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

МЕТОДИЧНІ ВКАЗІВКИ

до виконання лабораторних робіт з дисципліни

«Теорія ймовірностей і математична статистика»

для напряму підготовки 6.030601 «**Менеджмент**»

Київ – 2016

Методичні вказівки до виконання лабораторних робіт (комп'ютерного практикуму) з дисципліни «Теорія ймовірностей і математична статистика» для студентів напряму підготовки 6.030601 «Менеджмент» студ. Видавн.-поліграф. ін.-ту / Укл. О.І. Кушлик-Дивульська, Б.Р. Кушлик − К.: НТУУ «КПІ». − 2016. – 205с.

Рекомендовано Вченою радою
Фізико-математичного факультету НТУУ «КПІ»
(Протокол № 8 від 24 листопада 2015 р.)

Н а в ч а л ь н е в и д а н н я
Доповнене, виправлене

МЕТОДИЧНІ ВКАЗІВКИ
до виконання лабораторних робіт (комп'ютерного практикуму)
з дисципліни
«Теорія ймовірностей і математична статистика»
для напряму підготовки 6.030601
«Менеджмент»

для студентів Видавничо-поліграфічного інституту

Укладачі:
Кушлик-Дивульська Ольга Іванівна
Кушлик Богдан Ростиславович

Відповідальний редактор *С. Д. Івасишен*, д-р фіз.-мат. наук, проф.

Рецензент *А.Б. Ільєнко,* канд. фіз.-мат. наук, доц.

Передмова

Теорія ймовірностей та математична статистика – математичні науки, які вивчають закономірності в масових випадкових явищах, і вони є складовою теоретичною основою викладання багатьох економічних, соціологічних та спеціальних дисциплін.

Мета комп'ютерного практикуму полягає в закріпленні знань, одержаних студентами під час вивчення дисципліни «Теорія ймовірності і математична статистика», їх застосуванні для вирішення конкретних практичних завдань із використанням теоретичних знань та на практиці можливостей пакету Microsoft Excel. Виконання лабораторних робіт сприяє формуванню самостійності у аналізі проведених обчислень, дослідженні практичних задач, які є необхідною складовою підвищення технічного та наукового рівня підготовки студента для вивчення фахових дисциплін [3], подальшого вміння використовувати отримані знання для реальних економічних задач, зокрема в поліграфічній галузі [10], [11]. Зміст і структура методичних вказівок відображають новітні тенденції у питаннях навчання та підготовки кваліфікованих спеціалістів.

Лабораторні роботи (комп'ютерний практикум) дисципліни містять короткі теоретичні відомості, приклади розв'язування задач, також індивідуальні завдання для виконання тематичної лабораторної роботи та перелік основних теоретичних питань, які також є обов'язковими. Протокол практичної роботи оформлюється у вигляді роздрукованих сторінок формату А4 та електронного файлу.

Типова структура виконаної практичної роботи містить:

титульний аркуш;

основна частина (короткі теоретичні відомості та розв'язані задачі);

додатки − виконані задачі за допомогою пакету Microsoft Excel із описанням та їх електронний файл.

Лабораторні роботи виконуються в 3-му семестрі стаціонарної форми навчання за програмою напряму підготовки 6.030601 «Менеджмент» освітньокваліфікаційного рівня бакалавр спеціальності 7.03060101 «Менеджмент організацій і адміністрування».

Ознайомлення, опрацювання за темами відповідних лабораторних робіт дає можливість застосовувати алгоритми Excel для практичного використовуваних статистичних задач, які є трудомісткими через значну кількість математичних обчислень. Тому дані вказівки розраховані на широку кількість користувачів, які мають доступ до програмного забезпечення і початкові необхідні знання теорії ймовірностей.

Методичні вказівки є виправленими, доповненими до [8], оскільки із змінами викладання навчальної дисципліни немає навчальних годин для практичних занять, але збільшено кількість годин для лабораторних робіт. Навчальне видання також сприятиме більш повному опануванню теоретичної частини матеріалу навчального посібника «Теорія ймовірності та математична статистика» [7], рекомендованого Міністерством освіти і науки як навчальний посібник для студентів технічних та економічних спеціальностей вищих навчальних закладів.

Автори вдячні співробітникам кафедри математичної фізики, кафедри математичного аналізу та теорії ймовірності фізико-математичного факультету НТУУ «КПІ» та всім читачам, що висловили свої побажання та зауваження. Також укладачі вдячні студентам груп економічної спеціальності Видавничо-поліграфічного інституту за співпрацю в підготовці задач для лабораторних робіт.

Лабораторна робота № 1

Елементи комбінаторики. Класичне означення ймовірності

***Мета роботи*:** Вивчення можливостей пакету Microsoft Excel для розв'язання задач з теорії ймовірностей з використанням елементів комбінаторики, класичного означення ймовірностей та основних теорем ймовірностей.

Теоретичні відомості

1.1. Елементи комбінаторики

Як відомо, згідно з класичним означенням ймовірностей, щоб обчислити ймовірність тієї чи іншої випадкової події, необхідно вміти обчислити кількість n усіх елементарних подій (розмір простору елементарних подій Ω) і число m елементарних подій, які сприяють появі випадкової події.

У багатьох випадках для обчислення чисел n і m використовуються наступні елементи комбінаторики: ***перестановки, розміщення*** та ***комбінації,*** які є окремим видом ***сполук*** елементів певної множини.

Сполуками називаються різні підмножини, утворені з елементів універсальної множити (простору елементарних подій Ω), що відрізняються елементами або порядком цих елементів.

Для обчислення цих сполук в пакеті Excel вбудовані наступні функції: **ФАКТР, ЧИСЛОКОМБ, ПЕРЕСТ**, перші дві з яких знаходяться в категорії функцій ***математичні,*** а третя − ***статистичні.***

Нехай M – множина, що містить n елементів.

Означення. ***Розміщенням*** *і*з n елементів по k називають довільну впорядковану підмножину з k елементів із множини M.

Два розміщення вважаються різними не тільки тоді, коли складаються з різних елементів, але й тоді, коли вони складаються з однакових елементів, але відрізняються їхнім порядком.

Кількість розміщень з n елементів по k $(k \le n)$ позначають A_n^k **і** обчислюють за формулою

$$A_n^k = \frac{n!}{(n-k)!} = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1). \qquad (1.1)$$

Означення. ***Розміщенням з повтореннями*** з n елементів по k називають довільну впорядковану підмножину з k елементів із множини M (елементи не обов'язково різні).

Кількість розміщень із повтореннями з n елементів по k обчислюють за формулою:

$$\overline{A_n^k} = n^k. \qquad (1.2)$$

Зауваження. $\overline{A_n^k}$ – це кількість способів, якими можна розкласти k різних предметів в n ящиків.

Означення. Розміщення з n елементів по n називаються ***перестановками.***

Різні перестановки відрізняються лише порядком елементів. Кількість перестановок із n елементів дорівнює добутку всіх натуральних чисел від 1 до n, тобто

$$P_n = 1 \cdot 2 \cdot \ldots \cdot n = n! \qquad (1.3)$$

Означення. ***Перестановками з повтореннями*** називають розміщення з n елементів, які мають однотипні елементи.

Кількість різних перестановок із повторенням, які можна утворити з n елементів, серед яких є k_1 елементів першого типу, k_2 елементів другого типу, …, k_m елементів m- того типу, дорівнює

$$P_n(k_1, k_2, \cdots, k_m) = \frac{n!}{k_1! \cdot k_2! \cdots k_m!}. \qquad (1.4)$$

Означення. ***Комбінацією*** (сполукою) з n елементів по k називають довільну підмножину з k елементів із множини M, яка відрізняється одна від одної хоча б одним елементом.

Порядок елементів у комбінаціях несуттєвий. Кількість комбінацій з n елементів по k $(k \le n)$ позначають C_n^k та обчислюють за формулою

$$C_n^k = \frac{n!}{k! \cdot (n-k)!} \qquad (1.5)$$

Означення. ***Комбінацією з повтореннями*** з n елементів по k називають довільну множину з k елементів із множини M (елементи не обов'язково різні).

Кількість комбінацій з повтореннями з n елементів по k обчислюють за формулою

$$\overline{C_n^k} = C_{n+k-1}^k = \frac{(n+k-1)!}{k! \cdot (n-1)!}. \qquad (1.6)$$

Зауваження. $\overline{C_n^k}$ – це кількість способів, якими можна розкласти k однакових предметів по n ящиках.

Значна кількість комбінаторних задач розв'язується з використанням двох основних правил комбінаторики: ***правила суми*** та ***правила добутку***.

Принцип суми. Нехай $n(A)$, $n(B)$– кількості елементів скінченних неперетинних множин A та B відповідно. Тоді для множини $C = A \bigcup B$ кількість елементів

$$n(C) = n(A) + n(B). \qquad (1.7)$$

Принцип добутку. Нехай потрібно послідовно виконати k дій, причому першу дію можна виконати n_1 способом, після чого другу дію – n_2 способами і т.д. до k –тої дії, яку можна виконати n_k способами. Тоді всі k дій можна виконати $n_1 \cdot n_2 \cdot \ldots \cdot n_k$ способами.

Наведені вище правила можуть бути розповсюджені на довільне число сумісно здійснюваних виборів.

Розв'язуючи комбінаторну задачу, перш за все потрібно відповісти на запитання − з яким із основних понять в даній ситуації ми маємо справу? Для цього відповідаємо на два питання:

1. Усі елементи множини використовуються чи ні? Якщо використовуються всі, то це перестановка.

2. Важливий порядок розміщення елементів або ні? Якщо порядок важливий, то це розміщення. У протилежному випадку − комбінація.

1.2. Класичне означення ймовірностей

Означення. ***Ймовірністю*** події *A* називають відношення кількості результатів випробування, сприятливих для *A*, до кількості всіх рівноможливих і попарно несумісних наслідків випробування:

$$P(A) = \frac{m}{n}. \tag{1.8}$$

Аксіоми ймовірності:

1. Для кожної події $A \subseteq \Omega$ *справджується нерівність:*

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

2. Імовірність достовірної події дорівнює 1: $P(\Omega) = 1.$

3.Імовірність неможливої події дорівнює 0: $P(\varnothing) = 0.$

Приклади виконання завдань

Приклад 1. Скільки перестановок можна утворити із трьох букв А, Б та В?

Розв'язання. Можна утворити шість перестановок: АБВ, АВБ, ВАБ, ВБА, БАВ, ВБА, якщо букви не повторюються, тоді маємо 3!=6.

Факторіал можна обчислити, використовуючи функцію Excel **ФАКТР**(число), яка активізується за допомогою команд

Вставка⇒Функція⇒Математические⇒ФАКТР.

Відкриється діалогове вікно (рис. 1.1),

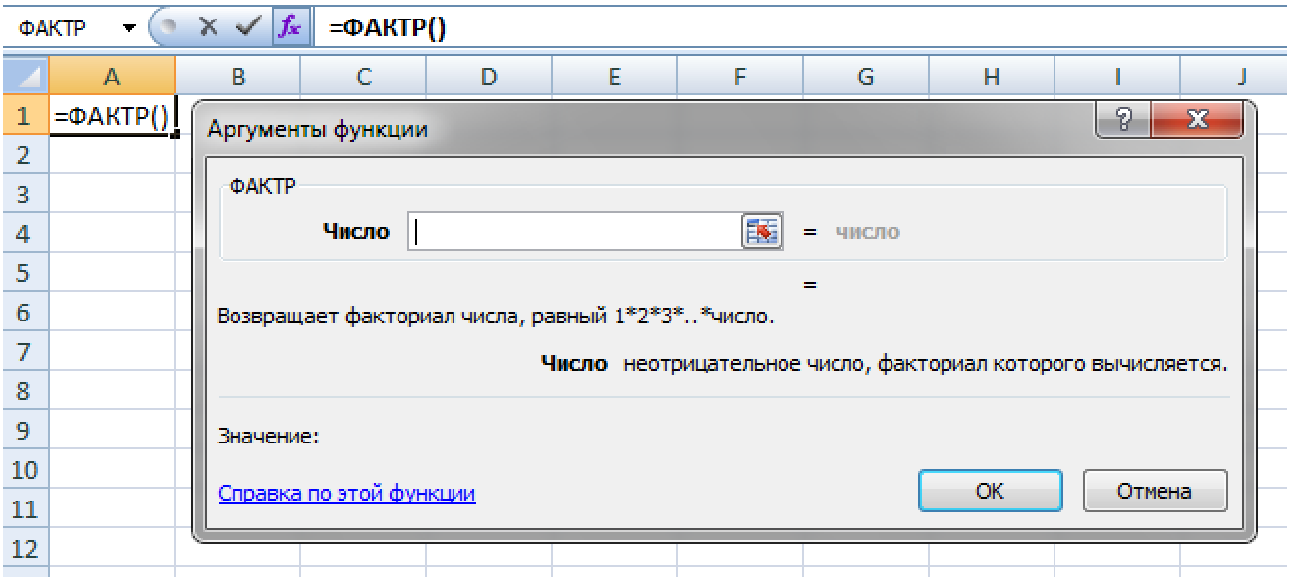


Рис.1.1. Обчислення кількості перестановок.

де в поле «число» — потрібно вказати невід'ємне ціле число (n), факторіал якого обчислюється.

Приклад 2. Скільки перестановок можна скласти із трьох букв А, Б і В?

Розв'язання. Із трьох букв А, Б і В можна скласти шість перестановок по дві букви: АБ, БА, АВ, В А, БВ, ВБ. За формулою (1.1) знаходимо $A_n^k = 3\cdot(3-1) = 6$.

Обчислення числа розміщень можна здійснити за допомогою функції **ПЕРЕСТ(**число; число_выбранных), яку викликають за допомогою команд **Вставка⇒ Функція⇒Статистические⇒ПЕРЕСТ**.

Відкриється діалогове вікно (рис.1.2), де в поле «число» — необхідно вказати невід'ємне ціле число, яке задає кількість елементів (n), а в поле «число_выбранных» − ціле число, що задає кількість елемент у кожному розміщенні (k).

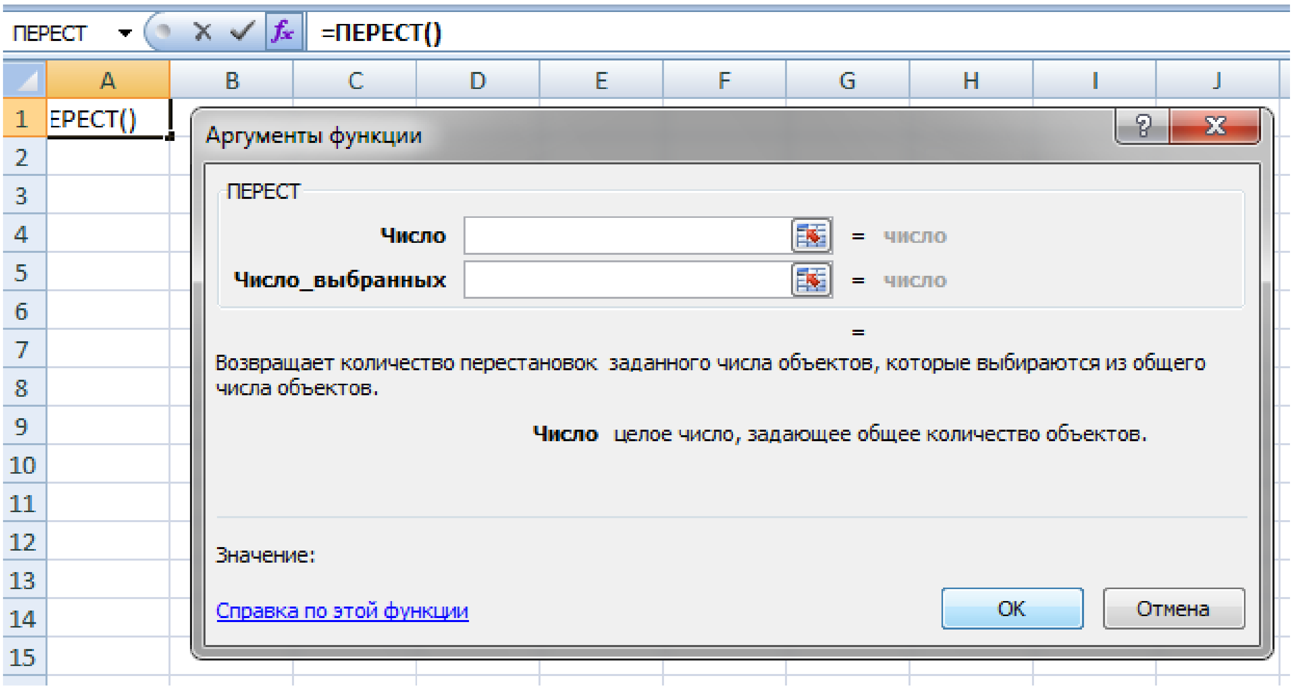


Рис. 1.2. Обчислення розміщень.

Приклад 3. Скільки комбінацій по дві букви можна скласти із трьох букв А, Б і В?

Розв'язання. Із трьох букв А, Б і В по дві букви, без врахування порядку розташування, можна скласти всього три комбінації: АБ, АВ, БВ. За формулою (1.5) знаходимо: $C_3^2 = \frac{3!}{2!(3-2)!} = 3$.

Обчислення числа комбінацій можна обчислити за допомогою функції **ЧИСЛОКОМБ** (число; число_выбранных) (рис. 1.3), яку викликають послідовністю команд

Вставка⇒ Функція⇒ Математические⇒ЧИСЛОКОМБ.

В діалоговому вікні «Аргументы функции» в поле «число» − вписують невід'ємне ціле число, яке задає кількість елементів (n), в поле «число_выбранных» − ціле число, що задає кількість елементів у кожній комбінації (k)

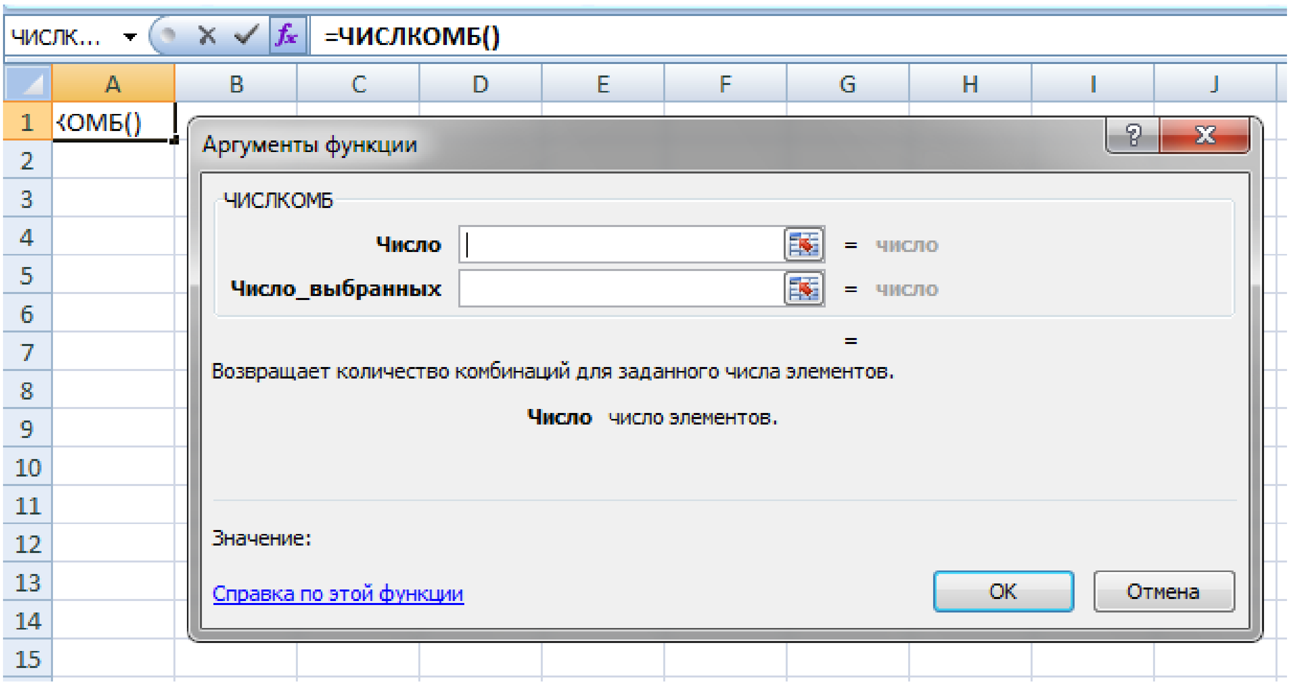


Рис. 1.3. Обчислення кількості комбінацій.

Приклад 5. Маємо дев'ять однакових карток, на кожній з яких записано одну з цифр: 1, 2, 3, 4, 5, 6, 7, 8, 9. Навмання беруть чотири картки і розкладають в один рядок. Яка ймовірність того, що при цьому дістанемо 1, 9, 7, 3?

Розв'язання. Кількість елементарних подій множини Ω. буде $n = A_9^4$. Кількість елементарних подій, що сприяють появі 1, 9, 7, 3(подія *B)* дорівнює одиниці $(m=1)$. Тоді $P(B) = \dfrac{m}{n} = \dfrac{1}{A_9^4} = \dfrac{1}{3024}$.Число n обчислено за допомогою функції **ПЕРЕСТ**(9;4).

Приклад 7. У шухляді міститься 10 однотипних деталей, 6 із яких є стандартними, а решта − бракованими. Навмання із шухляди беруть чотири деталі. Обчислити ймовірність таких випадкових подій:

A − усі чотири деталі виявляться стандартними;

B − усі чотири деталі виявляться бракованими;

C − із чотирьох деталей виявляться дві стандартними і дві бракованими.

Розв'язання.

Кількість усіх елементарних подій

$$n = C_{10}^4 = \text{ЧИСЛОКОМБ}(10;4) = 210;$$

кількість елементарних подій, що сприяють події A:

$$m_1 = C_6^4 = \text{ЧИСЛОКОМБ}(6;4) = 15;$$

кількість елементарних подій, що сприяють події B:

$$m_2 = C_4^4 = \text{ЧИСЛОКОМБ}(4;4) = 1;$$

кількість елементарних подій, що сприяють події *C:*

$$m_3 = C_6^4 \cdot C_4^2 = \text{ЧИСЛОКОМБ}(6;2) \cdot \text{ЧИСЛОКОМБ}(4;2) = 15 \cdot 6 = 90.$$

Імовірності цих подій будуть:

$$P(A) = \frac{m_1}{n} = \frac{15}{210} = \frac{1}{14},\ P(B) = \frac{m_2}{n} = \frac{1}{210},\ P(C) = \frac{m_3}{n} = \frac{90}{210} = \frac{3}{7}.$$

Виконання лабораторної роботи

Завдання до теми

Завдання 1.

Користуючись елементами комбінаторики та програмою Excel розв'язати наступні задачі (номер варіанту згідно порядкового номера за списком групи) двома способами:

– за допомогою формул комбінаторики;

– із використанням функцій Excel.

1.1. Серед n лотерейних квитків k виграшних. Навмання купили m квитків. Визначити ймовірність того, що серед них l виграшних.

Варіант	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
n	10	10	10	10	11	11	11	12	12	12	9	9	9	8	8
l	2	2	3	3	2	3	3	3	2	2	2	3	2	2	2
m	4	3	5	5	5	4	5	8	8	5	4	5	3	4	5
k	6	6	7	6	7	8	7	5	3	4	6	6	7	5	4

1.2. У ліфт k-поверхового будинку сіли n пасажирів ($n<k$). Кожний незалежно від інших із однаковою ймовірністю може вийти на довільному (починаючи з другого) поверсі. Визначити ймовірність того, що:

1) усі вийшли на різних поверхах;

2) хоча б двоє вийшли на одному поверсі.

Варіант	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
k	6	7	8	9	10	11	12	13	14	13	12	11	10	9	8
n	4	4	5	5	6	4	4	3	3	4	3	3	4	3	4

1.3. Завдання 1.1 виконати із використанням функції пакету Excel **ГИПЕРГЕОМЕТ** *(Число_успехов_в_выборке; Размер_выборки; Число_успехов_в_совокупности; Размер_совокупности).* Порівняти отримані результати.

Завдання 2.

Розв’язати наступні задачі із використанням функцій Excel

2.1. Серед 25 фірм, з яких 10 українських, а інші російські, розігрується 5 урядових контрактів. Вважається, що кожна фірма має рівні шанси для отримання контракту. Знайти ймовірність того, що принаймні дві українські фірми виграють контракт.

