.4, 0),$$

$$T = \begin{bmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.4 & 0.4 \end{bmatrix}.$$

Відповідно

$$p^{(1)} = (0.6, 0.4, 0) \times \begin{bmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.4 & 0.4 \end{bmatrix} = (0.42, 0.44, 0.14)$$

та ймовірність, що в понеділок погода буде ясною дорівнює 0,42.

Нехай $p_1^{(2)}$ – ймовірність того, що погода у вівторок буде ясною, $p_2^{(2)}$ – ймовірність того, що погода у вівторок буде помірно похмурою, $p_3^{(2)}$ – ймовірність того, що погода у вівторок буде похмурою.

Нехай $p^{(2)} = (p_1^{(2)}, p_2^{(2)}, p_3^{(2)})$.

$$p^{(2)} = (0.42, 0.44, 0.14) \times \begin{bmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.4 & 0.4 \end{bmatrix} = (0.37, 0.444, 0.186).$$

Тобто ймовірність, що у вівторок погода буде помірно похмурою дорівнює 0,444.

Нехай $p_i^{(m)}$ – ймовірність того, що результатом m-го експерименту буде стан X_i та

$$p^{(m)} = (p_1^{(m)}, p_2^{(m)}, p_3^{(m)}, \dots, p_n^{(m)}).$$

Теорема: Для довільного цілого числа m справедливо

$$p^{(m)} = p^{(0)} \times T^{(m)}. \tag{8.2}$$

Ланцюги Маркова також застосовуються в оцінці майбутніх продажів. Наприклад, зробивши опитування серед покупців тієї чи іншої марки автомобіля про їхній наступний вибір, можна скласти матрицю T.

Приклад 8.5. Оцінка майбутніх продаж

У процесі опитування власників автомобілів трьох американських марок: марки А, марки В, марки С, їм було поставлено питання про те, яку торгову марку вони вибрали б для наступної покупки.

1. Серед власників автомобілів марки А 20% сказали що виберуть знову цю ж марку, 50% сказали, що вони перейшли б на марку В, а 30% заявили, що віддали б перевагу марці С.

2. Серед власників автомобілів марки В 20% сказали, що перейдуть на марку А, натомість 70% заявили, що придбали б знову автомобіль марки В, а 10% заявили, що наступного разу віддали б перевагу марці С.

3. Серед власників автомобілів 30% відповіли, що перейшли б на марку А, 30% сказали, що перейшли б на марку В, а 40% заявили, що залишилися б вірними тій самій марці С.

Питання 1: Якщо хтось придбав автомобіль марки, то яка ймовірність, що його другою машиною буде автомобіль марки С?

Питання 2: Якщо при покупці першої машини покупець підкинув монету, вибираючи між автомобілями марки і , то яка ймовірність, що його третьою машиною стане автомобіль марки В?

Розв'язання. Матриця переходу має вигляд:

$$\begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0.2 & 0.7 & 0.1 \\ 0.3 & 0.3 & 0.4 \end{bmatrix}.$$

Для відповіді на перше питання матимемо $p^{(0)} = (1, 0, 0)$, тому:

$$p^{(1)} = (1, 0, 0) \times \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0.2 & 0.7 & 0.1 \\ 0.3 & 0.3 & 0.4 \end{bmatrix} = (0.2, 0.5, 0.3).$$

Ймовірність, що друга машина буде марки С дорівнює 0.3.

Для відповіді на друге питання потрібно знайти

$$T^{(2)} = \begin{bmatrix} 0.23 & 0.54 & 0.23 \\ 0.21 & 0.62 & 0.17 \\ 0.24 & 0.48 & 0.28 \end{bmatrix}.$$

Для другого питання $p^{(0)} = (0, 0.5, 0.5)$.

$$p^{(2)} = (0, 0.5, 0.5) \times \begin{bmatrix} 0.23 & 0.54 & 0.23 \\ 0.21 & 0.62 & 0.17 \\ 0.24 & 0.48 & 0.28 \end{bmatrix} = (0.225, 0.55, 0.225)$$

тому ймовірність, що другий автомобіль буде марки А, дорівнює 0.225.

