Академія праці і соціальних відносин
Федерації профспілок України

Кафедра інформаційних технологій та математичних методів

ЛАБОРАТОРНИЙ ПРАКТИКУМ

«Теорія ймовірностей та математична статистика»

для студентів факультету соціального управління

всіх форм навчання

Київ-2005

Рекомендовано до друку вченою радою Академії праці і соціальних відносин ФПУ (протокол № від 2005 р.)

Рецензенти:

Дзюбенко Г.А. – канд. фіз.-мат. наук, ст. наук. співр. міжнародного математичного центру НАН України;
Круглова С.П. – канд. фіз.-мат. наук, *доцент к-ри інформаційних технологій та математичних методів АПСВ*

Теорія ймовірностей та математична статистика / Укл. С.М.Коваленко, В.В.Листопад. - К., АПСВ, 2005. – 98с.

ЗМІСТ

ВСТУП

Математичні методи є досить ефективним апаратом аналізу соціально-економічних процесів і проблем сучасного суспільства. Оволодіння ними та застосування їх на практиці приводить до суттєвої економії ресурсів і дає досить високу точність і надійність отриманих результатів. Значення та поширення цих методів ще більше зросло у зв”язку з розвитком комп”ютерної техніки та застосуванням розроблених програмних комплексів. У зв”язку з цим вивчення курсу „Теорія ймовірностей та математична статистика” студентами спеціальностей 7.040202 – „Соціальна робота” та 6.040.201 – „Соціологія” є досить актуальним.

Мета курсу „Теорія ймовірностей та математична статистика” – сформувати у студентів знання, вміння і навички, необхідні для майбутньої практичної діяльності спеціаліста в галузі дослідження соціально-економічних проблем суспільства.

Цей практикум спрямований на засвоєння основних тем курсу “Теорія ймовірностей та математична статистика” й підготовки до вивчення дисципліни „Математичні методи дослідження соціальних процесів”.

При проведенні лабораторних занять пропонується використання комп”ютерної техніки, програмного забезпечення загального користування (EXEL) та спеціалізоване (STATISTICA 5.0, STATGRAFICS, SAS, SPSS, OSA та ін.).

Табличні значення *($\varphi(x)$, $\Phi(x)$, $t(\gamma,n)$, ...)* можна знаходити з будь-якого підручника з теорії ймовірностей, або користуватися необхідними функціями з Microsoft Excel.

Після виконання лабораторних робіт студенти пишуть контрольну роботу.

ЛАБОРАТОРНА РОБОТА 1

Тема: *Елементи комбінаторики*

I. Деякі теоретичні відомості

Основні терміни: *множина, підмножина, упорядкована та неупорядкована множина, сполука, розміщення, розміщення з повтореннями, перестановка, комбінація (сполучення), факторіал.*

Основні означення, твердження і формули

Множина – одне з основних неозначуваних понять математики, яке можна описати як сукупність однотипних об'єктів. Позначають множини, як правило, великими латинськими літерами *(A, B, C, ...)*, а їх елементи малими *(a, b, c, ...)*.

Підмножиною деякої множини A називається така множина B, кожен елемент якої належить також множині A. Позначення: $B \subset A$.

Упорядкованою називається множина, якщо про кожні два її елементи можна говорити, що один із них передує іншому (інакше кажучи, якщо на множині задане відношення порядку).

Сполука – це будь-яка непорожня підмножина даної множини.

Розміщення із n по k – це упорядкована сполука (підмножина), що містить k різних елементів, із заданої множини n елементів.

Розміщеннями з повтореннями із n по k називаються такі розміщення, які можуть містити також однакові елементи.

Перестановки із множини n елементів – це розміщення, які містять по n елементів (інакше: перестановки це упорядковані підмножини, що містять n елементів із заданої множини n елементів).

Комбінація (сполучення) із n по k – це неупорядкована сполука (підмножина), що містить k різних елементів, із заданої множини n елементів.

Факторіалом натурального числа n називається добуток послідовних натуральних чисел від 1 до n включно

$n!=1 \cdot 2 \cdot \ldots \cdot n; \qquad 0!=1.$

Тв. 1. Кількість усіх розміщень із n по k (A_n^k) обчислюється за формулами:

$$A_n^k = n(n-1)(n-2)\cdot\ldots\cdot(n-k+1)$$

$$A_n^k = \frac{n!}{(n-k)!}.$$

Власт. Якщо $n,\ k \in N$ і $k<n$, то $A_n^{k+1} = A_n^k (n-k)$.

Тв. 2. Кількість усіх розміщень з повтореннями із n по k ($\overline{A}_n^k$) рівна: $\overline{A}_n^k = n^k$.

Тв. 3. Кількість усіх перестановок із множини з n елементів (P_n) рівна: $P_n = n!$.

Тв. 4. Кількість усіх комбінацій із n по k (C_n^k) обчислюється за формулами:

$$C_n^k = \frac{n(n-1)(n-2)\cdot\ldots\cdot(n-k+1)}{k!}$$

$$C_n^k = \frac{n!}{(n-k)!\,k!}$$

Власт. Якщо $n,\ k \in N$ і $k<n$, то

$$C_n^k = C_n^{n-k},\ A_n^0 = A_0^0 = 1,\ C_{n+1}^{k+1} = C_n^{k+1} + C_n^k,\ k<n$$

Правило добутку. Якщо об'єкт А можна вибрати k способами і після кожного з цих виборів об'єкт В у свою чергу можна вибрати l способами, то вибір А і В можна здійснити $k \times l$ способами.

II. Тестове завдання

1. *Серед наведених далі множин вибрати впорядковані*

Варіант 1

а) множина точок координатної площини;
б) множина простих чисел;
в) множина старост академічних груп вузу.

Варіант 2

а) множина парних натуральних чисел;
б) множина жителів міста, що знаходиться на квартобліку;
в) множина точок простору R^3.

(по 1 балу)

2. *Чи правильне твердження?*

Варіант 1

а) перестановка є упорядкованою множиною;
б) розміщення є неупорядкованою множиною;
в) два різні розміщення, утворені з однієї множини, відрізняються лише порядком розташування елементів;
г) комбінації {A; B; C} та {B; A; C} є різними.

Варіант 2

а) комбінація є упорядкованою множиною;
б) перестановка не є упорядкованою множиною;
в) дві різні комбінації, утворені з однієї множини, можуть відрізнятися кількістю елементів, що входять до їх складу;
г) розміщення {-1; 2; 4} та {4; 2; -1} є однаковими.

(по 1 балу)

3. *Із множини непарних натуральних чисел, не більших 5, утворити усі*

Варіант 1: розміщення по два елементи;
Варіант 2: комбінації по два елементи.

(по 2 бали)

4. *Підрахувати*

Варіант 1		*Варіант 2*	
а) C_5^2	***(1 бал)***;	а) A_6^3	***(1 бал)***;
б) $A_5^3 \cdot P_2$	***(2 бали);***	б) $C_8^4 : A_4^1$	***(2 бали)***.

5. *Який із виразів більший? У скільки разів?*

а) *x!* чи *(x-3)!*; а) *(x+1)!* чи *(x-1)!*;
б) A_x^3 чи C_x^3. б) C_5^x чи A_5^x.

(по 2 бали)

6. *Чи правильна наступна рівність?*

а) $C_5^2 = C_5^3$; *а)* $A_n^1 = C_n^1$;

б) $P_7 = A_7^7$; *б)* $C_4^3 = A_4^3$;

в) $A_7^4 = 4A_7^3$. *в)* $A_8^4 = C_8^4 P_4$.

(по 2 бали)

7. *Знайти x методом підбору, якщо*

$A_4^x = 12;$ $\qquad\qquad$ $C_5^x = 10.$

(по 3 бали)

III. Приклади розв'язування задач

Задача 1. Скільки тризначних чисел можна скласти із 10 карток, на яких написані цифри від 0 до 9?

Розв'язування. Відзначимо перш за все, що цифри в представленні чисел повторюватись не можуть (тому що є лише 10 карток із різними цифрами), крім того, перша цифра не може бути 0. Таким чином, усі тризначні числа є упорядкованими підмножинами, що складаються з трьох елементів множини цифр {0; 1; 2; 3; 4; 5; 6; 7; 8; 9}, тобто розміщеннями із 10 по 3, з яких вилучено всі ті, які починаються нулем.

Кількість усіх розміщень із 10 по 3 рівна

$$A_{10}^3 = \frac{10!}{(10-3)!} = \frac{10!}{7!} = 8 \cdot 9 \cdot 10 = 720$$

Кількість розміщень із 10 по 3, у яких на першому місці стоїть 0, рівна кількості розміщень із 9 (без цифри 0) по 2.

$$A_9^2 = \frac{9!}{(9-2)!} = \frac{9!}{7!} = 8 \cdot 9 = 72$$

Отже, шукана кількість тризначних чисел рівна

$$A_{10}^3 - A_9^2 = 720 - 72 = 648$$

Задача 2. Студенти одного з курсів вивчають 8 навчальних дисциплін. Скількома способами можна скласти розклад занять на понеділок, якщо в цей день слід запланувати три лекції з різних предметів?

Розв'язування. Кількість таких способів дорівнює числу розміщень з 8 предметів по 3: $A_8^3 = 8 \times 7 \times 6 = 336\cdot$

Задача 3. Команда з “Клубу знавців” у складі 6 осіб займає місця за круглим столом. Скільки існує можливих варіантів розміщення гравців? Скільки таких варіантів у випадку, коли два гравці команди повинні сісти поруч?

Розв'язування. У першому випадку кількість варіантів розміщення гравців дорівнює числу перестановок з 6 елементів, тобто

$P_6 = 6! = 720$. У другому випадку для двох виділених осіб є 6 різних сусідніх пар місць, на кожному з яких ці дві особи можуть сісти двома способами. Отже, посадити їх поруч можна 12 способами. На місця, що залишились, решту можна розсадити $P_4 = 4!$ способами. За правилом добутку дістаємо кількість усіх варіантів розміщень: $12\times4!=288$.

Задача 4. Скільки потрібно купити карток лотереї "Спортлото 6 із 36", щоб напевне вгадати усі 6 номерів майбутнього тиражу? Скільки грошей на це треба витратити, якщо одна картка коштує 0,5 грн.?

Розв'язування. Варіанти заповнень карток лотереї є неупорядкованими (порядок слідування вгаданих номерів ролі не грає) підмножинами по 6 елементів із множини 36 елементів, тобто є комбінаціями із 36 по 6. Щоб напевне вгадати усі 6 номерів лотереї, потрібно перебрати усі можливі комбінації із 36 по 6 і заповнити відповідні картки. Таких комбінацій, а отже, і карток, буде

$$C_{36}^{6} = \frac{36!}{(36-6)!6!} = \frac{36!}{30!6!} = \frac{31\cdot32\cdot33\cdot34\cdot35\cdot36}{1\cdot2\cdot3\cdot4\cdot5\cdot6} = 1947792$$

Для придбання такої кількості карток потрібно $1947792 \times 0,5 = 973896$ грн.

Задача 5. Розв'язати рівняння $A_7^x = x\cdot A_7^{x-1}$

Розв'язування. Оскільки $A_7^x = \dfrac{7!}{(7-x)!}$;

$A_7^{x-1} = \dfrac{7!}{(7-(x-1))!} = \dfrac{7!}{(8-x)!}$, то рівняння запишеться у вигляді

$\dfrac{7!}{(7-x)!} = \dfrac{x\cdot7!}{(8-x)!}$, або $7!(8-x)! = x\cdot7!(7-x)!$;

$(8-x)! = x(7-x)!$; $\quad 1\cdot2\cdot\ldots\cdot(7-x)\cdot(8-x) = x\cdot1\cdot2\cdot\ldots\cdot(7-x)$;

$8-x = x$; $\quad 2x = 8$; $\quad x = 4$.

IV. Завдання для самостійного виконання

Завдання 1. Обчислити значення виразу

1.1 а) $\dfrac{A_5^2 - C_5^2}{3!}$; б) $\dfrac{A_n^k (n-k)!}{(n-1)!}$; в) $\dfrac{A_8^4 + C_8^4}{A_8^4 - P_4}$

1.2 а) $\dfrac{A_8^4 + P_6}{C_4^2}$; б) $A_5^3 : P_2$; в) $\dfrac{C_n^k (n-k)!(k-1)!}{(n-1)!}$

1.3 а) $(C_6^4 + C_3^2) \cdot A_3^2$; б) $C_{100}^{98} - P_3$; в) $\dfrac{A_n^m (n-m+1)!}{n!}$

Завдання 2

2.1. Скільки існує трицифрових чисел, що записуються цифрами 0; 2; 4; 6; 8, і в запису кожного з них використовуються різні цифри?

2.2. Скільки існує трицифрових чисел, які діляться на 5?

2.3. Скільки існує чотирицифрових чисел, які записуються різними цифрами і починаються дев'яткою?

Завдання 3

3.1. Скільки різних комбінацій існує для кодового цифрового замка, що складається із трьох барабанів, на кожному з яких нанесені числа від 0 до 8?

3.2. При проведенні турніру з волейболу, де команди зустрічалися між собою один раз, було зіграно всього 66 матчів. Скільки команд взяло участь у турнірі?

3.3. У шаховому турнірі, де учасники зустрічаються між собою один раз, три шахісти вибули через хворобу, зігравши відповідно одну, дві і три партії. Скільки шахістів почали турнір, якщо всього було зіграно 84 партії?

Завдання 4

4.1. У групі 30 студентів. Скількома способами можна обрати старосту і профорга групи за умови, що кожен студент може бути обраний лише одну з цих посад .

4.2. На одній із секцій студентської наукової конференції повинно виступити сім доповідачів. Скількома способами можна розмістити їх у списки ораторів ?

4.3. Команди вищої ліги з футболу провели за сезон у двох колах розиграшу 240 матчів. Скільки команд у вищій лізі ?

4.4. До профкому надійшло 8 заяв від студентів з проханням виділити їм туристські путівки. Скількома способами можна розділити три наявні путівки якщо вони: а) різні; б) однакові.

4.5. Скількома способами можна розділити чотири однакові папки у три письмові ящики столу, якщо кожен ящик може вмістити всі папки?

Завдання 5. Розв’язати рівняння:

5.1. $(n+4)! = 30(n-k)!\,A_{n+2}^{k+2}$

5.2. $A_n^5 = 20A_{n-1}^4$

5.3. $xC_x^2 = C_{x+1}^2 + 7x$

ЛАБОРАТОРНА РОБОТА 2

Тема: *Основні поняття теорії ймовірності. Класичне і статистичне означення ймовірності. Геометрична ймовірність*

I. Деякі теоретичні відомості

Основні терміни: *випробування, випадок, випадкова подія, частота та відносна частота події, ймовірність неможливої та достовірної події, практично неможлива та практично достовірна події, сумісні та несумісні, рівноможливі події, повна група подій, елементарний випадок, складена подія, геометрична ймовірність.*

Основні означення, твердження і формули

Випробування – здійснення цілком певної сукупності умов.

Випадок – результат випробування.

Випадкова подія (A, B, C, ...) – подія, яка при випробуванні може здійснитися чи ні.

Частотою N (абсолютною частотою) події A називається кількість здійснення події A при M випробуваннях.

Відносна частота W(A) події A – це відношення кількості випробувань, за яких подія A здійснилась (M), до загальної кількості випробувань (N)

$$W(A) = \frac{M}{N}$$

Достовірною називається подія (позначається **D**), яка обов'язково відбудеться при кожному випробуванні.

Неможливою називається подія (позначається $\varnothing$), яка ніколи не відбудеться.

Несумісними називаються події (дві або більше), якщо поява однієї з них виключає появу інших. У протилежному випадку події називаються *сумісними.*

Попарно несумісними називаються події, якщо кожні дві з них несумісні.

Рівноможливими називаються події, якщо є підстави вважати, що жодна з них не є більш можливою, ніж інша.

Кілька подій утворюють *повну групу*, якщо у результаті випробування обов'язково здійсниться хоча б одна з них (одна або більше).

Елементарний випадок (або елементарна подія) – це кожний із можливих результатів випробування.

Складними називаються події, які можна розкласти на елементарні події.

Статистичне означення ймовірності

Ймовірністю P(A) події А називається граничне значення відносної частоти W(A) події А при необмеженому збільшенні кількості випробувань N.

$$P(A) = \lim_{N\to\infty} W(A) = \lim_{N\to\infty} \frac{M}{N}$$

Класичне означення ймовірності

Ймовірністю P(A) події А називається відношення кількості m сприятливих цій події випадків до загальної кількості *n* усіх рівноможливих несумісних елементарних випадків, що утворюють повну групу.

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

Тв. 1. Ймовірність достовірної події *D* рівна 1, *(P(D)=1)*.

Тв. 2. Ймовірність неможливої події ∅ рівна 0, *(P(∅)=0)*.

Практично достовірною називається подія, ймовірність якої близька або рівна одиниці.

Практично неможливою називається подія, ймовірність якої близька або рівна нулю.

Тв. 3. Дві події А і В рівноможливі тоді і тільки тоді, коли їх ймовірності рівні *(P(A)=P(B))*.

Геометрична ймовірність – це ймовірність попадання в деяку частину заданої області (відрізок, частину площини, частину простору):

$$P(A) = \frac{mes(g)}{mes(G)},$$

де *mes(g)* – міра частини заданої області *g* (довжина частини відрізка, площа частини площини, об'єм частини простору), а *mes(G)* – міра області *G*.

II. Тестове завдання

1. *Випробування проводилось 754 (825) рази. Подія А відбулася в 16 (24) випадках. Чому дорівнює*
   а) абсолютна частота події А;
   б) відносна частота події А.

(по 1 балу)

2. *Випробування проводилось М разів. Подія А відбулася в N випадках. Чи правильно, що:*

Варіант 1

а) $P(A)=\dfrac{N}{M}$;

б) $\dfrac{N}{M}\to P(A)$ при $M\to\infty$;

в) $P(A)=\lim\limits_{M\to\infty}\dfrac{N}{M}$;

г) $\dfrac{M}{N}\to P(A)$ при $N\to\infty$;

д) $P(A)=\lim\limits_{N\to\infty}\dfrac{N}{M}$.

Варіант 2

а) $P(A)=\lim\limits_{M\to\infty}\dfrac{M}{N}$;

б) $P(A)=\dfrac{M}{N}$;

в) $P(A)\to\dfrac{M}{N}$ при $M\to\infty$;

г) $P(A)\to\dfrac{N}{M}$ при $M\to\infty$;

д) $\dfrac{N}{M}\to P(A)$ при $N\to\infty$.

(по 1 балу)

3. *Чи правильно, що ймовірність Р(А) випадкової події А є число таке, що*

а) $P(A)>0$;

б) $P(A)<10$;

в) $0\le P(A)\le 1$;

г) $P(A)\le 1$;

д) $-1<P(A)\le 1$.

а) $P(A)<1$;

б) $P(A)\ge 0$;

в) $P(A)>-1$;

г) $0\le P(A)<0{,}5$;

д) $-1\le P(A)\le 1$.

(по 1 балу)

4. *Вважатимемо, що N – загальна кількість випробувань, M – кількість випробувань, за яких подія A здійснилася. Чи правильно, що ймовірність достовірної (неможливої) події P(A) рівна:*

а) $P(A) = 0$; б) $P(A) = \dfrac{M}{N}$; в) $P(A) = \lim\limits_{N \to \infty} \dfrac{M}{N}$

г) $P(A) = 1$; ґ) $P(A) = -1$; д) $P(A) = \dfrac{N}{M}$;

е) $P(A) = \lim\limits_{M \to \infty} \dfrac{N}{M}$; є) $P(A) = 0{,}5$.

(по 1 балу)

5. *Підкидається гральний кубик.*

Подія A_i полягає в тому, що випадає число "*i*" ($i = \overline{1,6}$);

B – випадає парне число;

C – випадає непарне число;

E – випадає число не менше 2;

F – випадає число більше 5;

N – випадає число не більше 1.

5.1. *Чи утворюють повну групу події*

а) A_i, $i = \overline{1,6}$;	а) A_1 і E;
б) B і C;	б) C; A_2, A_4, A_6;
в) B і F;	в) C, N, F;
г) E і N.	г) B, C і N.

(по 2 бали)

5.2. *Чи рівноможливі події*

а) A_1 і A_6;	а) A_5 і N;
б) A_2 і B;	б) C і E;
в) B і C;	в) B і E;
г) E і N.	г) N і F.

(по 1 балу)

5.3. *Чому дорівнює ймовірність події (за класичним означенням)*

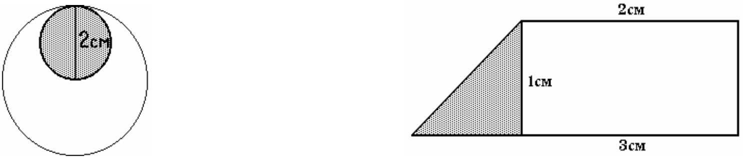
а) A_2	а) A_4;
б) B;	б) C;
в) E.	в) F.

(по 2 бали)

6. *На зображену фігуру навмання ставиться точка. Знайти ймовірність попадання цієї точки у заштриховану частину фігури*

(по 2 бали).

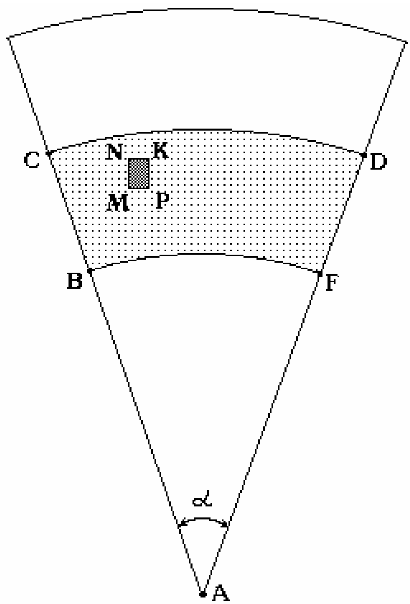
а) 

б) 

III. Приклади розв'язування задач

Задача 1. Відомо, що зона обстрілу артилерійської установки має форму частини сектора з кутом α=15° між колами радіусами 1500 і 1700м (див. мал. 1). Знайти ймовірність поразки цілі, яка знаходиться у зоні обстрілу і має форму прямокутника 2×3м при пострілі одним снарядом.