2.2. У фірмі десять співробітників (6 чоловіків і 4 жінки) претендують на заняття трьох вакансій. Вважають, що всі кандидатури мають рівні шанси на заняття цих вакансій. Знайти ймовірність того, що жінки не займуть жодної вакансії.

2.3. У групі з 12 бізнесменів тільки 8 мають досвід роботи у запропонованій новій галузі. Для проекту треба відібрати 4 особи. В припущенні, що відбір претендентів ведеться навмання, знайти ймовірність того, що в команду з чотирьох чоловік потраплять всі, хто має досвід роботи.

2.4. Комплект містить 7 виробів 1-го сорту, 6 – 2-го сорту і 2 вироби – 3-госорту. Випадковим способом відбирають 5 виробів. Знайти ймовірність того, що серед них не виявиться виробів 3-го сорту.

2.5. Із 15 рейсів, що виконуються з аеропорту протягом доби, 60% рейсів виконуються на власному літаковому парку. Знайти ймовірність того, що із вибраних навмання 5 рейсів рівно 3 виконуються на власному парку.

2.6. Із наявних 12 виробів, серед яких 4 нестандартних, випадковим способом вибирають 6. Знайти ймовірності того, що: а) серед них буде 2 нестандартних вироби; б) усі нестандартні вироби будуть в другій частині виробів.

2.7. У конкурсі газети бере участь 12 чоловіків та 8 жінок. Є два призових місця. Яка ймовірність того, що обидва місця займуть жінки?

2.8. З 10 літаків, що прибувають в аеропорт протягом доби, 80% мають повне комерційне завантаження. Знайти ймовірність того, що серед 5 випадковим способом узятих літаків тільки 4 мають повне завантаження.

2.9. В авіакасі було 15 квитків, серед яких 6 квитків до пункту *А*. До кінця зміни продано 8 квитків. Знайти ймовірність того, що в касі не залишилося квитків до пункту *А,* якщо ймовірність продажу кожного квитка однакова.

2.10. Партія з 30 виробів містить 10% браку. Знайти ймовірність того, що серед 7 виробів, узятих випадково: а) тільки 2 бракованих; б) жодного бракованого.

2.11. Комплект містить 12 виробів, 5 з яких коштують по 3 гривні кожний, інші – по 1 гривні. Знайти ймовірність того, що взяті навмання 4 вироби коштують разом 10 гривень.

2.12. Шість пасажирів придбали квитки на літак в одному ряді крісел із 6 місць і випадковим способом зайняли ці місця. Знайти ймовірність того, що кожний пасажир зайняв своє крісло.

2.13. Комплект із 40 виробів містить 30% нестандартних виробів, серед яких 50% − браковані. Знайти ймовірність того, що серед узятих випадковим способом чотирьох виробів: а) один бракований; б) усі браковані.

2.14. У коробці є 100 яєць, серед яких 5 неякісних. Знайти ймовірність того, що навмання взяте яйце виявиться неякісним.

2.15. У конверті серед 20 фотознімків знаходиться один, що розшукується. Навмання вийнято з конверта 10 знімків. Знайти ймовірність того, що серед них буде потрібний.

Завдання 3. Дати відповіді на подані нижче теоретичні запитання (згідно порядкового номера за списком групи парні та непарні номери відповідають на відповідно парні та непарні питання).

Теоретичні запитання до теми

1. Що називається сполуками елементів певної множини?
2. Які сполуки елементів певної множини називаються перестановками, розміщеннями та комбінаціями?
3. За якими формулами обчислюються перестановки, розміщення та комбінації?
4. За якими функціями програми Excel обчислюються перестановки, розміщення та комбінації?
5. Які є два основні правила комбінаторики? Навести приклади.
6. Як визначити, з яким із основних понять комбінаторики в даній ситуації ми маємо справу?
7. Які події називаються достовірними, неможливими, випадковими? Навести приклади.
8. Яка подія називається елементарною; складеною елементарною подією? Навести приклади.
9. Які події називаються сумісними, несумісними, рівноможливими ?
10. Дати класичне означення ймовірності випадкової події.
11. Що називають простором елементарних випадкових подій? Навести приклади.
12. Означити геометричну ймовірність та пояснити використання формули її обчислення на прикладах.

Оформлення звіту та порядок захисту

Лабораторна робота виконується на аркушах А4, в ній стисло відображаються всі формули теоретичної частини, зміст, хід роботи та отримані результати. При захисті студент повинен розуміти зміст роботи, порівнювати отримані результати проведених обчислень, а також знати відповіді на теоретичні запитання.

Лабораторна робота № 2

Основні теореми теорії ймовірностей. Послідовні незалежні випробування. Формула Бернуллі

***Мета роботи*:** Вивчення можливостей пакету Microsoft Excel для розв'язання задач, пов’язаних із основними теоремами теорії ймовірностей та серією незалежних випробувань

Теоретичні відомості

2.1. Теорема додавання для несумісних подій

Означення. Дві події називаються ***несумісними***, якщо вони не можуть одночасно настати в одному досліді, або, іншими словами, настання однієї події виключає можливість настання іншої.

Означення. Події $A_1, A_2, \ldots, A_k$ утворюють ***повну групу*** несумісних подій, якщо $A_i \cap A_j = \varnothing$ при $i \neq j \; (i,\; j = 1, 2, \ldots, k)$, $\quad \bigcup\limits_{i=1}^{k} A_i = \Omega$.

Теорема додавання для несумісних подій. *Якщо події A і B несумісні (A ⋂ B= Ø), причому відомі їх ймовірності P(A) і P(B), то ймовірність суми цих подій дорівнює сумі їх ймовірностей*

$$P(A \cup B) = P(A) + P(B). \tag{2.1}$$

Наслідок 1. Якщо події $A_1, A_2, \ldots, A_k$ попарно несумісні, то ймовірність суми подій дорівнює сумі ймовірностей цих подій:

$$P\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} P(A_i).$$

Наслідок 2. Імовірність протилежної до A події $\overline{A}$:

$$P(\overline{A}) = 1 - P(A). \tag{2.2}$$

Наслідок 3. Сума ймовірностей подій $A_1, A_2, \ldots, A_k$, що утворюють повну групу подій, дорівнює одиниці:

$$P(A_1) + P(A_2) + \ldots + P(A_k) = 1.$$

Приклад 1. В партії з 20 деталей є 16 стандартних. Знайти ймовірність того, що серед навмання взятих трьох деталей виявиться принаймні одна стандартна.

Розв'язання. Нехай подія A_1: виявиться точно одна стандартна; подія A_2: виявиться дві стандартні; подія A_3: виявиться три стандартні деталі. Ці події попарно несумісні.

Нехай подія A: серед навмання взятих трьох деталей виявиться принаймні одна стандартна. Отже, $A = A_1 \cup A_2 \cup A_3$ і за формулою (2.1) маємо

$$P(A) = P(A_1) + P(A_2) + P(A_3).$$

Обчислимо ймовірності подій A_1, A_2, A_3:

$$P(A_1) = \frac{C_{16}^1 \cdot C_4^2}{C_{20}^3} = \frac{8}{95};\ P(A_2) = \frac{C_{16}^2 \cdot C_4^1}{C_{20}^3} = \frac{40}{95};\ P(A_3) = \frac{C_{16}^3}{C_{20}^3} = \frac{28}{57}.$$

Отже, $P(A) = \dfrac{8}{95} + \dfrac{40}{95} + \dfrac{28}{57} = \dfrac{284}{285}$.

Зауваження. При розв'язуванні задач часто буває зручно переходити до протилежної події. Так, якщо подія A_0: не виявиться жодної стандартної деталі, тоді подія A_0 є протилежною до події A, і події A_0, A_1, A_2, A_3 утворюють повну групу попарно несумісних подій. Отже, будемо використовувати формулу (2.2). Обчисливши ймовірність $P(A_0) = \dfrac{C_4^3}{C_{20}^3} = \dfrac{1}{285}$, отримаємо $P(A) = 1 - \dfrac{1}{285} = \dfrac{284}{285}$.

Теорема додавання для сумісних подій

Нехай дві події A і B сумісні, причому відомі ймовірності цих подій $P(A)$, $P(B)$ та ймовірність їх сумісної появи $P(A \cap B)$.

Теорема. *Ймовірність появи принаймні однієї з двох сумісних подій A і B дорівнює сумі ймовірностей цих подій без ймовірності їх спільної появи*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{2.3}$$

Зауваження. Теорема може бути узагальнена на довільну скінченну кількість сумісних подій. Наприклад, для трьох сумісних подій

$$P(A\bigcup B\bigcup C)=P(A)+P(B)+P(C)-P(A\bigcap B)-P(A\bigcap C)-P(B\bigcap C)+$$
$$+P(A\bigcap B\bigcap C).$$

Приклад 2. В ящику 10 червоних і 6 синіх ґудзиків. Навмання виймають два ґудзики. Знайти ймовірність того, що ґудзики будуть одного кольору.

Розв'язання. Нехай подія A – ґудзики одного кольору, подія B – ґудзики червоні, подія C – ґудзики сині. Очевидно, $A=B\bigcup C$, і події B і C несумісні, тому за теоремою 1 $P(A)=P(B)+P(C)$. Знайдемо $P(B)$ та $P(C)$:

$$P(B)=\frac{C_{10}^2}{C_{16}^2}=\frac{10\cdot 9}{1\cdot 2}\cdot\frac{1\cdot 2}{16\cdot 15}=\frac{3}{8};\ P(C)=\frac{C_6^2}{C_{16}^2}=\frac{6\cdot 5}{1\cdot 2}\cdot\frac{1\cdot 2}{16\cdot 15}=\frac{1}{8}.$$

Отже, $P(A)=\frac{3}{8}+\frac{1}{8}=\frac{1}{2}.$

2.2. Теореми множення ймовірностей

Вище було сказано, що в основі означення ймовірності випадкової події лежить сукупність деяких певних умов. Якщо ж ніяких інших обмежень, крім цих умов, при обчисленні ймовірності $P(A)$ не накладається, то така ймовірність називається ***безумовною***. Якщо ж поява деякої події A відбувається за умови, що відбулась інша подія B, причому $P(B)>0$, то ймовірність появи події A називають ***умовною*** і обчислюють за формулою

$$P\left(A|B\right)=\frac{P(A\bigcap B)}{P(B)},\ (P(B)>0). \tag{2.4}$$

Означення. Ймовірність настання події A за умови, що подія B вже відбулася, називають ***умовною*** і позначають $P\left(A|B\right)$.

Тоді

$$P\left(A|B\right)=\frac{N(A\bigcap B)}{N(B)}=\frac{\frac{N(A\bigcap B)}{n}}{\frac{N(B)}{n}}=\frac{P(A\bigcap B)}{P(B)},\quad P(B)>0,$$

де $n-$ кількість наслідків стохастичного експерименту, $N(B), N(A\bigcap B)-$ кількості наслідків стохастичного експерименту, що сприяють відповідно події B та перетину події A з подією B.

Означення. ***Добутком*** двох подій *A* і *B* називають подію *C*, яка полягає в сумісній появі цих подій.

Означення. Подію *A* називають ***незалежною*** від події *B*, якщо ймовірність події *A* не залежить від того, відбулася, чи ні, подія *B* і має місце рівність

$$P\left(A|B\right)=P(A). \tag{2.5}$$

Означення. Подію *A* називають ***залежною*** від події *B*, якщо ймовірність події *A* залежить від того, відбулася чи ні подія *B*, тобто $P\left(A|B\right)\neq P(A)$.

Теорема множення ймовірностей залежних подій.

Розглянемо дві залежні події A і B, причому відомі ймовірності $P(A)$ і $P(B/A)$. Імовірність суміщення цих подій обчислюється за теоремою:

Імовірність сумісної появи двох залежних подій дорівнює добуткові ймовірності однієї з них на умовну ймовірність іншої, обчисленої за умови, що перша відбулася

$$P(AB)=P(A)\cdot P\left(B|A\right) \text{ або } P(AB)=P(B)\cdot P\left(A|\mathrm{B}\right). \tag{2.6}$$

Наслідок. Якщо події A_i $(i=\overline{1,n})$ залежні , то

$$P\left(\bigcap_{i=1}^{n}A_i\right)=P(A_1)\cdot P(A_2/A_1)\cdot P\left(A_3/\left(A_1\cap A_2\right)\right)\cdots P\left(A_n/\left(A_1\cap A_2\cap\cdots\cap A_{n-1}\right)\right),$$

тобто ймовірність сумісної появи декількох залежних подій дорівнює добуткові ймовірності однієї з них на умовні ймовірності всіх решти, причому ймовірність кожної наступної події обчислюється в припущенні, що всі попередні події відбулися.

Зокрема, для трьох залежних подій *A, B, C* маємо

$$P\left(A\cap B\cap C\right)=P(A)\cdot P(B|A)\cdot P(C|A\cap B). \tag{2.7}$$

Теорема множення ймовірностей незалежних подій. *Для двох незалежних подій A і B*

$$P(A\cap B)=P(A)\cdot P(B). \tag{2.8}$$

Імовірність настання хоча б однієї події

Нехай в результаті *n* послідовних випробувань можуть відбутися *n* подій $A_1, A_2, \ldots, A_n$, незалежних в сукупності, ймовірності яких відомі $p_1, p_2, \ldots, p_n$. Тоді ймовірності протилежних подій $\overline{A_1}, \overline{A_2}, \ldots, \overline{A_n}$ дорівнюють $q_1 = 1 - p_1, q_2 = 1 - p_2, ..., q_n = 1 - p_n$.

Теорема. *Ймовірність настання принаймні однієї з подій* $A_1, A_2, \ldots, A_n$, *незалежних в сукупності, обчислюють за формулою*

$$P(A) = 1 - q_1 q_2 \ldots q_n. \tag{2.9}$$

Наслідок. Якщо події $A_1, A_2, \ldots, A_n$ мають однакову ймовірність *p*, то ймовірність настання принаймні однієї з цих подій дорівнює

$$P(A) = 1 - q^n.$$

Приклад 3. Знайти:

1) імовірність влучити в мішень один раз при трьох пострілах − подія *A*, якщо ймовірність влучити в мішень при першому пострілі (подія B) становить 0,7, при другому (подія C) – 0,8, а при третьому (подія D) – 0,85.

2) імовірність влучити в мішень принаймні один раз при вказаних трьох пострілах.

Розв'язання.

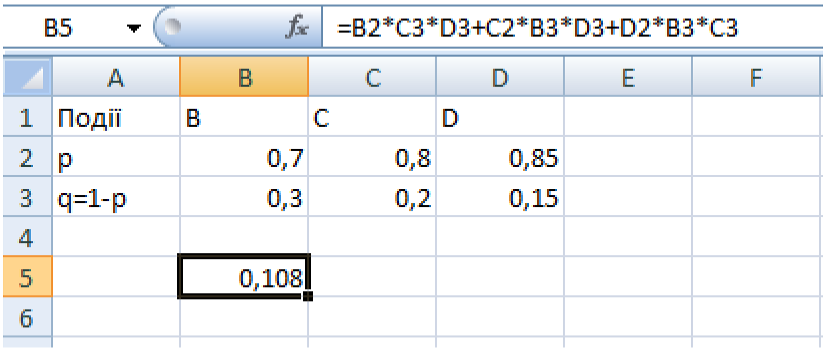
1) Події *A* – одне влучення при трьох пострілах відповідає один із варіантів: один із стрільців влучає, а два інші не влучають, тому за формулами (2.1) – теорема додавання для несумісних подій та (2.8) – теорема множення ймовірностей для незалежних подій , маємо

$$P(A) = P(B) \cdot P(\overline{C}) \cdot P(\overline{D}) + P(\overline{B}) \cdot P(C) \cdot P(\overline{D}) + P(\overline{B}) \cdot P(\overline{C}) \cdot P(D),$$

тобто

$$P(A) = 0,7 \cdot 0,2 \cdot 0,15 + 0,3 \cdot 0,8 \cdot 0,15 + 0,3 \cdot 0,2 \cdot 0,85 = 0,108.$$

Обчислення за допомогою програмного пакету Microsoft Excel виконують, ввівши відповідні ймовірності події та їй протилежної в відповідні комірки та задавши аналог функції за допомогою арифметичних дій множення та додавання (рис. 2.1).



B5 f_x =B2*C3*D3+C2*B3*D3+D2*B3*C3

	A	B	C	D	E	F
1	Події	B	C	D		
2	p	0,7	0,8	0,85		
3	q=1-p	0,3	0,2	0,15		
4						
5		0,108				
6						

Рис. 2.1. Вигляд побудованої функції до задачі.

2) Перейдемо до протилежної події – стрілець не влучив жодного разу в мішень із трьох пострілів (подія $\overline{A}$). Тоді $\overline{A} = \overline{B}\overline{C}\overline{D}$. Оскільки події *B*, *C*, *D*, а разом з ними відповідні протилежні події – незалежні, тому

$$P(\overline{A}) = P(\overline{B}) \cdot P(\overline{C}) \cdot P(\overline{D}) = 0,3 \cdot 0,2 \cdot 0,15 = 0,009,$$

звідки і знаходимо ймовірність події *A*:

$$P(A) = 1 - P(\overline{A}) = 1 - 0,009 = 0,991.$$

Відповідь. 1) 0,108; 2) 0,991.

2.3. Формули повної ймовірності та Байєса

Якщо подія *A* має складну структуру, тоді пряме знаходження її ймовірності може виявитись надто важким завданням (наприклад, за класичним означенням з використанням комбінаторного аналізу). Тому виникає питання, чи не можна подію *A* розглянути з іншими подіями, ймовірності яких відомі, і на основі розглянутих подій знайти ймовірність події *A*? Виявляється, що в багатьох випадках це є можливим і відповідь на поставлене питання дає формула повної ймовірності.

Теорема *(формула повної ймовірності)*. *Нехай випадкова подія A може відбуватись лише сумісно з однією із несумісних між собою подій* $H_1, H_2, \ldots, H_n$, *що утворюють повну групу* $\left(\bigcup_{i=1}^{n} H_i = \Omega, H_i \bigcap H_j = \varnothing \text{ при } i \neq j\right)$, *які називають гіпотезами,* $P(H_i) > 0$, $(i = 1, 2, \ldots, n)$. *Тоді ймовірність події A обчислюють за формулою*

$$P(A) = \sum_{i=1}^{n} P(H_i) P\left(A | H_i\right). \qquad (2.10)$$

Отже, ймовірність події *A* дорівнює сумі добутків ймовірностей гіпотез на умовні ймовірності події *A* при цих гіпотезах.

Якщо до випробування відомі апріорні ймовірності гіпотез $P(H_1), P(H_2), \ldots, P(H_n)$, а в результаті випробування сталася подія *A*, і $P(A) > 0$, то з урахуванням настання цієї події умовні ймовірності гіпотез обчислюють за **формулою Байєса**:

$$P\left(H_i \mid A\right) = \frac{P(H_i)P\left(A \mid H_i\right)}{P(A)}, \quad i = 1, 2, \ldots, n. \tag{2.11}$$

Приклад 4. Імовірність поразки команди групи технологів у матчі з командою групи економістів при дощовій погоді становить 0,5, а при відсутності дощу – 0,6. Імовірність дощу у день матчу становить 0,2. 1) Знайти ймовірність уникнення поразки командою групи технологів. 2) Команда групи технологів зазнала поразки. Яка ймовірність того, що матч відбувався при дощовій погоді?

Розв'язання. 1) Нехай подія *A* – поразка команди групи технологів. Утворимо дві гіпотези: H_1– під час матчу дощова погода; H_2– під час матчу не буде дощу. За умовою задачі

$$P(H_1) = 0,2; \quad P(H_2) = 0,8; \quad P\left(A \mid H_1\right) = 0,5; \quad P\left(A \mid H_2\right) = 0,6.$$

За формулою повної ймовірності (2.10)

$$P(A) = P(H_1) \cdot P\left(A \mid H_1\right) + P(H_2) \cdot P\left(A \mid H_2\right)$$

знаходимо ймовірність поразки команди групи технологів:

$$P(A) = 0,2 \cdot 0,5 + 0,8 \cdot 0,6 = 0,58.$$

Тоді ймовірність уникнення поразки становить:

$$P(\overline{A}) = 1 - P(A) = 1 - 0,58 = 0,42.$$

Зауваження. Обчислення ймовірності за формулою (2.10) можна виконати в за допомогою вбудованої математичної функції **СУММПРОИЗВ** (Массив1, Масив 2, Масив 3) (рис. 2.2),

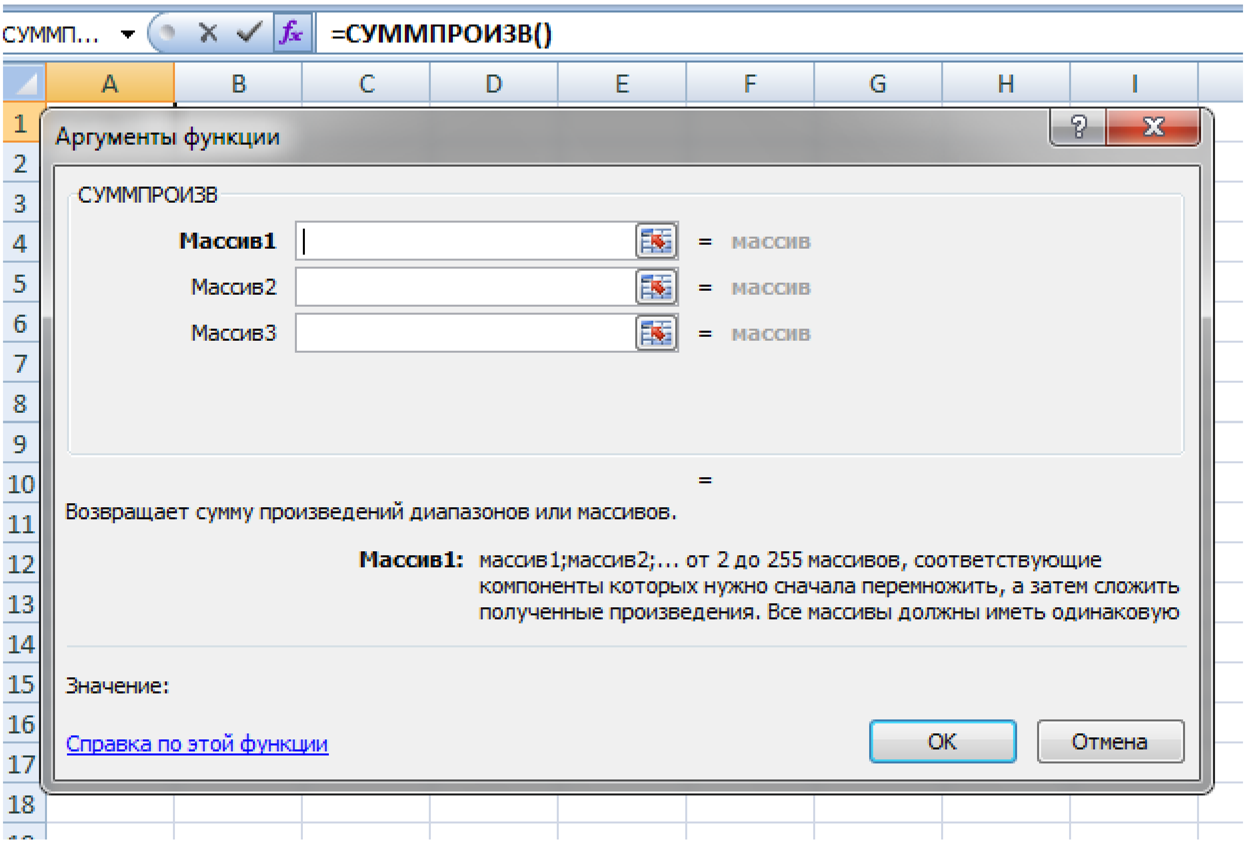


Рис. 2.2. Вигляд функції на екрані.

задаючи для даної задачі тільки 2 масиви даних, один з них є ймовірностями прийнятих гіпотез, а другий – відповідними умовними ймовірностями (рис. 2.3).