8.4. Система масового обслуговування

Системою масового обслуговування (СМО) називається кожна система, яка застосовується при обслуговуванні будь-якого потоку заяв. Наприклад, ремонтні майстерні, магазини, телефоні станції, білетні каси, тощо.

СМО діляться на *системи з відмовами і системи з чеканням*. В системі з відмовами заява, яка поступає до системи в момент, коли всі канали обслуговування зайняти, отримує відмову і залишає систему. В системі з чеканням така заява не залишає систему, а стає в чергу, поки не звільнено буде який-небудь канал. Час чекання може бути як необмеженим, так і обмеженим.

Система масового обслуговування називається пуассонівською, якщо всі потоки подій, які переводять її зі стану до стану, є пуассонівськими.

Докладніше висвітлено в Павлов О.А., Гавриленко О.В. «Теорія ймовірностей, імовірнісні процеси та математична статистика. Курс лекцій. Частина 3»

ДОМАШНЄ ЗАВДАННЯ 8

Варіанти 1-5

1. Скласти рівняння Колмогорова для графу станів (рис. 8.9) і знайти граничні ймовірності станів системи.

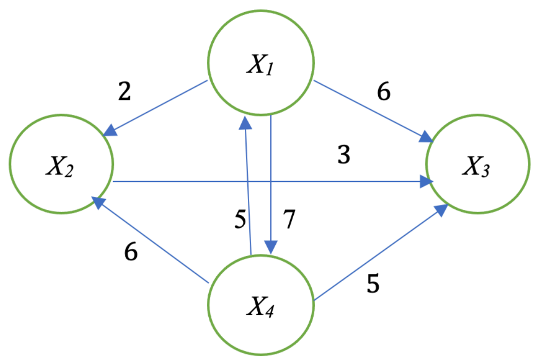


Рис. 8.9

2. Побудувати граф станів такого випадкового процесу: систему утворено з двох автоматів з продажу газованої води, кожний з яких може бути зайнятим або вільним.
3. Ймовірності переходу за один крок у ланцюгу Маркова задано матрицею:

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$

Потрібно: 1) знайти кількість станів; 2) побудувати граф, що відповідає матриці *P*.

4. Середнє число заяв на таксі, що поступають на диспетчерський пункт за 1 хвилину, дорівнює 4. Знайти ймовірність, що за 2 хвилини поступить 5 викликів.

5. В їдальні самообслуговування обід видають чотири повара. Кількість місць достатня для розміщення осіб, отримавши обід. Середній час обслуговування (видача обіду) складає 4 хвилини, щільність потоку біля двох людей на хвилину, а черга не може перевищувати 20 відвідувачів. У середньому відвідувач стоїть в черзі 10 хвилин, після чого уходить з їдальні. Визначити ймовірність , що відвідувач пообідає, скільки часу в середньому йому буде на це потрібно, якщо на обід він у середньому витрачає біля 10 хвилин.

6. Дана матриця Р перехідних ймовірностей ланцюга Маркова зі стану i ($i=1,2$) в стан j ($j=1,2$) за один крок. Знайти матрицю Р(2) зі стану i в стан j за два кроки.

$$P = \begin{pmatrix} 0{,}1 & 0{,}9 \\ 0{,}2 & 0{,}8 \end{pmatrix}$$

Варіанти 6-10

1. Середнє число заяв на таксі, що поступають на диспетчерський пункт за 1 хвилину, дорівнює 4. Знайти ймовірність, що за 2 хвилини поступить принаймні один.

2. Потік машин, які рухаються в одному напрямку, є найпростіший потік з щільністю λ. Людина виходить на шосе, щоб зупинити першу машину, яка попадеться і буде рухається даному напрямку. Знайти закон розподілу часу T, яке йому прийдеться чекати, Визначити математичне сподівання $M(t)$ і середнє квадратичне відхилення $\sigma(t)$.