Розв'язування. Позначимо через Z подію, яка полягає у попаданні в ціль при пострілі одним снарядом. За умовою задачі ∠BAF=15°, AB=AF=1500м, AC=AD=1700м, NK=MP=2м, MN=KP=3м. Використовуючи означення геометричної ймовірності можна записати, що $P(Z)=\dfrac{S_{MNKP}}{S_{BCDF}}$; $S_{MNKP}=2\cdot 3=6$ (м^2); $S_{BCDF}=S_{ACD}-S_{ABF}$.



Мал.1

Як відомо, площа сектора круга радіуса R із центральним кутом α рад обчислюється за формулою

$$S = \frac{\alpha R^2}{2}.$$

Звідси, оскільки $15° = \frac{2\pi}{360} \cdot 15 = \frac{\pi}{12}$ рад, маємо:

$$S_{ACD} = \frac{\pi(1700)^2}{12 \cdot 2} = \frac{289 \cdot 10^4 \pi}{24};$$

$$S_{ABF} = \frac{\pi(1500)^2}{12 \cdot 2} = \frac{225 \cdot 10^4 \pi}{24};$$

Тому

$$S_{BCDF} = \frac{289 \cdot 10^4 \pi}{24} - \frac{225 \cdot 10^4 \pi}{24} = \frac{64 \cdot 10^4 \pi}{24} = \frac{8}{3} 10^4 \pi,$$

а отже,

$$P(Z) = \frac{6}{\frac{8}{3} \cdot 10^4 \pi} = \frac{9}{4 \cdot 10^4 \pi} \approx 7{,}16 \cdot 10^{-5}.$$

Задача 2. Учасники жеребкування беруть із ящика навмання жетони з номерами від 1 до 100. Знайти ймовірність того, що номер першого навмання взятого жетону не містить цифри 5.

Розв'язування. Серед жетонів містять цифру 5 наступні 5, 15, 25, 35, 45, 50, 51, 52, ... 59, 65, 75, 85, 95. Їх 19. А тих, що не містять цифри 5 – 81. Отже, $P = \frac{81}{100} = 0{,}81$.

Задача 3. Серед 17 студентів групи, з яких 8 дівчат, розігрується 7 білетів. Яка ймовірність того, що серед власників білетів виявиться 4 дівчини?

Розв'язування. Число всіх можливих способів розділити 7 білетів між 17 студентами дорівнює числу сполучень із 17 елементів по 7, тобто C_{17}^7. Число способів відібрати чотирьох дівчат із восьми дорівнює C_8^4. Кожна четвірка дівчат може попадати з кожною трійкою хлопців із 9, а число таких трійок дорівнює C_9^3. Отже, число наслідків такого розподілу 7 білетів, що виявиться чотири дівчини, дорівнює $C_8^4 \times C_9^3$, а ймовірність шуканої події: $P = \frac{C_8^4 \times C_9^3}{C_{17}^7} = \frac{735}{2431} \approx 0{,}302$.

Задача 4. При страхуванні життя для розрахунків використовуються таблиці смертності, які дають розподіл по роках смертних випадків для деякої кількості людей одного віку. Наприклад, скорочена таблиця вимірювання 100 тис. людей десятилітнього віку така:

Вік до	10р.	20р.	30р.	40р.	50р.	60р.	70р.	80р.	90р.	100р.
Число доживаючих	100000	96061	89685	82277	72295	58842	37977	13987	1273	4

Користуючись таблицею знайти ймовірність, що студент, якому 20 р., проживе ще 20 років.

Розв'язування. Від 20 до 40 років доживає 96061-82277=13784. Від 10 до 40 років доживає 100000-82277=17723. Тоді шукана ймовірність: $P = \frac{13784}{17723} \approx 0{,}778$.

IV. Завдання для самостійного виконання

Завдання 1

1.1 Підкинути гральний кубик. Знайти ймовірність, що випаде:
   а) парна кількість очок;
   б) число не менше 2;
   в) число не більше 2;
   г) число більше 1 і менше 4.

1.2 Із 60 питань екзаменаційних білетів студент підготував 50. Яка ймовірність того, що витягнутий навмання білет, що має два питання, складатиметься із підготовлених питань?

Завдання 2

2.1 Яка ймовірність того, що навмання поставлена в даний круг точка буде всередині вписаного в нього квадрата?

2.2 Куб, всі грані якого покрашені, розрізано на 1000 кубиків однакового розміру, які потім ретельно змішали. Знайти ймовірність того, що навмання взятий кубик матиме викрашених граней:
   а) одну;
   б) дві;
   в) три.

2.3 На кожній із дев’яти однакових карток надрукована одна із наступних літер: Р, А, Д, Л, О, З, К, М, Т. Карточки ретельно перемішують. Знайти ймовірність того, що на семи витягнутих по одній і розташованих в “одну лінію” картках, можна буде прочитати слово “РОЗКЛАД”.

2.4 Із ретельно перемішаного повного набору 28 кісток доміно навмання вибирається одна. Знайти ймовірність того, що другу навмання вилучену кістку можна приставити до першої, якщо перша:
   а) є дублем;
   б) не буде дублем.

Завдання 3

3.1 Бібліотека складається із 10 різних книг, причому п’ять із них коштують по 4 грн. кожна, три книги – по 1 грн. і дві книги – по

3 грн. Знайти ймовірність того, що взяті навмання дві книги коштують 5 грн.

3.2 Круглий диск діаметрами АВ і СД розбито на 4 сектора. Дуги АС і ВД по довжині дорівнюють радіусу. Сектори АОС і ВОД зеленого кольору – інші синього. Диск повертають дуже швидко. Яка ймовірність того, що попадання буде в сектор зеленого кольору?

3.3 При стрільбі із гвинтівки відносна частота влучення в мішень дорівнює 0,85. Скільки буде влучень, якщо проведено 120 пострілів? Скільки потрібно провести пострілів, щоб влучити 119 разів?

3.4 Багатократно проводилося підкидання монети. Результати занесені в таблицю. Заповнити порожні клітинки.

Число підкидань	Число появ герба	Відносна частота
4040	2048	?
?	6019	0,5016
2400	?	0,5005

Як близько до ймовірності появи герба відносна частота в кожному з випадків?

ЛАБОРАТОРНА РОБОТА 3

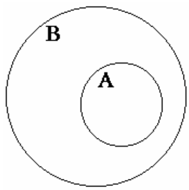
Тема: *Дії над подіями. Ймовірність суми та добутку подій. Формула повної ймовірності та формула Байєса*

I. Деякі теоретичні відомості

Основні терміни*: частинний випадок події, рівносильні події, протилежні події, залежні та незалежні події, сума, добуток та різниця подій, умовна ймовірність.*

Основні означення, твердження і формули

Подія А називається *частинним випадком* події В (А тягне за собою В), якщо при кожному випробуванні, при якому відбувається подія А, відбувається і подія В. Позначення: **А⊂В**



Рівносильні події – це такі події А і В, які при кожному випробуванні одночасно відбуваються чи не відбуваються. Позначення: **А=В**

Тв.1. Події А і В рівносильні тоді і тільки тоді, коли А є частинним випадком В і, водночас, В є частинним випадком А

$$\mathbf{A=B \Leftrightarrow A\subset B \text{ і } B\subset A}$$

Тв.2. Якщо події А і В рівносильні, то вони є рівноможливими

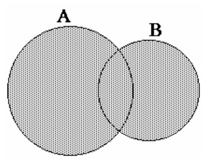
$$\mathbf{A=B \Rightarrow P(A)=P(B)}$$

Обернене твердження неправильне.

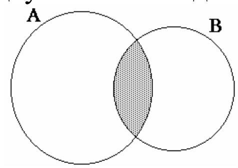
Протилежною до події A називається така подія $\overline{A}$, яка відбувається тільки тоді, коли A не відбувається і, навпаки, не відбувається тільки тоді, коли A відбувається.

Подія А називається ***незалежною (залежною)*** від події В, якщо поява події В не впливає (впливає) на появу події А.

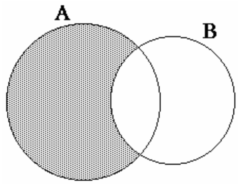
Сумою двох подій А і В називається така подія **С=А+В**, яка відбувається тоді, коли відбувається хоча б одна із подій А чи В (відбувається чи А чи В, чи обидві одночасно).



Добутком двох подій А і В називається така подія **С=А·В**, яка відбувається тоді, коли відбуваються обидві події А і В одночасно.



Різницею двох подій А і В називається така подія **С=А-В**, яка відбувається тоді, коли подія А відбувається і, водночас, подія В не відбувається.



Умовною ймовірністю події А за умови В називається ймовірність Р(А/В) події А, обчисленої за умови, що подія В відбулася.

Властивості дій

1. (Комутативний закон додавання і множення)
   $A + B = B + A; \quad A \cdot B = B \cdot A;$
2. (Асоціативний закон додавання і множення)
3. $A + (B + C) = (A + B) + C; \quad A \cdot (B \cdot C) = (A \cdot B) \cdot C;$
4. (Дистрибутивний закон) $(A + B) \cdot C = A \cdot C + B \cdot C;$
5. $A + A = A; \quad A \cdot A = A;$
6. $A + A \cdot B = A;$ ×× ·
7. $\overline{(\overline{A})} = A;$
8. $A + \overline{A} = D; \quad A \cdot \overline{A} = \varnothing;$
9. $A - B = A \cdot \overline{B}$.

Тв.1. Складну подію *А* можна представити у вигляді суми деяких елементарних подій.

Тв.2. Сума усіх подій $H_1, H_2, \ldots, H_n$, які утворюють повну групу подій, є достовірною подією: $H_1 + H_2 + \ldots + H_n = D$.

Тв.3. Ймовірність суми двох подій А і В обчислюється за формулою $P(A+B)=P(A)+P(B)-P(A\bullet B)$.

Якщо ж події А і В несумісні, то $P(A+B)=P(A)+P(B)$.

Тв.4. $P(A+\overline{A})=P(A)+P(\overline{A})=1$.

Тв.5. Ймовірність добутку двох подій А і В рівна

$$P(A\bullet B)=P(A)\bullet P(B/A)=P(B)\bullet P(A/B)$$

Якщо ж події A і B незалежні, то

$$P(A/B)=P(A),\ P(B/A)=P(B)$$
$$P(A\bullet B)=P(A)\bullet P(B)$$

У загальному випадку

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

Якщо ж $\boldsymbol{A_1}$, $\boldsymbol{A_2}$, ..., $\boldsymbol{A_n}$ – незалежні, то

$$P(A_1\bullet A_2\bullet\ldots\bullet A_n)=P(A_1)\bullet P(A_2)\bullet P(A_3)\bullet\ldots\bullet P(A_n)$$

Тв.6. (Формула повної ймовірності). Нехай випадкова подія А може відбутися тільки разом з однією із попарно несумісних подій $H_i\,(i=\overline{1,n})$, які утворюють повну групу, тоді

$$P(A)=\sum_{i=1}^{n}P(H_i)\cdot P(A/H_i)$$

Тв.7. (Формула Байєса). Якщо виконуються умови *Тв.6*, то

$$P(H_i/A)=\frac{P(H_i)\cdot P(A/H_i)}{\sum_{j=1}^{n}P(H_j)\cdot P(A/H_j)},\ i=1,2,\ldots,n.$$

II. Тестове завдання

1. Чи правильне твердження?

Варіант 1

а) якщо події А і В рівносильні, то Р(А)=Р(В);
б) якщо $A\subset B$ і $B\subset A$, то Р(А)=Р(В);
в) якщо $A\subset B$, то Р(А)>Р(В);
г) якщо у результаті випробування подія А відбулася, а В - ні, то А протилежна до В.
д) якщо події А і В несумісні, то А-В=А;
е) якщо події А і В несумісні, то $A\bullet B=A$;
є) події А і В сумісні, коли $A-B=\varnothing$;

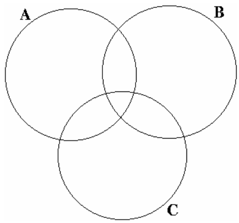
ж) події A і B сумісні, коли A+B=A.

Варіант 2

а) якщо P(A)=P(B), то події A і B рівносильні;
б) якщо P(A)=P(B), то $A \subset B$;
в) якщо $A \subset B$, то P(A)<P(B);
г) якщо A і B протилежні, і в результаті випробування A відбулася, то B не відбудеться;
д) якщо події A і B несумісні, то $A \bullet B = \varnothing$;
е) якщо події A і B несумісні, то $A + B = \varnothing$;
є) події A і B сумісні, коли A•B=B;
ж) події A і B сумісні, коли $A \bullet B = \varnothing$.

(по 1 балу)

2. Дано події A, B і C.



Зобразити на малюнку події:

Варіант 1
а) (A-B)•C;
б) B•(C+A);
в) A•C-B.

Варіант 2
а) (C-A)•B;
б) (B-A)•(C+A);
в) (A+B)-(A•B•C).

(по 2 бали)

3. Чи може бути, що
а) A-B=A;
б) A+B=A•B.

а) A•B=B;
б) B-A=B+A.

(по 2 бали)

4. Події A і $\overline{A}$ – взаємопротилежні. Обчислити

$P(A)$, якщо $P(\overline{A}) = \frac{1}{3}$;

$P(\overline{A})$, якщо $P(A) = 0{,}7$.

(по 1 балу)

5. Знайти

Варіант 1

а) $P(A)$, якщо $P(A+B)=\frac{1}{2}$, $P(A\cdot B)=\frac{1}{3}$, $P(B)=\frac{2}{3}$;

б) $P(A\cdot B)$, якщо Р(А)=0,1, Р(В)=0,4, Р(А/В)=0,2, Р(В/А)=0,8.

Чи будуть події А і В залежними?

Варіант 2

а) $P(B)$, якщо $P(A)=0{,}5$, $P(A+B)=0{,}9$, $A\cdot B=\varnothing$;

б) $P(A\cdot B)$, якщо

$P(A)=\frac{1}{3}$, $P(B)=\frac{2}{3}$, $P(A/B)=\frac{1}{3}$, $P(B/A)=\frac{2}{3}$.

Чи будуть події А і В незалежними?

(по 2 бали)

6. Порівняти вирази

Варіант 1

а) $P(A+B)$ *і* $P(A)+P(B)$;

б) $P(A\cdot B)$ *і* $P(B)$.

Варіант 2

а) $P(A\cdot B)$ *і* $P(B/A)$;

б) $P(A+B)+P(A\cdot B)$ *і* $P(A)+P(B)$.

(по 2 бали)

7. В урні є 4 білі і 6 чорних кульок. Навмання вибирається одна кулька. Подія А полягає у тому, що вийнята кулька біла, а В – чорна:

а) за допомогою дій над подіями записати достовірну подію;

(3 бали)

б) чи будуть події А і В рівноможливими; ***(2 бали)***

в) виписати повну групу подій, якщо з урни вибираються одночасно дві кульки; ***(3 бали)***

г) обчислити $P(A-B)$; ***(3 бали)***

8. Є 4 карти різних мастей. Навмання вибирається одна із них. Подія *A* полягає у тому, що вийнята карта – піка, *B* – хрест, *C* – чирва, *E* – бубна:
   а) за допомогою дій над подіями *A, B, C, E* записати неможливу подію; ***(3 бали)***
   б) чи будуть події А і Е рівноможливими; ***(2 бали)***
   в) виписати повну групу подій, якщо із наявних карт навмання вибираються лише дві карти; ***(3 бали)***
   г) обчислити $P(A+B+C)$ ***(3 бали)***

ІІІ. Приклади розв'язування задач

Задача 1. Ймовірність влучення в ціль одного стрілка 0,8, а другого – 0,7. Стрілки незалежно один від одного зробили по одному пострілу. Яка ймовірність того, що принаймні один стрілок влучить у ціль?

Розв'язування. Введемо у розгляд такі події:

А – принаймні один стрілок влучить у ціль;

В – перший стрілок влучить у ціль;

С – другий стрілок влучить у ціль.

Подія *A* складна, її можна представити у вигляді суми подій *B* і *C*. Отже, *A=B+C.*

За формулою знаходження ймовірності суми подій маємо:

$$P(A)=P(B+C)=P(B)+P(C)-P(B\cdot C) \qquad (*)$$

За умовою *P(B)=0,8; P(C)=0,7.* Оскільки події В і С незалежні, то

$$P(B\cdot C)=P(B)\cdot P(C)=0{,}8\cdot 0{,}7=0{,}56.$$

У результаті із (*) маємо:

$$P(A)=0{,}8+0{,}7-0{,}56=1{,}5-0{,}56=0{,}94.$$

ІІ спосіб. Ймовірність появи принаймні однієї із подій В або С дорівнює:

$$P(A)=1-P(\overline{B})\times P(\overline{C})=1-0{,}2\times 0{,}3=0{,}94.$$

Задача 2. Три команди A_1, A_2, A_3 першого курсу змагаються з трьома командами B_1, B_2, B_3 другого курсу. Ймовірність того, що команди першого курсу виграють зустрічі у команд другого курсу рівні: при зустрічі A_1 з B_1 – 0,8; A_2 з B_2 – 0,4; A_3 з B_3 – 0,4. Для перемоги необхідно виграти не менше двох ігор із трьох (нічиї до уваги не беруться). Перемога якого із курсів найімовірніше?

Розв'язування. Нехай подія A_1 це команда A_1 виграє в B_1, A_2 – команда A_2 виграє в B_2 і A_3 – команда A_3 виграє в B_3. Подія A – виграно не менше двох матчів командою першого курсу.

$$A = A_1 A_2 \overline{A}_3 + A_1 \overline{A}_2 A_3 + \overline{A}_1 A_2 A_3 + A_1 A_2 A_3 .$$

Оскільки всі події незалежні між собою, то ймовірність суми дорівнює сумі ймовірностей, а ймовірність добутку дорівнює добутку ймовірностей. Отже,

$$P(A) = P(A_1) \cdot P(A_2) \cdot P(\overline{A}_3) + P(A_1) \cdot P(\overline{A}_2) \cdot P(A_3) +$$
$$+ P(\overline{A}_1) \cdot P(A_2) \cdot P(A_3) + P(A_1) \cdot P(A_2) \cdot P(A_3)$$

Підставляючи дані задачі, отримаємо: $P(A) = 0{,}544 > \frac{1}{2}$.

Найімовірніше перемога першого курсу.

Задача 3. Стрільба проводиться по п'яти мішенях типу А, трьом – типу В і двом типу – С. Ймовірність влучення в мішень типу А дорівнює 0,4; типу В – 0,1; типу С – 0,15. Знайти ймовірність влучення в мішень при одному пострілі (може бути влучення в будь-яку з мішеней).

Розв'язування. Нехай подія А – це мішень типу А; В – мішень типу В і С – мішень типу С. Тоді: P(A)=0,5; P(B)=0,3; P(C)=0,2. Нехай подія Е – влучення в мішень. За умовою задачі $P_A(E)=0{,}4$; $P_B(E)=0{,}1$; $P_C(E)=0{,}15$. Тоді за формулою повної ймовірності

$$P(E) = P_A(E) \cdot P(A) + P_B(E) \cdot P(B) + P_C(E) \cdot P(C) .$$

Підставляючи дані, отримаємо:

$$P(E) = 0{,}5 \cdot 0{,}4 + 0{,}3 \cdot 0{,}1 + 0{,}2 \cdot 0{,}15 = 0{,}26 .$$

Задача 4. На трьох фабриках виробляється відповідно 5000, 8000, 2000 електроламп. Відомо, що на першій фабриці питома вага браку у виготовленій продукції становить 0,3%, на другій – 0,2%, на третій – 0,5%. Визначити, на якій із фабрик виготовлена куплена бракована лампа (з найбільшою ймовірністю).

Розв'язування. Позначимо через H_1, H_2, H_3 випадкові події, які полягають у тому, що куплена лампа виготовлена на першій, другій, третій фабриках відповідно, а В - куплена лампа бракована. Події H_1, H_2, H_3 попарно несумісні, вони утворюють повну групу. За класичним означенням ймовірності маємо:

$$P(H_1) = \frac{5000}{5000+8000+2000} = \frac{5}{15};$$
$$P(H_2) = \frac{8000}{5000+8000+2000} = \frac{8}{15};$$
$$P(H_3) = \frac{2000}{5000+8000+2000} = \frac{2}{15},$$
$$P(B/H_1) = 0{,}003; \quad P(B/H_2) = 0{,}002; \quad P(B/H_3) = 0{,}005.$$

За формулою Байєса

$$P(H_i/B) = \frac{P(H_i)\cdot P(B/H_i)}{\sum_{j=1}^{3} P(H_j)\cdot P(B/H_j)}, \quad i = 1,2,3$$

$$P(H_1/B) \approx 0{,}37; \quad P(H_2/B) \approx 0{,}39; \quad P(H_3/B) \approx 0{,}24.$$

Отже, найбільш ймовірно, що бракована лампа виготовлена на другій фабриці.

IV. Завдання для самостійного виконання

Завдання 1

1.1 Нехай ймовірність того, що покупцю необхідно взуття 41-го розміру дорівнює 0,2. Знайти ймовірність того, що п’ять перших покупців вимагатимуть взуття 41-*го* розміру.

1.2 Ймовірність влучення в мішень першим стрілком при одному пострілі дорівнює 0,8, а другим стрілком – 0,6. Знайти ймовірність того, що в ціль влучить один стрілок.