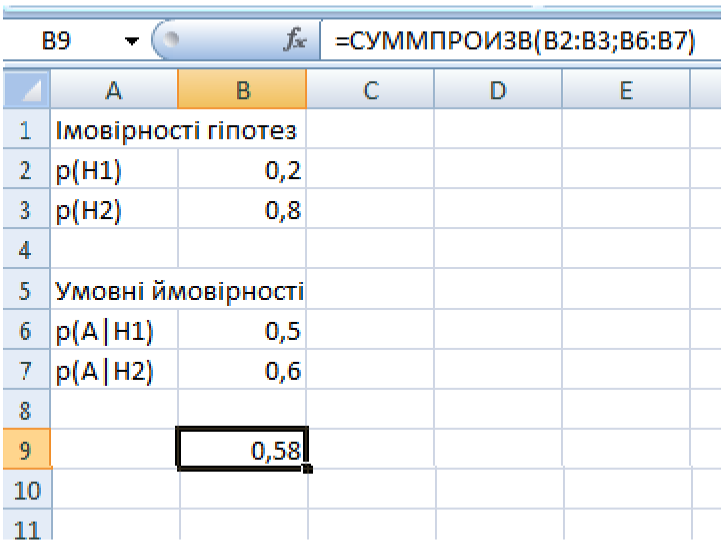
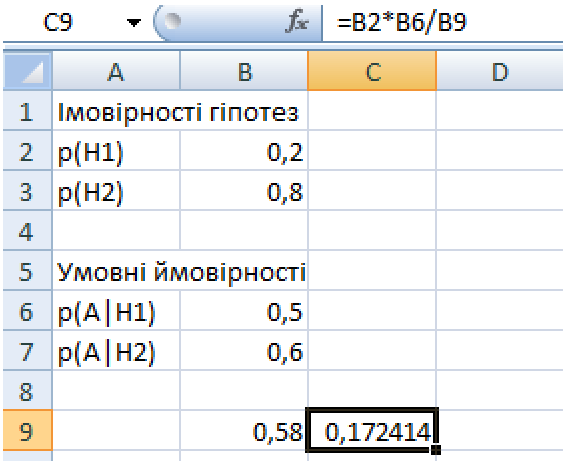


Рис. 2.3. Обчислення за формулою повної ймовірності.

2) За формулою Байєса знаходимо ймовірність того, що йшов дощ під час матчу, в якому команда групи технологів зазнала поразки:

$$P(H_1 \mid A) = \frac{P(H_1) \cdot P(A \mid H_1)}{P(A)} = \frac{0,2 \cdot 0,5}{0,58} = 0,17.$$

Обчислення за допомогою описаної функції в Excel подано на рис. 2.4.



C9 =B2*B6/B9

	A	B	C	D
1	Імовірності гіпотез			
2	p(H1)	0,2		
3	p(H2)	0,8		
4				
5	Умовні ймовірності			
6	p(A|H1)	0,5		
7	p(A|H2)	0,6		
8				
9		0,58	0,172414	

Рис. 2.4. Обчислення за формулою Байєса.

Відповідь. 1) 0,58; 2) 0,17.

2.4. Послідовні незалежні випробування. Формула Бернуллі

Означення. Якщо n незалежних випробувань проводити в однакових умовах і ймовірність появи події A в усіх випробуваннях однакова та не залежить від настання або ненастання A в інших випробуваннях, то таку послідовність випробувань називають ***схемою Бернуллі.***

Імовірність того, що подія $A\left(P(A)=p\right)$ в n спробах з′явиться рівно k разів, а в решті $n-k$ спроб з′явиться протилежна подія $\overline{A}$, $P(\overline{A})=q$ $(p+q=1)$, за теоремою множення ймовірностей незалежних подій дорівнює $p^k \cdot q^{n-k}$. При цьому подія A в n спробах може з′явитися рівно k разів в різних комбінаціях, кількість яких C_n^k. Оскільки всі комбінації подій є подіями несумісними і не важливо, в якій послідовності з′явиться подія A або подія $\overline{A}$, то, застосовуючи теорему додавання ймовірностей несумісних подій, отримаємо ***формулу Бернуллі***

$$P_n(k) = C_n^k p^k q^{n-k}. \tag{2.12}$$

Отже, ймовірність $P_n(k)$ того, що подія A настане k разів в n випробуваннях $(0 \le k \le n)$ обчислюють за ***формулою Бернуллі***:

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

Число k_0, при якому ймовірність $P_n(k_0)$ найбільша, називають ***найімовірнішим*** числом настання події A. Його визначають співвідношенням

$$np - q \le k_0 \le np + p \text{ або } (n+1)p - 1 \le k_0 \le (n+1)p. \qquad (2.13)$$

Число k_0 повинно бути цілим. Якщо $(n+1)p$ – ціле число, тоді найбільше значення ймовірність $P_n(k)$ має при двох числах $k_1 = (n+1)p - 1$ та $k_2 = (n+1)p$.

Зауваження. Імовірність появи події A в n випробуваннях схеми Бернуллі ***менш ніж*** m ***разів*** знаходять за формулою

$$P_n(k < m) = P_n(0) + P_n(1) + \cdots + P_n(m-1).$$

Імовірність появи події A ***не менш ніж*** m ***разів*** знаходять за формулою $P_n(k \ge m) = P_n(m) + P_n(m+1) + ... + P_n(n)$ або за такою формулою

$$P_n(k \ge m) = 1 - \sum_{k=0}^{m-1} P_n(k).$$

Імовірність появи події A ***хоча б один раз у n випробуваннях*** доцільно обчислювати за формулою $P_n(1 \le m \le n) = 1 - q^n$.

Приклад 5. Обчислити ймовірність народження 6-ти дівчаток в багатодітній сім’ї, яка налічує 10 дітей, якщо ймовірність народження дівчинки дорівнює 0,48.

Розв’язання. Імовірність народжених дівчаток із 10 діток обчислюємо за формулою Бернуллі

$$P_{10}(6) = C_{10}^6 \cdot 0,48^6 \cdot 0,52^4 \approx 0,18.$$

Використаємо функцією Excel категорії «Статистические» **БИНОМРАСП** (*число_успехов; число_испытаний; вероятность_успеха;*

интегральная): *число_успехов*= 6; *число_испытаний* = 10; *вероятность_успеха* = 0,48; *интегральная* – 0 (рис. 2.5).

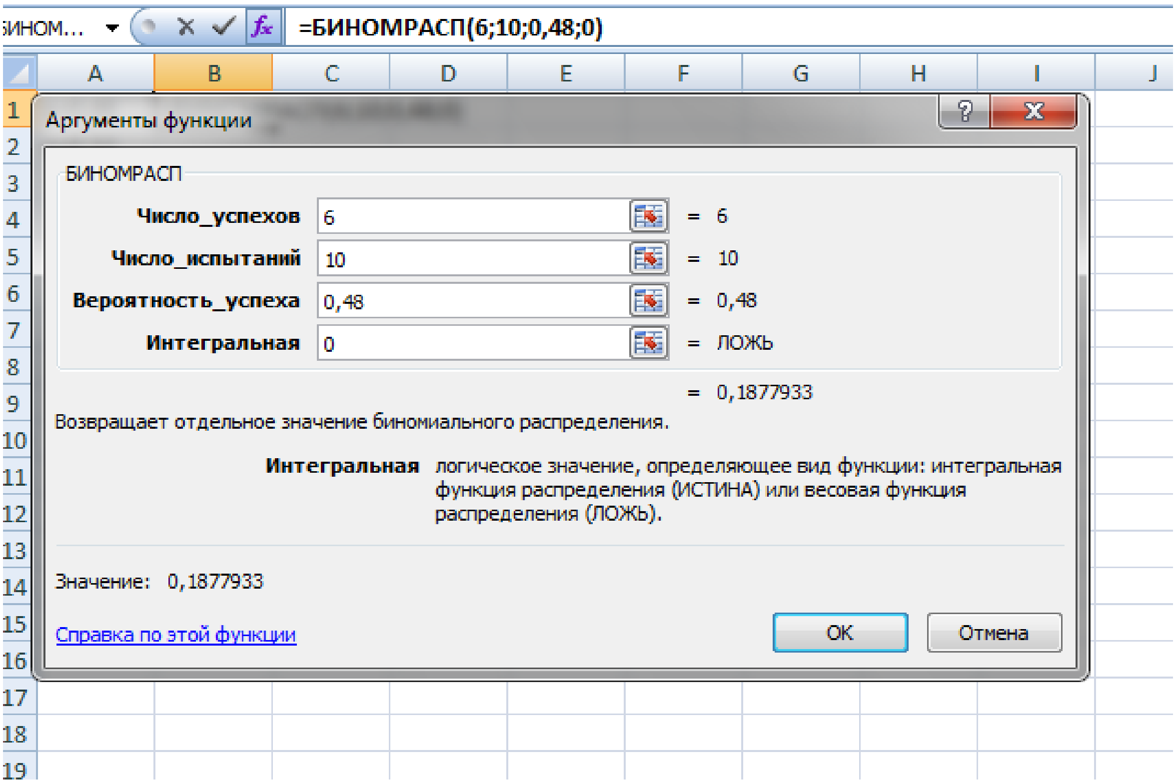


Рис. 2.5. Використання формули Бернуллі.

Відповідь. $P_{10}(6) = 0{,}1878$.

Виконання лабораторної роботи

Завдання до теми

Користуючись теоретичними знаннями та засобами програмного забезпечення Excel, розв’язати завдання 1-4.

Завдання 1. У двох партіях k_1 та k_2% (відсотків) доброякісних виробів відповідно. Навмання вибирають по одному виробу з кожної партії. Яка ймовірність виявити серед них:

1) хоча б один бракований виріб;
2) два браковані вироби;
3) один доброякісний та один бракований виріб?

Варіант	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
k_1	71	78	87	72	79	86	73	81	85	74	82	84	75	83	75
k_2	47	39	31	46	38	32	45	37	33	44	36	34	43	35	42

Завдання 2.

2.1. Фірма має можливість отримати два контракти. Імовірність отримання першого контракту дорівнює 0,9, а другого – 0,8. Вважаючи ці події незалежними, знайти ймовірності подій: а) фірма отримає обидва контракти; б) фірма отримає принаймні один контракт.

2.2. Надійність лінії зв'язку між об'єктами (ймовірність безвідмовної роботи за певний час) дорівнює 0,75. Для підвищення якості зв'язку встановлена резервна лінія з надійністю 0,65. Визначити надійність зв'язку з резервною лінією.

2.3. Через метеорологічні умови літак було відправлено на запасний аеродром, при наближенні до якого на борту літака залишалося палива на 3 заходи на посадку. Ймовірність посадки літака при першому заході дорівнює 0,8, при другому – 0,95, при третьому – 0,995. Знайти ймовірність благополучної посадки літака.

2.4. Ймовірність виготовлення виробу вищого сорту на першому верстаті становить 0,7, на другому – 0,8. На першому верстаті виготовлено 2 вироби, на другому – 3. Знайти ймовірність того, що всі вироби належать до вищого сорту.

2.5. Відомо, що в деякому регіоні 40% компаній мають у штаті юриста і 80% компаній мають у штаті економіста. Вважаємо, що ці дві події незалежні. Знайти ймовірність того, що фірма має в штаті економіста і юриста.

2.6. Радіостанція аеропорту надсилає 3 повідомлення для літака. Ймовірність прийому літаком першого повідомлення дорівнює 0,6; другого – 0,65; третього – 0,7. Знайти ймовірність того, що літаком прийняті: а) тільки два повідомлення; б) усі три повідомлення.

2.7. З опитаних бізнесменів 80 % віддають перевагу зберігати гроші в банку (подія A), 60 % вкладає гроші в цінні папери (подія B), 50 % одночасно тримають гроші в банку та у вигляді цінних паперів. Результати опитування подані таблицею.

	A	$\overline{A}$	Разом
B	0,5		0,5
$\overline{B}$			
Разом	0,8		1

Заповнити таблицю до кінця. Знайти ймовірність того, що навмання обраний бізнесмен тримає гроші в банку або у вигляді цінних паперів.

2.8. Власник кафе звернув увагу на те, що 75% відвідувачів беруть після обіду десерт, 60% – соки, 40% – і те, і інше. Знайти ймовірність того, що навмання вибраний відвідувач, який замовив десерт, візьме ще і сік.

2.9. Менеджером з'ясовано, що 40% співробітників чудово справляються зі своєю роботою, 30% активно використовують в своїй роботі комп'ютер, 20% задовольняють обом вимогам. Знайти ймовірність того, що навмання обраний співробітник, використовує комп'ютер при умові, що відмінно справляється зі своєю роботою.

2.10. Знайти ймовірність того, що навмання взятий виріб виявиться першосортним, якщо 4% усіх виробів є браком, а першосортні вироби складають 75% від усіх небракованих.

2.11. Відділ технічного контролю перевіряє вироби. Ймовірність того, що виріб виявиться стандартним, дорівнює 0,9. Знайти ймовірність того, що: а) із трьох перевірених виробів тільки один буде нестандартним; б) четвертий з перевірених виробів виявиться нестандартним.

2.12. Імовірність одного влучення у ціль при одному залпі з двох гармат дорівнює 0,38. Знайти ймовірність влучення у ціль при одному пострілі з першої гармати, якщо відомо, що для другої гармати ймовірність такої події дорівнює 0,8.

2.13. Три економісти незалежно один від одного заповнюють одні і ті ж статистичні дані. Імовірність того, що перший економіст зробить помилку дорівнює 0,1; для другого і третього такі ймовірності відповідно дорівнюють 0,15 та 0,2. Знайти ймовірність того, що хоча б один із економістів припустить помилку; два економісти помиляться.

2.14. Пристрій містить два елементи, які працюють незалежно. Імовірності відмов елементів дорівнюють 0,05 та 0,08. Знайти ймовірність відмови пристрою, якщо для цього досить, щоб відмовив хоча б один елемент.

2.15. Для сигналізації про аварію встановлено два незалежно працюючих пристрої. Імовірність того, що при аварії пристрій спрацює дорівнює 0,95 для першого пристрою та 0,9 для другого. Знайти ймовірність того, що при аварії спрацює лише один пристрій.

Завдання 3.

3.1. Серед виробів, які випускаються заводом, 96 % відповідають стандарту. Спрощена схема контролю визнає доброякісною стандартну продукцію з ймовірністю 0,98 і нестандартну – з ймовірністю 0,05. Знайти: а) ймовірність того, що навмання взятий виріб пройде спрощений контроль; б) за умови проходження спрощеного контролю ймовірність того, що він відповідає стандарту.

3.2. Інвестор вкладає свої гроші в акції великого ризику, в акції сприйнятливого ризику та акції без ризику в пропорціях 10%, 30%, 60%. Ймовірність отримання прибутку від цих акцій складають 0,6; 0,75 та 1 відповідно. Знайти: а) ймовірність того, що інвестор отримав прибуток; б) якщо інвестор отримав прибуток, то яка ймовірність того, що він отриманий від акцій без ризику?

3.3. Економічний факультет провів обстеження працевлаштування своїх випускників. Імовірність того, що людина, яка працює в сфері бізнесу та має прибуток вище N грн., складає 0,9, а поза сферою бізнесу – 0,3. З'ясовано, що 80 % випускників працюють в сфері бізнесу. Знайти: а) ймовірність того, що навмання обраний випускник має прибуток вище N грн.; б) за умови отримання прибутку випускником вище N грн. якою є ймовірність, що він працює в сфері бізнесу?

3.4. Задачу розв'язують самостійно 2 відмінники, 3 посередні студенти та 5 студентів, що вчаться добре. Ймовірність розв'язання задачі відмінником дорівнює 0,9, добрим студентом – 0,8 та посереднім – 0,5. До дошки навмання викликається один із студентів. 1) Знайти ймовірність того, що він розв'язав задачу. 2) Викликаний студент розв'язав задачу. Знайти ймовірність того, що він є: а) відмінником; б) посереднім студентом.

3.5. Підприємець тримає свої гроші на валютному та гривневому рахунках у пропорції 3:2. Ймовірність падіння курсу валют в наступному місяці складає 0,02, а гривні – 0,06. Знайти: а) ймовірність того, що гроші підприємця девальвувалися; б) відомо, що гроші підприємця девальвувалися і яка ймовірність того, що це трапилося через падіння курсу гривні?

3.6. Фабрика виготовляє однотипну продукцію на трьох конвеєрних лініях, які мають однакову продуктивність. На першій лінії виробляється продукція тільки першого сорту, на другій лінії продукція першого сорту становить 90 %, а на третій – 85 %. а) Знайти ймовірність того, що випадковим способом узятий виріб буде першосортним; б) Випадково взятий виріб виявився першосортним. Знайти ймовірність того, що він виготовлений на третій лінії.

3.7. Менеджер з інвестицій передбачає в наступному році три варіанти розвитку економічної ситуації: високий зріст, відсутність зростання та спад зростання. Ймовірності цих подій складають 0,6; 0,3 та 0,1 відповідно. Очікується отримання прибутку з наявного активу. Ймовірність отримання прибутку складає: випадок високого зростання – 0,8; випадок відсутності зростання – 0,6; випадок спаду – 0,1. а) Знайти ймовірність того, що отримано прибуток з наявного активу; б) Виявилось, що отримано прибуток з наявного активу і якою є ймовірність того, що цей прибуток отримано в умовах високого зростання економіки?

3.8. Два робітники виготовили по однаковій кількості деталей. Ймовірність бракованої продукції, виготовленої першим робітником, складає 5%, а другим – 1%. а) Знайти ймовірність того, що узята навмання деталь буде бракованою. б) Відділ технічного контролю виявив браковану деталь. Якою є ймовірність того, що вона виготовлена першим робітником?

3.9. На фабриці перша машина виробляє 40 %, а друга – 60 % усієї продукції. У середньому 9 із 1000 одиниць продукції, виготовленої першою машиною, виявляється браком, а для другої машини брак становить 2 одиниці на 500 одиниць продукції. а) Знайти ймовірність браку продукції, виготовленої на фабриці. б) Деяка одиниця продукції, яка вибрана випадковим способом із даної продукції фабрики, виявилася браком. На якій з машин вона ймовірніше за все виготовлена?

3.10. З першого верстата-автомата на складання приладів надходить 40 %, з другого – 30%, з третього − 20% і з четвертого – 10% однакових деталей, виготовлених у цеху. Серед деталей, виготовлених на першому, другому, третьому та четвертому верстатах, трапляється відповідно 2%, 1%, 0,5% і 0,2% браку. а) Знайти ймовірність того, що випадково взята деталь, що надійшла на складання приладів, не є бракованою. б) Надійшла не бракована деталь і на якому верстаті вона ймовірніше за все виготовлена?

3.11. Для участі математичній олімпіаді з груп № 101, 102 і 103 запрошено відповідно 4, 5 і 6 студентів. Ймовірність того, що переможцем олімпіади стане студент із першої, другої чи третьої групи, відповідно дорівнює 0,9; 0,8 і 0,7. а) Знайти ймовірність перемоги на олімпіаді студента однієї зі згаданих груп. б) Один із студентів зазначених груп виявився переможцем. До якої групи він ймовірніше за все належить?

3.12. Податкові інспектори роблять перевірку діяльності підприємств. Перший інспектор обслуговує 40 підприємств, серед яких 25% не мають заборгованостей, другий – 60 підприємств, із них 40% без заборгованостей. Знайти ймовірність того, що навмання обране підприємство: а) не має заборгованості; б) підприємство, що не має заборгованості, перевіряв перший інспектор.

3.13. До каси підприємства надійшли банкноти у пачках від банків: 50 пачок від першого банку і 70 – від другого. Ймовірності помилки касирів першого банку становлять 0,05, а другого – 0,02. Знайти: а) ймовірність того, що навмання обрану пачку сформовано без помилок; б) пачку без помилок сформували касири другого банку.

3.14. У збірній команді Видавничо-поліграфічного інституту 40% складають студенти кафедри технології, 45% – кафедри репрографії і 15% – студенти-економісти. Після першого дня змагань невдача трапилась відповідно з 5%, 4% та 2% студентів. Нехай випадково зустрінутий студент виявивсь невдахою. Яка ймовірність того, що він: а) студент-репрограф; б) студент-економіст?

3.15. У піраміді встановлено 10 гвинтівок, з яких 4 мають оптичний приціл. Імовірність того, що стрілець влучить у мішень при пострілі з гвинтівки з оптичним прицілом, дорівнює 0,95; для гвинтівки без оптичного прицілу ця ймовірність дорівнює 0,8. Стрілець влучив у мішень з навмання взятої гвинтівки. Що є більш імовірним: стрілець робив постріл із гвинтівки з оптичним прицілом чи без нього?

Завдання 4.

Імовірність виграшу в лотерею на один квиток дорівнює p. Куплено n квитків. Знайти ймовірність виграшу на k квитків.

Варіант	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
p	0.7	0.6	0.2	0.7	0.9	0.8	0.4	0.3	0.1	0.5	0.4	0.3	0.5	0.3	0.2
n	7	9	10	6	8	9	5	7	30	14	6	4	13	25	20
k	3	5	4	3	6	4	2	5	10	5	5	2	4	11	5

Завдання 5. Дати відповіді на подані нижче теоретичні запитання (для звіту дати письмово відповіді на 3 питання).

Теоретичні запитання до теми

1. Означити та пояснити сумісні та несумісні події.
2. Пояснити на прикладах залежні та незалежні події та обчислення їх ймовірностей.
3. Навести приклади повної групи несумісних подій.

4. Сформулювати теореми додавання ймовірностей для сумісних та несумісних подій. Навести приклади їх використаня.

5. Сформулювати теореми множення ймовірностей для залежних та незалежних подій. Навести приклади їх використаня.

6. Сформулювати теорему обчислення повної ймовірності, пояснити на прикладі.

7. Навести приклад використання математичної функції Excel для обчислень за формулою повної ймовірності.

8. Записати формулу Байєса та пояснити її використання.

9. Яку послідовність незалежних випробувань називають схемою Бернуллі?

10. Записати і пояснити на прикладі формулу Бернуллі.

Оформлення звіту та порядок захисту

Лабораторна робота виконується на аркушах А4, в ній стисло відображаються формули теоретичної частини, хід роботи та отримані результати. При захисті студент повинен розуміти зміст роботи, порівняти отримані результати проведених обчислень, а також знати відповіді на теоретичні запитання. Також студент здає електронний варіант проведених обчислень практичних задач.

Лабораторна робота № 3

Формула Бернуллі, граничні теореми формули Бернуллі

***Мета роботи*:** Вивчення можливостей пакету Microsoft Excel для розв'язання задач, пов’язаних із серією незалежних випробувань

Теоретичні відомості

3.1. Граничні теореми формули Бернуллі

Як відомо, імовірність $P_n(k)$ того, що подія *A* настане *k* разів в серії незалежних *n* випробувань $(0 \le k \le n)$ обчислюють за ***формулою Бернуллі*:**

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

Якщо проводяться випробування за схемою Бернуллі та числа n і k великі, то обчислення ймовірності за формулою Бернуллі викликає певні труднощі. У таких випадках для обчислення цих імовірностей застосовують асимптотичні (наближені) формули, які випливають із ***локальної та інтегральної теорем Муавра-Лапласа*** та ***граничної тереми Пуассона.*** Назва «гранична» в обох випадках пов’язана з тим, що ці теореми встановлюють поведінку ймовірності $P_n(k)$ або $P_n(k_1 \le k \le k_2)$ за певних умов, до яких обов’язково входить умова $n \to \infty$.

З огляду на це, за досить великих значень *n* та за достатньо малих p замість формули Бернуллі часто використовують наближені асимптотичні формули.

3.1.1.Формула Пуассона

Теорема (теорема Пуассона). *Якщо* $n \to \infty$ *і* $p \to 0$ *так, що* $np \to \lambda, 0 < \lambda < \infty$, *то*

$$P_n(k) = C_n^k p^k q^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda} \qquad (3.1)$$

для будь-якого постійного $k = 0, 1, 2, \cdots$.

Наслідок. Ймовірність появи події A k разів в n випробуваннях схеми Бернуллі знаходять за наближеною *формулою Пуассона*

$$P_n(k) \approx \frac{\lambda^k}{k!} e^{-\lambda}, \tag{3.2}$$

де $\lambda = np$.

Ця формула дає досить точне наближення при значеннях *p*, близьких до нуля $(p < 0{,}1)$, тобто для подій, що рідко трапляються і для достатньо великих *n* $(npq \le 9)$.

3.1.2. Локальна теорема Муавра-Лапласа

Теорема (локальна теорема Муавра-Лапласа). *Якщо ймовірність* p *появи події* A *в кожній спробі стала і така, що* $0 < p < 1$*, то ймовірність числа* k *появ події в* n *спробах обчислюється за наближеною рівністю*

$$P_n(k) \approx \frac{1}{\sqrt{npq}} \varphi\left(\frac{k - np}{\sqrt{npq}}\right), \text{ де } \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{3.3}$$

Формула (3.3) дає добре наближення, якщо *n* достатньо велике, *p* та *q* не дуже близькі до нуля, $npq > 9$.