3. Знайти граничні ймовірності для «гибелі та розмноження» (рис. 8.10)

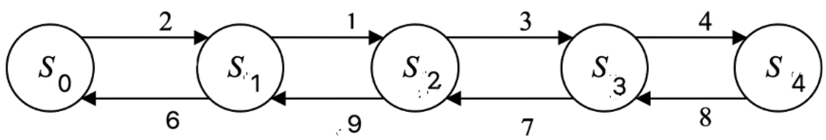


Рис.8.10

4. Погода на деякому острові через певні проміжки часу стає чи дощовою (Д), чи сонячною (С). Імовірності щоденних змін задано матрицею:

$$P = \begin{array}{c|cc} & Д & С \\ \hline Д & 0{,}7 & 0{,}3 \\ С & 0{,}4 & 0{,}6 \end{array}.$$

Якщо в середу погода дощова, то яка ймовірність того, що вона буде дощовою в найближчу п'ятницю?

5. Побудувати граф станів системи: електричний прилад, якій може бути включеним, або виключеним, або зламаним.

6. Дана матриця Р перехідних ймовірностей ланцюга Маркова зі стану i ($i=1,2$) в стан j ($j=1,2$) за один крок. Знайти матрицю Р(2) зі стану i в стан j за два кроки.

$$P = \begin{pmatrix} 0{,}3 & 0{,}7 \\ 0{,}4 & 0{,}6 \end{pmatrix}$$

Варіанти 11-15

1. Побудувати граф станів системи: електричний прилад, якій може бути включеним, або виключеним, або зламаним, або знищеним.

2. Погода на деякому острові через певні проміжки часу стає чи дощовою (Д), чи похмурою (П), чи сонячною (С). Ймовірності щоденних змін задано матрицею:

	Д	П	С
Д	0,2	0,4	0,4
П	0,1	0,4	0,5
С	0,7	0,2	0,1

Якщо в середу очікується дощова погода з ймовірністю 0,3, а похмура з ймовірністю 0,4, то яка ймовірність того, що вона буде дощовою в найближчу п'ятницю?

3. Знайти граничні ймовірності для циклічного процесу (рис. 8.11)

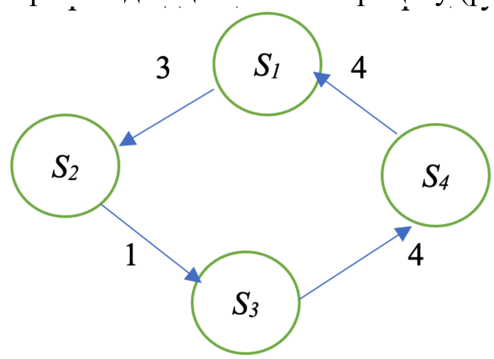


Рис. 8.11

4. Середнє число заяв на таксі, що поступають на диспетчерський пункт за 1 хвилину, дорівнює 4. Знайти ймовірність, що за 2 хвилини не поступить жодного виклика.

5. Одна плавуча база обслуговує 10 траулерів. Середній час плавання траулера дорівнює 3 доби. На базі є один причал, середній час обслуговування траулера – 8 один. Визначити середню довжину черзі, середній час простою траулера, середній час чекання заявки в черзі.

6. Дана матриця P перехідних ймовірностей ланцюга Маркова зі стану i ($i=1,2$) в стан j ($j=1,2$) за один крок. Знайти матрицю P(2) зі стану i в стан j за два кроки.

$$P = \begin{pmatrix} 0,2 & 0,8 \\ 0,3 & 0,7 \end{pmatrix}$$

Варіанти 16-20

1. Погода на деякому острові через певні проміжки часу стає чи дощовою (Д), чи похмурою (П), чи сонячною (С). Ймовірності щоденних змін задано матрицею:

	Д	П	С
Д	0,3	0,4	0,3
П	0,2	0,3	0,5
С	0,6	0,2	0,2

Якщо в середу очікується дощова погода з ймовірністю 0,5, а сонячна з ймовірністю 0,3, то яка ймовірність того, що вона буде дощовою в найближчу п'ятницю?