1.3 У мішку змішані нитки, серед яких 30% білих, а інші – червоні. Визначити ймовірність того, що вийняті навмання дві нитки будуть:
   а) одного кольору;
   б) різних кольорів.

1.4 Нехай ймовірність того, що людина помре на 71-му році життя дорівнює 0,04. Яка ймовірність, що із трьох людей 70-літнього віку через рік будуть жити:
   а) всі;
   б) принаймні один.

Завдання 2

2.1 У телевізійному ательє є 4 кінескопи. Ймовірність того, що кожен з них витримає гарантійний термін відповідно дорівнюють 0,8; 0,85; 0,9; 0,95. Знайти ймовірність того, що взятий навмання кінескоп витримає гарантійний термін.

2.2 Для участі в студентських відбіркових спортивних змаганнях виділено з першого курсу – 4, з другого – 5, з третього – 6 студентів. Ймовірність того, що студент першого, другого або третього курсу попаде в збірну Академії відповідно рівні 0,9; 0,7 і 0,8. Навмання вибраний студент у результаті змагань попав у збірну. До якого курсу найімовірніше він належав?

2.3 У групі спортсменів 20 лижників, 6 велосипедистів та 4 бігуни. Ймовірність виконати кваліфікаційну норму для лижника 0,9, для велосипедиста – 0,8, і для бігуна – 0,75. Знайти ймовірність того, що навмання вибраний спортсмен виконає норму.

2.4 Стрільба проводиться по п’яти звичайних мішенях, трьох – рухомим і двом – летючим. Ймовірність попадання у звичайну мішень дорівнює 0,4, у рухому – 0,1, у летючу – 0,15. Постріл в одну із мішеней дав влучення. Знайти ймовірності того, що влучення було в звичайну, рухому, летючу мішень.

2.5 Ймовірність двом близнюкам бути хлопчиками дорівнює 0,32, дівчатками – 0,28. Припускаючи, що подія: α) перша дитина хлопчик, а друга – дівчинка; β) перша дитина - дівчинка, а друга – хлопчик – рівноймовірні, знайти безумовні ймовірності народження:

а) хлопчика;

б) дівчинки.

ЛАБОРАТОРНА РОБОТА 4

Тема: *Формули Бернуллі, Лапласа та Пуассона. Інтегральна теорема Лапласа. Найімовірніше число появ події у незалежних випробуваннях*

I. Деякі теоретичні відомості

***Основні терміни**: незалежні повторні випробування, ймовірність події в незалежному повторному випробуванні, ймовірність того, що в n випробуваннях подія відбудеться k разів (n>k), відхилення відносної частоти від постійної ймовірності в незалежних випробуваннях, найімовірніше число появ події в незалежних випробуваннях.*

Основні означення, твердження і формули

У цій роботі розглядаємо незалежні випробування, в кожному із яких ймовірність появи події дорівнює p $(0<p<1)$.

Тв.1 (Формула Бернуллі). Ймовірність того, що в n незалежних випробуваннях, у кожному з яких ймовірність появи події постійна і рівна числу p $(0<p<1)$, подія відбудеться рівно k разів ($k \le n$, в будь-якій послідовності) дорівнює

$$P_n(k) = C_n^k p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k}, де\ q = 1 - p .$$

Тв.2 (Локальна теорема Лапласа). Якщо n і k великі числа, то доцільно застосувати формулу Лапласа

$$P_n(k) = \frac{1}{\sqrt{npq}} \cdot \varphi(x), \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \text{ де } x = \frac{k - np}{\sqrt{npq}} .$$

Зауваження 1. Таблиці для обчислення значень $\varphi(x)$ дані наприкінці підручника, або скористатися функцією НОРМРАСП Microsoft Excel.

Зауваження 2. Якщо $x<0$, то треба використати парність функції $\varphi(x)$, тобто $(\varphi(-x) = \varphi(x))$.

Тв.3 (Інтегральна теорема Лапласа). Ймовірність того, що в n незалежних випробуваннях, у кожному з яких ймовірність появи події дорівнює p $(0<p<1)$, подія відбудеться не менше k_1 разів і не більше k_2 разів, наближено дорівнює:

$$P_n(k_1, k_2) = \Phi(x_2) - \Phi(x_1) ,$$

де $\Phi(x)=\dfrac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$ – функція Лапласа, $x_1=\dfrac{k_1-np}{\sqrt{npq}}, x_2=\dfrac{k_2-np}{\sqrt{npq}}$.

Зауваження. Таблиця значень для $\Phi(x)\,(0\le x\le 5)$ наведена наприкінці підручника [1]; для значень *x>5 Ф(x)=0,5*; для *x<0* користуються тою ж таблицею, враховуючи непарність функції *Ф(x),* $(\Phi(-x)=-\Phi(x))$, або скористатися функцією НОРМРАСП Microsoft Excel.

Абсолютним відхиленням відносної частоти від ймовірності називається модуль різниці $\left|\dfrac{m}{n}-p\right|$.

Тв. 4. Ймовірність того, що в *n* незалежних випробуваннях, у кожному з яких ймовірність появи події дорівнює *p (0<p<1)*, абсолютне відхилення не перевищує числа *ε* наближено дорівнює:

$$P\left(\left|\frac{m}{n}-p\right|\le\varepsilon\right)\approx 2\Phi\left(\varepsilon\sqrt{\frac{np}{q}}\right)$$

Тв. 5. (Теорема Пуассона). Якщо в кожному випробуванні ймовірність появи події постійна і мала, а число випробувань достатньо велике, то $P_n(k)\approx\dfrac{\lambda^k}{k!}e^{-\lambda}$ де $\lambda=np$

Число k_0 називається найімовірнішим числом появи події в n-незалежних випробуваннях, якщо ймовірність того, що подія відбудеться k_0 разів не менше ймовірності можливих наслідків випробувань.

Тв. 6. Найімовірніше число появ події k_0 визначається із нерівності: $np-q\le k_0<np+p$

Зауваження. Якщо число

а) $(np-q)$ – дробове, то існує одне найімовірніше число k_0;

б) $(np-q)$ – ціле, то існує два найімовірніших числа k_0 і k_0+1;

в) np – ціле, то найімовірніше число $k_0=np$.

II. Приклади розв'язування задач

Задача 1. Два рівносильних суперники грають у шахи. Що ймовірніше виграти:

а) одну партію із двох чи дві із чотирьох?

б) не менше двох партій із чотирьох чи не менше трьох із п'яти?

Нічиї не враховуються.

Розв'язування. Оскільки грають рівносильні суперники, то ймовірність виграшу $p = \frac{1}{2}$ і ,отже, поразки $q = \frac{1}{2}$. Оскільки в усіх партіях ймовірність виграшу постійна і все рівно в якій послідовності будуть виграні партії, то можна застосувати формулу Бернуллі.

а) Обчислимо ймовірність того, що виграна одна партія з двох:

$$P_2(1) = C_2^1 p^1 q^1 = \frac{2!}{1! \cdot 1!}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^1 = \frac{1}{2}.$$

Обчислимо ймовірність того, що дві партії з чотирьох будуть виграні: $P_4(2) = C_4^2 p^2 q^2 = \frac{4!}{2! \cdot 2!}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = \frac{6}{16} = \frac{3}{8}$.

Оскільки $P_2(1) > P_4(2)$, то ймовірніше виграти одну партію з двох.

б) Обчислимо ймовірність виграти не менше двох партій з чотирьох:

$$P_4(2) + P_4(3) + P_4(4) = 1 - (P_4(0) + P_4(1)) = \frac{11}{16}.$$

Обчислимо ймовірність виграти не менше трьох партій з п'яти:

$$P_5(3) + P_5(4) + P_5(5) = \frac{8}{16}.$$

Отже, ймовірніше виграти не менше двох партій з чотирьох.

Задача 2. Ймовірність народження хлопчика дорівнює 0,515. Знайти ймовірність того, що з 200 новонароджених дітей хлопчиків і дівчаток буде порівну.

Розв'язування. У даному випадку *n=200*, *k=100*, *p=0,515*, *q=1-p=0,485*. Оскільки *n* і *k* великі, то застосовуємо локальну теорему Лапласа.

$$\sqrt{npq} = \sqrt{200 \cdot 0{,}515 \cdot 0{,}485} \approx 7{,}068,$$

$$x = \frac{k - np}{\sqrt{npq}} = \frac{100 - 200 \cdot 0{,}515}{7{,}068} = -0{,}424.$$

Оскільки $\varphi(-0{,}424) = \varphi(0{,}424) = 0{,}3647$, то отримаємо

$$P_{200}(100) = \frac{(-0{,}424)}{7{,}068} \approx 0{,}052.$$

Задача 3. Знайти ймовірність того, що серед 1000 новонароджених хлопчиків буде від 465 до 550 включно, якщо народження хлопчика та дівчинки рівноможливі події.

Розв'язування. Оскільки народження хлопчика та дівчинки рівноможливі події, то *p=q=0,5*. Застосовуємо інтегральну теорему Лапласа

$$x_2 = \frac{550 - 1000 \cdot 0{,}5}{\sqrt{1000 \cdot 0{,}5 \cdot 0{,}5}} \approx 3{,}16;\ x_1 = \frac{465 - 1000 \cdot 0{,}5}{\sqrt{1000 \cdot 0{,}5 \cdot 0{,}5}} \approx -2{,}21$$

$$P_{1000}(465{,}550) = \Phi(3{,}16) - \Phi(-2{,}21) \approx 0{,}49931 + 0{,}4861 = 0{,}98541.$$

Задача 4. Французький вчений Бюфорон (XVIII ст.) підкидав монету 4040 разів, причому герб випав 2048 разів. Знайти ймовірність того, що при повторенні досліду Бюфорона відносна частота появи герба відхилиться від ймовірності появи герба по абсолютній величині не більше як у досліді Бюфорона.

Розв'язування. Перш як застосувати Тв.4, знайдемо *ε*. Оскільки *n=4040, m=2048, p=0,5, q=0,5*, то $\left|\frac{m}{n} - p\right| = \left|\frac{2048}{4040} - 0{,}5\right| = 0{,}0069 \cdot$

Тоді

$$P\left(\left|\frac{m}{n} - p\right| \le 0{,}0069\right) = 2\Phi\left(0{,}0069\sqrt{\frac{4040}{0{,}25}}\right) \approx 2\Phi(0{,}877) \approx 0{,}6196.$$

Задача 5. Підручник видано тиражем 100 000 примірників. Ймовірність того, що підручник сформований неправильно дорівнює 0,0001. Знайти ймовірність того, що тираж має рівно 5 бракованих книг.

Розв'язування. За умовою *n=100 000, p=0,0001, k=5*. Оскільки події, книга сформована неправильно, незалежні, число *n* велике, а ймовірність мала, то скористаємося теоремою Пуассона:

$$P_n(k) \approx \frac{\lambda^k}{k!} e^{-\lambda} = \frac{10^5 e^{-10}}{5!} \approx \frac{10^5 \cdot 0{,}000045}{120} \approx 0{,}0375.$$

Задача 6. Нехай ймовірність того, що пасажир запізниться на поїзд дорівнює 0,02. Знайти найімовірніше число пасажирів, які запізнилися із 855 пасажирів.

Розв'язування. За умовою $n=855$, $p=0,02$, $q=0,98$. Знайдемо найімовірніше число m із подвійної нерівності $np-q \le k_0 \le np+p$. Підставивши сюди дані задачі отримаємо:

$855 \cdot 0,02 - 0,98 \le k_0 < 855 \cdot 0,02 + 0,02$, або $16,12 \le k_0 < 17,12$.

Оскільки m ціле число, то $k_0=17$.

III. Завдання для самостійного виконання

Завдання 1

1.1 Два рівносильних суперники грають у шахмати. Що ймовірніше: виграти дві партії з чотирьох, чи три партії з шести (нічиї не враховуються).

1.2 У сім'ї п'ятеро дітей. Знайти ймовірність того, що серед цих дітей:
   а) два хлопчики;
   б) не більше двох хлопчиків;
   в) більше двох хлопчиків;
   г) не менше двох і не більше трьох хлопчиків.

   Вважати ймовірність народження хлопчика рівною 0,51.

1.3 Ймовірність влучення в мішень при одному пострілі дорівнює 0,8. Знайти ймовірність того, що при 100 пострілах буде 75 влучень у мішень.

1.4 Ймовірність появи події у кожному із 100 незалежних випробувань постійна і рівна p=0,8. Знайти ймовірність того, що подія відбудеться:
   а) не менше 75 разів і не більше 80;
   б) не менше 75 разів;
   в) не більше 74 разів.

1.5 Ймовірність появи події у кожному незалежному випробуванні дорівнює 0,8. Скільки треба провести випробувань, щоб з ймовірністю 0,9 можна було б чекати, що подія з'явиться не менше 75 разів?

Завдання 2

2.1 Нехай ймовірність того, що покупцю необхідно взуття $41^{го}$ розміру, дорівнює 0,2. Знайти ймовірність того, що із 750 покупців не більше 120 потрібно взуття $41^{го}$ розміру.

2.2 Ймовірність появи події у кожному із незалежних випробувань дорівнює 0,5. Знайти число випробувань n, при якому з ймовірністю 0,7698 можна чекати, що відносна частота появи події відхилиться від її ймовірності, по абсолютній величині, не більше як на 0,02.

2.3 Чому повинно бути рівне число n-підкидань гранчастого кубика, щоб ймовірність нерівності $\left|\frac{m}{n} - p\right| \leq 0{,}1$ була не менше як ймовірність протилежної ймовірності, де m – число появ 1 в n підкиданнях?

Завдання 3

3.1 Ймовірність смерті на 21-у році життя дорівнює 0,006. Застрахована група в 1000 осіб у віці 20 років. Яка ймовірність того, що протягом року помре 5 застрахованих?

3.2 Застосовуючи а) формулу Бернуллі; б) локальну теорему Лапласа і в) формулу Пуассона знайти ймовірність того, що серед 200 осіб виявиться четверо лівшів, якщо у середньому лівші становлять 1%. Пояснити результати обчислень.

3.3 Скільки треба посіяти зерен, сходження яких 70%, щоб найімовірніше число зерен, які не зійшли, дорівнює 60?

3.4 Гральний кубик кидають 80 разів. Знайти з ймовірністю 0,99973 межі, в яких лежатиме m число появ шестірки.

Завдання 4

4.1 Товарознавець оглядає 24 зразки товару. Ймовірність того, що кожен із зразків буде придатний до продажу, дорівнює 0,6. Знайти найімовірніше число зразків, які товарознавець визначить придатними до продажу.

4.2 Батарея провела 6 пострілів по воєнному об’єкту. Ймовірність влучення в ціль при одному пострілі дорівнює 0,3. Знайти:
   а) найімовірніше число влучень;
   б) ймовірність найімовірнішого числа влучень;
   в) ймовірність того, що об’єкт буде зруйнований, якщо для цього достатньо принаймні два влучення.

4.3 Застраховано на один рік: а) 1000 і б) 4000 осіб 20-літнього віку. Страховий внесок кожного 1,5 грн. У випадку смерті застрахованого родичі отримають 120 грн. Яка ймовірність того, що до кінця року страхова установа не буде банкротом, якщо ймовірність смерті на 21-у році життя для кожного дорівнює 0,006?

ЛАБОРАТОРНА РОБОТА 5

Тема: Одновимірні дискретні випадкові величини та їх числові характеристики. Нормальний розподіл

I. Деякі теоретичні відомості

Основні терміни: *випадкова величина, дискретна випадкова величина, закон розподілу ймовірностей, можливі значення випадкових величин, ймовірності можливих значень, числові характеристики випадкових величин, математичне сподівання (властивості), дисперсія (властивості), середнє квадратичне відхилення, щільність нормального розподілу.*

Основні означення, твердження і формули

Випадковою величиною називається змінна, яка залежно від наслідків випробування може приймати різні значення. Позначення: $X, У, Z, A, B, C, X_1, X_2, X_3$, і т.д.

Дискретною називають випадкову величину, можливі значення якої є окремі ізольовані числа, котрі ця величина приймає з певною ймовірністю.

Законом розподілу дискретної випадкової величини називають перелік її значень та відповідних їм ймовірностей.

Закон розподілу дискретної випадкової величини X може бути заданий у вигляді таблиці, перший рядок якої складається з можливих значень x_i випадкової величини X, а друга – ймовірності p_i:

X	x_1	x_2	…	x_n
P	p_1	p_2	…	p_n

де $\sum_{i=1}^{n} p_i = 1$.

Закон розподілу дискретної випадкової величин також може бути заданий за допомогою формули (аналітично) та графіка (многокутник розподілу).

Характеристикою розташування середнього значення випадкової величини є математичне сподівання.

Математичним сподіванням дискретної випадкової величини X називається сума добутків усіх її можливих значень на відповідні ймовірності:

$$M(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n.$$

Для математичного сподівання виконуються властивості:

1. $M(C)=C$.
2. $M(X_1 + X_2 + \ldots + X_n) = M(X_1) + M(X_2) + \ldots + M(X_n)$.
3. Математичне сподівання добутку попарно-незалежних випадкових величин дорівнює добутку їх математичних сподівань: $M(X_1 \cdot X_2 \cdot \ldots \cdot X_n) = M(X_1) \cdot M(X_2) \cdot \ldots \cdot M(X_n)$.
4. Математичне сподівання біномінального розподілу (для нього ймовірність знаходиться за формулою Бернуллі) дорівнює добутку числа випробувань на ймовірність появи події в одному випробуванні: $M(X) = np$.

Характеристиками розсіяння можливих значень випадкової величини навколо математичного сподівання є дисперсія та середнє квадратичне відхилення.

Дисперсією випадкової величини X називають математичне сподівання квадрата відхилення: $D(X) = M(X - M(X))^2$.

Дисперсію зручно обчислювати за формулою:

$$D(X) = M(X^2) - M^2(X)$$

Дисперсія випадкової величини має властивості:

1. Дисперсія сталої дорівнює нулю: $D(C) = 0$.
2. Сталий множник можна винести за знак дисперсії, підносять його до квадрату: $D(CX) = C^2 D(X)$.
3. Дисперсія суми попарно-незалежних випадкових величин дорівнює сумі дисперсій додатків: $D(X_1 + X_2 + \ldots + X_n) = D(X_1) + D(X_2) + \ldots + D(X_n)$.
4. Дисперсія біноміального розподілу дорівнює добутку числа випробувань на ймовірності появи і не появи події в одному випробуванні: $D(X) = npq$.

Середнім квадратичним відхиленням випадкової величини називають квадратний корінь із дисперсії: $\sigma(X) = \sqrt{D(X)}$.

Над випадковими величинами можна виконувати операції додавання, віднімання та множення.

Сумою випадкових величин X і У – це нова випадкова величина, яка приймає всі значення вигляду: $x_s + y_j$ з ймовірностями $P_{sj}\left(p_{sj} = p_s \times p_j\right)$ – якщо X і У незалежні.

Аналогічно визначаються різниця та добуток випадкових величин.

Добутком випадкової величини *X* та сталої *k* називають нову випадкову величину, значення якої рівне добутку kx_i, а ймовірності ті ж самі.

Квадратом випадкової величини *X*, тобто X^2 – нова випадкова величина, яка з тими ж ймовірностями, що й *X* приймає значення, рівні квадратам значень випадкової величини *X*.

Нормальним називається розподіл ймовірностей неперервної випадкової величини, який записується диференціальною функцією (функція щільності розподілу):

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

де $a = M(X)$ – математичне сподівання випадкової величини,

$\sigma = \sigma(x)$ – середнє квадратичне відхилення нормального розподілу.

Нормованим називають нормальний розподіл з параметрами $a = 0$ і $\sigma = 1$.

Тв. 1. Якщо випадкова величина X має нормальний розподіл, то ймовірність того, що X набере значень, які належать інтервалу *(α,β)* дорівнює: $P(\alpha < X < \beta) = \int\limits_{\alpha}^{\beta} f(x)dx = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right)$,

де $\Phi(X) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{z^2}{2}} dz$ – інтегральна функція Лапласа.

Зауваження. Функція Лапласа непарна. $\Phi(-X) = -\Phi(X)$. Значення функції Лапласа знаходяться з таблиць наприкінці підручника, або скористатися функцією НОРМРАСП Microsoft Excel.

Тв. 2. Ймовірність того, що відхилення нормально розподіленої випадкової величини X по абсолютній величині менше заданого числа δ дорівнює: $P\left(|x-a|<\delta\right)=2\Phi\left(\frac{\delta}{\sigma}\right)$.

Зауваження. Для $a=0$ $P\left(|x|<\delta\right)=2\Phi\left(\frac{\delta}{\sigma}\right)$

Тв. 3 (Правило трьох сигм). Ймовірність того, що відхилення по абсолютній величині буде менше потроєного середньоквадратичного відхилення, дорівнює 0,9973.

$$P\left(|x-a|<3\sigma\right)=2\Phi(3)=0{,}9973 .$$

На практиці правило трьох сигм застосовується так: якщо розподіл досліджуваної величини невідомий, але умова, вказана в твердженні, виконується, то є підстава вважати, що досліджувана величина розподілена нормально; у протилежному випадку вона не має нормального розподілу.

II. Тестове завдання

1. Чи може наступна таблиця задавати закон розподілу ймовірностей випадкової величини?

Варіант 1

а)

X	0	1	7	7,5	8
P	0,1	1/2	0,15	0,05	0,2

б)

X	-5	-4	0	4	5
P	0,05	0,05	0,7	0,15	0,15

в)

X	-100	-55	1,5	2	7
P	-0,1	0,2	-1,5	1,5	0,9

Варіант 2

а)

X	-5	-4	2	7	25
P	2/3	1/25	3/7	1/3	1/3

б)

X	0	0,5	1,5	2,7	18
P	0,05	0,05	0,01	0,09	0,8

в)

X	27	28	31	33	40
P	-1/2	0,5	1	2/27	1/9

(по 2 бали)

2. Заповнити пусті клітинки у таблиці, яка задає закон розподілу випадкової величини

Варіант 1

X	-7	0	7	13	20
P	0,125		0,3	0,145	0,03

Варіант 2

X	1,5	1,75	2	2,25	2,5
P	2/7	1/14	3/7		1/7

(по 2 бали)

3. Побудувати закон розподілу випадкової величин *X*, якщо

Варіант 1: X – це кількість очок, які випадають у результаті одного кидання грального кубика.