3.1.3. Інтегральна теорема Муавра-Лапласа

Теорема (інтегральна теорема Муавра-Лапласа). *Імовірність того, що в n незалежних випробуваннях схеми Бернуллі, в яких подія A може відбутись з ймовірністю p, подія A відбудеться не менше* k_1 *та не більше* k_2 *разів, наближено рівна:*

$$P_n(k_1 \le k \le k_2) \approx \Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right), \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt. \tag{3.4}$$

Формула (3.4) дає добре наближення, якщо *n* достатньо велике, *p* та *q* не дуже близькі до нуля, $npq > 9$. Для всіх значень $x \ge 5$ можна вважати $\Phi(x) \approx 0{,}5$.

Користуючись формулою Бернуллі, її граничними теоремами та програмою Excel розв'язати наступні задачі:

Приклад 1. Імовірність влучити в мішень при одному пострілі дорівнює 0,8. Знайти найімовірніше число влучень із 6 пострілів та відповідну ймовірність.

Розв'язання. Знайдемо величину виразу $np + p = 0{,}8 \cdot 6 + 0{,}8 = 5{,}6$ – не ціле число. Тоді найбільше ціле число, яке не перевищує 5,6, дорівнює 5. Таким чином, найбільш імовірне ціле число влучень $k_0 = 5$. Імовірність п'яти влучень із шести пострілів обчислюємо за формулою Бернуллі

$$P_6(5) = C_6^5 \cdot 0{,}8^5 \cdot 0{,}2 = 0{,}39.$$

Використаємо функцією Excel категорії «Статистические» **БИНОМРАСП** (*число_успехов; число_испытаний; вероятность_успеха; интегральная*) : *число_успехов* = 5*; число_испытаний* = 6*; вероятность_успеха* = 0,8*; интегральная* −0.

Відповідь. $k_0 = 5$, $P_6(5) = 0{,}39$.

Приклад 2. Імовірність влучити в мішень при одному пострілі дорівнює 0,8. Знайти ймовірність того, що із 100 пострілів число влучень буде: а) точно 90; б) лежить у межах від 75 до 85.

Розв'язання. Згідно з умовою задачі $np = 80 > 5$ і $nq = 20 > 5$, тому можна скористатися формулами Муавра-Лапласа.

а) За локальною формулою Муавра-Лапласа (3.3) знаходимо:

$$x = \frac{90 - 0{,}8 \cdot 100}{\sqrt{100 \cdot 0{,}8 \cdot 0{,}2}} = \frac{10}{4} = 2{,}5.$$

Для $x = 2{,}5$ знаходимо (за таблицею значень $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$)

$\varphi(x) = 0{,}0175$ та, відповідно, $P_{100}(90) = \frac{0{,}0175}{4} = 0{,}004$.

Знаючи функцію Excel категорії «Статистические» **БИНОМРАСП** (*число_успехов; число_испытаний; вероятность_успеха; интегральная*), обчислення проводять набагато швидше (рис. 3.1).

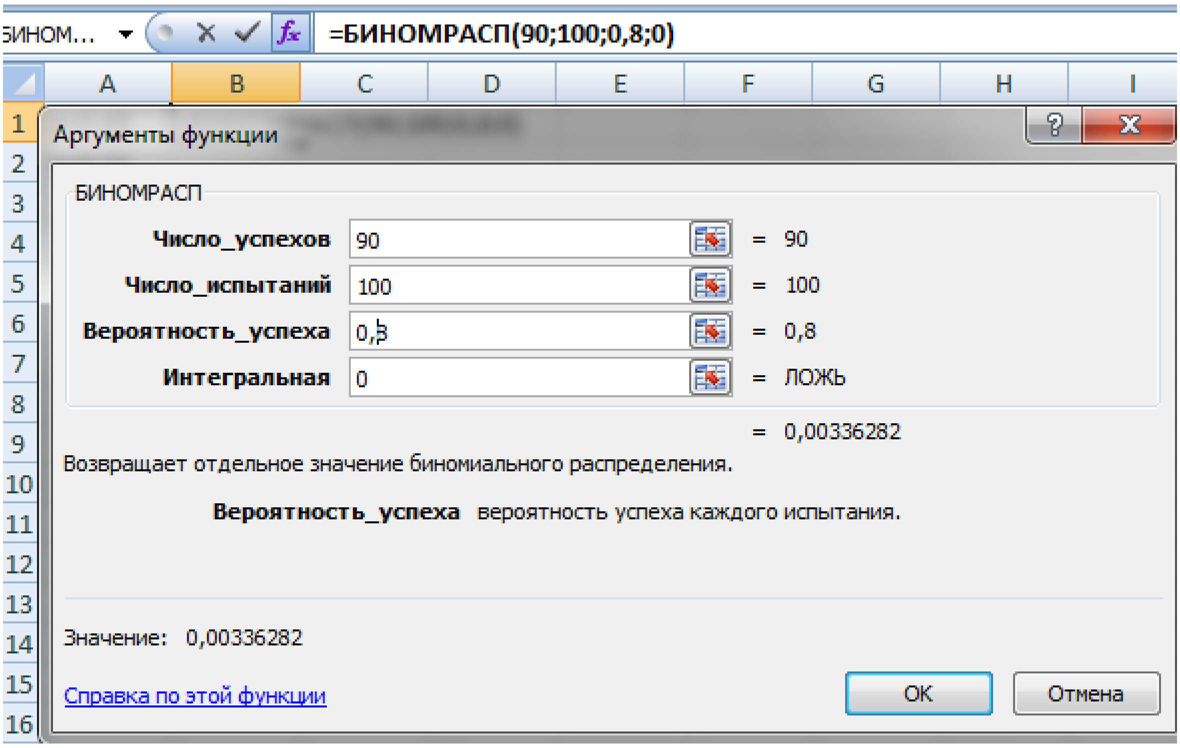


Рис. 3.1. Обчислення за локальною формулою Муавра-Лапласа.

б) За інтегральною формулою Муавра-Лапласа (3.4) знаходимо:

$$x_1 = \frac{75-80}{4} = -1{,}25; \quad x_2 = \frac{85-80}{4} = 1{,}25.$$

Для $x = 1{,}25$ знаходимо за таблицею значень функції Лапласа $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}dt}$:

$$\Phi(1{,}25) = 0{,}39435.$$

Тоді $\Phi(-1{,}25) = -0{,}39435$. Шукану ймовірність знаходимо як різницю

$$P_{100}(75 < k < 85) = \Phi(1{,}25) - \Phi(-1{,}25) = 0{,}39435 - (-0{,}39435) = 0{,}79.$$

Зауваження. Можна також провести обчислення із використанням тієї ж функції Excel категорії «Статистические» **БИНОМРАСП** (*число_успехов; число_испытаний; вероятность_успеха; интегральная*), де для пункту а) брали *интегральная* – 0, але для пункту б) слід взяти значення *интегральная* – 1 (повертає значення: кількість успішних випробувань не менше значення аргументу *число_испытаний* , тобто обчислюють два значення функції для $k = 75, k = 85$ та знаходять їх різницю (рис. 3.2).

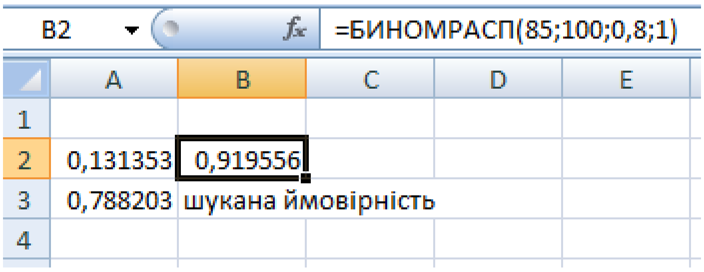


Рис. 3.2. Обчислення для інтегральної формули Муавра-Лапласа.

Відповідь. а) 0,004; б) 0,79.

Приклад 3. Словник має 1500 сторінок. Імовірність друкарської помилки на одній сторінці дорівнює 0,001. Знайти ймовірність того, що в словнику:

а) буде точно 3 помилки;

б) не буде жодної помилки;

в) буде хоча б одна помилка.

Розв'язання. а) За умовою ймовірність друкарської помилки на одній сторінці $p = 0{,}001 < 0{,}01$, а добуток

$$npq = 1500 \cdot 0{,}001 \cdot 0{,}999 = 1{,}49 < 5 \; (np = 1500 \cdot 0{,}001 = 1{,}5 < 20).$$

Тоді за формулою Пуассона (3.2) (за таблицею значень $p_k = \frac{\lambda^k}{k!} e^{-\lambda}$) обчислюємо $P_{1500}(3) = \frac{1{,}5^3}{3!} \cdot e^{-1{,}5} \approx 0{,}125$.

Скористаємось функцією Excel категорії «Статистические» **ПУАССОН** (x; *среднее; интегральная);* для таких параметрів: x − кількість сприятливих подій; *среднее* − середнє значення $\lambda = n \cdot p$; *интегральная* − 0 або 1. В умові задачі $x = 3$; *среднее* $\lambda = 1{,}5$; *интегральная* – 0, тобто обчислення точної кількості подій – 3 помилки для словника із 1500 сторінками (рис. 3.3).

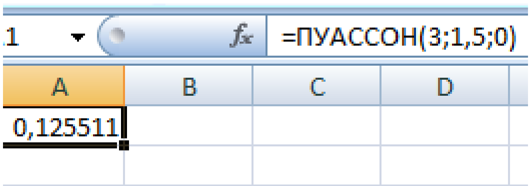


Рис.3.3. Обчислення за допомогою функції **ПУАССОН.**

б) Імовірність того, що в словнику не буде жодної помилки, тобто $k = 0$, обчислюємо за тією ж формулою Пуассона

$$P_{1500}(0) = \frac{1{,}5^0}{0!} \cdot e^{-1{,}5} = \frac{1}{1 \cdot \sqrt{e^3}} \approx \frac{1}{4{,}48} \approx 0{,}223.$$

в) подія A – у словнику буде хоча б одна помилка, є протилежною до події – у словнику немає жодної помилки. Тому

$$P(A) = 1 - P_{1500}(0) = 1 - 0{,}223 = 0{,}777.$$

Зауваження. Якщо аргумент *«интегральная»* приймає значення 1(«істина»), тоді функція **ПУАССОН** повертає ймовірність того, що кількість винадкових подій була в діапазоні від 0 до x включно.

Відповідь. а) 0,125; б) 0,223; в) 0,777.

Виконання лабораторної роботи

Завдання до теми

Користуючись теоретичними знаннями та засобами програмного забезпечення Excel, розв’язати завдання 1-2.

Завдання 1.

Імовірність виграшу в лотерею на один квиток дорівнює p. Куплено n квитків. Знайти найбільш імовірне число виграшних квитків та відповідну ймовірність.

Варіант	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
p	0.3	0.3	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4
n	10	14	13	12	11	15	11	13	14	10
Варіант	**11**	**12**	**13**	**14**	**15**					
p	0.4	0.4	0.5	0.4	0.5					
n	12	15	12	12	11					

Завдання 2.

Розв’язати наступні задачі (номери задач за варіантами у викладача).

Номер варіанту	Номери завдань	Номер варіанту	Номери завдань
1	1, 4, 12	9	6, 8, 19
2	2, 6, 15	10	9, 17, 14
3	3, 8, 14	11	3, 4, 15
4	5, 7, 16	12	6, 8, 19
5	9, 17, 20	13	7, 15, 18

6	10, 11, 19	14	1, 10, 5
7	3, 4, 15	15	9, 8, 15
8	13, 2, 18		

2.1. Серед великого числа виробів, що знаходяться в комплекті, 30% − нестандартні. Знайти ймовірності того, що серед 5 виробів, навмання взятих із комплекту, буде: а) тільки один нестандартний; б) принаймні один нестандартний.

2.2. Імовірність того, що кожен клієнт, який звернувся в авіакасу, замовить квиток до аеропорту N, дорівнює 0,1. Знайти ймовірності того, що із 100 клієнтів, що звернулись в касу, замовлять квиток до аеропорту N: а) менше 15 чоловік; б) від 5 до 12 чоловік; в) більше 20 чоловік.

2.3. На біржі виставлено 10 цінних паперів. Імовірність того, що вони подорожчають протягом одного дня, дорівнює 0,6. Знайти ймовірності того, що подорожчає: а) рівно 5 паперів; б) не більше ніж 4 папери; в) від 3 до 5 цінних паперів.

2.4. Авіакомпанія виконує протягом місяця 400 рейсів. Імовірність повного комерційного завантаження кожного рейсу дорівнює 0,8. Знайти ймовірності того, що протягом місяця з повним комерційним завантаженням буде виконано: а) не менше 300 рейсів; б) більша частина рейсів.

2.5. За статистичними даними у середньому 1% пасажирів відмовляється від рейсу. Знайти ймовірності того, що з 300 пасажирів, що мають квитки на рейс, відмовляться від польоту: а) не більше 5 пасажирів; б) не менше 3 пасажирів.

2.6. Інвестор укладає договір на фондовій біржі. Ймовірність укладання однієї угоди за день дорівнює 0,7. Виходячи із припущення, що за 10 робочих днів укладається не більше однієї угоди в день, знайти ймовірності подій: а) буде укладено 7 угод; б) буде укладено не менше 8 угод; в) жодної угоди не буде укладено.

2.7. Кількість помилок у рахунках торгових підприємств складає 5%. Аудитор перевіряє 10 навмання вибраних рахунків. Якщо не виявиться жодної помилки, то рахунки підприємства далі не перевіряються. Яка ймовірність того, що в 10 рахунках підприємства: а) не буде жодної помилки; б) буде 3 помилки; в) буде від 3 до 5 помилок.

2.8. Телефонна станція обслуговує 2000 абонентів. Імовірність того, що будь-який абонент зателефонує на станцію протягом години, дорівнює 0,001. Знайти ймовірності того, що протягом години на телефонну станцію зателефонують: а) 5 абонентів; б) не більше 3 абонентів.

2.9. Імовірність того, що інвестиційний проект принесе через рік прибуток, дорівнює 0,8. Знайти ймовірності того, що з 15 інвестиційних проектів: а) 10 проектів виявляться прибутковими; б) не менше 8 проектів виявляться прибутковими; в) від 5до 9 проектів будуть прибутковими.

2.10. Фабрика випускає 75% продукції першим сортом. Знайти ймовірність того, що з 300 виробів, виготовлених фабрикою, число першосортних виробів буде: а) не менше 250; б) від 220 до 235; в) не більше 200.

2.11. За статистичними даними 30% усіх затримок рейсів авіакомпанії відбувається з вини служби перевезень. Протягом тижня з різних причин із затримкою було виконано 12 рейсів. Знайти найбільш ймовірне число рейсів, затриманих із вини служби перевезень і обчислити відповідну ймовірність.

2.12. При перевезенні скляних виробів в середньому 0,05% від їх числа пошкоджується. Знайти ймовірності того, що при перевезенні 1000 виробів будуть пошкоджені: а) рівно 3 вироби; б) не більше 3 виробів; в) хоча б один виріб.

2.13. Близько 40% клієнтів банку використовують спеціальні кредитні картки. Знайти ймовірність того, що з 25 клієнтів банку, картки використовують: а) рівно 12 клієнтів; б) не менше 10 клієнтів; в) від 15 до 20 клієнтів.

2.14. За даними аеропорту в листопаді через метеорологічні умови відкладається 10% рейсів. Знайти ймовірності того, що з 400 рейсів, запланованих на листопад, буде відкладено: а) 50 рейсів; б) від 30 до 50 рейсів; в) не більше 30 рейсів.

2.15. Вибрали навмання 200 акцій. Імовірність того, що акція принесе збитки, дорівнює 0,025. Знайти ймовірність того, що серед акцій виявляться збитковими: а) не більше 3 акцій; б) жодної акції; в) більше 4 акцій.

2.16. Імовірність того, що в бухгалтерському звіті є помилка, дорівнює 0,04. Для перевірки аудитор бере 100 документів. Знайти ймовірність того, що при перевірці документів помилки будуть виявлені: а) в 50 документах; б) від 10 до 20 документах; в) від 40 до 60 документів.

2.17. На біржі виставлено 100 цінних паперів. Імовірність того, що вони подорожчають протягом одного дня, дорівнює 0,06. Знайти ймовірність того, що подорожчає: а) рівно 50 паперів; б) не більше ніж 40 паперів; в) від 30 до 60 цінних паперів.

2.18. Завод відправив на базу 500 виробів. Імовірність пошкодження кожного виробу при перевезенні дорівнює 0,001. Знайти ймовірності пошкодження при перевезенні: а) рівно 3 виробів; б) менше 3 виробів; в) принаймні одного виробу.

2.19. У фірмі по продажу комп’ютерів 0,5% деталей, наведених у каталозі, знаходяться на допоміжному складі фірми і треба кілька днів для їх доставки. Знайти ймовірність того, що з 1000 навмання замовлених деталей на допоміжному складі: а) знаходиться не більше 3 деталей; б) знаходиться не менше 6 деталей; в) немає жодної деталі.

2.20. В автопарку 12 автобусів. Імовірність виходу на лінію кожного з них дорівнює 0,8. Знайти ймовірність нормальної роботи автопарку в найближчий день, якщо для цього потрібно мати на лінії 8 машин.

Завдання 3. Дати відповіді на подані нижче теоретичні запитання (згідно порядкового номера за списком групи парні та непарні номери відповідають на відповідно парні та непарні питання).

Теоретичні запитання до теми

1. Яку послідовність незалежних випробувань називають схемою Бернуллі?

2. Записати і пояснити формулу Бернуллі.

3. Як визначають найімовірніше число настання події в серії незалежних випробувань?

4. У яких випадках використовують формулу Пуассона?

5. За яких умов локальна формула Муавра-Лапласа дає добре наближення?

6. Коли використовують інтегральну формулу Муавра-Лапласа?

7. Пояснити на прикладі задачі використання функції Excel **БИНОМРАСП**, вибір значення *интегральная*.

8. Пояснити на прикладі задачі використання функції Excel **ПУАССОН**, вибір значення *интегральная*.

Оформлення звіту та порядок захисту

Лабораторна робота виконується, як і попередні лабораторні роботи.

Лабораторна робота № 4

Дискретні випадкові величини, обчислення їх числових характеристик

***Мета роботи*:** Вивчення законів розподілу дискретних випадкових величин та застосування табличного процесору Microsoft Excel для обчислення основних числових характеристик

Теоретичні відомості

4.1. Закони розподілу дискретних випадкових величин, їх задання

Означення. ***Випадковою величиною*** називають таку величину, яка внаслідок випробування може набути лише одне числове значення, заздалегідь невідоме і обумовлене випадковими причинами.

Отже, ***випадковою величиною***, пов'язаною з даним дослідом, називають величину, яка при кожному здійсненні досліду може приймати те, чи інше числове значення, залежно від випадку.

Між випадковими подіями і випадковими величинами існує тісний зв'язок. Випадкова подія є якісною характеристикою випадкового результату досліду, а випадкова величина – його кількісною характеристикою. Випадкові величини за своїм характером поділяються на *дискретні* і *неперервні.*

Означення. ***Дискретною випадковою величиною (ДВВ)*** називають таку величину, яка внаслідок випробування набуває скінченну або злічену кількість значень з відповідними ймовірностями.

Всі можливі значення дискретної випадкової величини можуть бути перенумеровані

$$x_1,\, x_2, \ldots, x_n,\, \ldots\ .$$

Іншими словами, вона має властивість, що кожне з її можливих значень має окіл, який вже не містить жодного з інших значень цієї ж величини.

Означення. Співвідношення, яке встановлює зв'язок між можливими значеннями випадкової величини і ймовірностями, з якими приймаються ці значення, називається ***законом розподілу ймовірностей випадкової величини.***

Для дискретної випадкової величини X закон розподілу може бути заданий таблично або графічно.

В першому випадку закон розподілу називається ***рядом розподілу ймовірностей*** випадкової величини X.

Таблиця 4.1

X	x_1	x_2	$\dots$	x_n	$\dots$
P	p_1	p_2	$\dots$	p_n	$\dots$

В першому рядку таблиці записують всі можливі значення випадкової величини, а в другому − відповідні їм ймовірності. Оскільки події $\{X = x_1\}$, $\{X = x_2\}$, ..., $\{X = x_n\}$ становлять повну групу несумісних подій, то за теоремою додавання ймовірностей маємо

$$\sum_{k=1}^{\infty} p_k = 1,$$

тобто сума ймовірностей всіх можливих значень випадкової величини дорівнює одиниці.

Графічне зображення закону розподілу називається ***многокутником розподілу***: по осі абсцис відкладаємо можливі значення x_k випадкової величини X, а по осі ординат – ймовірності p_k цих значень; точки (x_k, p_k) послідовно з'єднуємо відрізками прямих.

Для кількісної оцінки закону розподілу випадкової величини (дискретної або неперервної) задають ***функцію розподілу ймовірностей випадкової величини,*** яку визначають як імовірність того, що випадкова величина X прийме значення, менше деякого фіксованого числа x і позначають $F(x) = P(X < x)$ або $F(x) = P(-\infty < X < x)$.

Функцію розподілу $F(x)$ називають ***інтегральною функцією*** розподілу ймовірностей випадкової величини.

Серед дискретних випадкових величин особливе місце в теорії ймовірностей посідають такі, які набувають лише цілих невід'ємних значень $X = x_k = 0,1,2,...$.

Ці випадкові величини називають цілочисловими.

4.2. Числові характеристики законів розподілу

Кожен закон розподілу випадкової величини доповнюється кількісними показниками, які називають ***числовими характеристиками*** цього розподілу. Вони узагальнено характеризують випадкову величину. Найбільш часто використовують три числові характеристики: ***математичне сподівання, дисперсію*** **та** ***середнє квадратичне відхилення.***

Означення. ***Математичним сподіванням дискретної випадкової величини*** X називають суму добутків всіх можливих значень випадкової величини та їх ймовірностей

$$M(X) = \sum_k x_k p_k . \tag{4.1}$$

Якщо при цьому множина можливих значень X є нескінченною, то накладається умова абсолютної збіжності ряду.

Математичне сподівання називають центром розсіювання випадкової величини і воно відповідає середньому значенню випадкової величини.

Означення. ***Дисперсією*** випадкової величини називають математичне сподівання квадрата різниці випадкової величини і її математичного сподівання:

$$D(X) = M(X - M(X))^2 . \tag{4.2}$$

Теорема. *Дисперсія випадкової величини дорівнює різниці між математичним сподіванням квадрата цієї величини і квадратом її математичного сподівання:*

$$D(X) = M(X^2) - \left(M(X)\right)^2 . \tag{4.3}$$

Формула для розрахунку дисперсії дискретної випадкової величини має такий вигляд:

$$D(X) = \sum_{i=1}^{n} \left(x_i - M(X)\right)^2 p_i . \tag{4.4}$$

Означення. Числову характеристику закону розподілу випадкової величини

$$\sigma(X) = \sqrt{D(X)}\,, \tag{4.5}$$

називають ***середньоквадратичним відхиленням або стандартним відхиленням***.

Приклад 1. По мішені проводяться чотири незалежні постріли. Ймовірність влучення при одному пострілі дорівнює 0,25. Скласти ряд розподілу випадкової величини X − кількості влучень в мішень та обчислити його основні числові характеристики. Визначити функцію розподілу $F(x)$ та побудувати її графік.

Розв'язання. Випадкова величина X − кількість попадань в мішень може приймати значення 0, 1, 2, 3, 4. Оскільки розглядувані випробування задовольняють схемі Бернуллі, то X має біномний закон розподілу. У даному випадку $n = 4$, $p = 0{,}25$, $q = 1 - p = 0{,}75$.

Для складання ряду розподілу використаємо функцію Excel категорії «Статистические» **БИНОМРАСП** (*число_успехов; число_испытаний; вероятность_успеха; интегральная*) із такими параметрами: *Число_успехов* :k − змінна величина, яка приймає значення: 0, 1, 2, 3, 4; *Число_испытаний* :4 − кількість незалежних випробувань; *Вероятность_успеха:* 0,25 − ймовірність успіху у кожному випробуванні; *Интегральная:*0 − для знаходження ймовірності випадкової події $P(X = k)$.

Відповідні ймовірності, знайдені за допомогою даної функції **БИНОМРАСП**, запишемо у вигляді ряду розподілу

k	0	1	2	3	4	Сума
P	0,3164	0,4219	0,2109	0,0469	0,0039	1

В останньому стовпчику знайдена сума $\sum_{k=0}^{4} p_k$ для перевірки умови нормування.

Знайдемо основні числові характеристики розподілу даної випадкової величини: математичне сподівання, дисперсію, середнє квадратичне відхилення.