2. На автозаправочній станції (АТС) працюють 4 заправочних колонки (n=4). Заправка однієї машини триває в середньому 3 хвилини. В середньому на АЗС кожну хвилини під'їжає машина для заправки паливом. Число місць в черзі необмежено. Всі машини стають в чергу. Визначити середній час з моменту прибуття до моменту заправки, середнє число зайнятих колонок, середнє число машин в черзі, середній час простою колонки між заправками.

3. Знайти граничні ймовірності для системи, графи яких зображено на рис.8.12

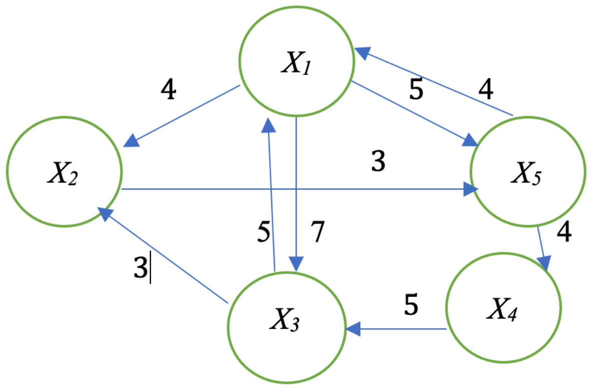


Рис.8.12

4. Середня кількість літаків, що прибувають до літовища за хвилину дорівнює 4. Знайти ймовірність того, що за 5 хвилин прибуде 6 літаків. При цьому вважати потік найпростішим.

5. Матриця ймовірностей переходу ланцюга Маркова має вид:

$$P = \begin{pmatrix} 0{,}1 & 0{,}5 & 0{,}4 \\ 0{,}6 & 0{,}2 & 0{,}2 \\ 0{,}3 & 0{,}4 & 0{,}3 \end{pmatrix}$$

Розподіл по станам в момент $t = 0$ визначається вектором (0,7; 0,2; 0,1). Знайти розподіл по станам в момент $t = 2$.

6. Дана матриця Р перехідних ймовірностей ланцюга Маркова зі стану *i* (*i=1,2*) в стан *j* (*j=1,2)* за один крок. Знайти матрицю Р(2) зі стану *i* в стан *j* за два кроки.

$$P = \begin{pmatrix} 0{,}4 & 0{,}6 \\ 0{,}5 & 0{,}5 \end{pmatrix}$$

Варіанти 21-25

1. Погода на деякому острові через певні проміжки часу стає чи дощовою (Д), чи похмурою (П), чи сонячною (С). Ймовірності щоденних змін задано матрицею:

	Д	П	С
Д	0,4	0,3	0,3
П	0,5	0,3	0,2
С	0,5	0,2	0,3

Якщо в середу очікується дощова погода з ймовірністю 0,5, а сонячна з ймовірністю 0,2, то яка ймовірність того, що вона буде дощовою в найближчу п'ятницю?

2. Матриця ймовірностей переходу ланцюга Маркова має вид:

$$P = \begin{pmatrix} 0{,}1 & 0{,}5 & 0{,}4 \\ 0{,}6 & 0{,}2 & 0{,}2 \\ 0{,}3 & 0{,}4 & 0{,}3 \end{pmatrix}$$

Розподіл по станам в момент $t = 0$ визначається вектором (0,6; 0,2; 0,2). Знайти розподіл по станам в момент $t = 2$.

3. Профілактичний огляд автомашин відбувається на пункті з однією групою проведення огляду (один канал), що працює цілодобово. На огляд і встановлення дефектів на кожну машину витрачається в середньому 0,5 години. На огляд прибуває в середньому 15 машин на добу. Потік заяв і обслуговування є найпростіший. Якщо канал зайнято. Машина від'їжає не обслугованою. Визначити ймовірність станів і характеристики обслуговування пункту огляду.