Варіант 2: X – це кількість випадання копійки при двократному киданні монети.

(по 3 бали)

4. Дискретна випадкова величина *X* задана законом розподілу

Варіант 1

X	-1	0	1
P	0,1	0,7	0,2

Варіант 2

X	1	3
P	0,6	0,4

Знайти

а) математичне сподівання $M(x)$;

б) дисперсію $D(x)$;

в) середнє квадратичне відхилення $\sigma(x)$;

г) моду $M_0(x)$.

III. Приклади розв'язування задач

Задача 1. Написати біноміальний закон розподілу дискретної випадкової величини X – числа появ герба при двох підкиданнях монети та побудувати многокутник розподілу.

Розв'язування. При двох підкиданнях монети герб може появитися 0, 1 або 2 рази. Знайдемо ймовірності появи герба, користуючись формулою Бернуллі.

$$P_2(0) = C_2^0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^2 = \frac{2!}{0!\cdot 2!} \cdot 1 \cdot \frac{1}{4} = \frac{1}{4},$$

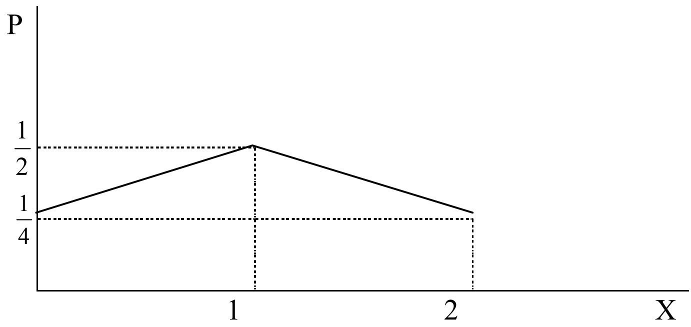
$$P_2(1) = C_2^1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^1 = \frac{2!}{1!\cdot 1!} \cdot \frac{1}{4} = \frac{1}{2},$$

$$P_2(2) = C_2^2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^0 = \frac{2!}{0!\cdot 2!} \cdot \frac{1}{4} = \frac{1}{4}.$$

Отже, закон розподілу матиме вигляд:

X	0	1	2
P	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Побудуємо многокутник розподілу



Задача2. Дано закони розподілу незалежних випадкових величин X і У.

X	-1	0	1
P	0,2	0,3	0,5

У	0	1	2	3
P	0,1	0,2	0,3	0,4

Скласти закони розподілу для (X+У), (X·У) та обчислити M(X), M(У), D(X), D(У), M(X+У), M(X·У), D(X+У), D(X·У).

Розв'язування. Складемо допоміжну таблицю.

№з з/п	**X**	**У**	**X+У**	**X·У**	**Ймовірність відповідного результату**
1	-1	0	-1	0	0,2·0,1=0,02
2	-1	1	0	-1	0,2·0,2=0,04
3	-1	2	1	-2	0,2·0,3=0,06
4	-1	3	2	-3	0,2·0,4=0,08
5	0	0	0	0	0,3·0,1=0,03
6	0	1	1	0	0,3·0,2=0,06
7	0	2	2	0	0,3·0,3=0,09
8	0	3	3	0	0,3·0,4=0,12
9	1	0	1	0	0,5·0,1=0,05
10	1	1	2	1	0,5·0,2=0,1
11	1	2	3	2	0,5·0,3=0,15
12	1	3	4	3	0,5·0,4=0,2

Запишемо тепер розподіли (X+У) та (X•У)

X+У	-1	0	1	2	3	4

Р	0,02	0,07	0,17	0,27	0,27	0,2

Х•У	-3	-2	-1	0	1	2	3
Р	0,08	0,06	0,04	0,37	0,1	0,15	0,20

Знайдемо числові характеристики

М(Х)=0,3 М(Х+У)=2,3

М(У)=2 М(Х•У)=0,6

D(Х)=0,61 D(Х+У)=1,61

D(У)=1 D(Х•У)=3,14.

Задача 3. Екзаменатор задає студенту додаткове запитання. Ймовірність того, що студент відповість на будь-яке запитання – 0,9. Викладач припиняє екзамен, якщо виявляє незнання відповіді на поставлене запитання. Знайти:

а) скласти закон розподілу дискретної випадкової величини X – кількості додаткових запитань, які даватиме викладач студенту;

б) найімовірніше число k_0 поставлених додаткових запитань.

Розв'язування. а) Дискретна випадкова величина Х – число заданих додаткових запитань приймає такі значення: $x_1=1,\ x_2=2,\ x_3=3,\ \ldots,\ x_k=k,\ \ldots$

Знайдемо ймовірності цих можливих значень. Величина X прийме значення $x_1=1$ (викладач дасть тільки одне запитання), якщо студент не відповість на перше запитання. Ймовірність цього значення дорівнює

$1\text{-}0,9=0,1$. Таким чином, $P(X_1=1)=0,1$.

Величина X набуде можливого значення $x_2=2$ (екзаменатор поставить тільки два запитання), якщо студент відповідатиме на перше запитання і не відповідатиме на друге. Таким чином, $P(X=2)=0,9\bullet 0,1=0,09$.

Аналогічно знайдемо:

$P(X=3)=0,9\bullet 0,9\bullet 0,1=0,081,\ \ldots,\ P(X=k)=0,9^{k-1}\bullet 0,1,\ \ldots$

Отже, шуканий закон розподілу має вигляд:

Х	1	2	3	...	k	...
Р	0,1	0,09	0,081	...	$0,9^{k-1}\bullet 0,1$	...

б) Найімовірніше число k_0 заданих запитань (найімовірніше можливе значення Х), тобто число поставлених викладачем запитань, яке має найбільшу ймовірність, як бачимо із закону розподілу, дорівнює одиниці.

Задача 4. Дано перелік можливих значень дискретної випадкової величини X: $x_1=-1$, $x_2=0$, $x_3=1$, а також дано математичне сподівання випадкової величини та її квадрати: $M(X)=0,1$, $M(X^2)=0,9$. Знайти ймовірності p_1, p_2, p_3, які відповідають можливим значенням x_1, x_2, x_3.

Розв'язування. Користуючись тим, що сума ймовірностей можливих значень X дорівнює одиниці, а також враховуючи, що $M(X)=0,1$, $M(X^2)=0,9$, отримаємо систему рівнянь:

$$\begin{cases} p_1 + p_2 + p_3 = 1 \\ -p_1 + p_3 = 0{,}1 \\ (-1)^2 p_1 + (-1)^2 p_3 = 0{,}9 \end{cases}$$

Розв'язуючи цю систему методом Гауса, Крамера або матричним, отримаємо: $p_1=0,4$, $p_2=0,1$, $p_3=0,5$.

Задача 5. Знайти математичне сподівання та дисперсію дискретної випадкової величини X – числа появ події А в п'яти незалежних випробуваннях, якщо ймовірність появи події А в кожному випробуванні дорівнює 0,2.

Розв'язування. Використовуючи властивість 4 для математичного сподівання та дисперсії отримаємо: $M(X)=n \cdot p=5 \cdot 0,2=1$, $D(X)=npq=5 \cdot 0,2 \cdot 0,8=0,8$.

III. Завдання для самостійного виконання

Завдання 1. Дано закони розподілу незалежних дискретних випадкових незалежних величин X та $У$.

X	x_1	x_2	x_3	x_4	x_5
P	a	a	$5a$	a	$2a$

$У$	y_1	y_2	y_3	y_4
P	0,4	0,3	0,1	0,2

Знайти:

а) a;

б) закони розподілу $2X$, $X+У$, $XУ$, $X-У$;

в) $M(X)$, $D(X)$, $\sigma(X)$.

Дані x_1, x_2, x_3, x_4, x_5, *та* y_1, y_2, y_3, y_4 задані у табл. 1.

Таблиця 1

№ варіанта	x_1	x_2	x_3	x_4	x_5	y_1	y_2	y_3	y_4
1	-4	-3	-2	0	1	0	1	2	3
2	-3	-2	-1	0	1	1	2	3	4
3	-2	-1	0	1	2	-10	-9	-8	-7
4	-1	0	1	2	3	-9	-8	-7	-6
5	0	1	2	3	4	-8	-7	-6	-5
6	1	2	3	4	5	-7	-6	-5	-4
7	2	3	4	5	6	-6	-5	-4	-3
8	-10	-9	-8	-7	-6	-5	-4	-3	-2
9	-9	-8	-7	-6	-5	-4	-3	-2	-1
10	-8	-7	-6	-5	-4	-3	-2	-1	0
11	-7	-6	-5	-4	-3	-2	-1	0	1
12	-6	-5	-4	-3	-2	-1	0	1	2
13	-5	-4	-3	-2	-1	0	2	4	6
14	-10	-8	-6	-4	-2	1	3	5	7
15	-9	-7	-5	-3	-1	-4	-2	0	2

Завдання 2. Задана нормально розподілена випадкова величина X. Математичне сподівання і дисперсія X рівні $M(X)$ і $D(X)$ відповідно:

- записати щільність розподілу ймовірностей $y = f(x)$ і побудувати графік щільності;
- обчислити ймовірність попадання випадкової величини X в інтервал $(\alpha;\beta)$;
- знайти ймовірність того, що абсолютна величина відхилення $|x-a|$ не перевищує заданого δ, тобто $P(|x-a|<\delta)$.

Дані $M(X)$, $D(X)$, α, β і δ задано в таблиці 2.

Таблиця 2

№ варіанта	$M(X)$	$D(X)$	α	β	δ
1	3	4	1	2	10
2	0	4	2	3	15
3	-1	1	0	14	4
4	-3	3	-1	1	10
5	4	4	0	5	5
6	-2	9	7	9	1
7	-1	4	7	8	2
8	-3	1	-2	0	3
9	4	1	7	9	1
10	3	4	-2	0	7
11	7	4	-2	3	6
12	5	1	0	7	1
13	-5,5	2	3	5	5
14	-7	16	-3	5	10
15	4	4	-4	4	4

ЛАБОРАТОРНА РОБОТА 6

Тема: Двовимірні дискретні випадкові величини та їх числові характеристики

I. Деякі теоретичні відомості

Основні терміни: *двовимірна випадкова величина, закон розподілу дискретної двовимірної випадкової величини, умовні закони розподілу складових системи дискретних випадкових величин, умовне математичне сподівання, кореляційний момент, коефіцієнт кореляції.*

Основні означення, твердження і формули

Крім одновимірних випадкових величин, вивчають величини, можливі значення яких визначаються двома, трьома, ..., n числами. Такі величини називають відповідно двовимірними, тривимірними, ..., n-вимірними.

Закон розподілу дискретної двовимірної випадкової величини називають перелік можливих значень цієї величини (тобто пар чисел (x_i, y_j)) та їх ймовірностей $P(x_i, y_j)(i = 1, 2, ..., n;\ j = 1, 2, ..., m)$. Найчастіше закон розподілу задають таблицею:

X / $У$	X_1	X_2	. . .	X_i	. . .	X_n
$У_1$	$P(x_1,y_1)$	$P(x_2,y_1)$	. . .	$P(x_i,y_1)$	. . .	$P(x_n,y_1)$
. . .			. . .		. . .	
$У_j$	$P(x_1,y_j)$	$P(x_2,y_j)$	. . .	$P(x_i,y_j)$	. . .	$P(x_n,y_j)$
. . .			. . .		. . .	
$У_m$	$P(x_1,y_m)$	$P(x_2,y_m)$	. . .	$P(x_i,y_m)$	. . .	$P(x_n,y_m)$

Знаючи закон розподілу двовимірної дискретної випадкової величини, можна знайти закони розподілу кожної із складових X і $У$. Наприклад, для складової X будуть значення першого рядка – варіантами, а ймовірності знаходяться як сума ймовірностей відповідного стовпця (для x_1 – першого, для x_2 – другого та ін.).

Позначимо умовну ймовірність того, що $X=x_1$, за умови, що $У=y_1$ через $P(x_1/y_1)$. Вона, як бачимо, не рівна $P(x_1)$.

Умовним розподілом складової X при $У=y_j$ називається сукупність умовних ймовірностей:

$$P(x_1/y_j), P(x_2/y_j), \ldots, P(x_n/y_j) \ .$$

Знаючи закон розподілу двовимірної дискретної випадкової величини, можна легко визначити закони розподілу складових. Наприклад, умовний закон розподілу *X*, у припущенні, що подія $У=y_1$, відбулася, може бути знайдений за формулою:

$$P(x_i/y_1) = \frac{P(x_i/y_1)}{P(y_1)}, i = 1, 2, \ldots, n \ .$$

Сума ймовірностей умовного закону розподілу дорівнює 1.

Умовним математичним сподіванням дискретної випадкової величини У при $X=x$ (x – фіксоване можливе значення X) називають добуток можливих значень *У* на їх умовні ймовірності:

$$P(У/X = x) = \sum_{j=1}^{m} y_j P(y_j/x).$$

Кореляційним моментом μ_{xy} випадкових величин *X* і *У* називають математичне сподівання добутку відхилень цих величин:

$$\mu_{xy} = M\left[(X - M(X))(У - M(У))\right] = M(XУ) - M(X)\cdot M(У).$$

Для обчислення кореляційного моменту дискреційних величин користуються формулою:

$$\mu_{xy} = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i - M(X))(y_j - M(У))\cdot P(x_i, y_j).$$

Твердження. Кореляційний момент двох незалежних випадкових величин *X* і *У* дорівнює нулю.

Коефіцієнтом кореляції r_{xy} випадкових величин *X* і *У* називають відношення кореляційного моменту до добутку середньоквадратичних відхилень цих величин: $r_{xy} = \dfrac{\mu_{xy}}{\sigma_x \sigma_y}$.

Коефіцієнт кореляції визначає тісноту зв’язку між Х і У. Він лежить у межах: $-1 \le r_{xy} \le 1$.

Дві випадкові величини *X* і *У* називаються корельованими, якщо їх кореляційний момент (або те саме, що коефіцієнт кореляції) відмінний від нуля; *X* і *У* називаються некорельованими, якщо їхній кореляційний момент дорівнює 0.

II. *Приклади розв'язування задач*

Задача 1. Дискретна двовимірна випадкова величина задана таблицею.

$У \backslash X$	$x_1=1$	$x_2=3$	$x_3=4$	$x_4=8$
$y_1=3$	0,15	0,06	0,25	0,04
$y_2=6$	0,30	0,1	0,03	0,07

Знайти:

а) закони розподілу складових;

б) μ_{xy}, r_{xy};

в) умовні закони розподілу $У / X = x_2, X / У = y_1$, та $M(У / X = x_2), M(X / У = y_1)$.

Розв'язування

а)

X	1	3	4	8
P	0,45	0,16	0,28	0,11

$У$	3	6
P	0,5	0,5

б) *M(X)=2,93, M(У)=4,5, M(XУ)=13,185*

$$\mu_{xy} = \mathrm{M}(\mathrm{XУ}) - M(X) \cdot M(У) = 0$$

Отже, *X* і *У* незалежні випадкові величини. Тому $r_{xy} = 0$.

в) знайдемо умовний розподіл ймовірностей величини *У* при $X=x_2=3$:

$$P(y_1/x_2) = \frac{P(x_2, y_1)}{P(x_2)} = \frac{0{,}06}{0{,}16} = 0{,}375$$

$$P(y_2/x_2) = \frac{P(x_2, y_2)}{P(x_2)} = \frac{0{,}1}{0{,}16} = 0{,}625$$

Тоді математичне сподівання буде рівне:

$$M(У/X = x_2) = \sum_{j=1}^{2} y_j \cdot P(y_j/x_2) = 3 \cdot 0{,}375 + 6 \cdot 0{,}625 = 4{,}875$$

Аналогічно:

$$P(x_1/y_1) = \frac{P(x_1, y_1)}{P(y_1)} = \frac{0{,}15}{0{,}5} = 0{,}3$$

$$P(x_2/y_1) = \frac{P(x_2, y_1)}{P(y_1)} = \frac{0{,}06}{0{,}5} = 0{,}12$$

$$P(x_3/y_1) = \frac{P(x_3, y_1)}{P(y_1)} = \frac{0{,}25}{0{,}5} = 0{,}5$$

$$P(x_4/y_1) = \frac{P(x_4, y_1)}{P(y_1)} = \frac{0{,}04}{0{,}5} = 0{,}08$$

$$M(X/У = y_1) = \sum_{i=1}^{4} x_i \times P(x_i/y_1) = 1\times 0{,}3 + 3\times 0{,}12 + 4\times 0{,}5 + 8\times 0{,}08 = 3{,}3 .$$

III. Завдання для самостійного виконання

а) Запитання для самоконтролю:

1. Що таке двовимірні дискретні випадкові величини?
2. Як подається закон розподілу двомірної дискретної випадкової величини?
3. Як знайти безумовні та умовні закони розподілу дискретної двовимірної випадкової величини?
4. Написати формули для обчислення *M[x], D[x], M[y], D[y], M[x/y], M[y/x]*, $\sigma[x]$, $\sigma[y]$, μ_{xy}, r_{xy}.
5. Як впливає r_{xy} на зв'язок між *X* і *У*?

б) Даний закон системи двох дискретних випадкових величин

У	***X***			
	x_1	x_2	x_3	x_4
y_1	0,05	0,05	0,05	0,05
y_2	0,05	0,05	0,05	0,15
y_3	0,10	0,05	0,05	0,05
y_4	0,10	0,05	0,05	0,05

Обчислити $M[x]$, $D[x], \sigma\left[x\right]$, $M[y]$, $D[y], \sigma\left[y\right]$, μ_{xy} , r_{xy}. Чи будуть X і $У$ залежними? Побудувати умовні закони розподілу $У/_{X = x_1}$, $X/_{У = y_4}$ і знайти $M\left(У/_{X = x_1}\right), M\left(X/_{У = y_4}\right)$.

x_1, x_2, x_3, x_4 *та* y_1, y_2, y_3, y_4, – дані в таблиці 3.

Таблиці 3.

Варіант	Дані							
	x_1	x_2	x_3	x_4	y_1	y_2	y_3	y_4
1	1	2	3	4	-10	-9	-8	-7
2	5	6	7	8	-6	-5	-4	-3
3	9	10	11	12	-2	-1	0	1
4	5	10	15	20	-1	0	1	-1
5	4	6	18	10	-3	5	-7	6
6	1	3	5	7	-8	-6	-4	-2
7	2	4	6	8	-7	-5	-3	-1
8	3	5	7	9	-15	-13	-11	-9
9	4	8	12	16	-14	-12	-10	-8
10	5	7	9	11	-13	-11	-9	-7
11	6	8	10	12	-12	-9	-6	-3
12	7	9	11	13	-11	-7	-3	0
13	8	10	12	14	-10	-9	-8	-7
14	9	11	13	15	-9	-7	-5	-3
15	10	12	14	16	-8	-6	-4	-2

ЛАБОРАТОРНА РОБОТА 7

Тема: *Закон великих чисел*

I. Деякі теоретичні відомості

Основні терміни*: відхилення випадкової величини, ймовірність протилежної нерівності, середнє арифметичне випадкових величин, обмежені дисперсії.*

Основні означення, твердження і формули

Тв. 1. (нерівності Маркова). Якщо значення випадкової величини X невід'ємні, то ймовірність того, що вона прийме значення більше числа $\varepsilon>0$, не більше $\frac{a}{\varepsilon}$, тобто $P(X>\varepsilon)\le\frac{a}{\varepsilon}$, де $a=M(X)$, а $P(X\le\varepsilon)\ge 1-\frac{a}{\varepsilon}$.

Зауваження. Друга нерівність є наслідком рівності $P(X>\varepsilon)+P(X\le\varepsilon)=1$.

Відхиленням випадкової величини X від її математичного сподівання називається різниця X-a.

Тв. 2. (нерівності Чебишова). Для ймовірності абсолютного відхилення випадкової величини X виконуються нерівності:

$$P(|X-a|<\varepsilon)\ge 1-\frac{D(X)}{\varepsilon^2},$$

$$P(|X-a|\ge\varepsilon)\le\frac{D(X)}{\varepsilon^2}, де\ a=M(X)$$

Наслідок. Користуючись нерівністю Чебишова можна оцінити ймовірність того, що випадкова величина X відхилиться від свого математичного сподівання менше як на три середньоквадратичних відхилення: $P(|X-a|<3\sigma)\ge\frac{8}{9}=0,9973$.

Тв. 3. (теорема Чебишова). Якщо послідовність попарно-незалежних випадкових величин $X_1, X_2, ..., X_n, ...$ має математичне сподівання скінченне число і дисперсії цих величин рівномірно обмежені (не перевищують C=*const*), то середнє арифметичне випадкових величин сходиться по ймовірності до середнього арифметичного їх математичних сподівань, тобто для будь-якого $\varepsilon>0$:

$$\lim_{n\to\infty} P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n}-\frac{a_1+a_2+\ldots+a_n}{n}\right|<\varepsilon\right)=1,$$

$$\text{де } a_1=M(X_1), a_2=M(X_2), \ldots, a_n=M(X_n).$$

Наслідок. Якщо виконуються всі умови теореми Чебишова і $M(X_1)=M(X_2)=\ldots=M(X_n)=a$, то

$$\lim_{n\to\infty} P\left(\left|\frac{\sum\limits_{i=1}^{n} X_i}{n}-a\right|<\varepsilon\right)=1.$$

На теоремі Чебишова базується вибірковий метод у статистиці, суть якого полягає у тому, що за порівняно невеликою вибіркою роблять висновок про генеральну сукупність об'єктів.