Математичне сподівання дискретної випадкової величини обчислюється за формулою (4.1):

$$M(X) = \sum_{k=0}^{4} x_k p_k = 0 \cdot 0{,}3164 + 1 \cdot 0{,}4219 + 2 \cdot 0{,}2109 + 3 \cdot 0{,}0469 + 4 \cdot 0{,}0039 = 1.$$

Для обчислення дисперсії знайдемо $M(X^2)$

$$M(X^2) = \sum_{k=0}^{4} (x_k)^2 p_k = 0 \cdot 0{,}3164 + 1 \cdot 0{,}4219 + 4 \cdot 0{,}2109 + 9 \cdot 0{,}0469 + 16 \cdot 0{,}0039 = 1{,}75$$

Звідси

$$D(X) = M(X^2) - (M(X))^2 = 1{,}75 - 1^2 = 0{,}75.$$

Середнє квадратичне відхилення обчислимо за формулою

$$\sigma(X) = \sqrt{D(X)} = \sqrt{0{,}75} = 0{,}866025.$$

Всі основні числові характеристики зручно обчислювати в Excel, власне для математичного сподівання використати функцію категорії «Математические» **СУММПРОИЗВ** (*массив 1, массив 2, массив 3*), для дисперсії обчислити за цією ж функцією $M(X^2)$ і задати формулу (4.3), для середнього квадратичного відхилення використати функцію **КОРЕНЬ** (*число*).

Для функції розподілу випадкової величини скористаємось функцією **БИНОМРАСП** (*число_успехов; число_испытаний; вероятность_успеха; интегральная*) з параметрами: *Число_успехов*: k − змінна величина, яка приймає значення 0, 1, 2, 3, 4; *Число_испытаний*: 4 − кількість незалежних випробувань; *Вероятность_успеха:* 0,25 − ймовірність успіху кожного випробування; *Интегральная:* 1 − для знаходження функції розподілу $F(x)$. Відповідні значення, знайдені за допомогою даної функції, наведено в таблиці 4.2.

Таблиця 4.2

X	0	1	2	3	4
$F(x)$	0,3164	0,7383	0,9492	0,9961	1

Зауваження.

При використані функції **БИНОМРАСП** для побудови функції розподілу $F(x)$ потрібно врахувати, що **БИНОМРАСП** повертає значення $P(X \le x)$, а не $P(X < x)$. Результат застосування табличного процесора Microsoft Excel для розв'язування прикладу 1 наведено на рисунку 4.1.

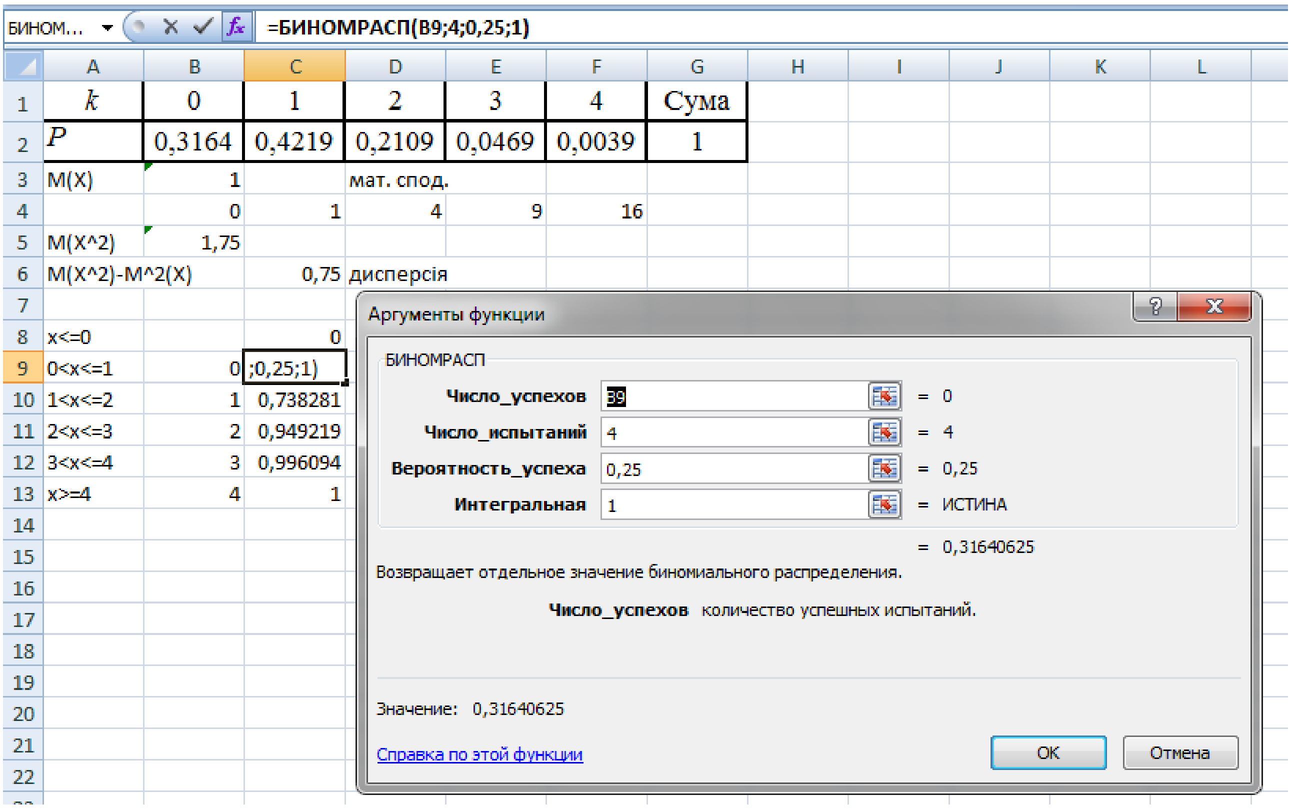


Рис. 4.1. Вигляд обчислень на екрані.

Функція розподілу $F(x)$ має вигляд

$$F(x) = \begin{cases} 0, & \text{якщо} \quad x \le 0; \\ 0{,}3164, & \text{якщо} \quad 0 < x \le 1; \\ 0{,}7383, & \text{якщо} \quad 1 < x \le 2; \\ 0{,}9492, & \text{якщо} \quad 2 < x \le 3; \\ 0{,}9961, & \text{якщо} \quad 3 < x \le 4; \\ 1, & \text{якщо} \quad x > 4. \end{cases},$$

а її графік зображено нижче:

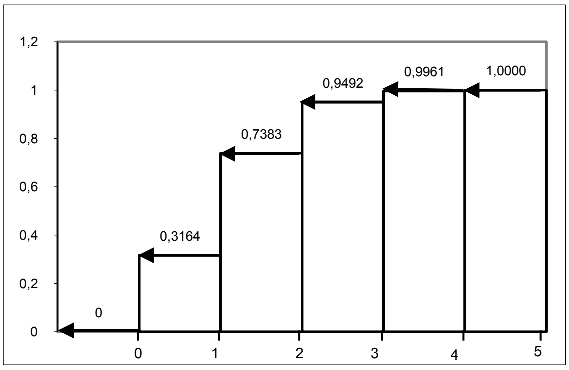


Рис. 4.2. Функція розподілу.

Приклад 2. Середнє число відвідувачів магазину протягом 10-ти хвилинного інтервалу, дорівнює 2. Поява відвідувачів у магазині відбувається випадково і незалежно один від одного. Потрібно:

1) Скласти ряд розподілу числа відвідувачів магазину протягом 10 хвилин і побудувати його графік.

2) Знайти числові характеристики цього розподілу.

3) Обчислити функцію розподілу числа відвідувачів магазину протягом 10 хвилин.

4) Обчислити ймовірність того, що протягом 10 хвилин число відвідувачів магазину виявиться менше 3 і не менше 3.

Розв'язання. Нехай випадкова величина $X -$ кількість відвідувачів магазину протягом 10 хвилин. Дана дискретна випадкова величина може приймати значення *X:* $0,1,2,\ldots,n$, середнє значення, якої дорівнює 2. Отже, дана величина має закон розподілу Пуассона з $\lambda = 2$ і $X = k = 0,1,2,\ldots,n.$

1) Для складання ряду розподілу скористаємось функцією Excel **ПУАССОН** (x; *среднее; интегральная);* за таких параметрів: $x = k$ − змінна величина, яка приймає значення: $0,1,2,\ldots,n$; *среднее:* 2 − середнє значення $\lambda = n \cdot p$; *интегральная:* 0 − для знаходження ймовірності випадкової події $X = k$.

Відповідні ймовірності, знайдені за допомогою даної функції, наведено в табл. 4.3, а відповідний графік − на рис. 4.3.

Таблиця 4.3

k	0	1	2	3	4	5	6	7	8	9
p_k	0,1353	0,2707	0,2707	0,1804	0,0902	0,0361	0,0120	0,0034	0,0009	0,0002

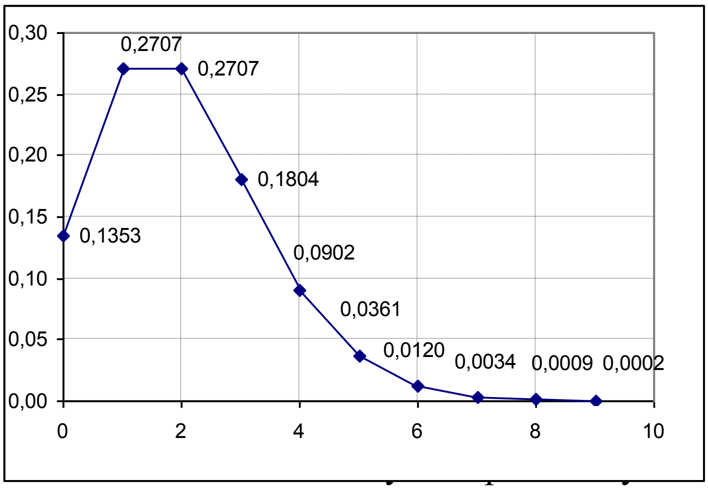


Рис. 4.3. Многокутник розподілу.

Для побудови ймовірнісного многокутника (многокутника розподілу) використано команду **«Диаграмма»** ***(Вставка ⇒ Диаграмма ⇒ Точечная…)*** з відповідними параметрами.

2) В даному прикладі числові характеристики обчислюються за відповідними формулами :

$$M(X) = 2, \quad D(X) = 2, \quad \sigma(X) = \sqrt{2} = 1{,}4142, \; M(X) = \lambda = D(X).$$

3) Функцію розподілу можна обчислити за допомогою функції **ПУАССОН** (*x ; среднее; интегральная*) для параметрів « *x* » та «среднее» як в завданні 1) та «*итегральная*» =1. Результати обчислень наведені в таблиці 4.4. (*Врахувати зауваження в прикладі 1*).

Таблиця 4.4

k	0	1	2	3	4	5	6	7	8	9
$F(x)$	0,1353	0,4060	0,6767	0,8571	0,9473	0,9834	0,9955	0,9989	0,9998	1,0000

4) Імовірність того, що протягом 10 хвилин число відвідувачів

магазину виявиться менше 3 можна обчислити за формулою

$$P(X<3)=P(X=0)+P(X=1)+P(X=2)=0{,}1353+0{,}2707+0{,}2707=0{,}6767,$$

де числові дані взяті з таблиці. Цей самий результат можна одержати зразу, за $X \le 2$ (рис. 4.4).

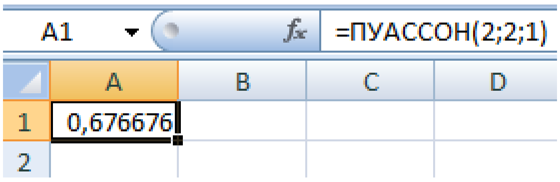


Рис.4.4. Використання функції **ПУАССОН**.

Імовірність того, що протягом 10 хвилин число відвідувачів магазину виявиться не менше 3, тобто 3 і більше, можна обчислити за формулою

$$P(X \ge 3)=1-P(X<3)=1-0{,}6767=0{,}3133.$$

Виконання лабораторної роботи

Завдання до теми

Завдання 1. За даним законом розподілу дискретної випадкової величини X побудувати многокутник розподілу. Знайти $M(X), D(X)$ і $\sigma(X)$.

1.

X	-2	-1	0	4	5	7
p	0,12	0,18	0,2	0,3	0,17	0,03

2.

X	-3	-1	0	1	3	5
p	0,13	0,17	0,2	0,3	0,18	0,02

3.

X	-2	-1	0	1	3	4
p	0,15	0,2	0,25	0,2	0,15	0,05

4.

X	2	4	7	9	12	15
p	0,05	0,15	0,35	0,2	0,15	0,1

5.

X	3	4	7	9	12	14
p	0,1	0,3	0,2	0,05	0,15	0,2

6.

X	1	4	8	9	12	13
p	0,13	0,17	0,2	0,3	0,18	0,02

7.

X	-5	-4	0	1	2	4
p	0,15	0,2	0,25	0,2	0,15	0,05

8.

X	-2	-1	0	1	4	6
p	0,12	0,28	0,22	0,18	0,12	0,08

9.

X	-7	-5	-2	1	5	9
p	0,13	0,17	0,2	0,3	0,18	0,02

10.

X	-8	-6	-2	1	5	6
p	0,12	0,28	0,22	0,18	0,12	0,08

11.

X	-3	-2	0	1	2	4
p	0,13	0,17	0,2	0,3	0,18	0,02

12.

X	-6	-5	-2	3	5	7
p	0,13	0,17	0,2	0,3	0,18	0,02

13.

X	-8	-4	-2	2	5	8
p	0,12	0,18	0,2	0,3	0,17	0,03

14.

X	-10	-5	-2	1	5	10
p	0,12	0,28	0,22	0,18	0,12	0,08

15.

X	-9	-5	-1	1	5	7
p	0,05	0,15	0,35	0,2	0,15	0,1

Завдання 2.

Проводять k незалежних випробувань, в яких імовірність успіху для кожного з них дорівнює p. Скласти ряд розподілу випадкової величини X – кількості успішних випробувань та обчислити основні числові характеристики.

Варіант	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
p	0.2	0.55	0.3	0.4	0.6	0.2	0.45	0.3	0.25	0.6	0.4	0.3	0.35	0.3	0.45
k	3	2	4	3	5	4	3	5	3	2	4	5	4	2	6

Завдання 3.

3.1. Завод випускає 96% виробів першого сорту та 4% − виробів другого сорту. Навмання відібрали партію з 100 виробів. Побудувати закон розподілу ймовірностей дискретної випадкової величини X − кількості виробів другого сорту в цій вибірці, зобразити многокутник розподілу, знайти функцію розподілу ймовірностей $F(x)$; обчислити $M(X)$, $D(X)$, $\sigma(X)$.

3.2. Радіотелефонна станція отримує цифровий текст. Внаслідок атмосферних завад імовірність спотворення цифри в середньому дорівнює 0,001. Було отримано текст, що налічує 2000 цифр. Записати закон розподілу дискретної випадкової величини X – кількості спотворених цифр в отриманому тексті, функцію розподілу ймовірностей $F(x)$; обчислити $M(X)$, $D(X)$, $\sigma(X)$ та побудувати многокутник розподілу.

3.3. Серед 12 однотипних телевізорів 8 відповідають вимогам стандарту, а решта − ні. Побудувати закон розподілу дискретної випадкової величини X – кількості телевізорів, що відповідають вимогам стандарту серед 4 навмання взятих. Обчислити $M(X)$, $D(X)$ та побудувати многокутник розподілу.

3.4. Телефонна станція обслуговує 1000 абонентів. Імовірність того, що протягом години абонент розмовлятиме по телефону дорівнює, в середньому, 0,002. Знайти $M(X)$, $D(X)$, $\sigma(X)$ дискретної випадкової величини X – кількості абонентів, що розмовляють протягом години; функцію розподілу ймовірностей та побудувати многокутник розподілу.

3.5. Серед 10 однотипних планшетів 8 відповідають вимогам стандарту, а решта − ні. Побудувати закон розподілу дискретної випадкової величини X – кількості планшетів, що відповідають вимогам стандарту серед 6 навмання взятих. Обчислити $M(X)$, $D(X)$ та побудувати многокутник розподілу.

3.6. Побудувати ряд розподілу випадкової величини X – кількості попадань м'ячем у кошик при двох киданнях, якщо ймовірність попадання дорівнює 0,4. Знати математичне сподівання, дисперсію та середньоквадратичне відхилення, побудувати многокутник розподілу.

3.7. На шляху руху автомобіля стоять п'ять світлофорів, кожний із яких з імовірністю 0,5 дозволяє або забороняє рух. Побудувати закон розподілу ймовірностей дискретної випадкової X – кількості світлофорів, що їх автомобіль промине без затримки, та обчислити $M(X)$, $D(X)$, $\sigma(X)$.

3.8. Чотири прилади потрібно перевірити на надійність. Імовірність того, що прилад витримає перевірку на надійність, для кожного дорівнює 0,8. Побудувати закон розподілу дискретної випадкової величини X – кількості приладів, які пройшли випробування.

3.9. На заводі можуть одночасно працювати три однотипних верстати, які вмикаються незалежно. Ймовірність того, що в даний момент працює перший, другий чи третій верстат дорівнює 0,2; 0,5; 0,3 відповідно. Записати ряд розподілу для дискретної випадкової величини X – кількості одночасно працюючих верстатів та обчислити $M(X)$, $D(X)$, $\sigma(X)$.

3.10. Під час виготовлення деталі робітникові необхідно виконати чотири незалежні між собою технологічні операції. Імовірність того, що при виконанні першої операції робітник не допустить дефекту, дорівнює 0,95; для другої, третьої і четвертої операцій ця ймовірність становить відповідно 0,9; 0,85; 0,8. Побудувати закон розподілу дискретної випадкової величини X – кількості операцій, під час виконання яких робітник не допустить браку.

3.11. Серед 15 однакових мобільних телефонів телевізорів 14 відповідають вимогам стандарту. Побудувати закон розподілу дискретної випадкової величини X – кількості телефонів, що відповідають вимогам стандарту серед 2 навмання взятих.

Обчислити $M(X)$, $D(X)$ та побудувати функцію розподілу.

3.12. Побудувати ряд розподілу випадкової величини X – суми числа очок, які можуть з'явитися при киданні двох гральних кубиків. Обчислити $M(X)$ та $D(X)$, $\sigma(X)$.

3.13. Пристрій складається із чотирьох приладів, які працюють незалежно один від одного. Ймовірності відмови приладів наступні: $p_1 = 0,3$; $p_2 = 0,4$; $p_3 = 0,5$; $p_4 = 0,6$. Знайти закон розподілу випадкової величини X – кількості приладів, які відмовили та обчислити $M(X)$, $D(X)$, $\sigma(X)$.

3.14. Двічі кидають монету. Нехай дискретна випадкова величина X – кількість випадань герба. Знайти розподіл ймовірностей випадкової величини X, ,функцію розподілу ймовірностей $F(x)$; обчислити $M(X)$, $D(X)$, $\sigma(X)$ та побудувати многокутник розподілу.

3.15. Імовірність того, що футболіст реалізує одинадцятиметровий штрафний удар дорівнює 0,9. Футболіст виконав три такі удари. Побудувати закон розподілу ймовірностей дискретної випадкової величини X – кількості реалізованих штрафних. Обчислити $M(X)$, $D(X)$, $\sigma(X)$.

Завдання 4. Дати відповіді на подані нижче теоретичні запитання (для звіту дати письмово відповіді на 4 питання).

Теоретичні запитання до теми

1. Означити випадкову величину та навести приклади.
2. Пояснити відмінність ДВВ від НВВ.
3. Означити закон розподілу ДВВ та пояснити побудову многокутника розподілу.
4. Пояснити на прикладі побудову функції розподілу ДВВ.
5. Записати числові характеристики дискретної випадкової величини та формули їх обчислення.
6. Пояснити на прикладі обчислення математичного сподівання за формулою та за допомогою функцій Excel.

7. Пояснити на прикладі обчислення дисперсії за формулою за допомогою функцій Excel.

8. Навести приклади використання функції **БИНОМРАСП .**

9. Навести приклади використання функції **ПУАССОН.**

10. Пояснити використання «*итегральная*» 0 чи 1 для обчислень у функціях **БИНОМРАСП, ПУАССОН.**

Оформлення звіту та порядок захисту

Лабораторна робота виконується на аркушах А4, в ній стисло відображаються формули теоретичної частини, розв’язання завдань самостійної роботи та отримані результати (студент здає електронний варіант проведених обчислень практичних задач).

При захисті студент повинен розуміти зміст роботи, порівнювати отримані результати проведених обчислень, також знати відповіді на теоретичні запитання.

Лабораторна робота № 5

Неперервні випадкові величини та їх числові характеристики

***Мета роботи*:** Використання можливостей пакету Microsoft Excel для дослідження законів розподілу неперервних випадкових величин, побудови графіків інтегральної та диференціальної функцій розподілу, обчислення основних числових характеристик.

Теоретичні відомості

5.1. Неперервні випадкові величини, їх задання

Означення. ***Неперервною випадковою величиною*** *(НВВ)* називають величину, якщо сукупність її можливих значень цілком заповнює деякий проміжок числової осі, який може бути скінченним або нескінченним. Наприклад, випадкова величина X − час безвідмовної роботи приладу є неперервною, оскільки її можливе значення $t > 0$.

Закон розподілу неперервної випадкової величини може бути заданий графічно або аналітично $y = p(x) = f(x)$ (за допомогою формули). Табличне задання неможливе, оскільки ймовірність отримати будь-яке певне значення неперервної величини дорівнює нулеві, що пов'язане не з неможливістю самої події (попадання в певну точку на числовій осі), а з нескінченно великою кількістю можливих випадків.

Тому для неперервних випадкових величин (як, зрештою, і для дискретних) визначають ймовірність попадання в деякий інтервал числової осі.

Імовірність попадання випадкової величини X в інтервал $[a, b]$ визначають як ймовірність події $P(a \le X < b)$.

Для кількісної оцінки закону розподілу випадкової величини (дискретної або неперервної) задають ***функцію розподілу ймовірностей випадкової величини,*** яку визначають як імовірність того, що випадкова величина X

прийме значення, менше деякого фіксованого числа x, і позначають $F(x)=P(X<x)$ або $F(x)=P(-\infty<X<x)$.

Функцію розподілу $F(x)$ називають ***інтегральною функцією*** розподілу ймовірностей випадкової величини.

Знаючи функцію розподілу $F(x)$, можна обчислити ймовірність попадання випадкової величини в деякий інтервал $[a, b)$:

$$P(a<X<b)=F(b)-F(a). \tag{5.1}$$

Оскільки ймовірність того, що випадкова величина X набуде конкретного можливого значення, завжди дорівнює нулю, тобто $P(X=x_k)=0$, то мають місце рівності

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

Означення. ***Диференціальною функцією розподілу*** або ***щільністю розподілу ймовірностей*** неперервної випадкової величини називають

$$f(x)=F'(x). \tag{5.2}$$

Теорема. *Ймовірність того, що неперервна випадкова величина* X *набуде будь-яке значення з проміжку* (a,b) *обчислюється за формулою*

$$P\left(X\in(a,b)\right)=\int_a^b f(x)dx. \tag{5.3}$$

Наслідок. Функцію розподілу випадкової величини визначають через її функцію щільності $f(x)$ таким чином:

$$F(x)=\int_{-\infty}^{x} f(t)dt. \tag{5.4}$$

Умову нормування через функцію щільності записують таким чином:

$$\int_{-\infty}^{\infty} f(t)dt=1.$$

Якщо неперервна випадкова величина X визначена лише на проміжку (a,b), то умова формування має вигляд $\int_a^b f(t)dt=1.$

5.2. Числові характеристики неперервних випадкових величин

У випадку неперервних випадкових величин математичне сподівання, дисперсія та середнє квадратичне відхилення мають такий самий зміст та властивості, як і для дискретних випадкових величин, але обчислюють їх за іншими формулами.