4. Ймовірності переходу за один крок у ланцюгу Маркова задано матрицею:

$$P = \begin{pmatrix} 0{,}8 & 0{,}1 & 0{,}1 \\ 0{,}5 & 0{,}3 & 0{,}2 \\ 0{,}6 & 0{,}2. & 0{,}2 \end{pmatrix}$$

Потрібно: 1) знайти кількість станів; 2) побудувати граф, що відповідає матриці *P*.

5. Знайти граничні ймовірності для циклічного процесу (рис. 8.13)

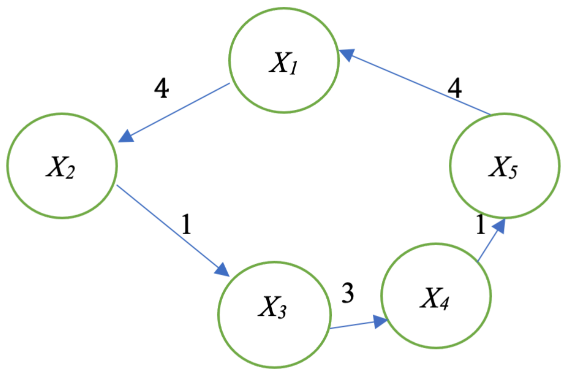


Рис. 8.13

6. Дана матриця Р перехідних ймовірностей ланцюга Маркова зі стану *i* (*i=1,2*) в стан *j* (*j=1,2)* за один крок. Знайти матрицю Р(2) зі стану *i* в стан *j* за два кроки.

$$P = \begin{pmatrix} 0{,}6 & 0{,}4 \\ 0{,}7 & 0{,}3 \end{pmatrix}$$

Варіанти 25-30

1. Скласти рівняння Колмогорова для графу станів (рис. 8.14) і знайти граничні ймовірності станів системи.

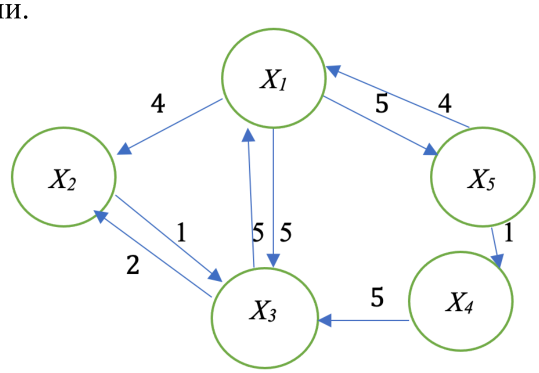


Рис. 8.14

2. Побудувати граф станів такого випадкового процесу: систему утворено з трьох автоматів з продажу газованої води, кожний з яких може бути зайнятим або вільним.

3. Ймовірності переходу за один крок у ланцюгу Маркова задано матрицею:

$$P = \begin{pmatrix} 0{,}1 & 0{,}5 & 0{,}4 \\ 0{,}6 & 0{,}2 & 0{,}2 \\ 0{,}3 & 0{,}4 & 0{,}3 \end{pmatrix}$$

Потрібно: 1) знайти кількість станів; 2) побудувати граф, що відповідає матриці *P*.

4. Середнє число заяв на таксі, що поступають на диспетчерський пункт за 1 хвилину, дорівнює 4. Знайти ймовірність, що за 2 хвилини поступить 6 викликів.

5. Потік машин, що рухаються в даному напряму є регулярним з щільністю λ. Людина виходить на шосе, щоб зупинити першу машину, яка попадеться і буде рухається даному напрямку. Знайти закон розподілу часу T, яке йому прийдеться чекати. Визначити математичне сподівання $M(t)$ і середнє квадратичне відхилення $\sigma(t)$.