Тв. 4. (теорема Бернуллі). Якщо в кожному із *n* незалежних випробувань ймовірність *p* появи події *A* постійна, то близька до одиниці ймовірність того, що відхилення відносної частоти від ймовірності *p* за абсолютною величиною буде як завгодно малим, якщо число випробувань достатньо велике. Тобто, якщо $\varepsilon>0$ мале число, то $\lim\limits_{n\to\infty} P\left(\left|\frac{m}{n}-p\right|<\varepsilon\right)=1$.

Зауваження. Із теореми Бернуллі не слідує $\lim\limits_{n\to\infty}\frac{m}{n}=p$ статистичне означення ймовірності.

II. Приклади розв'язування задач

Задача 1. Математичне сподівання швидкості вітру біля Землі в даному населеному пункті дорівнює 8 м/с. Оцінити ймовірність того, що в даному пункті швидкість вітру буде: а) більше 30 м/с; б) не перевищить 32 м/с.

Розв'язування. Нехай випадкова величина Х – швидкість вітру біля Землі в даному населеному пункті. Застосовуючи по черзі нерівності Маркова для а=8 м/с, ε=30, отримаємо:

а) $P(X>30)\le\frac{8}{30}\approx 0{,}267$;

б) $P(X \le 32) \ge 1 - \dfrac{8}{32} = 0{,}75$.

Задача 2. Вважаючи для спрощення підрахунків ймовірність народження хлопчика 0,5 оцінити за допомогою нерівності Чебишова ймовірність того, що серед 1200 новонароджених хлопчиків буде від 550 до 650 включно.

Розв'язування. Число хлопчиків серед 1200 новонароджених є випадковою величиною, розподіленою за біноміальним законом. Математичне сподівання і дисперсія її дорівнює

$$M(X) = np = 1200 \cdot 0{,}5 = 600, \text{ а } D(X) = npq = 1200 \cdot 0{,}5 \cdot 0{,}5 = 300.$$

Оскільки 500 та 650 межі допустимих значень випадкової величини – симетричні відносно математичного сподівання, рівного 600 (одна менша, а друга більша від нього на 50), нерівність 550≤X≤650 можна замінити еквівалентним $|X - 600| \le 50$. Тому згідно з нерівністю Чебишова при *ε=50* отримаємо:

$$P(550 \le X \le 650) = P(|X - 600| \le 50) \ge 1 - \frac{300}{50^2} = 0{,}88.$$

Отже, ймовірність шуканої події не менше 0,88.

Задача 3. Ймовірність появи події у кожному випробуванні постійна і рівна 0,3. Застосовуючи нерівність Чебишова, знайти число випробувань, необхідних для того, щоб ймовірність відхилення частоти події від її ймовірності була по абсолютній величині не більше 0,01, більше 0,99.

Розв'язування. Згідно з нерівністю $P\left(\left|\dfrac{m}{n} - p\right| \le \varepsilon\right) \ge 1 - \dfrac{pq}{n\varepsilon^2}$, яка є наслідком нерівності Чебишова для біноміального розподілу. Якщо при відомих значеннях *p, q і ε, n* буде випробувано таким, що різниця $1 - \dfrac{pq}{n\varepsilon^2}$ виявиться не менше даної в умові ймовірності *P*, то отримаємо нерівність $1 - \dfrac{pq}{n\varepsilon^2} \ge P$, яка і визначить шукане найменше число випробувань *n*.

Ця нерівність при *p=0,3, q=0,7, ε=0,01 і P=0,99* дає

$$1 - \frac{0{,}3 \times 0{,}7}{n \times 0{,}01^2} \ge 0{,}99 \text{ або } 0{,}01 \ge \frac{0{,}21}{n \times 0{,}01^2} \Rightarrow n \ge 210000.$$

Найменше число випробувань рівне 210 000.

Задача 4. Математичне сподівання швидкості вітру на даній висоті дорівнює 25 км/г, а середньоквадратичне відхилення дорівнює 4,5 км/г. Які швидкості вітру можна очікувати з ймовірністю, не меншою 0,9?

Розв'язування. Із нерівності Чебишова виходить, що якщо при даній дисперсії D(X) ε буде вибрано таким чином, що $1-\frac{D(X)}{\varepsilon^2}=0{,}9$, то умова задачі виконуватиметься. Отримаємо рівність, з якої знайдемо *ε*:

$$1-\frac{4{,}5^2}{\varepsilon^2}=0{,}9 \Rightarrow \varepsilon=14{,}2 \text{, або}$$

$$|X-25|\le 14{,}2 \text{, тобто } 10{,}8\le X\le 39{,}2 .$$

Задача 5. Дисперсія кожної із 3000 незалежних випадкових величин не перевищує 6. Оцінити ймовірність того, що відхилення середнього арифметичного цих випадкових величини від середнього арифметичного їх математичних сподівань не перевищить 0,3.

Розв'язування. Нерівність Чебишова для *n* незалежних випробувань має вигляд

$$P\left(\left|\frac{X_1+X_2+...+X_n}{n}-\frac{a_1+a_2+...+a_n}{n}\right|\le\varepsilon\right)\ge 1-\frac{C}{n\varepsilon^2} ,$$

де C – число, яким обмежена дисперсія.

Застосуємо цю нерівність при ε=0,3, C=6, n=3000, отримаємо:

$$P\left(\left|\frac{X_1+X_2+...+X_n}{n}-\frac{a_1+a_2+...+a_n}{n}\right|\le 0{,}3\right)\ge 1-\frac{6}{3\times 0{,}3^2}\approx \mathit{0{,}978} .$$

Отже, ймовірність шуканої події не менше 0,978.

Задача 6. Чи можна застосувати до послідовності випадкових величин X_1, X_2, ..., X_n, ... закон великих чисел, якщо випадкова величина X_n має розподіл:

а)

X_n	-100n	0	100n
P_n	$\frac{1}{2n^2}$	$1-\frac{1}{n^2}$	$\frac{1}{2n^2}$

б)

X_n	-2^n	2^n
P_n	$\frac{1}{2}$	$\frac{1}{2}$

Розв'язування. Перевіримо виконання умови теореми Чебишова про обмеженість дисперсій даних випадкових величин. Оскільки

а) $$M(X_n) = -100n \cdot \frac{1}{2n^2} + 0 \cdot \left(1 - \frac{1}{n^2}\right) + 100n \cdot \frac{1}{2n^2} = 0,$$ то

дисперсія $D(X_n) = M(X_n^2) = \frac{10000n^2}{2n^2} + 0^2 \cdot \left(1 - \frac{1}{n^2}\right) + \frac{10000n^2}{2n^2} = 10000.$

Отже, $M(X_n)=0$ і дисперсії рівні 10000, то до цієї послідовності можна застосувати закон великих чисел.

б) $M(X_n) = -2^n \cdot \frac{1}{2} + 2^n \cdot \frac{1}{2} = 0$, то дисперсія

$$D(X_n) = M(X_n^2) = (-2^n)^2 \frac{1}{2} + (2^n)^2 \frac{1}{2} = 2^{2n+1}.$$

Оскільки $D(X_n) \to \infty \ (n \to \infty)$, то до даної послідовності не можна застосовувати закон великих чисел (дисперсії необмежені).

III. Завдання для самостійного виконання

Завдання 1

1.1 Середня швидкість вітру біля Землі в даному населеному пункті дорівнює 16 км/г. Оцінити ймовірність того, що швидкість вітру не перевищуватиме 80 км/г.

1.2 Середня кількість студентської групи 28 осіб. Знайти ймовірність того, що у деякій групі не менше 32 студентів.

1.3 Середня кількість пасажирів, що проходять через станцію метрополітену становить 800 осіб. Оцінити ймовірність того, що в один день через цю станцію пройде не більше 1200 осіб.

1.4 Середня кількість студентів, що приймаються на перший курс в Академію праці і соціальних відносин становить 240 осіб. Використовуючи нерівність Маркова, оцінити ймовірність того, що в наступному році на перший курс буде прийнято не більше 260 студентів.

Завдання 2

2.1 Добова потреба електроенергії в даному населеному пункті є випадковою величиною, математичне сподівання якої дорівнює 2000 квт./год., а дисперсія становить 20 тис. Оцінити ймовірність того, що в найближчий день витрати електроенергії у цьому населеному пункті будуть від 1500 до 2500 квт/год.

2.2 Нехай ймовірність того, що складений годинник має хід у межах стандарту дорівнює 0,97. Оцінити ймовірність того, що серед 1000 годинників частина тих, які мають хід у межах норми відхилятиметься (по абсолютній величині) від ймовірності 0,97 не більше як на 0,02.

2.3 Середня кількість пасажирів тролейбуса становить 35 осіб, а дисперсія 1,4. Використовуючи нерівність Чебишова, оцінити ймовірність того, що в деякому тролейбусі пасажирів виявиться не менше 33 і не більше 37.

2.4 Середня кількість студентської групи становить 29 осіб, а дисперсія 0,8. Використовуючи нерівність Чебишова оцінити ймовірність того, що в деякій групі виявиться студентів не менше 27 і не більше 31.

Завадння3

3.1 Дисперсія кожної із 2500 незалежних випадкових величин не перевищує 5. Оцінити ймовірність того, що абсолютна величина відхилення середнього арифметичного цих випадкових величин від середнього арифметичного їх математичних сподівань не перевищує 0,4.

3.2 Оцінити ймовірність того, що абсолютна величина відхилення середнього зросту 1000 осіб від математичного сподівання величини, що виражає зріст кожного чоловіка, не перевищить 0,5 см, припускаючи, що середньоквадратичне відхилення кожної із цих випадкових величин дорівнює 2,5 см.

3.3 Чи можна застосувати до послідовності випадкових величин $X_1, X_2, \ldots, X_n, \ldots$ закон великих чисел, якщо випадкова величина X_n задана законом розподілу:

а)

X_n	-10n	0	10n
P_n	$\frac{1}{2^n}$	$1-\frac{1}{2^{n-1}}$	$\frac{1}{2^n}$

б)

X_n	-7^n	7^n
P_n	$\frac{1}{2}$	$\frac{1}{2}$

ЛАБОРАТОРНА РОБОТА 8

Тема: *Зведення та групування статистичних даних. Ряди розподілу*

I. Деякі теоретичні відомості

Основні терміни*: генеральна та вибіркова сукупність, статистичний розподіл вибірки, емпірична функція розподілу, полігон, гістограма, вибіркова середня та вибіркова дисперсія, міжгрупова, всередині групова, загальна медіана, мода, розмах, коефіцієнт варіації варіаційного ряду.*

Основні означення, твердження і формули

Вибірковою сукупністю або просто вибіркою називається сукупність випадково відібраних об’єктів.

Генеральною сукупністю називається сукупність об’єктів, з якої проводиться вибірка.

Об’ємом сукупності (вибіркової або генеральної) називають число об’єктів цієї сукупності.

Статистичним розподілом вибірки називають перелік варіант (послідовність інтервалів) та відповідних їм частот.

Емпіричною функцією розподілу називають функцію $F^*(x)$, яка визначає для кожного x відносну частоту події $X<x$. Тобто $F^*(x)=\frac{n_x}{n}$, де n_x – число варіант, менших n, n – об’єм вибірки.

Полігоном частот (відносних частот) називають ламану, відрізки якої сполучають точки $(x_1, n_1), (x_2, n_2), \ldots, (x_k, n_k)$ ($(x_1, W_1), (x_2, W_2), \ldots, (x_k, W_k)$, $W_i=\frac{n_i}{n}$ – відносна частота).

Гістограмою частот називають ступеневу фігуру, яка складається з прямокутників, основами яких є довжини інтервалів h, а висоти дорівнюють відношенню $\frac{n_i}{h}$.

Вибірковою середньою називають середнє арифметичне значень ознаки вибіркової сукупності:

а) якщо значення ознаки $x_1, x_2, \ldots, x_n$ різні, то

$$\bar{x}_в=\frac{x_1+x_2+\ldots+x_n}{n};$$

б) якщо варіанти $x_1, x_2, \ldots, x_k$ мають відповідні частоти $n_1, n_2, \ldots, n_k$, причому $n_1 + n_2, + \ldots + n_k = n$, то

$$\overline{x}_в = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_k x_k}{n} \text{ або } \overline{x}_в = \frac{\sum_{i=1}^{k} n_i x_i}{n}.$$

Аналогічні означення можна дати для даних генеральної сукупності.

Вибірковою дисперсією $Д_в$ називається середнє арифметичне квадратів відхилень спостережуваних значень ознаки від їх середнього значення $\overline{x}_в$ (якщо виконуються властивості з попереднього означення):

а) $$Д_в = \frac{\sum_{i=1}^{n} \left(x_i - \overline{x}_в\right)^2}{n};$$

б) $$Д_в = \frac{\sum_{i=1}^{k} n_i \left(x_i - \overline{x}_в\right)^2}{n}.$$

Вибірковим середнім квадратичним відхиленням називається квадратний корінь із вибіркової дисперсії: $\sigma_в = \sqrt{Д_в}$.

Тв. 1. Дисперсія дорівнює середньому квадратів значень ознаки без квадрата загальної середньої: $Д = \overline{x^2} - \left(\overline{x}\right)^2$.

Всерединігруповою дисперсією називається середнє арифметичне групових дисперсій зважену за об’ємами груп:

$$Д_{всгр} = \frac{\sum N_j D_{ігр}}{n}, \text{ де } N_j \text{ – об’єм групи } j.$$

Міжгруповою дисперсією називають дисперсію групових середніх відносно загальної середньої:

$$Д_{міжгр} = \frac{\sum N_j \left(\overline{x}_j - \overline{x}\right)^2}{n}, \text{ де } \overline{x}_j \text{ – групова середня групи j;}$$

N_j – об’єм групи j;

$\overline{x}$ – загальна середня;

$$n = \sum_{j=1}^{k} N_j \ \text{– об'єм усієї сукупності.}$$

Тв. 2. Якщо сукупність складається із кількох груп, то загальна дисперсія дорівнює сумі всерединігрупової та міжгрупової дисперсії:

$$Д_{заг} = Д_{всгр} + Д_{міжгр}.$$

Модою варіаційного ряду називають варіанту, яка має найбільшу частоту.

Медіаною називають варіанту, яка ділить ряд на дві частини, рівні за кількістю варіант. Якщо кількість варіант парна *n=2k*, то

$$m_e = \frac{x_{k+1} + x_k}{2}.$$

Розмахом варіації *R* називають різницю між найбільшою і найменшою варіантами.

Коефіцієнтом варіації *V* називають виражене в % відношення вибіркового середньоквадратичного відхилення до вибіркової середньої:

$$V = \frac{\sigma_в}{\bar{x}_в} \cdot 100\%.$$

Коефіцієнт варіації використовують для порівняння величини розсіяння щодо вибіркової середньої двох варіаційних рядів.

II. Приклади розв'язування задач

Задача 1. Нехай маємо умовні дані про робітників підприємства:

Стаж роботи, років	1	2	1	1	2	3	2	2	3	6
Вік, років	23	23	25	25	26	29	30	30	30	33
Стаж роботи, років	2	5	1	4	6	6	3	3	3	4
Вік, років	34	34	35	35	35	35	36	36	36	36

а) Згрупувати робочих за двома заданими ознаками, побудувавши для одної ознаки дискретний ряд, а для другої інтервальний.

б) Побудувати для одного розподілу полігон, а для другого – гістограму частот, та функцію розподілу.

в) Для одного із побудованих розподілів обчислити характеристики розподілу:

1) показники центру розподілу: середнє, моду, медіану;
2) показники варіації: розмах варіації, середньоквадратичне відхилення, дисперсію, коефіцієнт варіації.

Розв'язування

а) Побудуємо варіаційний ряд за ознакою стажу роботи та інтервальний ряд за ознакою – вік. Варіантами першого ряду буде стаж роботи – у роках.

Стаж роботи, років (x)	1	2	3	4	5	6
Частота (n_i)	4	5	5	2	1	3

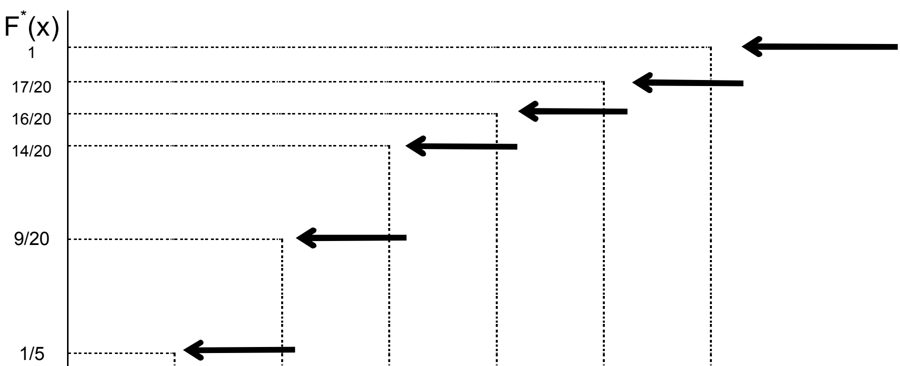
Для побудови інтервального ряду вік розіб'ємо на п'ять рівних інтервалів. Отримаємо:

Вік $x_{i-1} - x_i$	23-26	26-29	29-32	32-35	35-38
Кількість робітників з віком у даному інтервалі	4	1	4	3	8

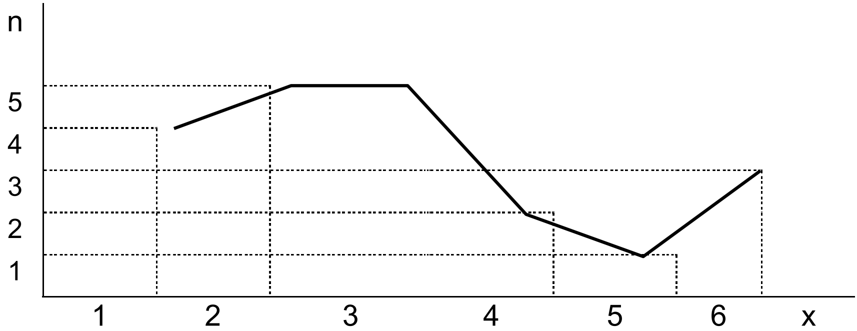
б) Побудуємо функцію розподілу для першого розподілу:

$$F^*(x) = \begin{cases} 0, \ \text{при} \ \le 1, \\ \frac{1}{5}, \ \text{при} \ 1 < x \le 2, \\ \frac{9}{20}, \ \text{при} \ 2 < x \le 3, \\ \frac{14}{20}, \ \text{при} \ 3 < x \le 4, \\ \frac{16}{20}, \ \text{при} \ 4 < x \le 5, \\ \frac{17}{20}, \ \text{при} \ 5 < x \le 6, \\ 1, \ \text{при} \ x > 6, \end{cases}$$

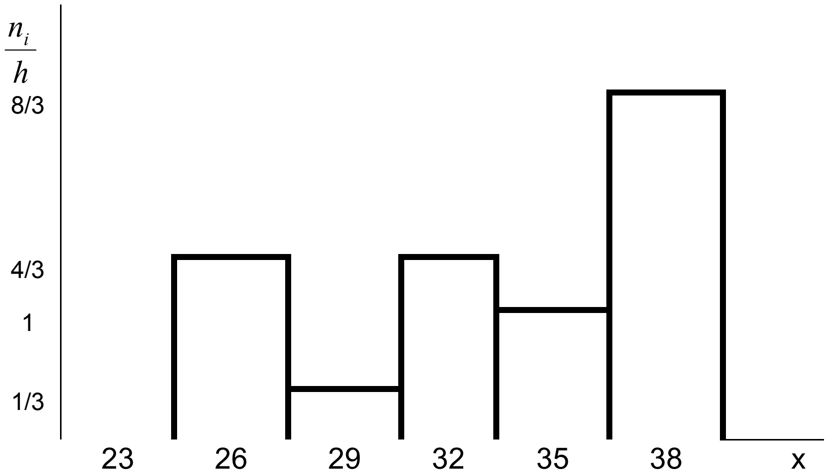
та її графік



Побудуємо полігон частот для першого розподілу та гістограму частот для другої. Для цього в системі координат нанесемо точки (1,4), (2,5), (3,5), (4,2), (5,1), (6,3) та сполучимо їх ламаною лінією.



Частинні інтервали довжини h=5	23-26	26-29	29-32	32-35	35-38
Сума частот варіант даного інтервалу n_i	4	1	4	3	8
Густина частоти $\frac{n_i}{h}$	4/3	1/3	4/3	1	8/3



в)

1) Обчислимо показники центру розподілу для першого ряду:

середнє $\bar{x}=\frac{1\cdot4+2\cdot5+3\cdot5+4\cdot2+5\cdot1+6\cdot3}{20}=\frac{60}{20}=3$;

мода $Mo=\frac{2+3}{2}=2{,}5$ (тому, що два варіанти мають найбільшу частоту, рівну 5); медіана $Me=\frac{3+4}{2}=3{,}5$ (тому, що в ряду парна кількість варіант).

2) розмах варіації $R=6\text{-}1=5$, дисперсія:

$$\overline{x^2}=\frac{1^2\cdot4+2^2\cdot5+3^2\cdot5+4^2\cdot2+5^2\cdot1+6^2\cdot3}{20}=\frac{234}{20}=11{,}70;$$

$$Д(x)=\overline{x^2}-(\bar{x})^2=11{,}7-9=2{,}7;$$

середньоквадратичне відхилення $\sigma(x)=\sqrt{Д(x)}=\sqrt{2{,}7}\approx1{,}64$;

коефіцієнт варіації $V=\frac{1{,}64}{3}\cdot100\%\approx54{,}67\%$.

III. Завдання для самостійного виконання

Нехай маємо умовні дані про робітників підприємства (у стовпчику стать 1 – це чоловіча і 2 – жіноча).