Означення. ***Математичним сподіванням*** *неперервної випадкової величини X*, яка задана щільністю розподілу $f(x)$, називають

$$M(X)=\int_{-\infty}^{+\infty} xf(x)dx, \tag{5.5}$$

якщо цей інтеграл абсолютно збіжний. Зокрема, якщо можливі значення неперервної випадкової величини X належать проміжку $[a,b]$, тоді

$$M(X)=\int_{a}^{b} xf(x)dx. \tag{5.6}$$

Для обчислення дисперсії використовувати формули:

$$D(X)=\int_{-\infty}^{+\infty} x^2 f(x)dx-(M(X))^2; \tag{5.7}$$

якщо її можливі значення належать відрізку $[a,b]$, тоді

$$D(X)=\int_{a}^{b} x^2 f(x)dx-(M(X))^2. \tag{5.8}$$

Означення. Числову характеристику закону розподілу випадкової величини, як дискретної, так і неперервної, а саме величину

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

називають ***середньоквадратичним відхиленням*** або ***стандартним відхиленням.***

Приклад 1. Закон розподілу неперервної випадкової величини X задано формулою

$$F(x)=\begin{cases} 0, & \text{якщо} \quad x\le -1; \\ \dfrac{(x+1)^3}{64}, & \text{якщо} \quad -1<x\le 3; \\ 1, & \text{якщо} \quad x>3. \end{cases}$$

Знайти $f(x)$ та побудувати графіки функцій $f(x)$ і $F(x)$. Обчислити $P(0 < X < 2)$, скориставшись формулами (5.1) та (5.3).

Розв'язання. За означенням диференціальної функції (5.2) маємо

$$f(x) = F'(x) = \begin{cases} 0, & \text{якщо} \quad x \le -1; \\ \dfrac{3 \cdot (x+1)^2}{64}, & \text{якщо} \quad -1 < x \le 3; \\ 0, & \text{якщо} \quad x > 3. \end{cases}$$

Графіки функцій $f(x)$ і $F(x)$, побудовані за їх табличними значеннями, зображено відповідно на рис. 5.1 та рис. 5.2.

x	-1	0	1	2	3	4
$F(x)$	0	0,016	0,016	0,016	0,016	1
$f(x)$	0	0,047	0,188	0,422	0,750	0

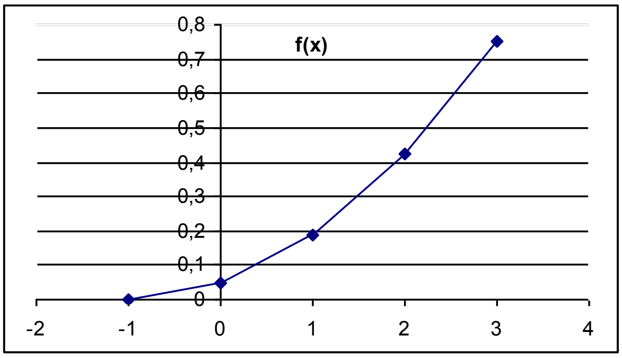


Рис.5.1. Диференціальна функція.

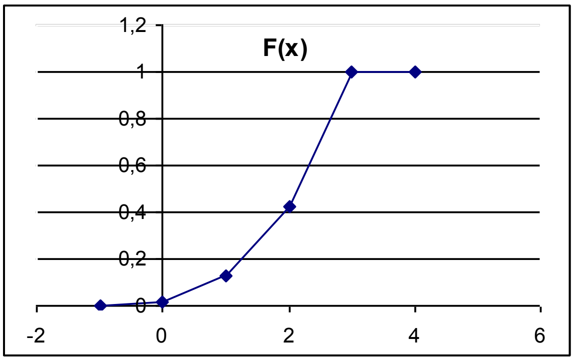


Рис.5.2. Інтегральна функція.

При виконанні обчислень в Excel інтегральну та диференціальну функцію задають вручну, вибравши за її виглядом відповідні значення незалежної змінної (рис. 5.3).

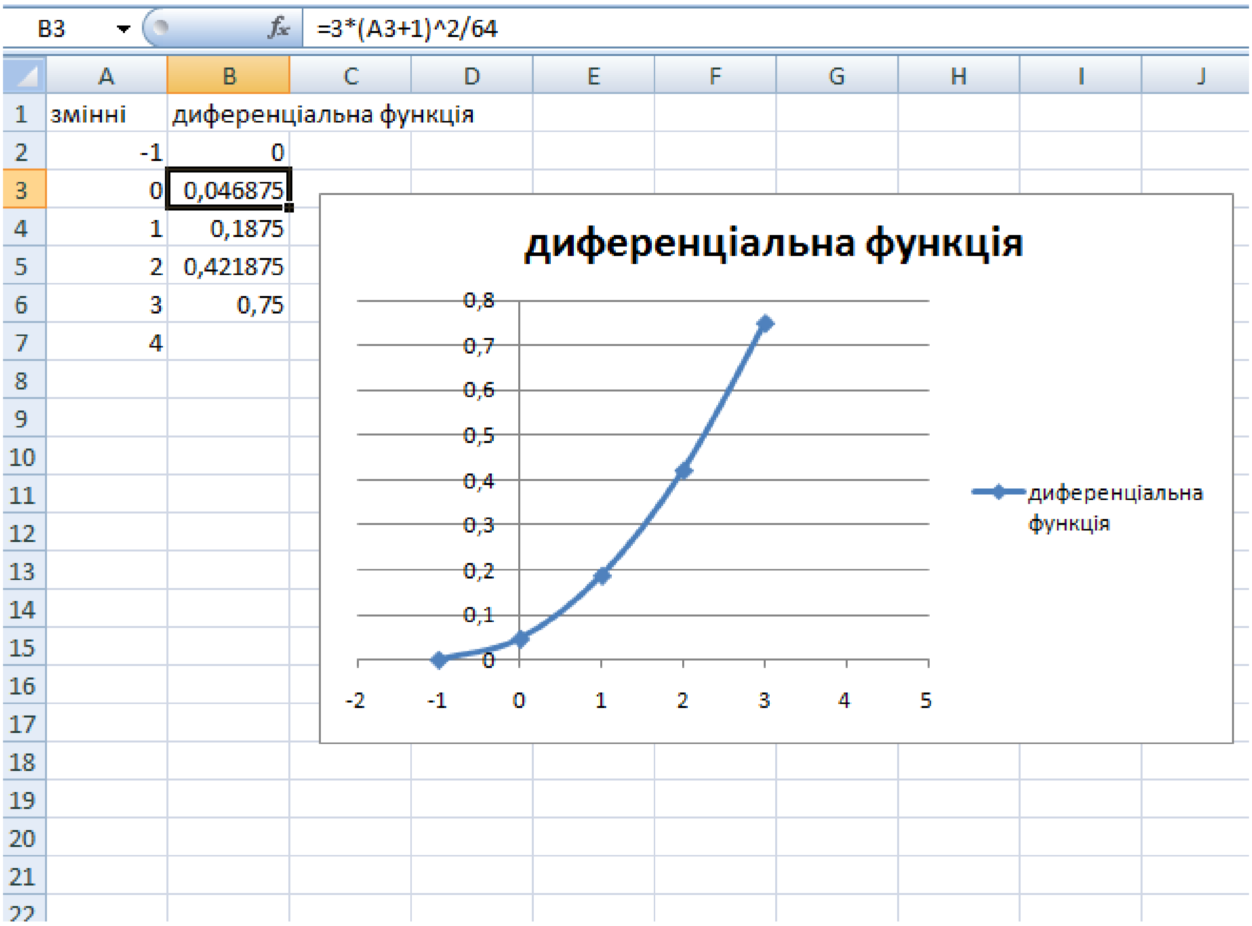


Рис. 5.3. Вигляд обчислень на екрані для побудови графіка.

Імовірність події $0 < X < 2$ обчислимо за формулою (5.1)

$$P(0 < x < 2) = F(2) - F(0) = \frac{27}{64} - \frac{1}{64} = \frac{13}{32}.$$

Те саме одержуємо за формулою (5.3)

$$P(0 < x < 2) = \int_0^2 \frac{3\cdot(x+1)^2}{64}dx = \left.\frac{(x+1)^3}{64}\right|_0^2 = \frac{27}{64} - \frac{1}{64} = \frac{13}{32}.$$

Приклад 2. Диференціальна функція розподілу випадкової величини X має вигляд

$$f(x) = \begin{cases} 0, & \text{якщо} \quad x \le 0; \\ \frac{1}{2}\sin x, & \text{якщо} \quad 0 < x \le \pi; \\ 0, & \text{якщо} \quad x > \pi. \end{cases}$$

Знайти $F(x)$ та побудувати графіки функцій $f(x)$ і $F(x)$. Обчислити $P\left(\frac{\pi}{6} < X < \frac{\pi}{2}\right)$.

Розв'язання. Згідно з формулою (5.4) маємо:

$$F(x) = \int_0^x \frac{1}{2}\sin t \cdot dt = \frac{1}{2}(-\cos t)\Big|_0^x = \frac{1}{2}(1 - \cos x).$$

Таким чином функція розподілу ймовірності має вигляд

$$F(x) = \begin{cases} 0, & \text{якщо} \quad x \le 0; \\ \frac{1}{2}(1 - \cos x), & \text{якщо} \quad 0 < x \le \pi; \\ 1, & \text{якщо} \quad x > \pi. \end{cases}$$

Графіки функцій $f(x)$ та $F(x)$, побудовані за їх табличними значеннями, зображено відповідно на рис.5.4 і рис.5.5.

x	0,00	0,52	1,05	1,57	2,09	2,64	3,14
$f(x)$	0,00	0,25	0,43	0,50	0,43	0,24	0,00
$F(x)$	0,00	0,07	0,25	0,50	0,75	0,94	1,00

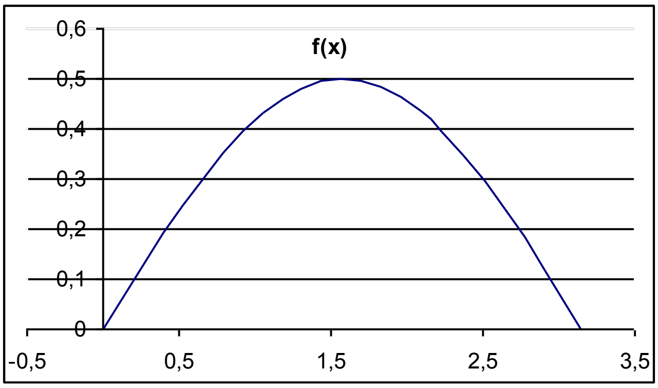


Рис.5.4. Диференціальна функція.

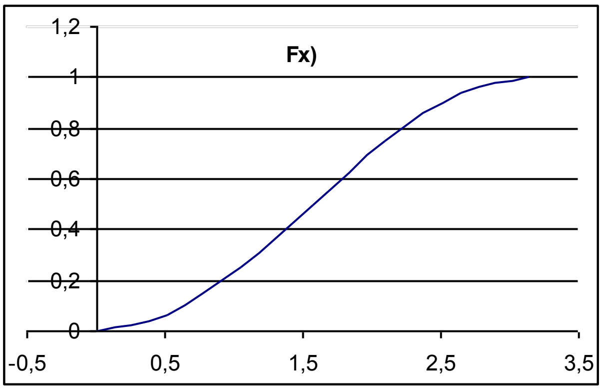


Рис 5.5. Інтегральна функція.

Ймовірність події $P\left(\frac{\pi}{6} < x < \frac{\pi}{2}\right)$ можна обчислити згідно за формулою (5.1) або (5.3). За формулою (5.1) маємо

$$P\left(\frac{\pi}{6} < x < \frac{\pi}{2}\right) = \frac{1}{2}\int_{\frac{\pi}{6}}^{\frac{\pi}{2}} \sin t dt = \frac{1}{2}(-\cos t)\Big|_{\frac{\pi}{6}}^{\frac{\pi}{2}} = -\frac{1}{2}\left(0 - \frac{\sqrt{3}}{2}\right) = \frac{\sqrt{3}}{4}.$$

За формулою (5.3) будемо мати $P\left(\frac{\pi}{6} < x < \frac{\pi}{2}\right) = F\left(\frac{\pi}{2}\right) - F\left(\frac{\pi}{6}\right) = \frac{\sqrt{3}}{4}.$

Приклад 3. Знайти числові характеристики випадкової величини X, яка задана функцією розподілу $F(x) = \begin{cases} 0, & x \le 0; \\ x^2/25, & 0 < x \le 5; \\ 1, & x > 5. \end{cases}$

Розв'язання. Спочатку знайдемо диференціальну функцію розподілу, тобто щільність розподілу ймовірності за формулою $f(x) = F'(x)$.

$$f(x) = \begin{cases} \frac{2}{25}x, & 0 < x \le 5, \\ 0, & x \notin (0, 5]. \end{cases}$$

За формулою $M(X) = \int_a^b xf(x)dx$ знаходимо математичне сподівання:

$$M(X) = \int_0^5 x \cdot \frac{2}{25}xdx = \frac{2}{25} \cdot \frac{x^3}{3}\Big|_0^5 = \frac{2}{75}\left(5^3 - 0\right) = \frac{10}{3}.$$

Для дисперсії використовуємо формулу $D(X) = \int_a^b x^2 f(x)dx - (M(X))^2$:

$$D(X) = \int_0^5 x^2 \cdot \frac{2}{25}xdx - \left(\frac{10}{3}\right)^2 = \frac{2}{25} \cdot \frac{x^4}{4}\Big|_0^5 - \frac{100}{9} = \frac{25}{18}.$$

Середнє квадратичне відхилення $\sigma(X) = \sqrt{D(X)} = \sqrt{\frac{25}{18}} = \frac{5}{3\sqrt{2}} = \frac{5\sqrt{2}}{6} \approx 1{,}17.$

Виконання лабораторної роботи

Завдання до теми

Завдання 1.

1. Задана функція розподілу неперервної випадкової величини *X*. Знайти коефіцієнт *A;* записати щільність розподілу $f(x)$; обчислити числові характеристики $M(X), D(X)$, а також $P(\alpha \le X \le \beta)$. Зробити креслення графіків функції розподілу та щільності розподілу.

1.1. $F(x)=\begin{cases}0, & x\le 0,\\ Ax^5, & 0<x\le 3,\\ 1, & x>3;\end{cases}$ $\alpha=0,25;\ \beta=0,75$.

1.2. $F(x)=\begin{cases}0, & x\le 0,\\ A\left(2x-\dfrac{x^2}{3}\right), & 0<x\le 3,\\ 1, & x>3;\end{cases}$ $\alpha=1;\ \beta=2$.

1.3. $F(x)=\begin{cases}0, & x\le 0,\\ A\left(-x^2+4x\right), & 0<x\le 2,\\ 1, & x>2;\end{cases}$ $\alpha=0,5;\ \beta=1,5$.

1.4. $F(x)=\begin{cases}0, & x\le 0,\\ Ax^2, & 0<x\le 1,\\ 1, & x>1;\end{cases}$ $\alpha=0,3;\ \beta=0,8$.

1.5. $F(x)=\begin{cases}0, & x\le 0,\\ Ax^3, & 0<x\le 2,\\ 1, & x>2;\end{cases}$ $\alpha=0,2;\ \beta=1,7$.

1.6. $F\left(x\right)=\begin{cases}0, & x\le\dfrac{\pi}{6},\\ A\cos 3x, & \dfrac{\pi}{6}<x\le\dfrac{\pi}{3},\\ 1, & x>\dfrac{\pi}{3};\end{cases}$ $\alpha=\dfrac{\pi}{6},\quad \beta=\dfrac{\pi}{4}$.

1.7. $F(x)=\begin{cases}0, & x\le 0,\\ Ax^2, & 0<x\le 2,\\ 1, & x>2;\end{cases}$ $\alpha=1,3;\ \beta=1,5.$

1.8. $F(x)=\begin{cases}0, & x\le 0,\\ Ax^3, & 0<x\le 3,\\ 1, & x>3;\end{cases}$ $\alpha=0,25;\ \beta=0,75.$

1.9. $F(x)=\begin{cases} 0, & x\le 0; \\ A\sin x, & 0<x\le \dfrac{\pi}{6}; \\ 1, & x>\dfrac{\pi}{6}; \end{cases}$ $\alpha=\dfrac{\pi}{12}$, $\beta=\dfrac{\pi}{6}$.

1.10. $F(x)=\begin{cases} 0, & x\le 0, \\ Ax^2, & 0<x\le 3, \\ 1, & x>3; \end{cases}$ $\alpha=0;\ \beta=2.$

1.11. $F(x)=\begin{cases} 0, x\le 3, \\ \dfrac{A\left(x^2-9\right)}{54}, 3<x\le 6, \\ 1, x>6. \end{cases}$ $\alpha=4;\ \beta=5.$

1.12. $F(x)=\begin{cases} 0, x\le 2, \\ \dfrac{A\left(x^2-4\right)}{21}, 2<x\le 5, \\ 1, x>5. \end{cases}$ $\alpha=3;\ \beta=5.$

1.13. $F(x)=\begin{cases} 0, x\le 1, \\ \dfrac{A\left(x^2-1\right)}{6}, 1<x\le 2, \\ 1, x>2. \end{cases}$ $\alpha=1;\ \beta=1{,}5.$

1.14. $F(x)=\begin{cases} 0, x\le 2, \\ A\left(x-2\right)^2, 2<x\le 3, \\ 1, x>3. \end{cases}$ $\alpha=2;\ \beta=2{,}5.$

1.15. $F(x)=\begin{cases} 0, x\le 2, \\ A\left(x-2\right)^2, 2<x\le 3, \\ 1, x>3. \end{cases}$ $\alpha=2;\ \beta=2{,}25.$

Завдання 2.

Неперервна випадкова величина задана щільністю розподілу $y = f(x)$. Записати інтегральну функцію розподілу $y = F(x)$; обчислити числові характеристики $M(X), D(X)$, а також $P(\alpha \le X \le \beta)$. Зробити креслення графіків диференціальної та інтегральної функцій розподілу.

1. $f(x) = \begin{cases} 0, \quad x \le 0, x > \pi/2, \\ \cos x, 0 < x \le \pi/2; \end{cases}, \quad \alpha = -\pi/4; \beta = \pi/4.$

2. $f(x) = \begin{cases} 0, \quad x \le 0, x > \pi/2, \\ \sin x, 0 < x \le \pi/2; \end{cases}, \quad \alpha = -\pi; \beta = \pi/4.$

3. $f(x) = \begin{cases} 0, \quad x \le \pi/6, x > \pi/3, \\ 3\sin 3x, \pi/6 < x \le \pi/3; \end{cases}, \quad \alpha = 0; \beta = \pi/4.$

4. $f(x) = \begin{cases} 0, & x \le 1, \\ 4(x-1)^3, & 1 < x \le 2, \\ 0, & x > 2; \end{cases} \alpha = -1; \beta = 1{,}5.$

5. $f(x) = \begin{cases} 0, & x \le 1, \\ x - \dfrac{1}{2}, & 1 < x \le 2, \\ 0, & x > 2; \end{cases} \alpha = 1; \beta = 1{,}5.$

6. $f(x) = \begin{cases} 0, & x \le 0, \\ \sin 2x, & 0 < x \le \dfrac{\pi}{2}, \\ 0, & x > \dfrac{\pi}{2}; \end{cases} \alpha = 0; \beta = \pi/4.$

7. $f(x)=\begin{cases} 0, & x\le 0, \\ 2\cos x, & 0<x\le \dfrac{\pi}{6}, \alpha=0;\ \beta=\pi/12. \\ 0, & x>\dfrac{\pi}{6}; \end{cases}$

8. $f(x)=\begin{cases} 0, & x\le 0, \\ 3x^2, & 0<x\le 1,\ \alpha=0;\ \beta=1/2. \\ 0, & x>1; \end{cases}$

9. $f(x)=\begin{cases} 0, & x\le 0, \\ \dfrac{1}{2}(2x-1), & 0<x\le 2;\ \alpha=1/2;\ \beta=2. \\ 0, & x>2. \end{cases}$

10. $f(x)=\begin{cases} 0, & x\le 0, \\ 6x(1-x), & 0<x\le 1,\ \alpha=-0,5;\ \beta=0,75. \\ 0, & x>1; \end{cases}$

11. $f(x)=\begin{cases} 0, & x\le 0, \\ \dfrac{1}{6}(2x+1), & 0<x\le 2,\ \alpha=1;\ \beta=1,25. \\ 0, & x>2; \end{cases}$

12. $f(x)=\begin{cases} 0, & x\le 0, \\ 3(x-1)^2, & 0<x\le 1, \alpha=-1;\ \beta=1,5. \\ 0, & x>1; \end{cases}$

13. $f(x)=\begin{cases} 0, & x\le 0, \\ \dfrac{4}{5}(x^3+1), & 0<x\le 1, \alpha=1;\ \beta=1,5. \\ 0, & x>1; \end{cases}$

14. $f(x)=\begin{cases} 0, & x \le 0, \\ \dfrac{2}{9}\left(x^2 - x\right), & 0 < x \le 3, \alpha = 1; \beta = 2. \\ 0, & x > 3; \end{cases}$

15. $f(x)=\begin{cases} 0, & x \le 0, \\ \dfrac{3}{4}\left(x^2 + 1\right), & 0 < x \le 1, \alpha = 0; \beta = 0{,}5. \\ 0, & x > 1; \end{cases}$

Зауваження. Вибирати числові дані для побудови функцій розподілу, щоб побачити їх вигляд, як на рис. 5.1, 5.2 або, на рисунках 5.4, 5.5.

Завдання 3. Дати відповіді на подані нижче теоретичні запитання (для звіту дати письмово відповіді на 2 питання).

Теоретичні запитання до теми

1. Означити неперервну випадкову величину, навести приклади.
2. Пояснити взаємозв'язок інтегральної та диференціальної функцій розподілу НВВ.
3. Які основні характеристики для неперервних випадкових величин?
4. Записати формули обчислення числових характеристик неперервних випадкових величин.
5. Пояснити побудову інтегральної та диференціальної функцій на практиці, їх графічне зображення засобами програмного забезпечення Excel.

Оформлення звіту та порядок захисту

Лабораторна робота виконується на аркушах А4, в ній стисло відображаються формули теоретичної частини, розв'язання завдань та отримані результати (студент здає електронний варіант проведених обчислень практичних задач). При захисті студент повинен розуміти зміст роботи, порівняти отримані результати проведених обчислень, знати відповіді на теоретичні запитання.

Лабораторна робота № 6

Основні закони розподілу випадкових величин та їх числові характеристики

***Мета роботи*:** Використання можливостей пакету Microsoft Excel для розв'язання задач теорії ймовірності з використанням основних законів розподілу дискретних та неперервних випадкових величин.

Теоретичні відомості

6.1. Основні закони розподілу дискретних випадкових величин

1. ***Біномний закон розподілу.*** Цілочислова випадкова величина X має біномний закон розподілу, якщо ймовірність її можливих значень обчислюється за формулою Бернуллі.

$$P(X=k)=C_n^k p^k q^{n-k}, \quad k=0,1,2,\ldots,n, \quad p+q=1. \qquad (6.1)$$

2. ***Закон розподілу Пуассона.*** Цілочислова випадкова величина X має закон розподілу Пуассона, якщо ймовірність її можливих значень обчислюється за формулою Пуассона

$$P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}, \quad k=0,1,2,\ldots, \qquad (6.2)$$

де $\lambda=np$.

Зауваження. Розподіл Пуассона визначає ймовірність того, що в серії з великої кількості $(n\to\infty)$ рідкісних випробувань кількість успіхів набуває значення $k(k=0,1,\ldots)$, $\lambda=np-$ параметр розподілу.

3. ***Геометричний закон розподілу.*** Цілочислова випадкова величина X має геометричний закон розподілу, якщо ймовірність її можливих значень обчислюється за формулою

$$P(X=k)=pq^{k-1}, \quad k=1,2,3,\ldots, \qquad (6.3)$$

де $p=P(A)$ – ймовірність появи події A в кожному випробуванні, $q=1-p$, $X=k$ – кількість випробувань до появи події A в серії незалежних повторних випробуваннях.

Зауваження. Можливий також випадок, коли перше значення випадкової величини є 0.

4. ***Гіпергеометричний закон розподілу.*** Цілочислова випадкова величина X має гіпергеометричний закон розподілу, якщо ймовірність її можливих значень обчислюється за формулою

$$P(X=m)=\frac{C_k^m\cdot C_{N-k}^{n-m}}{C_N^n},\quad m=0,1,2,\ldots,n,k\geq n. \qquad (6.4)$$

Він вказує ймовірність появи m елементів з певною властивістю серед n елементів, взятих із сукупності N елементів, яка містить k елементів саме такої властивості.

Цей розподіл використовують у багатьох задачах статистичного контролю якості.

Зауваження. Якщо обсяг вибірки n малий у порівнянні з обсягом N сукупності, тобто

$$\frac{n}{N}\leq 0{,}1;\quad \frac{n}{k}\leq 0{,}1;\quad \frac{n}{N-k}\leq 0{,}1,$$

то ймовірності у гіпергеометричному розподілі будуть близькими до відповідних ймовірностей біномного розподілу з $p=\frac{k}{N}$. У статистиці це означає, що розрахунки ймовірностей для вибірки без повторення будуть мало відрізнятися від розрахунків ймовірностей для повторної вибірки.