6. Дана матриця P перехідних ймовірностей ланцюга Маркова зі стану *i* (*i=1,2*) в стан *j* (*j=1,2)* за один крок. Знайти матрицю P(2) зі стану *i* в стан *j* за два кроки.

$$P = \begin{pmatrix} 0{,}2 & 0{,}8 \\ 0{,}4 & 0{,}6 \end{pmatrix}$$

ПРАКТИЧНЕ ЗАНЯТТЯ 9

Перевірочна контрольна робота 2

1. Задано вибірку X_1:

59	55	60	46	57	54	57	54	55	46
57	59	57	46	54	55	57	59	54	54
46	55	57	60	55	54	46	55	57	54
59	47	46	55	60	47	54	57	59	60

та вибірку X_2:

55	55	60	46	54	57	59	60	57	46
55	46	60	46	54	57	60	60	57	59
47	47	55	46	54	60	55	54	57	59
47	57	59	46	55	54	59	59	54	60

1) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність дисперсій нормальних генеральний сукупностей;
2) При рівні значущості $\alpha = 0{,}05$ для вибірки X_2 перевірити гіпотезу про рівність математичного сподівання нормальної генеральної сукупності гіпотетичному значенню: ;
3) При рівні значущості $\alpha = 0{,}05$ для вибірки X_1 перевірити гіпотезу про рівність виправленої вибіркової дисперсії гіпотетичному значенню $\sigma^2 = \sigma_0^2 = 0{,}15$;
4) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність математичних сподівань нормальних генеральних сукупностей;
5) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність нулю коефіцієнту кореляції.

2. Задано вибірку X_1:

6,80	6,50	6,53	7,30	6,95	6,90	6,93	6,72	6,54	6,79
6,80	6,75	6,62	6,65	7,00	6,83	6,70	6,69	6,54	6,97
6,82	6,67	6,91	6,93	6,76	6,71	6,66	6,76	6,79	6,82
6,71	6,91	6,90	7,00	6,53	6,72	6,66	6,80	6,82	6,81

1) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про нормальність розподілу генеральної сукупності;
2) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про біноміальність розподілу генеральної сукупності;
3) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про експоненційність розподілу генеральної сукупності;
4) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність математичних сподівань нормальних генеральних сукупностей за допомогою критерію Колмогорова;
5) При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про нормальність розподілу генеральної сукупності за допомогою методу спрямлюючих діаграм.

3. Погода на деякому острові через певні проміжки часу стає чи дощовою (Д), чи похмурою (П), чи сонячною (С). Ймовірності щоденних змін задано матрицею:

	Д	П	С
Д	0,2	0,4	0,4
П	0,1	0,4	0,5
С	0,7	0,2	0,1

Якщо в середу очікується дощова погода з ймовірністю 0,3, а похмура з ймовірністю 0,4, то яка ймовірність того, що вона буде дощовою в найближчу п’ятницю?

4. Скласти рівняння Колмогорова для графу станів (рис. 9.14) і знайти граничні ймовірності станів системи.

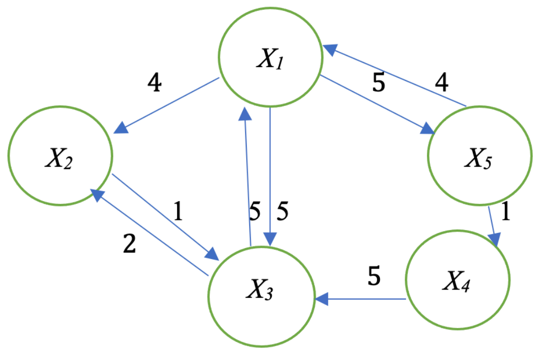


Рис. 9.1

5. Дана матриця P перехідних ймовірностей ланцюга Маркова зі стану *i* (*i*=1,2) в стан *j* (*j*=1,2) за один крок. Знайти матрицю P(2) зі стану *i* в стан *j* за два кроки.

6. Ймовірності переходу за один крок у ланцюгу Маркова задано матрицею:

$$P = \begin{pmatrix} 0{,}1 & 0{,}5 & 0{,}4 \\ 0{,}6 & 0{,}2 & 0{,}2 \\ 0{,}3 & 0{,}4 & 0{,}3 \end{pmatrix}$$

Потрібно: 1) знайти кількість станів; 2) побудувати граф, що відповідає матриці *P*.