№ з/п	Стаж роботи, р.	Вік, р.	Роз ряд	Зарплата, грн.	Стать	№ з/п	Стаж роботи, р.	Вік, р.	Роз ряд	Зарплата, грн.	Стать
1	4	36	3	160	1	55	6	52	5	210	1
2	5	36	3	180	1	56	6	52	4	210	1
3	4	36	4	160	1	57	6	52	4	210	1
4	6	36	4	210	1	58	9	52	4	300	1
5	3	37	2	130	1	59	9	53	6	300	1
6	4	37	4	160	1	60	10	54	6	300	1
7	3	38	2	130	1	61	10	55	6	300	1
8	3	38	3	130	1	62	10	55	6	300	1
9	3	38	4	140	1	63	9	56	6	300	1
10	4	38	4	160	1	64	10	58	6	300	1
11	4	39	3	160	1	65	10	59	6	300	1
12	4	39	3	160	1	66	1	22	6	300	1
13	4	39	3	160	1	67	1	24	1	110	1
14	5	39	3	180	1	68	1	26	1	110	1
15	4	39	4	160	1	69	4	28	1	110	1
16	4	39	4	160	1	70	2	28	2	150	1
17	4	39	4	160	1	71	2	29	3	130	1
18	4	40	4	170	1	72	3	30	2	130	1
19	2	40	3	130	1	73	2	32	3	130	1
20	3	40	3	130	1	74	1	33	3	130	1
21	4	40	3	160	1	75	2	33	1	110	1
22	4	40	4	160	1	76	3	35	3	130	1

№ з/п	Стаж роботи, р.	Вік, р.	Роз ряд	Зарплата, грн.	Стать	№ з/п	Стаж роботи, р.	Вік, р.	Роз ряд	Зарплата, грн.	Стать
23	4	40	4	170	1	77	3	35	3	130	1
24	5	40	4	190	1	78	5	35	3	140	1
25	5	40	4	190	1	79	5	35	3	180	1
26	5	40	4	190	1	80	3	35	3	180	1
27	5	41	4	190	1	81	3	35	4	140	2
28	3	41	4	140	1	82	5	35	4	140	2
29	4	41	4	160	1	83	5	36	4	190	2
30	4	41	4	170	1	84	3	37	4	190	2
31	4	41	4	170	1	85	3	37	3	130	2
32	4	41	4	170	1	86	5	37	3	130	2
33	4	41	4	170	1	87	5	38	4	190	2
34	4	41	4	170	1	88	3	39	3	180	2
35	5	42	4	190	1	89	5	39	2	130	2
36	2	42	3	130	1	90	4	41	3	180	2
37	5	42	4	190	1	91	4	41	4	160	2
38	5	43	4	190	1	92	4	41	4	160	2
39	5	44	3	180	1	93	4	41	4	170	2
40	5	44	3	190	1	94	4	41	4	170	2
41	5	44	3	190	1	95	4	42	5	180	2
42	5	44	3	180	1	96	4	42	4	160	2
43	3	44	4	140	1	97	4	42	4	160	2
44	5	44	4	190	1	98	4	42	4	160	2
45	5	44	4	190	1	99	5	42	4	160	2
46	5	44	4	190	1	100	5	42	4	190	2
47	5	45	4	190	1	101	5	42	4	190	2
48	5	47	4	190	1	102	5	42	4	190	2
49	6	48	4	210	1	103	5	42	4	190	2
50	6	49	5	230	1	104	5	42	4	190	2
51	6	50	4	210	1	105	5	42	4	190	2
52	6	51	5	230	1	106	5	43	4	190	2
53	6	51	5	230	1	107	5	44	5	180	2
54	6	51	5	230	1	108	5	44	4	190	2

Варіант	**рядки**	**стовпці**
1	1-27	2, 3
2	1-27	4, 5
3	1-27	5, 10
4	1-27	3, 5
5	28-54	2, 4
6	28-54	2, 5
7	28-54	3, 5
8	28-54	3, 11
9	55-81	8, 11
10	55-81	9, 11
11	55-81	10, 11
12	55-81	4, 11
13	82-108	8, 10
14	82-108	9, 10

15	82-108	10, 11

Завдання

1) Згрупувати робочих по двох заданих ознаках, побудувавши для однієї ознаки дискретний ряд, а для другої інтервальний.
2) Побудувати для дискретного розподілу полігон частот та функцію розподілу, а для інтервального ряду – гістограму частот.
3) Для дискретного розподілу обчислити характеристики:
   а) показники центру розподілу: середнє, моду, медіану;
   б) показники варіації: розмах варіації, середньоквадратичне відхилення, дисперсію, коефіцієнт варіації.

ЛАБОРАТОРНА РОБОТА 9

Тема: Точкові та інтервальні статистичні оцінки параметрів генеральної сукупності нормального розподілу

I. Деякі теоретичні відомості

***Основні терміни**: точкові та інтервальні оцінки, виправлена дисперсія, надійність, довірча ймовірність.*

Основні означення, твердження і формули

Точковою називають оцінку, яка визначається одним числом. До точкових оцінок належать середня вибіркова, вибіркова дисперсія, вибіркове середньоквадратичне відхилення.

Для оцінки дисперсії генеральної сукупності використовують виправлену дисперсію $S^2 = \frac{\sum_{i=1}^{k}\left(x_i - \overline{x}\right)^2 n_i}{n-1}$, яка відрізняється від звичайної дисперсії лише знаменником. Для великих значень n вона не відрізняється від вибіркової дисперсії. На практиці користуються виправленою дисперсією, якщо $n<30$.

Інтервальною називають оцінку, яка визначається двома числами – кінцями інтервалу. Інтервальні оцінки дають можливість встановити точність і надійність оцінок.

Точністю оцінки називають додатнє число $\delta>0$, для якого виконується нерівність $\left|\theta - \theta^*\right| < \delta$, де θ^* - оцінка невідомого параметра θ. Але статистичні методи не можуть категорично стверджувати, що виконується нерівність; можна лише говорити про ймовірність γ, з якою ця нерівність виконується.

Надійністю (довірчого ймовірністю) оцінки θ по θ^* називають ймовірність γ, з якою виконується нерівність $\left|\theta - \theta^*\right| < \delta$.

Довірчим називають інтервал $\left(\theta^* - \delta; \theta^* + \delta\right)$, який покриває невідомий параметр із заданою надійністю.

Довірчий інтервал для оцінки математичного сподівання нормального розподілу при відомому σ має вигляд:

$\left(\bar{x} - t\frac{\sigma}{\sqrt{n}}; \bar{x} + t\frac{\sigma}{\sqrt{n}}\right)$, де $\delta = t\frac{\sigma}{\sqrt{n}}$ – точність оцінки. Число t визначається із співвідношення $2\Phi(t) = \gamma$, або $\Phi(t) = \frac{\gamma}{2}$ (γ – надійність). Із формули, яка визначає точність вказаної оцінки, можна дійти висновків:

1) при зростанні об'єму вибірки n число δ спадає, а отже, точність оцінки збільшується;
2) збільшення надійності оцінки $\gamma = 2\Phi(t)$ приводить до збільшення t, а отже, і до зростання δ (іншими словами, збільшення надійності класичної оцінки тягне за собою зменшення її точності).

Довірчий інтервал для оцінки математичного сподівання нормального розподілу при невідомому σ має вигляд (з надійністю γ):

$\bar{x} - t_\gamma \frac{S}{\sqrt{n}} < a < \bar{x} + t_\gamma \frac{S}{\sqrt{n}}$, де S – зміщене середньоквадратичне відхилення, $t_\gamma = t(\gamma, n)$ – знаходимо з таблиць, n – об'єм вибірки, $\bar{x}$ – вибіркова середня.

Довірчі інтервали для оцінки середньоквадратичного відхилення σ нормального розподілу мають вигляд:

$S(1-q) < \sigma < S(1+q)$, якщо $q<1$,

$0 < \sigma < S(1+q)$, якщо $q>1$,

де $q=q(n,\gamma)$ – знаходимо з таблиць.

У теорії помилок прийнято точність вимірів характеризувати за допомогою середньоквадратичного відхилення σ випадкових помилок вимірювань. Для оцінки σ використовують "виправлене" середньоквадратичне відхилення S.

II. Приклади розв'язування задач

Задача 1. Знайти вибіркову і виправлену дисперсію варіаційного ряду, отриманого у результаті соціологічного опитування:

варіанта	1	2	5	8	9
частота	3	4	6	4	3

Розв'язування. Знайдемо спочатку середню вибіркову:

$$\bar{x}_в = \frac{3\cdot1+4\cdot2+6\cdot5+4\cdot8+3\cdot9}{3+4+6+4+3} = \frac{100}{20} = 5,$$

тоді дисперсія:

$$Д_в = \frac{\sum n_i(x_i-\bar{x}_в)^2}{n} = \frac{3\cdot(1-5)^2+4\cdot(2-5)^2}{20} +$$

$$+\frac{6\cdot(5-5)^2+4\cdot(8-5)^2+3\cdot(9-5)^2}{20} = 8{,}4$$

Виправлена дисперсія:

$$S^2 = \frac{\sum n_i(x_i-\bar{x}_в)^2}{n-1} = \frac{3\cdot(1-5)^2+4\cdot(2-5)^2}{19} +$$

$$+\frac{6\cdot(5-5)^2+4\cdot(8-5)^2+3\cdot(9-5)^2}{19} = 8{,}84.$$

Задача 2. У результаті соціологічного дослідження отримали вибірку об'єму $n=10$, яка має нормальний розподіл та для неї підрахували $\sigma=2$, $\bar{x}_в=5{,}4$. Знайти довірчий інтервал для оцінки невідомого математичного сподівання генеральної сукупності з надійністю $\gamma=0{,}95$.

Розв'язування. Використаємо формулу для оцінки математичного сподівання при відомому σ:

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

З таблиць знаходимо $\Phi(t)=0{,}475$, $t=1{,}96$.

Тоді $\bar{x} - t\dfrac{\sigma}{\sqrt{n}} = 5{,}4 - \dfrac{1{,}96\cdot2}{\sqrt{10}} \approx 4{,}16$,

$$\bar{x}_в + t\frac{\sigma}{\sqrt{n}} = 5{,}4 + \frac{1{,}96\cdot2}{\sqrt{10}} \approx 6{,}64.$$

Відповідь: $P(4{,}16 < a < 6{,}64) = 0{,}95.$

III. Завдання для самостійного виконання

Досліджується якісний показник успішності студентів певної групи (відсоток позитивних оцінок від загальної кількості оцінок). Для цього було проведено зріз (атестацію) 212 студента по предметах, які виносяться на заліки та екзамени, та було підраховано якість для кожного студента. Дані подані нижче (у %):

$$70\frac{N}{N+1};72\frac{N}{N+1};75\frac{N}{N+1};75\frac{N}{N+1};77\frac{N}{N+1};78\frac{N}{N+1};80\frac{N}{N+1};$$

$$81\frac{N}{N+1};82\frac{N}{N+1};82\frac{N}{N+1};83\frac{N}{N+1};83\frac{N}{N+1};83\frac{N}{N+1};84\frac{N}{N+1};$$

$$84{,}5\frac{N}{N+1};85\frac{N}{N+1};85\frac{N}{N+1};85\frac{N}{N+1};87\frac{N}{N+1};89\frac{N}{N+1};98\frac{N}{N+1},$$

N – номер варіанту.

Необхідно:

а) згрупувати отримані дані в сім рівних інтервалів;

б) побудувати гістограму частот;

в) з надійністю $\gamma=0{,}95$ побудувати довірчі інтервали для оцінки математичного сподівання a і середньоквадратичного відхилення σ відсоткового показника якості для факультету, вважаючи, що він розподілений за нормальним законом.

ЛАБОРАТОРНА РОБОТА 10

Тема : Перевірка гіпотези про нормальний розподіл генеральної сукупності за критерієм Пірсона

I. Деякі теоретичні відомості

Основні терміни: *варіаційний ряд, інтервальний варіаційний ряд, рівень значущості, генеральна сукупність, нормальний розподіл, локальна та інтегральна функції Лапласа.*

Нехай вибірка задана варіаційним рядом з рівновіддаленими варіантами:

1)

Таблиця 10.1

x_i	x_1	x_2	…..	$x_к$
n_i	n_1	n_2	…..	n_k

Або інтервальним варіаційним рядом:

2)

Таблиця 10.2

x_i-x_{i+1}	x_1-x_2	x_2-x_3	…..	$x_к$-$x_{к+1}$
n_i	n_1	n_2	…..	n_k

Потрібно з рівнем значущості α перевірити гіпотезу про те, що генеральна сукупність X має нормальній розподіл. Для цього необхідно:

1. Обчислити теоретичні частоти $n_i^`$ для варіант вибірки. Згідно з класичним означенням ймовірності $P_i=\dfrac{n_i^`}{n}$, $n_i^`=p_i n$, i=1,2,….,к.

Отже, для знаходження теоретичних частот потрібно знайти ймовірності: $P_i=P(x=x_i)$ або $p_i=P(x_i<x<x_i+1)$ відповідно для таблиці 10.2.

Ймовірність $p_к=P(x=x_к)$ для таблиці 10.1 можна знайти, використовуючи локальну функцію Лапласа φ(x) та дані вибірки за формулою:

$$P_i = P(x=x_i) = \frac{h}{\sigma_в}\varphi(u_i) \qquad (10.1)$$

де $h = x_{i+1} - x_i$; $u_i = \frac{x_i - \overline{x_в}}{\sigma_в}$; $\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$ варіанти x_i рівновіддалені.

Ймовірність $p_i = P(x_i < x < x_{i+1})$ для таблиці 10.2. можна знайти, якщо використати інтегральну функцію Лапласа $Ф(x)$ за формулою:

$$Pi = P(x_i < x < x_{i+1}) = Ф\left(\frac{x_{i+1} - \overline{x_в}}{\sigma_в}\right) - Ф\left(\frac{x_i - \overline{x_в}}{\sigma_в}\right), \qquad (10.2)$$

$\overline{x_в}$ і $\sigma_в$ у формулі (10.1) обчислені за табл. 10.1, а у (10.2) – за табл. 10.2.

2. Обчислити розрахункове значення критерію χ^2 за формулою:

$$\chi^2_{розр} = \sum_{i=1}^{к} \frac{(n_k - n_k^`)^2}{n_k^`}$$

3. За таблицею критичних точок розподілу χ^2, за заданим рівнем значущості α і числом ступенів волі m=к–3 (к – число груп вибірки) знаходимо критичну точку $\chi^2_{кр}(\alpha, m)$ правосторонньої критичної області, або функція ХИ2ОБР.

Якщо $\chi^2_{розр} < \chi^2_{кр}$ – то X має нормальний розподіл, у протилежному випадку – ні. Іншими словами: емпіричні та теоретичні частоти відрізняються суттєво ($\chi^2_{розр} > \chi^2_{кр}$), або несуттєво ($\chi^2_{розр} < \chi^2_{кр}$).

Зауваження: Малочисельні частоти (n_i <5) потрібно об’єднати; при цьому і відповідні теоретичні частоти додаються. При визначенні числа ступенів волі потрібно враховувати, що число груп зменшилося. Для інтервального розподілу ця процедура аналогічна.

II. Приклади розв’язування задач

Приклад 10.1. Для даних про відсоток вільних місць у готелі протягом 100 днів (у %) побудувати інтервальний варіаційний ряд з довжиною інтервала 3% та використовуючи критерій Пірсона при рівні значущості α=0,05, встановити, чи узгоджується з даними цієї вибірки гіпотеза про те, що випадкова величина *X* розподілена за нормальним законом.

Таблиця 10.3.

1	3	27	18	13	20	17	18	10	6
21	13	15	22	16	13	20	15	16	23
13	4	14	10	19	12	17	9	14	13
19	18	7	19	30	25	6	16	17	9
10	21	17	14	5	11	16	12	15	16
13	14	12	24	17	21	23	18	17	22
17	3	15	16	21	27	4	6	14	10
19	15	20	7	15	11	30	13	12	20
16	10	18	22	12	23	8	21	19	20
8	14	13	911	14	24	15	12	16	17

Розв'язування. Згрупуємо дані в інтервальний ряд з кроком h=3.

Таблиця 10.4

x_i-x_{i+1}	1-4	4-7	7-10	10-13	13-16	16-19	19-22	22-25	25-30
n_i	5	6	10	17	21	17	13	6	5

Дві останні частоти об'єднали так, як вони були менші п'яти. Всі інші обчислення зведемо в табл. 10.5

Таблиця 10.5

x_i-x_{i+1}	n_i	$x_{i+\frac{1}{2}}$	$x_i n_i$	x_i	$\frac{x_i-\bar{x}}{\sigma_в}=z_i$	$Ф=(z_i)$	$Ф=(z_{i+1})$-$Ф=(z_i)$	$n_i^{`}=p_i n$	$\frac{(n_i-n_i^{`})^2}{n_i^{`}}$
				1	-2,268	-0,4884			
1-4	5	2,5	12,5	4	-1,772	-0,4616	0,0268	2,68	2,0084
4-7	6	5,5	33	7	-1,277	-0,3997	0,0619	6,19	0,0058
7-10	10	8,5	85	10	-0781	-0,2823	0,1174	11,74	0,2579
10-13	17	11,5	195,5	13	-0,286	-0,1141	0,1682	16,82	0,0019
13-16	21	14,5	304,5	16	0,21	+0,0832	0,1973	19,73	0,0817
16-19	17	17,5	297,5	19	0,705	+0,2611	0,1779	07,79	0,0351
19-22	13	20,5	266,5	22	1,201	0,3849	0,1238	12,38	0,0311
22-25	6	23,5	141	25	1,696	0,4554	0,0705	7,05	0,1564
25-30	5	27,5	137,5	30	2,522	0,4941	0,0387	3,87	0,3299
	100		1473						2,9082

Для даної вибірки :

$$\bar{x} = \frac{\sum x_i n_i}{n} = \frac{1473}{100} = 14{,}73, \quad \text{D(x)} = 36{,}6571;$$

σ (x) = $\sqrt{D(x)} = \sqrt{36{,}6571} \approx 6{,}05$.

Використовуючи ці дані та таблиці для

$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz$ із табл. 10.5 отримаємо: $\chi^2_{розр}$=2,9082

Із таблиць $\chi^2_{кр.} = \chi^2(0{,}05;6) \approx 12{,}6$ $\chi^2_{розр} < \chi^2_{табл}$, отже *X* має нормальний розподіл.

Приклад 10.2. Для даних за кількістю народжених чотирьох близнюків протягом 100 років необхідно побудувати варіаційний ряд та встановити за допомогою критерію Пірсона (при рівні значущості α=0,05) чи узгоджується з даними цієї вибірки гіпотеза про те, що випадкова величина X матиме нормальний розподіл.

Таблиця 10.6

0	1	1	0	1	1	0	0	0	1
2	1	1	4	1	6	1	1	0	5
2	0	2	1	2	0	1	0	3	1
2	3	2	2	1	3	2	1	1	2
2	2	2	2	1	1	1	5	1	1
0	0	0	1	0	1	0	1	1	1
0	2	1	1	3	0	3	4	0	3
0	3	1	1	1	0	3	0	0	1
0	2	2	2	1	0	3	1	0	2
3	2	2	2	4	3	1	1	4	2

Розв'язування: Побудуємо дискретний розподіл і оскільки для значень 5 і 6 частоти малі, то об'єднаємо ці значення з 4. Отримаємо розподіл:

Таблиця 10.7

x_i	0	1	2	3	4
n_i	24	36	22	11	7

$$\bar{x} = \frac{\sum x_i n_i}{n} = \frac{141}{100} = 1{,}41, \quad \mathrm{D(x)} = \overline{x^2} - (\bar{x})^2 = 1{,}3619$$

$$\sigma(x) = \sqrt{D(x)} = \sqrt{1{,}3619} \approx 1{,}17 .$$

Всі інші обчислення зведемо в таблицю 10.8.

Таблиця 10.8

№ з/п	x_i	n_i	$u_i = \frac{x_i - \bar{x}}{\sigma}$	$\varphi(u_i)$	$\frac{h\varphi(u_i)}{\sigma} = p_i$	$n_i^` = np_i$	$\frac{(n_i - n_i^`)^2}{n_i^`}$
1	0	24	-1,2082	0,1923	0,1648	16,476	3,436
2	1	36	-0,3513	0,3751	0,3214	32,139	0,464
3	2	22	0,5056	0,3511	0,3008	30,084	2,172
4	3	11	1,3625	0,1577	0,1351	13,513	0,467
5	4	7	2,2194	0,034	0,0291	2,9125	5,736

Із таблиці 10.8 знаходимо суму останнього стовпця, отримаємо $\chi^2_{розр} = 12{,}28$ $\chi^2_{кр} = \chi^2(0{,}05;5-3) = 5{,}99$.

Оскільки $\chi^2_{розр} > \chi^2_{кр}$, то X не буде нормально розподілена.

ІІІ. Завдання для самостійного виконання

а) Запитання для самоконтролю:

1. Як задається варіаційний ряд ?
2. Що таке рівень значущості (довіри) гіпотези?
3. Що таке генеральна та вибіркова сукупність ?
4. Який розподіл називається нормальним ?
5. Як перевірити гіпотезу про нормальний розподіл генеральної сукупності за критерієм Пірсона.

б) ***Завдання 1-5.***

Наведені нижче дані за кількістю народження трьох близнюків протягом 50 років. Нееобхідно:

1) побудувати варіаційний ряд;

2) знайти емпіричну функцію розподілу випадкової величини X={кількість народжених трьох близнюків} і побудувати її графік.
3) побудувати полігон частот та відносних частот;
4) знайти вібіркову середню, вибіркову дисперсію та вибіркове середнє квадратичне відхилення;
5) використовуючи критерії Пірсона при рівні значущості 0, 0N, встановити, чи узгоджується з даними цієї вибірки гіпотеза про те, що випадкова величина X розподілена за нормальним законом.

1.