Справедливі формули обчислення основних числових характеристик для дискретних законів розподілу ймовірностей

Біномний закон розподілу:

$$M(X)=np,\ D(X)=npq,\ \sigma(X)=\sqrt{npq}. \qquad (6.5)$$

***Закон розподілу Пуассона*:**

$$M(X)=np=\lambda,\ \ D(X)=\lambda,\ \ \sigma(X)=\sqrt{\lambda}. \qquad (6.6)$$

Геометричний закон розподілу:

$$M(X)=\frac{1}{p},\ D(X)=\frac{q}{p^2},\ \sigma(X)=\frac{\sqrt{q}}{p}. \qquad (6.7)$$

Гіпергеометричний закон розподілу:

$$M(X)=\frac{Ms}{N};\ D(X)=\frac{Ms}{N}\left(1-\frac{M}{N}\right)\left(1-\frac{s-1}{N-1}\right);\ \sigma(X)=\sqrt{D(X)}. \qquad (6.8)$$

В Excel для обчислення ймовірності окремого значення біномного розподілу або значення випадкової величини за заданою ймовірністю використовуються функції **БИНОМРАСП** і **КРИТБИНОМ**. Функція **БИНОМРАСП** використовується для обчислення ймовірності в задачах із фіксованим числом випробувань, коли результатом будь-якого випробування може бути тільки «успіх» або «невдача»; випробування незалежні; ймовірність «успіху» – стала величина протягом усього експерименту. Ця функція використовує параметри і має запис: **БИНОМРАСП** (*число_успехов; число_испытаний; вероятность_успеха; интегральная*) з параметрами: *Число_успехов*: k − змінна величина, яка приймає значення 0, 1, 2, 3, …, *Число_испытаний*: − кількість незалежних випробувань; *Вероятность_успеха:* − ймовірність успіху в кожному випробуванні; *Интегральная:* 0 або 1 − логічне значення, яке визначає форму успіху. Якщо ця величина дорівнює 1 (ИСТИНА), функція **БИНОМРАС**П повертає інтегральну функцію розподілу, тобто обчислює ймовірність того, що кількість успішних випробувань не менше, ніж значення аргументу *«число_успехов»*; якщо логічне значення дорівнює 0 (ЛОЖЬ), тоді обчислюється значення функції щільності ймовірності, тобто ймовірність того, що кількість успішних випробувань в точності дорівнює значенню аргументу *«число_успехов»*.

Функція **КРИТБИНОМ** обчислює найменше значення кількості успішних випробувань випадкової величини, для якого інтегральний біномний розподіл більший або дорівнює заданій величині (критерію). Ця функція використовується в задачах, пов’язаних з контролем якості.

Аналогічно, можна використовувати також для відповідних розподілів випадкових величин функції **ПУАССОН** та **ГИПЕРГЕОМЕТ** із їх параметрами задання.

Приклад 1. По мішені проводяться чотири незалежні постріли. Ймовірність влучення при одному пострілі дорівнює 0,25. Обчислити основні числові характеристики розподілу випадкової величини X − кількості влучень в мішень.

Розв'язання. Випадкова величина X − кількість попадань в мішень може приймати значення 0, 1, 2, 3, 4. Оскільки розглядувані випробування задовольняють схемі Бернуллі, то X має біноміний закон розподілу. У даному випадку $n = 4$, $p = 0{,}25$, $q = 1 - p = 0{,}75$.

Знайдемо основні числові характеристики розподілу даної випадкової величини: математичне сподівання, дисперсію, середнє квадратичне відхилення.

Оскільки, у даному випадку, маємо справу із дискретною випадковою величиною, яка має ***біномний розподіл***, то основні числові характеристики можна обчислити за формулами (6.1):

$$M(X) = np = 4 \cdot 0{,}25 = 1, \; D(X) = npq = 4 \cdot 0{,}25 \cdot 0{,}75 = 0{,}75,$$

$$\sigma(X) = \sqrt{0{,}75} = 0{,}866025.$$

Приклад 2. Середнє число відвідувачів магазину протягом 10-ти хвилинного інтервалу, дорівнює 2. Поява відвідувачів у магазині відбувається випадково і незалежно один від одного. Потрібно знайти числові характеристики розподілу числа відвідувачів магазину протягом 10 хвилин.

Розв'язання. В даному прикладі числові характеристики обчислюються за формулами (6.6):

$$M(X) = np = \lambda = 2, \quad D(X) = \lambda = 2, \quad \sigma(X) = \sqrt{\lambda} = \sqrt{2} = 1{,}4142.$$

Приклад 3. Спортсмен стріляє зі спортивної рушниці по одній і тій самій мішені. Ймовірність влучити в мішень при одному пострілі є величиною сталою і дорівнює 0,8. Стрільба по мішені ведеться до першого влучення. Скласти таблицю розподілу, побудувати імовірнісний многокутник та визначити *M(X),* *D(X)*, $\sigma(X)$ дискретної випадкової величини $X-$ кількості витрачених спортсменом набоїв.

Розв'язання. Випадкова величина X є цілочисловою з *геометричним законом* розподілу ймовірностей. За умовою задачі: $p=0{,}8,\ \ q=0{,}2$. Розподіл випадкової величини *X* наведено в таблиці 6.1, де значення p_k обчислюються за формулою (6.3), а імовірнісний многокутник − на рисунку 6.1.

Таблиця 6.1

k	1	2	3	4	5	6	7	8
p_k	0,80000	0,16000	0,03200	0,00640	0,00128	0,00026	0,00005	0,00001

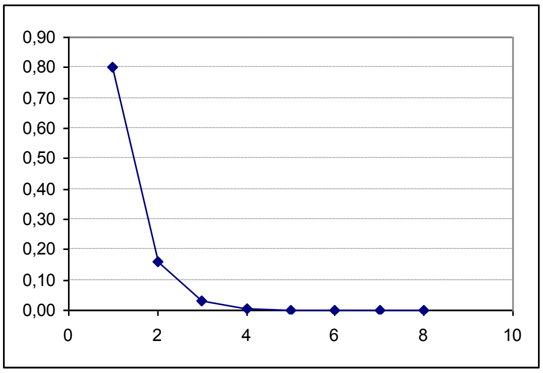


Рис.6.1. Многокутник розподілу.

За формулами (6.7) обчислюємо числові характеристики:

$$M(X)=\frac{1}{p}=\frac{1}{0{,}8}=\frac{5}{4},\ D(X)=\frac{q}{p^2}=\frac{0{,}2}{0{,}64}=\frac{5}{16},\ \sigma(X)=\frac{\sqrt{q}}{p}=\frac{\sqrt{5}}{4}.$$

Приклад 4. Серед дев'яти однотипних виробів п'ять відповідають стандарту, а решта − ні. Навмання береться s виробів. Визначити закон розподілу цілочислової випадкової величини $X-$ кількості появи числа виробів, що відповідають стандарту і обчислити для цієї величини числові характеристики, якщо їх максимальна кількість $s=4$.

Розв'язання. У даному випадку випадкова величина X задовольняє гіпергеометричному закону розподілу при $N=9,\ \ M=5,\ \ N-M=4$ та s, що може приймати різні значення від 0 до 4.

Використаємо формулу (6.4) $p_k = P(X = k) = \dfrac{C_5^k C_4^{s-k}}{C_9^s}$.

Якщо $s = 4,\ \ p_k = \dfrac{C_5^k C_4^{4-k}}{C_9^s}$, де $k = s-(N-M),\ldots,s = 0,1,2,3,4.$

Для побудови функції розподілу скористаємося функцією пакету Excel. **ГИПЕРГЕОМЕТ***(Число_успехов_в_выборке;Размер_выборки; Число_успехов_в_совокупности; Размер_совокупности),*
де *Число_успехов_в_выборке*: k − змінна величина, яка приймає значення: 0, 1, 2, 3, 4;
Размер_выборки: s − кількість навмання взятих виробів;
Число_успехов_в_совокупности: M − кількість виробів, які сприяють події; *Размер_совокупности*: N − загальна кількість виробів.

В таблиці 6.1 наведені гіпергеометричний закон та числові характеристики закону, отримані в Excel

Таблиця 6.1

X	0	1	2	3	4	Сума
p_k	0,007937	0,15873	0,47619	0,31746	0,039683	1
$k \cdot p_k$	0	0,15873	0,952381	0,952381	0,15873	$M(X)$ =2,2222
$(k-M(X))^2 \cdot p_k$	0,039193	0,237115	0,023516	0,192044	0,125416	$D(X)$ =0,6172
						$\sigma(X)$ =0,7856

Аналогічні результати можна одержати, використовуючи формули (6.8):

$$M(X) = \frac{Ms}{N} = \frac{5 \cdot 4}{9} = 2,2222;$$

$$D(X) = \frac{Ms}{N}\left(1-\frac{M}{N}\right)\left(1-\frac{s-1}{N-1}\right) = \frac{5 \cdot 4}{9}\left(1-\frac{5}{9}\right)\left(1-\frac{4-1}{9-1}\right) = 0,6173;$$

$$\sigma(X) = \sqrt{D(X)} = \sqrt{0,61728} = 0,7857.$$

6.2. Закони розподілу неперервних випадкових числових величин та їх числові характеристики

Найчастіше використовуються наступні закони розподілу.

1. ***Рівномірний закон розподілу.*** Розподіл ймовірностей називається ***рівномірним***, якщо на інтервалі, якому належать всі можливі значення випадкової величини, функція щільності розподілу дорівнює константі.

Випадкова величина *X*, розподілена ***рівномірно*** на проміжку $[a,b]$, має таку функцію щільності розподілу:

$$f(x)=\begin{cases} 0, & x \le a; \\ \dfrac{1}{b-a}, & a < x \le b; \\ 0, & x > b. \end{cases} \qquad (6.9)$$

Функція розподілу ймовірностей має вигляд

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

2. ***Показниковий розподіл.*** Випадкову величина X називають розподіленою за ***показниковим*** розподілом, якщо її щільність розподілу ймовірностей має вигляд

$$f(x)=\begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \le 0, \end{cases} \qquad (6.10)$$

де $\lambda > 0$ – параметр розподілу.

Показниковий розподіл задовольняють: час телефонної розмови, час ремонту техніки, час безвідмовної роботи комп’ютерної мережі.

Якщо випадкова величина X розподілена за показниковим розподілом, то її інтегральна функція розподілу має вигляд

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

відповідно

$$P(a < X < b)=\begin{cases} e^{-a\lambda}-e^{-b\lambda}, & a \ge 0, \\ 1-e^{-b\lambda}, & a < 0, \quad b > 0, \\ 0, & b < 0. \end{cases}$$

Графіки диференціальної та інтегральної функцій показникового розподілу зображено на рис. 6.1 а, б.

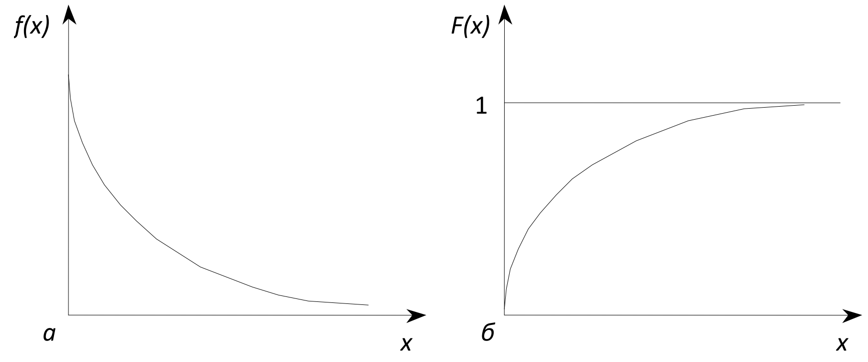


Рис. 6.1. Диференціальна та інтегральна функції показникового розподілу.

Показниковий розподіл часто зустрічається в теорії масового обслуговування, теорії надійності. Нехай t – час безвідмовної роботи деякого елемента, а λ – інтенсивність відмов (середнє число відмов за одиницю часу). Тоді час t роботи елемента можна вважати неперервною випадковою величиною, розподіленою за показниковим законом із функцією розподілу

$$F(x) = p(t < x) = 1 - e^{-\lambda x} \quad (\lambda > 0),$$

яка визначає ймовірність відмови елемента за час, який менший x.

Імовірність

$$R(x) = p(t \ge x) = e^{-\lambda x} \tag{6.11}$$

називають ***функцією надійності***, яка визначає ймовірність безвідмовної роботи елемента за час x.

3. ***Нормальний закон розподілу.***

Означення. ***Нормально розподіленою*** з параметрами a та σ називається випадкова величина X, функція щільності розподілу якої має вигляд:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}}, \ -\infty < x < +\infty, \ -\infty < a < +\infty, \ \sigma > 0. \tag{6.12}$$

Підпорядкування випадкової величини X нормальному закону розподілу з параметрами a та σ позначають $N(a;\sigma)$. Графік функції щільності нормального розподілу називають *кривою Гаусса* або *нормальною кривою.* Параметри a та σ впливають на форму кривої розподілу $M(X)=a$ та $D(X)=\sigma^2$, де a та $\sigma-$ параметри розподілу.

Інтегральна функція нормального розподілу має вигляд

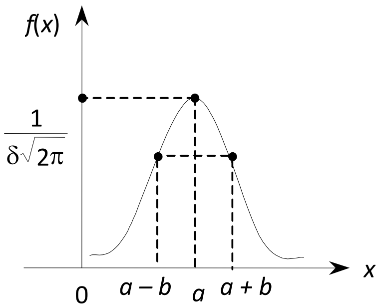
$$F(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}}dt. \qquad (6.13)$$

Розподіл $N(0;1)$ називають *стандартним нормальним розподілом*. В цьому випадку $f(x)=\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}}, -\infty<x<+\infty;$

$$F(x)=\frac{1}{\sqrt{2\pi}}\cdot\int_{-\infty}^{x} e^{-\frac{t}{2}}dt=0,5+\frac{1}{\sqrt{2\pi}}\cdot\int_{0}^{x} e^{-\frac{t}{2}}dt=0,5+\Phi(x),$$

де $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt-$ функція Лапласа.

Графіки функцій щільності ймовірності та розподілу наведено відповідно на рис. 6.2 та рис. 6.3.



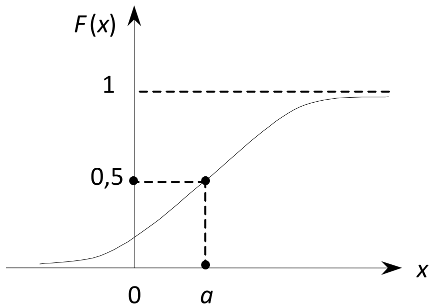


Рис. 6.2. Диференціальна функція *Рис. 6.3.* Інтегральна функція

Імовірність того, що нормально розподілена випадкова величина X прийме значення на проміжку (α,β) визначається за формулою:

$$P(\alpha<X<\beta)=\int_\alpha^\beta f(t)dt=\frac{1}{\sigma\sqrt{2\pi}}\cdot\int_\alpha^\beta e^{-\frac{(t-a)^2}{2\sigma^2}}dt=\Phi\left(\frac{\beta-a}{\sigma}\right)-\Phi\left(\frac{\alpha-a}{\sigma}\right). \qquad (6.14)$$

Правило «трьох сигм» для нормального закону. Коли $\delta=3\sigma,$ то

$$P\left(\left|x-a\right|<3\sigma\right)=2\Phi\left(\frac{3\sigma}{\sigma}\right)=2\Phi(3)=2\cdot 0,49865=0,9973. \qquad (6.15)$$

Звідси випливає

$$P\left(\left|x-a\right|>3\sigma\right)=1-P\left(\left|x-a\right|<3\sigma\right)=1-0,9973=0,0027,$$

тобто ймовірність того, що внаслідок проведення експерименту випадкова величина X, яка має розподіл $N(a;\sigma)$, не потрапить в проміжок $[a-3\sigma;a+3\sigma]$, дорівнює 0,0027.

Зміст ***«правила** 3σ **»***: *практично достовірно, що нормально розподілена випадкова величина може відхилятись від свого математичного сподівання не більше, ніж на потроєне середнє квадратичне відхилення.*

На практиці його використовують так: *якщо закон розподілу випадкової величини X невідомий, але $\left|X-a\right|<3\sigma$, тоді можна припустити, що X розподілена нормально.*

Нормальний закон розподілу широко застосовується в математичній статистиці. Головна особливість нормального закону полягає в тому, що він є граничним законом, до якого наближаються інші закони розподілу за типових умов.

Справедливі формули обчислення основних числових характеристик для неперервних законів розподілу ймовірностей

Рівномірний закон розподілу:

$$M(X)=\frac{a+b}{2},\quad D(X)=\frac{(b-a)^2}{12},\quad \sigma(X)=\frac{b-a}{2\sqrt{3}}. \qquad (6.16)$$

Показниковий розподіл:

$$M(X)=\frac{1}{\lambda},\quad D(X)=\frac{1}{\lambda^2}. \qquad (6.17)$$

Нормальний закон розподілу:

$$M(X)=a\ \ D(X)=\sigma^2. \qquad (6.18)$$

Приклад 3. Час t безперервної роботи планшета має показниковий розподіл. Знайти ймовірність його безвідмовної роботи протягом 600 годин, якщо середній час його роботи – 400 годин.

Розв'язання. За умовою задачі математичне сподівання випадкової величини t дорівнює 400, тому $\lambda = \frac{1}{M} = \frac{1}{400}$. За формулою (6.11) $R(x) = p(t \geq x) = e^{-\lambda x}$, маємо

$$R(x = 600) = e^{-\frac{600}{400}} = e^{-1,5} = 0,2231.$$

Отже, ймовірність того, що планшет працюватиме не менше 600 годин, дорівнює 22%.

Приклад 4. Знайти числові характеристики випадкової величини, розподіленої за законом

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

Розв'язання. Для даного показникового розподілу випадкової величини з параметром $\lambda = 4$ за формулами (6.17) маємо:

$$M(X) = \frac{1}{\lambda} = \frac{1}{4} = 0,25; \quad D(X) = \frac{1}{\lambda^2} = \frac{1}{16} = 0,0625.$$

Приклад 3. Нормально розподілена випадкова величина X має математичне сподівання $M(X) = 30$ та середньоквадратичне відхилення $\sigma(X) = 10$. Ставиться завдання:

1) Записати вигляд диференціальної функції розподілу $f(x)$, побудувати її графік та графік інтегральної функції розподілу $F(x)$.

2) Знайти ймовірність того, що внаслідок випробування випадкова величина набуде значення з інтервалів: $(10,50)$, $(a - 3\sigma, a + 3\sigma)$ (тобто знайти $P(10 < X < 50)$ і $P(a - 3\sigma < X < a + 3\sigma)$).

Розв'язання. 1) Згідно умови задачі $a = 30$, $\sigma = 10$, тому за формулою щільності розподілу (6.12)

$$f(x)=\frac{1}{10\sqrt{2\pi}}\cdot\int_{\alpha}^{\beta}e^{-\frac{(x-30)^2}{2\cdot 10^2}}dx=\frac{1}{10\sqrt{2\pi}}\cdot\int_{\alpha}^{\beta}e^{-\frac{(x-30)^2}{200}}dx.$$

Для побудови графіка функції $f\left(x\right)$ знайдемо значення у точках $x=5,\ 10,\ \ldots\ ,\ 50$, використовуючи функцією пакету Excel, яка має вигляд: **НОРМРАСП**(x; среднее; стандартное_откл; интегральная), де x− значення, для якого будується розподіл; *среднее* − математичне сподівання $M(X)$; *стандартное_откл* – середнє квадратичне відхилення $\sigma\left(X\right)$; *интегральная* − логічне значення 0 або 1. (Якщо «*интегральная*»: − 0, отримаємо значення диференціальної функції розподілу $f\left(x\right)$ в точці x, якщо «*интегральная*»: 1 − функція повертає значення інтегральної функції розподілу $F\left(x\right)$ в точці x).

Для побудови графіка функції $F\left(x\right)$ знайдемо її значення у тих же точках $x=5,\ 10,\ \ldots\ ,\ 50$, використовуючи функцію **НОРМРАСП**(x; $M(X)$; $\sigma\left(X\right)$; 1).

2) Використовуючи формулу (6.14), одержимо:

$$P(10<X<50)=\int_{10}^{50}f(t)dt=\Phi\left(\frac{50-30}{10}\right)-\Phi\left(\frac{10-30}{10}\right)=2\Phi(2)=0{,}9544,$$

де $\Phi(2)$ знаходимо за таблицею значень функції Лапласа.

Цю ймовірність набагато легше обчислити, якщо скористатись функцією **НОРМРАСП,** використовуючи різницю виразів **НОРМРАСП**(50;30;10;1) і **НОРМРАСП**(10;30;10;1), що дорівнює 0,9545. Результат застосування функції **НОРМРАСП** та графіки функції $f\left(x\right)$ і $F\left(x\right)$ наведено на рис. 4.8.

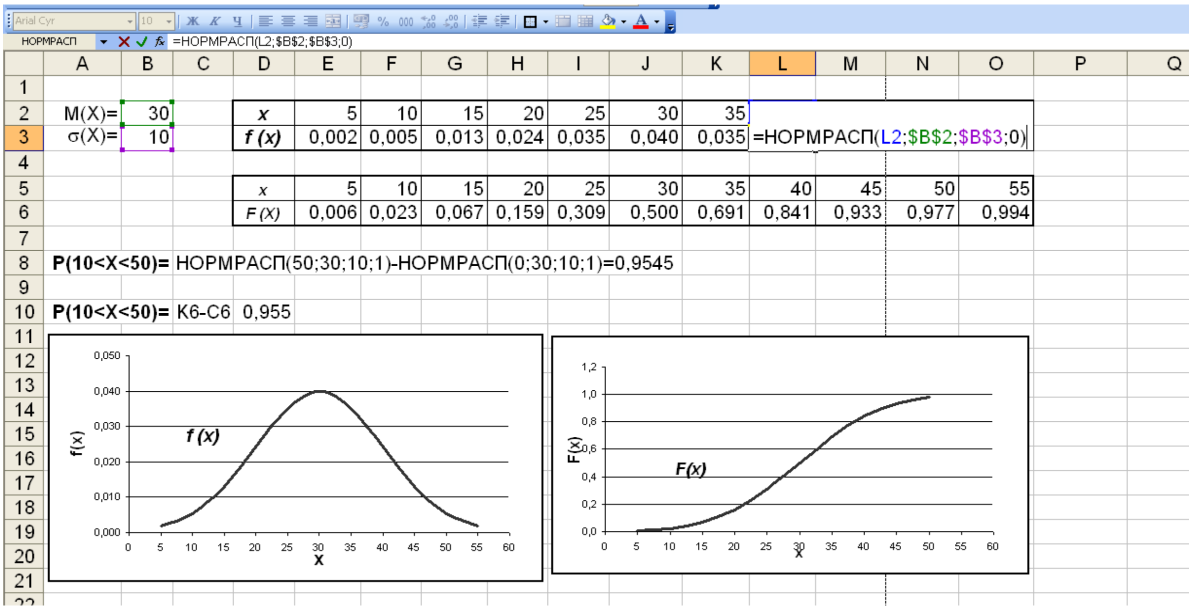


Рис.6.4. Обчислення за допомогою функції **НОРМРАСП.**

Для інтервалу $(a-3\sigma, a+3\sigma)=(0,60)$ шукана ймовірність обчислюється аналогічно за допомогою різниці виразів **НОРМРАСП**(60;30;10;1), **НОРМРАСП**(0;30;10;1) і дорівнює 0,9973, що співпадає також з результатами, одержаними за формулою (6.14).

Виконання лабораторної роботи

Завдання до теми

Завдання 1.

Проводять k незалежних випробувань, в яких імовірність успіху для кожного з них дорівнює p. Обчислити основні числові характеристики випадкової величини X – кількості успішних випробувань.

Варіант	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
p	0.2	0.55	0.3	0.4	0.6	0.2	0.45	0.3	0.25	0.6	0.4	0.3	0.35	0.3	0.45
k	3	2	4	3	5	4	3	5	3	2	4	5	4	2	6

Завдання 2.

Отримано партію k підручників з математики для однієї шкіл м. Києва. Ймовірність неправильного брошурування підручника дорівнює p. Записати закон розподілу випадкової величини X – кількості бракованих підручників та обчислити його основні характеристики.