РОЗРАХУНКОВО-ГРАФІЧНА РОБОТА

Завдання 1

- За вибіркою з нормальної генеральної сукупності критерієм χ^2 перевірити гіпотезу про нормальний розподіл;
- Побудувати довірчі інтервали для: математичного сподівання при відомій дисперсії; математичного сподівання при невідомій дисперсії; для дисперсії;
- Задати структуру та параметри (кожному індивідуально) ідеальної багатовимірної лінійної регресії, реалізувати віртуальний активний експеримент і за його результатами методом найменших квадратів знайти оцінки її коефіцієнтів.

Література:

1. Денисюк В.П., Бобков В.Н., Репета В.К. Вища математика, частина 4, Теорія ймовірностей та математична статистика. К. НАУ – друк, 2009, с. 200-214;
2. Конспект лекцій.

Завдання 2

- Задати (кожному індивідуально) для однорідного ланцюга Маркова матрицю переходів π_1 (містить нулі), що гарантує існування граничних ймовірностей та знайти їх значення;
- Задати (кожному індивідуально)
  матрицю Λ $(\forall\lambda_{ij}, i \neq j,$ існують $\lambda_{ij} = 0, i \neq j)$
  умовних інтенсивностей однорідного регулярного марківського процесу (ОРМП), що гарантує існування граничних безумовних ймовірностей станів ОРМП та побудувати граф переходів для ОРМП; записати систему диференціальних рівнянь Колмогорова та знайти граничні безумовні ймовірності станів ОРМП.

Література:

Конспект лекцій.

Завдання 3

- Задати (кожному індивідуально) параметри марківської однокольної СМО (λ, μ, m) з обмеженою чергою та по графу переходів знайти граничні ймовірності її станів, характеристики її роботи в стаціонарному режимі;
- Розв'язати аналогічну задачу для марківської багатоканальної СМО з обмеженою чергою (задавши кількість її каналів обслуговування).

Література:

Конспект лекцій.

Завдання 4

- Задати (кожному індивідуально) параметри $(p, \Phi_1, \dots, \Phi_p)$ стаціонарної моделі авторегресії порядку p та знайти значення коефіцієнтів кореляції $\rho_1, \dots, \rho_p$;

- Задати (кожному індивідуально) параметри моделі ковзуного середнього порядку q $(q, \theta_1, \dots, \theta_p)$, що гарантують розв’язання проблеми оберненої моделі та знайти значення коефіцієнтів кореляції $\rho_1, \dots, \rho_q$.

Література:
Конспект лекцій.

ПИТАННЯ ДЛЯ САМОКОНТРОЛЮ

1. Розподіл χ^2.
2. Розподіл Стьюдента.
3. Розподіл Фішера-Снедекора.
4. Розподіл Кочрена
5. Варіаційний ряд. Основні поняття, види, побудова.
6. Геометричні форми представлення виборок (полігон, гісторгама кумулята).
7. Емпірична функція розподілу.
8. Точкові оцінки параметрів виборок. Основні поняття (генеральне –вибіркове середнє, генеральна-вибіркова дисперсії, зміщені-незміщені оцінки, виправлені-невиправлені оцінк).
9. Метод найбільшої правдоподібності.
10. Метод моментів.
11. Метод найменших квадратів.
12. Побудова інтервальних оцінок для нормального розподілу.
13. Статистична перевірка гіпотез. Основні понятя.
14. Критерій Пірсона.
15. Критерій Колмогорова.
16. Критерій знаків.
17. Дисперсійний аналіз.
18. Задача регресії.
19. Марківськи процеси. Основні поняття. (Поняття Марківського процесу, задання за допомогою графу, система рівнянь Колмогорова, схеми “народження та загибелі”, “циклічна”).
20 Критерій Бартлета.
21. Критерій Кочрена.
22. Критерій Вілкоксона.
23 Критерії порівняння вибіркових дисперсій або вибіркових середніх.