1	1	2	1	1	2	2	1	1	2
1	2	3	2	2	5	2	2	1	4
1	3	2	1	1	2	4	4	1	3
2	2	2	2	3	4	2	1	2	1
0	1	0	0	2	3	1	1	4	1

2.

0	0	1	1	0	0	0	1	1	1
3	0	6	0	1	3	1	4	5	3
1	3	0	0	1	3	0	0	1	1
1	1	3	0	2	3	1	1	2	2
4	1	1	3	1	1	5	1	1	2

3.

3	5	2	2	4	6	2	3	3	1
0	5	5	2	3	3	1	1	1	2
2	5	0	3	0	2	2	6	2	4
1	4	1	2	1	2	1	0	1	0
3	2	3	2	3	4	3	4	3	0

4.

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

5.

2	2	3	4	5	1	1	4	4	2
1	0	2	3	5	2	3	3	4	0
2	2	2	1	0	3	4	0	2	4
2	1	2	2	1	1	4	1	1	2

1	2	2	1	1	2	1	1	0	1

Завдання 6-10.

Нижче наведені дані про відсоток вільних місць у готелі, що досліджувались протягом 50 днів. Використовуючи ці дані необхідно:

1) побудувати інтервальний варіаційний ряд з довжиною інтервала 4%;
2) знайти емпіричну функцію розподілу випадкової величини X={відсоток вільних місць у готелі} та побудувати її графік.
3) побудувати полігон частот і гістограму вибірки;
4) знайти вибіркову середню, вибіркову дисперсію та вибіркове середньоквадратичне відхилення;
5) використовуючи критерій Пірсона при рівні значущості 0, 0N, встановити, чи узгоджується з даними цієї вибірки гіпотеза про те, що випадкова величина X розподілена за нормальним законом.

6.

16	20	21	22	16	17	4	10	11	15
7	18	23	12	16	9	16	23	8	28
0	19	23	15	12	15	27	13	21	6
17	22	24	16	10	17	9	5	22	7
20	1	11	25	4	26	14	28	17	11

7.

18	21	16	24	13	26	10	27	10	23
20	14	2	20	3	13	18	29	6	12
19	14	12	4	8	17	14	16	21	15
15	25	13	3	13	5	18	30	19	18
7	17	7	23	20	16	25	9	19	18

8.

0	13	27	29	19	11	20	6	21	12
12	3	28	18	10	8	5	9	10	19
14	1	6	16	7	4	26	12	11	21
20	15	2	3	30	25	22	13	14	15
16	4	23	24	9	12	8	7	13	20

9.

21	5	6	23	11	15	14	14	19	16
21	22	11	10	10	26	18	18	17	13
20	12	7	24	19	16	17	15	17	16
13	22	15	9	23	16	17	24	15	25
17	14	8	15	25	16	18	11	18	17

10.

1	16	16	2	16	6	15	14	19	15
3	10	25	26	27	6	7	17	14	18
7	2	25	17	5	7	20	8	10	13
19	13	16	4	23	9	11	9	30	21
14	15	3	24	10	8	9	29	13	15

Завдання 11-15.

Нижче у випадковому порядку наведені дані про заробітну платну (в у.о.) 50 працівників підприємства. Виконати всі пункти завдань 6-10, вибравши довжину інтервала h=6.

11.

140	129	114	138	140	126	93	114	126	81
163	117	158	125	129	151	171	152	139	99
137	143	106	140	137	116	129	108	124	105
122	132	97	123	112	120	122	145	136	169
122	117	142	133	119	144	101	148	128	124

12.

124	108	90	83	86	125	170	138	100	80
138	144	137	111	128	163	109	100	125	160
165	156	152	115	104	87	111	130	99	109
88	94	92	132	125	111	107	131	124	162
152	107	134	157	101	112	150	102	82	113

13.

155	147	101	163	157	120	120	121	112	158
119	132	125	129	148	122	129	141	119	133
144	131	101	92	91	116	108	111	125	100
166	139	128	142	88	119	122	137	102	121
141	125	133	160	164	113	121	109	85	156

14.

135	113	126	156	144	132	130	116	79	146
117	124	90	133	140	114	145	107	132	112
138	84	106	136	123	102	116	98	150	78
114	93	126	96	118	158	131	113	127	115
90	108	130	117	162	146	122	145	141	112

15.

130	165	119	133	127	96	116	106	144	142
115	130	113	137	112	163	92	151	124	118
148	149	124	150	140	100	148	125	95	108
124	111	104	107	110	155	89	117	101	119

163	101	153	100	135	162	110	130	138	90

ЛАБОРАТОРНА РОБОТА 11

Тема: Коефіцієнти парної кореляції

I. Деякі теоретичні відомості

Основні терміни: *статистичний зв'язок, коваріація, коефіцієнт парної кореляції, прямий та обернений зв’язок, лінійна залежність, статистична значущість коефіцієнта кореляції.*

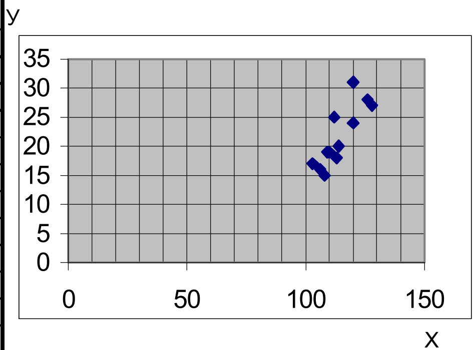
Часто виникають задачі з практики: як пов’язані між собою дві змінні у вибірковій сукупності? Наприклад: чи мають учні, які навчилися раніше читати, тенденцію до більш високої успішності? Щоб відповісти на це запитання треба провести спостереження по кожній змінній.

Поняття статичного зв’язку

Розглянемо дані, зібрані в результаті відповіді на кілька запитань (оцінка IQ Стенфорда-Біне), та оцінка тесту з хімії

(35 запитань).

№ п/п	Оцінка IQ Стенфорда-Біне (x)	Оцінка тесту з хімії (y)
1	120	31
2	112	25
3	110	19
4	120	24
5	103	17
6	126	28
7	113	18
8	114	20
9	106	16
10	108	15
11	128	27
12	109	19
Середнє значення	21,6	114,1



Щоб визначити тип зв’язку, зобразимо дані діаграмою розсіяння, відклавши по осі Ох оцінку IQ, а по Оу – оцінку тесту з хімії.

Із діаграми розсіяння видно, що узагальненої міри зв’язку немає. Положення об’єкта відносно інших у вибірці по Х і У проявляється у знаках відхилень $(x_i-\bar{x})$ і $(y_i-\bar{y})$. Якщо об’єкт має високий рівень по обох змінних, то добуток буде великим додатнім, (аналогічно, якщо відносно низький). Якщо x і y в основному пов'язані *прямо* (великі X з великими У, а малі X з малими У), то більшість добутків буде додатніми і $\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) > 0$.

Якщо *x* і *y* мають *обернений* зв'язок (великі X з малими У і навпаки), то добутки $(x_i - \overline{x})(y_i - \overline{y})$ будуть від'ємними і $\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) < 0$.

Якщо X і У не мають систематичного зв'язку (великі X з малими У так само часто, як і з великими У і аналогічно для малих X), то сума добутків приблизно балансує додатні та від'ємні члени і $\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \approx 0$.

Коваріація є мірою зв'язку між X і У і визначається формулою :

$$S_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n}, \tag{11.1}$$

де $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$, $\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}$, n – кількість спостережень. Коваріація X самого з собою (аналогічно для У) є дисперсією $-\infty < S_{xy} < +\infty$.

Щоб позбавити коваріацію від впливу стандартних відхилень поділимо її на S_x і S_y. У результаті отримаємо міру зв'язку між X і У – коефіцієнт кореляції.

$$r_{xy} = \frac{S_{xy}}{\sigma_x . \sigma_y} = \frac{n\sum x_i y_i - \sum x_i y_i}{\sqrt{\left(n\sum x_i^2 - (\sum x_i)^2\right)\left(n\sum y_i^2 - (\sum y_i)^2\right)}}, \tag{11.2}$$

$-1 \le r_{xy} \le 1$, якщо $r_{xy} > 0$ – прямий зв'язок, $r_{xy} < 0$ – обернений.

Для незгрупованих даних при обчисленні коефіцієнта кореляції користуються формулою:

$$r_{xy} = \frac{\sum n_{xy} xy - n\overline{x}\,\overline{y}}{n\sigma_x \sigma_y} \tag{11.3}$$

Для того, щоб при рівні значущості α, перевірити суттєвість зв'язку між Х і У, потрібно знайти розрахункове значення $t_{розр}=\dfrac{r_{xy}\sqrt{n-2}}{\sqrt{1-r^2}}$ та порівняти його з табличним $t_{xp}=t(\alpha, n-2)$. Якщо $t_{розр}>t_{табл}$, то зв'язок між Х і У суттєвий, у протилежному випадку ні.

II. Приклади розв'язування задач.

Приклад 11.1. Чи існує залежність між частотою користування автомобілем (У) та часом на дорогу від місця проживання до місця стоянки (Х) для вибіркових даних.

Таблиця 11.2.

Х (хв.)	1	2	3	4	5
У (разів)	12	13	14	15	16

Розв'язання. Допоміжні результати зведемо в таблицю:

Таблиця 11.3

i	y_i	x_i	$(x_i-\bar{x})$	$(y_i-\bar{y})$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$
1	12	1	-2	-2	4	4	4
2	13	2	-1	-1	1	1	1
3	14	3	0	0	0	0	0
4	15	4	1	1	1	1	1
5	16	5	2	2	4	4	4
$\sum$	70	15			10	10	10

З таблиці отримаємо $\bar{x}=\dfrac{\sum x_i}{n}=\dfrac{70}{5}=14$, $\bar{y}=\dfrac{\sum y_i}{n}=\dfrac{15}{5}=3$,

$$S_{xy}=\frac{10}{5}=2,\qquad D(x)=\frac{\sum (x_i-\bar{x})^2}{n}=\frac{10}{5}=2,\qquad \sigma_x=\sqrt{D(x)}=\sqrt{2},$$

$$D(y)=\frac{\sum (y_i-\bar{y})^2}{n}=\frac{10}{5}=2,\ \sigma_y=\sqrt{D(y)}=\sqrt{2},$$

і в результаті отримаємо: $r_{xy}=\dfrac{S_{xy}}{\sigma_x\sigma_y}=\dfrac{2}{\sqrt{2}\cdot\sqrt{2}}=1$.

Між відстанню від місця проживання до місця стоянки та частотою користування автомобілем існує тісний прямий зв'язок, а між У і Х у цьому випадку буде лінійна залежність ($\hat{y} = x + 11$).

Зауваження. Якщо r_{xy}= 0, то прямолінійного зв'язку немає, але може бути криволінійний, або функціональний.

Приклад 11.2. Нехай маємо результати вибіркового соціологічного обстеження.

Таблиця 11.4

x	-5	-4	-3	0	3	4	5
y	0	3	4	5	4	3	0

Визначити форму та величину зв'язку між *x* і *y*.

Розв'язання. Для визначення форми зв'язку побудувати діаграму розсіяння (самостійно).

Із діаграми розсіяння та даних задачі видно, що лінійного зв'язку між Х і У немає.

Обчислимо r_{xy}.

Таблиця 11.5

i	y_i	x_i	$(y_i - \bar{y})$	$(x_i - \bar{x})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})^2$
1	-5	0	-5	$-\frac{19}{7}$	$\frac{95}{7}$	25	$\frac{361}{49}$
2	-4	3	-4	$\frac{2}{7}$	$-\frac{8}{7}$	16	$\frac{4}{49}$
3	-3	4	-3	$\frac{9}{7}$	$-\frac{27}{7}$	9	$\frac{81}{49}$
4	0	5	0	$\frac{16}{7}$	0	0	$\frac{256}{49}$
5	3	4	3	$\frac{9}{7}$	$\frac{27}{7}$	9	$\frac{81}{49}$
6	4	3	4	$\frac{2}{7}$	$\frac{8}{7}$	16	$\frac{4}{49}$
7	5	0	5	$-\frac{19}{7}$	$-\frac{95}{7}$	25	$\frac{361}{49}$
$\sum$	0	19				100	$\frac{164}{7}$

$\bar{y}=\frac{0}{7}=0;\quad \bar{x}=\frac{19}{7}$, S_{xy}=0, $\quad \sigma_x=\sqrt{\frac{100}{7}}\approx 3{,}8$, $\quad \sigma_y=\sqrt{\frac{164}{49}}\approx 1{,}8$.

Отже, між Х і У зв'язку немає.

Але з діаграми розсіяння і з даних бачимо, що через вказані точки можна провести півколо і його рівняння у=$\sqrt{25-x^2}$ – функціональний зв'язок.

Приклад 11.3. Чи існує зв'язок між довжиною туристичного маршруту (*У* км.) та вартістю екскурсії (*Х* у.о.), для даних вибіркового розподілу таблиці:

Таблиця 11.6.

X\Y	100	200	300	400	500	600	n_x
5	1	2					3
6		2	4				6
7			5	20	4		29
8			2	3	3		8
9					3	1	4
n_y	1	4	11	23	10	1	50

Складемо кореляційну таблицю 11.7 перейшовши до нових *u* і *v*. Замість "уявних нулів" виберемо значення варіант, які мають найбільшу частоту.

C_1=7–"уявний нуль" варіант Х, а $u_i=\frac{(x_i-C_1)}{h_1}$

h_1 – різниця між сусідніми варіантами, h_1 =1.

C_2 = 400 –"уявний нуль" варіант У, а $v_i=\frac{(y_i-C_2)}{h_2}$,

де h_2 – різниця між сусідніми варіантами, h_2 =100.

Отримаємо розподіл:

Таблиця 11.7

*U**V*	-3	-2	-1	0	1	2	n_u
-2	1	2					3
-1		2	4				6
0			5	20	4		29
1			2	3	3		8
2					3	1	4

n_y	1	4	11	23	10	1	50

У цьому випадку вибірковий коефіцієнт кореляції:

$$r_{xy}=r_{uv}=\left(\sum n_{uv}u\cdot v-n\overline{u}\overline{v}\right)/(n\sigma_u\sigma_v) \qquad (11.4)$$

$$\overline{u}=\frac{\sum n_u u}{n}=\frac{3\cdot(-2)+6\cdot(-1)+0\cdot 29+1\cdot 8+2\cdot 4}{50}=\frac{4}{50}=0{,}08$$

$$\overline{v}=\frac{\sum n_v\cdot v}{n}=\frac{(-3)\cdot 1+(-2)\cdot 4+(-1)\cdot 11+0\cdot 23+1\cdot 10+2\cdot 1}{50}=\frac{10}{50}=-0{,}2$$

$\sigma_u=\sqrt{D(u)}$; D(u)=$\overline{u^2}-(\overline{u})^2$,

$$\overline{u}^2=\frac{(-2)^2\cdot 3+(-1)^2\cdot 6+0^2\cdot 29+1^2\cdot 8+2^2\cdot 4}{50}=\frac{42}{50}=0{,}84,$$

D(u)=0,84–(0,08)^2=0,8336; $\sigma_u\approx 0{,}91$.

$\sigma_v=\sqrt{D(v)}$; D(v)=$\overline{v^2}-(\overline{v})^2$;

$$\overline{v^2}=\frac{(-3)^2\cdot 1+(-2)^2\cdot 4+(-1)^2\cdot 11+0^2\cdot 23+1^2\cdot 10+2^2\cdot 1}{50}=\frac{50}{50}=1$$

D(v) = 1 – (–0,2)^2=0,96 , $\sigma_v\approx 0{,}98$

Суму $\sum n_{uv}uv$ обчислимо із таблиці 11.7.

$\sum n_{uv}uv$=1•(-2)·(-3)+2(-2)•(-2)+2·(-1)·(-2)+4(-1)·(-1)+5·0·(-1)+20·0·0+4·0·1+2·1·(-1)+3·1·0+3·1·1+3·2·1+1·2·2=33

Отже, вибірковий коефіцієнт кореляції дорівнює:

$$r_{xy}=\frac{33-50\cdot(0{,}08)\cdot(-0{,}2)}{50\cdot 0{,}91\cdot 0{,}98}\approx 0{,}76.$$

Оскільки коефіцієнт кореляції близький до 1, то між довжиною маршруту та собівартістю екскурсії існує тісний прямий зв’язок.

Перевіримо статистичну значущість отриманого коефіцієнта ($\alpha=0{,}05$).

$$T_{розр}=\frac{0{,}76\sqrt{50-2}}{\sqrt{1-0{,}76^2}}\approx 8{,}1$$

$T_{кр}(0{,}05;48)=2{,}01$

Оскільки 8,1>2,01, то зв’язок між Х і У суттєвий.

IV. Завдання для самостійного виконання

Завдання 5.1. Побудувати діаграму розсіяння та з'ясувати, чи існує зв'язок між X і У. Перевірити значущість отриманого коефіцієнта кореляції на 0,0N рівні.

1

X	Y
1	15
2	14
3	13
4	12
5	11

2

X	Y
0	10
1	11
2	13
3	14
4	15
5	16

3

X	Y
2	21
4	24
6	27
8	30
10	33
12	36

4

X	Y
-1	1
0	2
1	3
2	4

5

X	Y
-3	-4
-2	-3
-1	-2
0	-1
1	0
2	1
3	2

6

X	Y
2	3
4	5
6	7
8	9
0	11
12	13

7

X	Y
1	16
2	14
3	13
4	12
5	11
6	10

8

X	Y
1	62
2	120
3	160
4	200

9

X	Y
-3	6
-1	4
1	2
3	0
5	2

10

X	Y
6	10
5	11
4	12
3	13
2	14
1	15

11

X	Y
10	120
15	130
20	140
25	145
30	150
35	150

12

X	Y
-1	10
-2	9
-3	8
-4	7
-5	6
-6	5

13

X	Y
-3	2
2	-3
-2	1
1	-2
-1	0
0	1

14

X	Y
-1	10
0	11
1	12
2	13
3	14
4	15
5	16
6	17

15

X	Y
5	120
6	130
7	140
8	150
9	160
10	170
11	180
12	190
13	200

Завдання 11.2. Для наведених систем величин Х і У обчислити коефіцієнт кореляції та перевірити його значущість на N% рівні.

Варіанти 1-5.

Дано розподіл родин за їх річними доходами на людину (*X*, тис. грн.) та витратами на відпочинок (*У*, грн.):

X\У	100-N	150-N	200-N	250-N	300-N	350-N	Разом
1,2+N	1	2					3
1,8+N		5	3				8
2,4+N			2	20	2		24
3,0+N			1	4	4	2	11
3,6+N					2	2	4
Разом	1	7	6	24	8	4	50

Варіанти 6-10.

Дано розподіл готелів за загальною кількістю номерів (Х) і кількістю номерів категорії “Люкс” (У).

Х\У	60+N	80+N	100+N	120+N	140+N	160+N	Разом
500-N	3	1					4
600-N	1	4	5				10
700-N		2	6	12			20
800-N			2	4	7		13
900-N					1	2	3
Разом	4	7	13	16	8	2	50

Варіанти 11-15.

Дано розподіл закордонних турів за чисельністю туристів у групі (Х) і за знижками на загальну вартість подорожі (у%).

Х\У	15-N	17-N	19-N	21-N	23-N	25-N	Разом
10+N	3						3
20+N	1	2			1		4
30+N		4	16	3			23
40+N		1	7	5	3		16
50+N					2	2	4
Разом	4	7	23	8	6	2	50

ЛАБОРАТОРНА РОБОТА 12

Тема: Коефіцієнти рангової кореляції

I. Деякі теоретичні відомості та приклади

Основні терміни: *ранг, коефіцієнт рангової кореляції Спірмена, коефіцієнт рангової кореляції Кендала, значущість* r_s *і* τ .

До якісних ознак виміряних з допомогою шкал порядкового і номінального рівня для вимірювання тісноти зв'язку застосовуються коефіцієнти рангової кореляції.

Коефіцієнт рангової кореляції Спірмена визначається за формулою:

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \quad (-1 \le r_s \le 1) \tag{12.1}$$

де d_i різниця між рангами (порядковими номерами) порівнюваних рядів, n – число порівнюваних пар рангів.

Приклад 4.5. Визначити r_s між відповідями працівників у вибірковій сукупності на запитання: "Чи цікава у Вас робота?" і "Чому ви обрали цю спеціальність?"

n=14 груп за професіями, розташовуємо їх по спаданню частини тих людей, які відповіли, що у них "цікава робота".

У першому стовпці номера груп. У третьому – значення ознаки (у %) – частина тих людей, у кожній із груп, які мотивували свій вибір спеціальності тим, що вона цікава.

Таблиця 12.1

№ професії, груп	Відповідь (%)		Номера рангів		d_i=I–II	d_i^2
	"цікава робота" I	"цікава спеціальність" II	I	II		
1	100,0	100,0	3	1	2	4
2	100,0	77,0	3	2	1	1
3	100,0	75,0	3	3	0	0
4	100,0	66,5	3	4,5	-1,5	2,25
5	100,0	50,0	3	8,5	-5,5	30,25
6	83,5	66,5	6,5	4,5	2	4
7	83,5	41,5	6,5	10	-3,5	12,25
8	83,0	65,5	8	6	2	4
9	82,5	41,0	9	11	-2	4
10	71	53,5	10	7	3	9
11	55,5	0	11	13,5	-2,5	6,25
12	50	50,0	12	8,5	3,5	12,25
13	28,5	28,5	13	12	1	1
14	0	0	14	13,5	0,5	0,25
n=14					$\sum d_i$ =0	$\sum d_i^{\,2}$ =90,5

Кожній групі, починаючи з другого стовпця, присвоєно ранги (по спаданню). Якщо є однакові елементи, то беремо середнє арифметичне їх номерів (100, (1+2+3+4+5):5=3; 83,5 (6+7):2=6,5; 66,5; (4+5):2 = 4,5 та ін.) ранги другого стовпця

записуємо в четвертий, третього – у п’ятий.