ЛІТЕРАТУРА

1. Барковський В.В. та ін. Математика для економістів. Теорій ймовірностей та математична статистика. – К.:Національна академія управління, 1999.

2. Гіхман І. І., Скороход А. В., Ядренко М. І. Теория ймовірностей та математична статистика, Київ, Вища школа, 1979.

3.Дороговцев А. Я, Сільвестров Д. С., Скорохол А. В., Ядренко М. Й. Теорія ймовірностей (збірник задач), Київ, Вища школа, 1977.

4. Денисюк В.П., Бобков В.М., Погребецька Т.А., Репета В.К. Вища математика. Ч4. Теорій ймовірностей і математична статистика: К: вид-во «НАУ-друк», 2009. – 256 с.

5. Жлуктенко В.І., Наконечний С.І. Теорія ймовірностей та математична статистика Ч.1. Теорія ймовірностей. – К.:КНЕУ, 2000.

6. Жлуктенко В.І., Наконечний С.І. Теорія ймовірностей та математична статистика Ч.2. Математична статистика – К.:КНЕУ, 2000.

7. Карташов М. В. Імовірність, процеси, статистика: Посібник. – К: Видавничо-поліграфічний центр "Київський університет", 2008.

8. Коваленко І. М., Гнеденко Б. В. Теорія ймовірностей, Київ, Вища школа, 1990.

9. Листопад В.В., Островська О.В. Практикум з теорії ймовірностей із застосуванням інформаційно-комунікаційних технологій [Електронний ресурс]:навчальний посібник – К.: НУХТ, 2016. – 103 с.

10. Павлов О.А., Гавриленко О.В., Рибачук Л.В Навчальний посібник з дисципліни «Теорія ймовірностей, імовірнісні процеси та математична статистика». Курс лекцій. Частина 1 для студентів спеціальності 126 «Інформаційні системи та технології» [Електронний ресурс]:навчальний посібник. – Київ: КПІ, 2021. – 154 с.

11. Павлов О.А., Гавриленко О.В., Жданова О.Г. Навчальний посібник з дисципліни «Теорія ймовірностей, імовірнісні процеси та математична статистика». Курс лекцій. Частина 2 для студентів спеціальності 126 «Інформаційні системи та технології» [Електронний ресурс]:навчальний посібник. – Київ: КПІ, 2022. – 72 с.

12. Павлов О.А., Гавриленко О.В. Навчальний посібник з дисципліни «Теорія ймовірностей, імовірнісні процеси та математична статистика». Курс лекцій. Частина 3 для студентів спеціальності 126 «Інформаційні системи та технології» [Електронний ресурс]:навчальний посібник. – Київ: КПІ, 2022. – 111 с.

13. Гавриленко О.В. Навчальний посібник з дисципліни «Теорія ймовірностей, імовірнісні процеси та математична статистика». Практикум. Частина 1 для студентів спеціальності 126 «Інформаційні системи та технології» [Електронний ресурс]:навчальний посібник. – Київ: КПІ, 2021. – 140 с.

14. Теорія ймовірностей. Збірник задач. Під ред. Скорохода А.В. – К.: Вища школа, 1975.

22. Феллер В. Введение в теорию вероятностей и ее приложения, Т. 1, М., Мир, 1984.

15. Шефтель З.Г. Теорія ймовірностей. –К.:Вища школа, 1994.

16. Галицька І.Є., Жданова О.Г., Кузнєцов В.М. Методичні вказівки до виконання індивідуальних завдань з дисципліни «Ймовірнісні процеси і математична статистика в автоматизованих системах». – К.: Політехніка, 2002. – 44 с.

ІНФОРМАЦІЙНІ РЕСУРСИ

1. Moodle https://do.ipo.kpi.ua/course/view.php?id=6403
2. «Електронний Кампус КПІ» https://ecampus.kpi.ua
3. Google classroom: https://classroom.google.com/u/0/c/MjYyNzg5MjgyNDg4
4. https://studfile.net/preview/7140458/page:20/