Підставивши отримані результати у формулу (12.1), отримуємо:

$$r_s=1-\frac{6\cdot 90{,}5}{14(14^2-1)}\approx 0{,}8\,.$$

Між відповідями респондентів "цікава спеціальність" та "цікава робота" існує тісний зв'язок.

Формула (12.1) справедлива лише тоді, коли всім об’єктам присвоюються різні ранги. А в нашому прикладі деякі ранги однакові (тоді r_s завищується). У цьому випадку більше підходить формула:

$$r_s=1-\frac{6(\sum d_i^2+T_x+T_y)}{n(n^2-1)}, \qquad (12.2)$$

де T_x і T_y визначаються таким чином:

$$T_x=\frac{\sum(t_x^3-t_x)}{12},\quad T_y=\frac{\sum(t_y^3-t_y)}{12} \qquad (12.3)$$

t_x, t_y – число об'єднаних (рівних) рангів у відповідному ряді. Перерахуємо r_s для нашого прикладу: $T_x=\frac{(5^3-5)+(2^3-2)}{12}$=10,5

Об'єднані 5 і 2 рангів.

$T_y=\frac{(2^3-2)+(2^3-2)+(2^3-2)}{12}$=1,5

Об'єднані тричі по два ранги.

Тоді згідно з (12.2): $r_s= 1-\frac{6(90{,}5+10{,}5+1{,}5)}{14(14^2-1)}\approx 0{,}77<0{,}8$.

***Значущість** r_s*

Для того, щоб при рівні значущості α перевірити нульову гіпотезу про рівність нулю генерального коефіцієнта рангової кореляції r_s Спірмена при альтернативній гіпотезі H_1: $r_s^г\neq 0$, потрібно обчислити критичну точку $T_{кр}=t_{кр}(\alpha.,\kappa)\sqrt{\frac{1-r_s^2}{n-2}}$, де n – об'єм вибірки, r_s – вибірковий коефіцієнт рангової кореляції Спірмена, $t_{кр}(\alpha$,к)- критична точка двосторонньої критичної області (із таблиць по α і $к = n$-2). Якщо $|r_s|>T_{кр}$ – то нульову гіпотезу відкидають. Між якісними ознаками існує значущий ранговий кореляційний зв'язок, у протилежному випадку – нульову гіпотезу приймають.

Для нашого прикладу при 5% рівні значущості α=0,05, к=14-2=12, $t_{кр}$=2,18; $T_{кр}=2{,}18^*\sqrt{\frac{1-0{,}77^2}{12}}\approx$0,4

Оскільки $|0{,}77|$>0,4, то між відповідями “цікава робота” і “цікава спеціальність” існує значущий ранговий кореляційний зв'язок.

Коефіцієнт рангової кореляції Кендала

Він підраховується за формулою:

$$\tau = \frac{4S^{+}}{n(n-1)} - 1 = \frac{S}{1n(n-1)}, \ (-1 \le \tau \le 1) \qquad (11.4)$$

Порядок обчислення S розглянемо на прикладі.

Приклад 12.2. Визначити τ між відповідями робітників “цікава робота” та “освіта відповідає роботі”.

Таблиця 12.2.

№ професії, груп	Відповідь (%)		Номера рангів		S_1	S_2
	“цікава робота” I	“освіта відповідає роботі” II	I	II		
1	100,0	100,0	3	13	13	0
2	100,0	87,0	3	8	8	3
3	100,0	77,0	3	5	5	6
4	100,0	75,0	3	4	4	6
5	100,0	50,0	3	2	2	6
6	83,5	92,0	6,5	7	7	1
7	83,5	83,5	6,5	3	3	4
8	83,0	90,0	8	5	5	1
9	82,5	94,5	9	5	5	0
10	71,0	87,0	10	3	3	1
11	55,5	87,5	11	3	3	0
12	50,0	50,0	12	2	2	0
13	28,5	43,0	13	1	1	0
14	0	0	14	0	0	0
					S^{+}=61	S^{-}=28

У стовпчик S_1 заносимо число, яке вказує, скільки рангів другого ранжованого ряду перевищують ранги, які стоять на другому, третьому,…, чотирнадцятому місцях $S_1 \approx 13$; другим записано 5,5 (нижче його більших від 5,5 – 8) і т.д.S^{+} – це сума елементів стовпця S_1.

В стовпчик S_2 записуємо числа, які вказують, скільки рангів менших за даними знаходиться нижче від даного рангу. Менше від 1 нема, тому S_2=0, менше 5,5 – 3 (3,4,2), тому S_2 = 3 і т.д. S^{-} – це сума елементів стовпця S_2.

S^{+}=61, S^{-}=28, $S=S^{+}-S^{-}$=61-23=33

Тоді $\tau = \dfrac{33}{7 \cdot 13} \approx 0{,}36$ – слабкий прямий зв'язок.

Якщо ранжовані об'єкти мають зв'язані (однакові) ранги, то коефіцієнт рангової кореляції Кендала підраховується за формулою:

$$\tau = \frac{S}{\sqrt{\left(\frac{1}{2}n(n-1) - T_x\right)\left(\frac{1}{2}n(n-1) - T_y\right)}}, \tag{12.5}$$

де $T_x = \frac{1}{2}\sum t_x(t_x - 1)$, $T_y = \frac{1}{2}\sum t_y(t_x - 1)$, (12.6)

t_x, t_y – число рівних рангів у відповідному ряді.

Для нашої задачі:

$$T_x = \frac{1}{2}(5 \cdot (5-1) + 2(2-1)) = 11,$$

$$T_y = \frac{1}{2}(2(2-1) + 2(2-1)) = 2,$$

$$\tau = \frac{33}{\sqrt{(7 \cdot 13 - 11)(7 \cdot 13 - 2)}} \approx 0{,}39 > 0{,}36 .$$

Для перевірки значущості коефіцієнта Кендала обчислюють:

$T_{кр} = Z_{кр} \cdot \sqrt{\dfrac{2(2n+5)}{9n(n-1)}}$, n – об'єм вибірки, $Z_{кр}$ – критична точка двосторонньої критичної області, яку знаходять по таблиці функції Лапласа із рівності $\Phi(Z_{кр}) = (1-\alpha)/2$. Якщо $|\tau| < T_{кр}$, то ранговий кореляційний зв'язок між якісними ознаками незначний. Якщо $|\tau| > T_{кр}$ – то навпаки.

Для нашого прикладу (α =0,05):

$$\Phi(Z_{кр}) = \frac{1 - 0{,}05}{2} = 0{,}475 \Rightarrow Z_{кр} = 1{,}96$$

$$T_{кр} = 1{,}96\sqrt{\frac{2 \cdot (2 \cdot 14 + 5)}{9 \cdot 14 \cdot (14 - 1)}} \approx 0{,}393$$

Оскільки 0,391<0,393, то зв'язок між відповідями "цікава робота" і "освіта відповідає роботі" не значний.

ІІ. Завдання для самостійного виконання

Знайти вибірковий коефіцієнт рангової кореляції Спірмена для завдань 1-7 та перевірити його значущість на *N%* рівні. Проінтерпретувати результати.

1. Завдання десяти студентів перевірено за двома тестами А і Б. Оцінки за стобальною системою виявились такими:

А	95	90	86	84	75	70	62	60	57	50
Б	92	93	83	80	55	60	45	72	62	70

2. Два інспектори перевірили 12 водіїв на швидкість реакції і розташували їх у порядку зменшення реакції (у дужках порядкові номера водіїв з однаковою реакцією).

   І. 1 (2, 3, 4) 5 (6, 7, 8) 9 10 11 12

   ІІ. 3 1 2 6 4 5 7 9 10 12 0 11

3. Два контролери А і Б розташували зразки виробів, виготовлених дев'ятьма майстрами у порядку погіршення якості (у дужках номера виробів однакової якості).

   А 1 2 (3, 4, 5) (6, 7, 8, 9)

   Б 2 1 4 3 5 (6, 7) 8 9.

4. Спеціалісти двох заводів проаналізували 11 факторів, які впливають на хід технологічного процесу і отримали дві послідовності рангів:

x_i	1	2	3	4	5	6	7	8	9	10	11
y_i	1	2	3	5	4	9	8	11	6	7	10

5 Два викладачі оцінили знання 12 студентів за стобальною системою і виставили їм такі оцінки :

98	94	88	80	76	70	63	61	60	58	56	51
99	91	93	74	78	65	64	66	52	53	48	62

6. Тринадцять кольорових стрічок розташовані в порядку спадання кольорів від темних до світлих і кожній стрічці присвоєно ранг – порядковий номер. При перевірці здатності розрізняти відтінки досліджувана особа розташувала їх таким чином:

x_i	1	2	3	4	5	6	7	8	9	10	11	12	13
y_i	6	3	4	2	1	10	7	8	9	5	11	13	12

7. Результати міжсесійного контролю студентів підгрупи розташовані у порядку спадання (середній бал) і окремо результати з дисципліни “Кількісні методи дослідження соціальних процесів”.

Ат.	5	5	4,9	4,8	4,7	4,7	4,7	4,5	4,2	4	3,5	3	2	2
Км.	5	5	4	5	4	4	4	3	3	4	3	2	2	2

8. Три арбітри оцінили виступ десяти фігуристів, у результаті чого отримали три послідовності рангів:

I	1	2	3	4	5	6	7	8	9	10
II	3	10	7	2	8	5	6	9	1	4
III	6	2	1	3	9	4	5	7	10	8

Визначити пару арбітрів, оцінки яких найбільш узгоджуються, використовуючи коефіцієнт рангової кореляції Спірмена. Перевірити значущість отриманого r_s для $\alpha = 0{,}08$.

9-15. Знайти вибірковий коефіцієнт рангової кореляції Кендала, використовуючи умови задач 1-7 (9-1; 10-2; 11-3; і т.д.) та перевірити його значущість для α =0,0N. Проінтерпретувати результати.

ЛАБОРАТОРНА РОБОТА 13

Тема: Рівняння лінійної парної регресії

I. Деякі теоретичні відомості та приклади

Основні терміни: *рівняння регресії, лінійна парна регресія, метод найменших квадратів (МНК), система нормальних рівнянь, рівняння лінійної регресії для незгрупованих даних*.

Намітивши форму лінії регресії та записавши рівняння в загальному вигляді, далі необхідно знайти числові значення його параметрів.

Теоретичне рівняння лінійної парної регресії

$$y = a + bx + u \tag{13.1}$$

де y – залежна змінна, x – незалежна, u – випадкова складова;

$$\hat{y} = \hat{a} + \hat{b}x \tag{13.2}$$

розрахункове рівняння, $u = y_i - \hat{y}_i$.

Випадкова складова може бути як додатня, так і від'ємна (залежно від того, де знаходиться точка y_i – над прямою чи під нею).

При знаходженні параметрів рівняння регресії (13.2) із всіх можливих прямих вибирають ту, в якої сума квадратів відхилень буде найменшою. Це основний принцип методу найменших квадратів (МНК). Це значить, що

$$\sum_{i=1}^{n}(y_i - \hat{y}_i) = \min. \tag{13.3}$$

Заміняючи в (13.3) $\hat{y}_i$ його виразом із (13.2), отримаємо суму квадратів як функцію від невідомих параметрів, тобто

$$f(\hat{a},\hat{b}) = \sum_{i=1}^{n}(y_i - \hat{a} - \hat{b}x_i)^2 \tag{13.4}$$

для якої і знаходимо мінімум. Щоб знайти мінімум, обчислюємо частинні похідні $f(\hat{a},\hat{b})$ і прирівнюємо їх до нуля. Отримаємо так звану систему нормальних рівнянь:

$$\begin{cases} \hat{a}n + \hat{b}\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ \hat{a}\sum_{i=1}^{n} x_i + \hat{b}\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \end{cases} \tag{13.5}$$

Розв'язуючи цю систему отримаємо:

$$\begin{cases} \hat{b} = \dfrac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} \\ \hat{a} = \dfrac{\sum y_i - \hat{b}\sum x_i}{n} \end{cases}, \qquad (13.6)$$

Оскільки частинні другі похідні по $\hat{a}$ і $\hat{b}$ від функції f більше нуля, то в таких точках $\hat{a}$ і $\hat{b}$ буде мінімум.

Приклад 13.1. За даними обстеження семи власників автомобілів вивчається залежність частоти користування автомобілем (за тиждень) від віддаленості його місця стоянки від місця проживання.

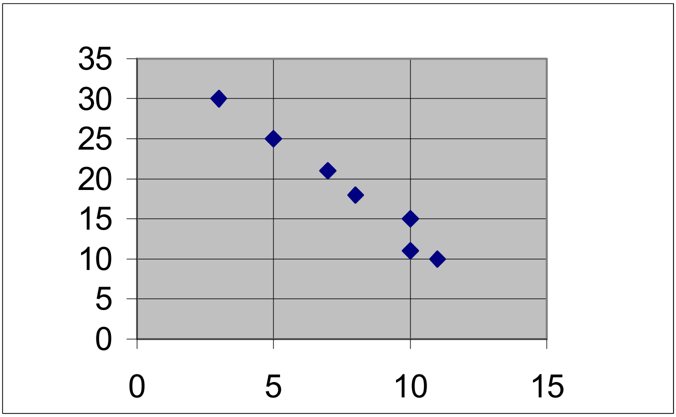
Дані обстеження в таблиці 13.1:

Таблиця 13.1

Частота користування автомобілем (за тиждень), разів ***У***	12	12	16	18	22	25	30
Час на дорогу від місця проживання до місця стоянки, (хв.) ***X***	11	12	10	8	7	5	3

1. Ідентифікуємо змінні. Нехай У – частота користування автомобілем (залежна змінна), Х – час на дорогу від місця проживання до місця стоянки, хв. (незалежна змінна).
2. Для специфікації моделі спочатку нанесемо у системі координат точки (x_i, y_i), тобто побудуємо діаграму розсіяння.

y



За діаграмою розсіяння бачимо, що через дані точки найкраще проводити пряму, тобто специфікуємо модель у лінійному вигляді.

y = a+bx + u – теоретичне рівняння,

$\hat{y} = \hat{a} + \hat{b}x$ – розрахункове рівняння.

3. Для визначення невідомих параметрів розрахункового рівняння потрібно скласти систему (13.5), а для цього заповнимо допоміжну таблицю 13.2:

Таблиця 13.2.

№ з/п	x_i	y_i	x_i^2	x_iy_i
1	11	12	121	132
2	12	12	144	144
3	10	16	100	160
4	8	18	64	144
5	7	22	49	154
6	5	25	25	125
7	3	30	9	90
Сума	56	135	512	949

Із останнього рядка цієї таблиці випишемо систему (13.5)

$$\begin{cases} 7\hat{a} + 56\hat{b} = 135 \\ 56\hat{a} + 512\hat{b} = 949 \end{cases}$$

Розв'язуючи останню систему, отримаємо:

$\hat{a} \approx 35{,}66$; $\hat{b} \approx -2{,}05$, тобто залежність частоти користування від віддаленості місця стоянки має вигляд:

$\hat{y} = 35{,}66 - 2{,}05x$.

Звідси, аналізуючи коефіцієнт регресії $\hat{b} = -2{,}05$ можна дійти висновку, що при збільшенні часу на дорогу від місця проживання до місця стоянки на 1 хв. частота користування У зменшується на 2,05 раза. Для перевірки правильності побудови рівняння в нього підставляють точку ($\overline{x}, \overline{y}$) (вона завжди належить лінії регресії).

$\overline{x} = \dfrac{\sum x_i}{n} = \dfrac{56}{7} = 8$, $\overline{y} = \dfrac{\sum y_i}{n} = \dfrac{135}{7} \approx 19.29$, $35{,}66 - 2{,}05 \cdot 8 \approx 19{,}29$.

Отже, рівняння побудоване правильно.

Для незгрупованих даних вибіркове рівняння лінійної регресії У на Х має вигляд:

$$\hat{y} - \overline{y} = r_y \frac{\sigma_y}{\sigma_x}(x - \overline{x}). \qquad (13.7)$$

де r_{yx} вибірковий коефіцієнт кореляції

$$r_x = \left(\sum n_{xy} x_y - n\overline{x}\overline{y}\right)/(n\sigma_x\sigma_y) = \left(\sum n_{uv} u \cdot v - n\overline{u}\overline{v}\right)/(n\sigma_u\sigma_v) \qquad (13.8)$$

де $u_i = (x_i - C_1)/h_1$,

$v_j = (y_j - C_2)/h_2$

(13.9)

де C_1 і C_2 – "уявні нулі" відповідно варіант Х і У h_1 і h_2 – крок, або різниця між двома сусідніми варіантами, відповідно Х і У (див.лабораторну роботу 11).

Приклад 13.2. Побудувати рівняння лінійної парної регресії для розподілу родин за їх річнимим доходами на людину(Х, тис.грн.) та витратами на відпочинок (У, грн.)

Таблиця 13.3

Х/У	100	150	200	250	300	350	n_x
1,2	1	2	-	-	-	-	3
1,8	-	5	3	-	-	-	8
2,4	-	-	2	20	2	-	24
3,0	-	-	1	4	4	2	11
3,6	-	-	-	-	2	2	4
n_y	1	7	6	24	8	4	50

Розв'язання: Для знаходження r_{xy}, σ_x, σ_y перейдемо до нової кореляційної таблиці, вибравши в якості "уявних нулів" C_1=2,4 (йому відповідає найбільша частота n_x=24) і C_2=250 (n_y=24). Кроки відповідно h_1=0,6=1,8–1,2 і h_2=250–200=50.

Таблиця 13.4

u	v						n_u
	-3	-2	-1	0	1	2	
-2	1	2	-	-	-	-	3
-1	-	5	3	-	-	-	8
0	-	-	2	20	2	-	24
1	-	-	1	4	4	2	11
2	-	-	-	-	2	2	4
n_v	1	7	6	24	8	4	n=50

Знайдемо $\overline{n}$ і $\overline{v}$:

$$\overline{u} = \frac{\sum u \cdot n_u}{n} = \frac{-2 \cdot 3(-1) \cdot 8 + 0 \cdot 24 + 1 \cdot 11 + 2 \cdot 4}{50} = \frac{5}{50} = \frac{1}{10} = 0{,}1;$$

$$\overline{v} = \frac{\sum v n_v}{n} = \frac{(-3) \cdot 1 + (-2) \cdot 7 + (-1) \cdot 6 + 0 \cdot 24 + 1 \cdot 8 + 2 \cdot 4}{50} = -\frac{7}{50} = -0{,}14.$$

Середньоквадратичні відхилення $\sigma_u = \sqrt{\overline{u^2} - (\overline{u})^2}$,

$\sigma_v = \sqrt{\overline{v^2} - (\overline{v})^2}$, тому:

$$\overline{u^2} = \frac{\sum u^2 n_u}{n} = \frac{(-2)^2 \cdot 3 + (-1)^2 \cdot 8 + 0^2 \cdot 24 + 1^2 \cdot 11 + 2^2 \cdot 4}{50} = \frac{47}{50} = 0{,}94$$

$$\overline{v^2} = \frac{\sum v^2 n_v}{n} = \frac{(-3)^2 \cdot 1 + (-2)^2 \cdot 7 + (-1)^2 6 + 0^2 \cdot 24 + 1^2 \cdot 8 + 2^2 \cdot 4}{50} =$$

$$= \frac{67}{50} = 1{,}34$$

Отже:

$$\sigma_u = \sqrt{0{,}94 - (0{,}1)^2} = \sqrt{0{,}93} \approx 0{,}96, \quad \sigma_v = \sqrt{1{,}34 - (-0{,}14)^2} \approx 1{,}15.$$

Знайдемо

$$\sum n_{uv} uv = 1\cdot(-2)\cdot(-3)+2\cdot(-2)\cdot(-2)+\ldots+2\cdot 2\cdot 2=46$$

Отже, вибірковий коефіцієнт кореляції із (13.8):

$r_{xy}= \dfrac{\sum n_{uv} u \cdot v - n\bar{u}\bar{v}}{n\sigma_u \cdot \sigma_v} = \dfrac{46 - 50 \cdot 0{,}1 \cdot (-0{,}14)}{50 \cdot 0{,}96 \cdot 1{,}15} \approx 0{,}85$ між витратами на відпочинок та річними доходами тісний прямий зв'язок.

Знайдемо далі $\bar{x}$ і $\bar{y}$ користуючись (13.9):

$\bar{x} = \bar{u} \cdot h_1 + C_1 = 0{,}1\cdot 0{,}6+2{,}4=2{,}46$;

$\bar{y} = \bar{v} \cdot h_2 + C_2 = -0{,}14 \cdot 50 + 250 = 243,$

а також σ_x і σ_y:

$\sigma_x = h_1 \cdot \sigma_u = 0{,}6\cdot 0{,}96=0{,}576$; $\sigma_y = h_2 \cdot \sigma_v = 50 \cdot 1{,}15 = 57{,}5$

Підставивши отримані значення в (13.7) :

$\hat{y} - 243 = 0{,}85 \cdot \dfrac{57{,}5}{0{,}576}(x - 2{,}46)$, або після перетворень:

$\hat{y} = 84{,}9x - 208{,}7$ – залежність між витратами на відпочинок та річним доходом на людину.

Висновок: При збільшенні річного доходу на одну людину на 1 тис.грн. витрати на відпочинок зростуть на 84,9 грн.

II. Завдання для самостійного виконання

Завдання 1.

Для даних лабораторної роботи № 11 (завдання 1, табл. 1-15), припустивши, що між у і x існує лінійна залежність, побудувати рівняння лінійної парної регресії $\hat{y} = \hat{a} + bx$ МНК. Проінтерпретувати отриманий коефіцієнт регресії.

Завдання 2.

Для даних лабораторної роботи № 11 (завдання 2, табл. 1-15) побудувати рівняння лінійної парної регресії для залежності y від x. Проінтерпретувати отриманий коефіцієнт регресії.