Варіант	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
p	0.0002	0.00055	0.0003	0.0004	0.0006	0.00012	0.0005	0.00013	0.00025
k	10000	20000	40000	30000	25000	14000	30000	50000	13000

10	**11**	**12**	**13**	**14**	**15**
0.0006	0.00024	0.00035	0.00045	0.0008	0.000075
20000	40000	15000	45000	20000	60000

Завдання 3.

Оператор фальцювальної машини здійснює запуск роботи на фальцювання у 1 згин. Ймовірність виконати точний фальц при запуску одного аркуша є величина стала і дорівнює p. Запуск по одному аркушу для фальцювання ведеться до першого точного потрапляння у розмір, після чого відбувається запуск автоматичного подавання всіх аркушів тиражу. Скласти таблицю розподілу, побудувати імовірнісний многокутник та визначити *M(X), D(X),* $\sigma(X)$ дискретної випадкової величини X – кількості витрачених на приладку аркушів.

Варіант	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
p	0.78	0.75	0.68	0.8	0.81	0.82	0.95	0.73	0.85	0.69	0.84	0.93	0.93	0.73	0.9

Завдання 4.

Випадкова величина X розподілена рівномірно на інтервалі $(a;b)$. Записати диференційну $f(x)$ та інтегральну $F(x)$ функції, знайти $M(X)$, $D(X)$, $\sigma(X)$, якщо

1) $a=1, b=6$; 2) $a=2, b=8$; 3) $a=3, b=8$;

4) $a=3, b=6$; 5) $a=3, b=7$; 6) $a=4, b=10$;

7) $a=-1, b=4$; 8) $a=2, b=9$; 9) $a=2, b=7$;

10) $a=3, b=9$; 11) $a=3, b=10$; 12) $a=-2, b=4$;

13) $a = 2, b = 12$;   14) $a = 12, b = 16$;   15) $a = 5, b = 10$.

Завдання 5.

Час t безперервної роботи електроприладу має показниковий розподіл. Знайти ймовірність його безвідмовної роботи протягом N годин, якщо середній час його роботи – T годин, записавши відповідну диференціальну функцію розподілу.

1) $N = 200, \quad T = 100$;   2) $N = 250, \quad T = 100$;   3) $N = 300, \quad T = 200$;

4) $N = 300, \quad T = 150$;   5) $N = 300, \quad T = 290$;   6) $N = 200, \quad T = 180$;

7) $N = 400, \quad T = 200$;   8) $N = 400, \quad T = 250$;   9) $N = 400, \quad T = 300$;

10) $N = 400, \quad T = 380$;   11) $N = 400, \quad T = 350$;   12) $N = 500, \quad T = 300$;

13) $N = 500, \quad T = 400$;   14) $N = 500, \quad T = 450$;   15) $N = 500, \quad T = 150$.

Завдання 6.

Неперервна випадкова величина X має розподіл $N(a;\sigma)$. Записати диференціальну функцію розподілу $f(x)$, побудувати її графік та графік інтегральної функції розподілу, знайти $P(\alpha < X < \beta)$, використовуючи функцією пакету Excel **НОРМРАСП**, якщо

1) $a = 3, \ \sigma = 4, \ \alpha = 0{,}5, \ \beta = 9$;   2) $a = 3, \ \sigma = 4, \ \alpha = 2, \ \beta = 9$;

3) $a = 0, \ \sigma = 12, \ \alpha = 3, \ \beta = 6$;   4) $a = 2, \ \sigma = 8, \ \alpha = 5, \ \beta = 9$;

5) $a = 2, \ \sigma = 1, \ \alpha = 4, \ \beta = 11$;   6) $a = 5, \ \sigma = 1, \ \alpha = 1, \ \beta = 5$;

7) $a = 4, \ \sigma = 2, \ \alpha = 2, \ \beta = 6$;   8) $a = 1, \ \sigma = 25, \ \alpha = 1, \ \beta = 3$;

9) $a = 2, \ \sigma = 25, \ \alpha = 3, \ \beta = 8$;   10) $a = 3, \ \sigma = 9, \ \alpha = 0{,}25, \ \beta = 10$;

11) $a = -1, \ \sigma = -1, \ \alpha = 0, \ \beta = 9$;   12) $a = -1, \ \sigma = 5, \ \alpha = 2, \ \beta = 7$;

13) $a = -2, \ \sigma = 9, \ \alpha = 0, \ \beta = 7$;   14) $a = -2, \ \sigma = 4, \ \alpha = 1, \ \beta = 5{,}25$;

15) $a = 12, \ \sigma = 4, \ \alpha = 5, \ \beta = 9$.

Зауваження. Вибирати числові дані для побудови функцій розподілу, щоб побачити їх вигляд, як на рис. 6.4.

Завдання 7. Дати відповіді на подані нижче теоретичні запитання (для звіту дати письмово відповіді на 2 питання).

Теоретичні запитання до теми

6. Записати основні закони розподілу дискретних випадкових величин, отримати для одного з них інтегральну функцію розподілу.

7. Записати основні закони розподілу неперервних випадкових величин, отримати для рівномірного розподілу інтегральну функцію.

8. Пояснити використання функції надійності на практиці.

9. Пояснити правило «трьох сигм» для нормального закону.

10. Навести приклад геометричного розподілу, коли перше значення випадкової величини є 0.

11. Як використовують функцією пакету Excel **НОРМРАСП**?

Оформлення звіту та порядок захисту

Лабораторна робота виконується на аркушах А4, в ній стисло відображаються формули теоретичної частини, розв’язання завдань самостійної роботи та отримані результати (студент здає електронний варіант проведених обчислень практичних задач).

При захисті студент повинен розуміти зміст роботи, порівняти отримані результати проведених обчислень, знати відповіді на теоретичні запитання.

Додаткове завдання – обчислення числових характеристик одного (на вибір) закону розподілу неперервної випадкової величини.

Лабораторна робота № 7

Система двох дискретних випадкових величин. Умовні закони розподілу

***Мета роботи*:** Дослідження законів розподілу системи двох дискретних випадкових величин, обчислення основних числових характеристик, використання можливостей пакету Microsoft Excel для опрацювання даних.

Теоретичні відомості

7.1. Дискретні двовимірні випадкові величини

Означення. Двовимірну випадкову величину називають ***дискретною,*** якщо множина значень, які вона може набути, є скінченною або зліченною.

Означення. ***Законом розподілу дискретної двовимірної випадкової величини*** називають перелік можливих значень цієї величини (x_i, y_k) та їх ймовірностей $p(x_i, y_k), i=1, 2, \ldots, m; k=1, 2, \ldots, n$.

Закон розподілу такої випадкової величини задається таблицею

Y	X				
	x_1	x_2	…	x_m	$\sum$
y_1	p_{11}	p_{21}	$\cdots$	p_{m1}	q_1
y_2	p_{12}	p_{22}	$\cdots$	p_{m2}	q_2
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
y_n	p_{1n}	p_{2n}	$\cdots$	p_{mn}	q_n
$\sum$	p_1	p_2	$\cdots$	p_m	1

причому $q_1 = \sum_{i=1}^{m} p_{i1}, q_2 = \sum_{i=1}^{m} p_{i2}, \ldots, q_n = \sum_{i=1}^{m} p_{in}$; $p_1 = \sum_{i=1}^{n} p_{1i}, p_2 = \sum_{i=1}^{n} p_{2i}, \ldots, p_m = \sum_{i=1}^{n} p_{mi}$.

Події $(X = x_i, Y = y_k), i=1, 2, \ldots, m; k=1, 2, \ldots, n$ утворюють повну групу, тому сума ймовірностей наведеної вище таблиці дорівнює 1. Додаючи ймовірності стовпчиків і рядків, одержують закони розподілу компонентів X і Y:

X	x_1	x_2	...	x_m
P	p_1	p_2	...	p_m

Y	y_1	y_2	...	y_n
P	q_1	q_2	...	q_n

Приклад 1. Дана дискретна двовимірна випадкова величина

Y	X			
	1	2	3	$\sum$
0	0,1	0	0,1	0,2
2	0	0,3	0,3	0,6
5	0,2	0	0	0,2
$\sum$	0,3	0,3	0,4	1

Знайти закони розподілу компонент X і Y.

Розв'язання. Додаючи стовпчики і рядки таблиці ймовірностей, отримують відповідні закони розподілу випадкових величин X і Y:

X	1	2	3
P	0,3	0,3	0,4

Y	0	2	5
P	0,2	0,6	0,2

7.2. Числові характеристики двовимірної випадкової величини. Коефіцієнт кореляції та його властивості

Нехай (X, Y) – дискретна двовимірна випадкова величина, для якої закон розподілу задано таблицею розподілу. Тоді, додаючи стовпчики і рядки таблиці ймовірностей, знаходять закони розподілу компонентів X і Y. Відповідно також обчислюють їхні числові характеристики.

Згідно з даними прикладу 1 можемо обчислювати числові характеристики кожного компонента.

$$M(X) = 1 \cdot 0{,}3 + 2 \cdot 0{,}4 + 3 \cdot 0{,}3 = 2,$$

$$D(X) = M(X^2) - (M(X))^2 = 1 \cdot 0{,}3 + 4 \cdot 0{,}4 + 9 \cdot 0{,}3 - 4 = 4{,}6 - 4 = 0{,}6;$$

$$\sigma(X) = \sqrt{D(X)} = \sqrt{0{,}6} = 0{,}7746.$$

Аналогічні обчислення для компонента Y:

$$M(Y) = 1 \cdot 0{,}5 + 2 \cdot 0{,}5 = 1{,}5;$$

$$D(Y) = M(Y^2) - (M(Y))^2 = 1 \cdot 0{,}5 + 4 \cdot 0{,}5 - 2{,}25 = 2{,}5 - 2{,}25 = 0{,}25;$$

$$\sigma(Y) = \sqrt{D(Y)} = \sqrt{0{,}25} = 0{,}5.$$

Означення. ***Коваріацією*** *(кореляційним моментом)* випадкових величин X і Y називається математичне сподівання добутку різниць випадкових величин і їх математичних сподівань:

$$K(X,Y) = \operatorname{cov}(X,Y) = M\big((X - M(X))(Y - M(Y))\big). \tag{7.1}$$

Після розкриття дужок

$$K(X,Y) = M(XY) - M(X)M(Y).$$

З останньої формули випливає, що для дискретної двовимірної випадкової величини (X,Y)

$$K(X,Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j p_{ij} - \sum_{i=1}^{m} x_i p_i \cdot \sum_{j=1}^{n} y_j q_j \tag{7.2}$$

Точка $(M(X), M(Y))$ називається *центром розсіювання* двовимірної випадкової величини.

Якщо випадкові величини X і Y незалежні, то $K(X,Y) = 0$. Коваріація є характеристикою *зв'язку* між величинами системи (X,Y), описує не тільки зв'язок між X і Y, а також і їх розсіювання. Дійсно, якщо хоча б одна з величин мало відрізняється від свого математичного сподівання, то коваріація буде малою, як би не були зв'язані між собою випадкові величини X і Y.

Істотним недоліком коваріації є те, що її розмірність збігається з добутком розмірностей випадкових величин. Тому замість коваріації використовують безрозмірну величину – *коефіцієнт кореляції*, який характеризує *тісноту зв'язку* між величинами X і Y.

Означення. ***Коефіцієнтом кореляції*** між випадковими величинами X і Y називається число

$$\rho = \rho(X,Y) = \frac{K(X,Y)}{\sigma(X)\cdot\sigma(Y)}. \qquad (7.3)$$

Властивості коефіцієнта кореляції

Теорема 1. *Для будь-яких випадкових величин* $-1 \le \rho(X,Y) \le 1$.

Теорема 2. *Коефіцієнт кореляції між випадковими величинами* X *і* Y *дорівнює* ± 1 *тоді і тільки тоді, коли* X *і* Y *зв'язані лінійною залежністю* $Y = aX + b$, *причому* $\rho(X,Y) = 1$ *при* $a > 0$, $\rho(X,Y) = -1$ *при* $a < 0$.

Теорема 3. *Якщо випадкові величини* X *і* Y *незалежні, то* $\rho(X,Y) = 0$.

Наслідок. Якщо $\rho(X,Y) \ne 0$, то X і Y залежні випадкові величини.

Означення. Випадкові величини X і Y, для яких $\rho(X,Y) = 0$, називаються ***некорельованими*** і для яких $\rho(X,Y) \ne 0$ називаються ***корельованими***.

Зауваження. З некорельованості двох випадкових величин, взагалі кажучи, не випливає їх незалежність. Для нормально розподілених випадкових величин некорельованість рівносильна незалежності.

Коефіцієнт кореляції служить для оцінки тісноти лінійного зв'язку між випадковими величинами X і Y: чим ближче модуль коефіцієнта кореляції до одиниці, тим зв'язок сильніший, чим ближче модуль коефіцієнта кореляції до нуля, тим зв'язок слабший.

7.3. Лінійна регресія

Означення. Пряма $y = \alpha_1 x + \alpha_0$ називається ***прямою регресії*** (або прямою середньоквадратичної регресії) Y на X, якщо коефіцієнти α_1 і α_0 вибрано так, щоб середнє квадратичне відхилення $\alpha_1 X + \alpha_0$ від Y було мінімальним

$$M(Y - \alpha_1 X - \alpha_0)^2 = \min.$$

Аналогічно визначається пряма регресії X на Y. З означення визначають коефіцієнти α_1 і α_0:

$$\alpha_1 = \rho \frac{\sigma_y}{\sigma_x}, \quad \alpha_0 = M(Y) - \alpha_1 M(X).$$

Коефіцієнт α_1 називається коефіцієнтом регресії Y на X. Рівняння прямої регресії Y на X має вигляд

$$y = \rho \frac{\sigma_y}{\sigma_x} x + M(Y) - \rho \frac{\sigma_y}{\sigma_x} M(X),$$

або $y - M(Y) = \rho \frac{\sigma_y}{\sigma_x}(x - M(X))$. Аналогічно пряма регресії X на Y має вигляд

$$x - M(X) = \rho \frac{\sigma_x}{\sigma_y}(y - M(Y)).$$

Обидві прямі регресії проходять через точку $\left(M(X), M(Y)\right)$ - центр розсіювання двовимірної випадкової величини. Прямі регресії збігаються тоді і тільки тоді, коли $\rho = \pm 1$.

Величина $(1-\rho^2)\sigma_y^2$ називається залишковою регресією Y на X. Вона характеризує величину похибки при заміні Y на $\alpha_1 X + \alpha_0$.

Приклад2. Двовимірна дискретна випадкова величина (X, Y) задана таблицею розподілу

Y	X			
	1	2	4	$\sum$
0	0,1	0	0,1	0,2
2	0	0,3	0,3	0,6
5	0,2	0	0	0,2
$\sum$	0,3	0,3	0,4	1

Знайти рівняння прямих регресії Y на X та X на Y.

Розв'язання.

Додаючи стовпчики і рядки таблиці ймовірностей, знаходимо закони розподілу випадкових величин X і Y:

X	1	2	4
p	0,3	0,3	0,4

Y	0	2	5
p	0,2	0,6	0,2

Тоді

$$M(X) = 1\cdot 0{,}3 + 2\cdot 0{,}3 + 4\cdot 0{,}4 = 2{,}5;$$

$$D(X) = M(X^2) - (M(X))^2 = 1\cdot 0{,}3 + 4\cdot 0{,}3 + 16\cdot 0{,}4 - 2{,}5^2 = 7{,}9 - 6{,}25 = 1{,}65;$$

$$\sigma_x = \sqrt{D(X)} = \sqrt{1{,}65} \cong 1{,}2845;$$

$$M(Y) = 0\cdot 0{,}2 + 2\cdot 0{,}6 + 5\cdot 0{,}2 = 2{,}2;$$

$$D(Y) = M(Y^2) - (M(Y))^2 = 4\cdot 0{,}6 + 25\cdot 0{,}2 - 2{,}2^2 = 7{,}4 - 4{,}84 = 2{,}56;$$

$$\sigma_y = \sqrt{D(Y)} = \sqrt{2{,}56} = 1{,}6;$$

$$M(XY) = 1\cdot 0\cdot 0{,}1 + 2\cdot 0\cdot 0 + 4\cdot 0\cdot 0{,}1 + 1\cdot 2\cdot 0 + 2\cdot 2\cdot 0{,}3 + 4\cdot 2\cdot 0{,}3 +$$
$$+ 1\cdot 5\cdot 0{,}2 + 2\cdot 5\cdot 0 + 4\cdot 5\cdot 0 = 4\cdot 0{,}3 + 8\cdot 0{,}3 + 5\cdot 0{,}2 = 4{,}6;$$

$$K(X,Y) = M(XY) - M(X)M(Y) = 4{,}6 - 2{,}5\cdot 2{,}2 = -0{,}9;$$

$$\rho(X,Y) = \frac{K(X,Y)}{\sigma_x\sigma_y} = \frac{-0{,}9}{1{,}2845\cdot 1{,}6} = -0{,}438;$$

$$\rho\cdot\frac{\sigma_y}{\sigma_x} = -0{,}438\cdot\frac{1{,}6}{1{,}2845} = -0{,}5456;\quad \rho\frac{\sigma_x}{\sigma_y} = -0{,}438\cdot\frac{1{,}2845}{1{,}6} = -0{,}3516.$$

Рівняння прямої регресії Y на X:

$$y - 2{,}2 = -0{,}5456(x - 2{,}5)$$ (пряма 1 на рис.7.1).

Рівняння прямої регресії X на Y:

$x - 2{,}5 = -0{,}3516(y - 2{,}2)$ або $x = -0{,}3516y + 3{,}2735$ (пряма 2 на рис.7.1).

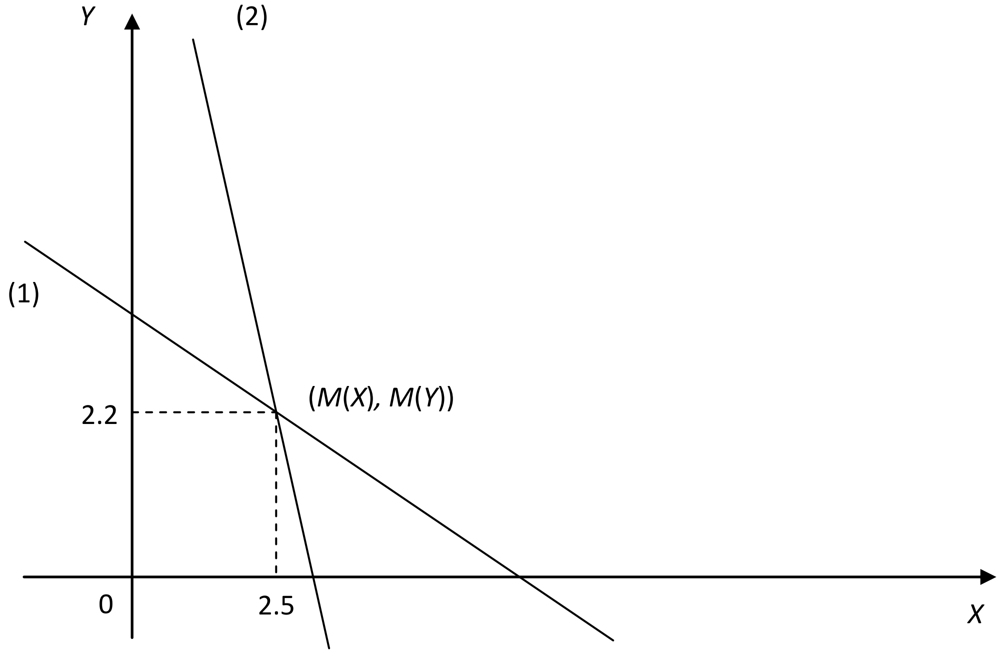
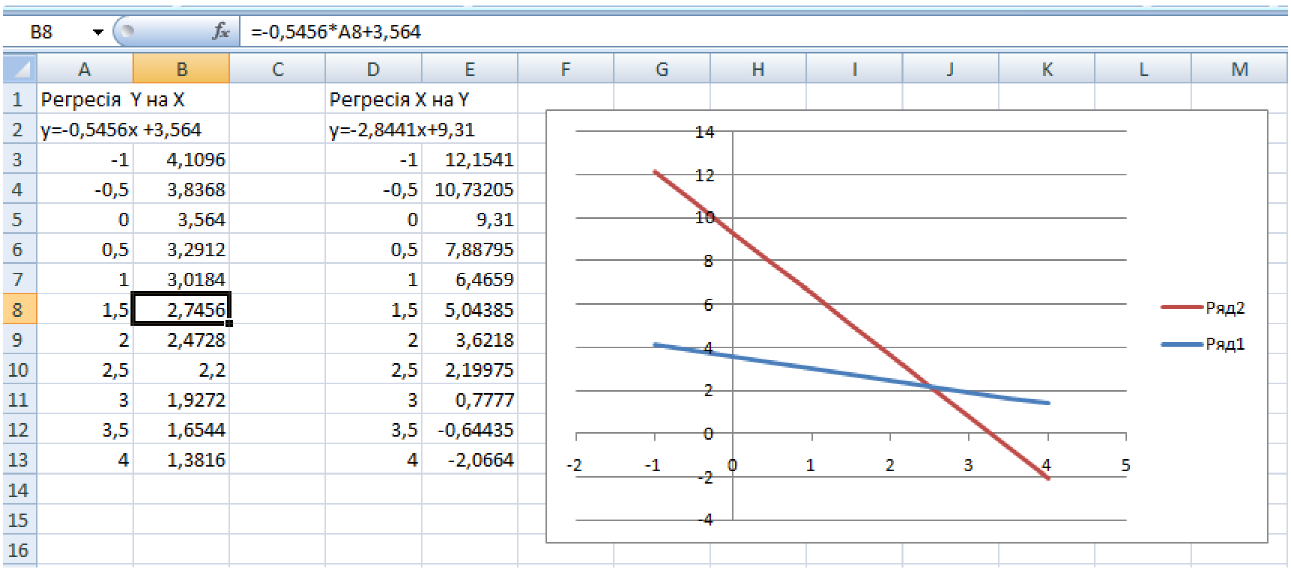


Рис.7.1. Рівняння прямих регресії.

Зручно будувати в Excel, виразивши для кожної прямої рівняння прямої у вигляді залежності y від x, задавши відповідну кількість значень незалежної змінної та обчисливши значення функції в цих точках (рис. 7.2).



	A	B	C	D	E
1	Регресія Y на X			Регресія X на Y	
2	y=-0,5456x +3,564			y=-2,8441x+9,31	
3	-1	4,1096		-1	12,1541
4	-0,5	3,8368		-0,5	10,73205
5	0	3,564		0	9,31
6	0,5	3,2912		0,5	7,88795
7	1	3,0184		1	6,4659
8	1,5	2,7456		1,5	5,04385
9	2	2,4728		2	3,6218
10	2,5	2,2		2,5	2,19975
11	3	1,9272		3	0,7777
12	3,5	1,6544		3,5	-0,64435
13	4	1,3816		4	-2,0664

Рис. 7.2. Побудова прямих, вигляд на екрані.

7.3. Умовні закони розподілу

Нехай (X, Y) – дискретна двовимірна випадкова величина, закон розподілу якої задається таблицею

Y	X				
	x_1	x_2	…	x_m	$\sum$
y_1	p_{11}	p_{21}	$\cdots$	p_{m1}	q_1
y_2	p_{12}	p_{22}	$\cdots$	p_{m2}	q_2
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
y_n	p_{1n}	p_{2n}	$\cdots$	p_{mm}	q_n
$\sum$	p_1	p_2	$\cdots$	p_m	1

Припустимо, що в результаті випробування величина Y набула значення $Y = y_j$, при цьому X набуде одне із своїх можливих значень $x_1, x_2, ..., x_m$. Позначимо умовну ймовірність того, що X прийме значення x_i, за умови , що $Y = y_j$ через $p\left(x_i \mid y_j\right)$.

Означення. ***Умовним законом розподілу*** *дискретної випадкової величини X при фіксованому значенні випадкової величини $Y = y_j$ називається множина можливих значень випадкової величини X та відповідних їм умовних ймовірностей, обчислених при фіксованому значенні $Y = y_j$, тобто розподіл*

X	x_1	x_2	$\cdots$	x_m
$p\left(X \mid y_j\right)$	$p\left(x_1 \mid y_j\right)$	$p\left(x_2 \mid y_j\right)$	$\cdots$	$p\left(x_m \mid y_j\right)$

Умовні ймовірності $p\left(x_i \mid y_j\right)$ обчислюють за формулою

$$p\left(x_i \mid y_j\right) = \frac{p_{ij}}{q_j} \quad (i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n). \tag{7.4}$$

Аналогічно визначається умовний закон розподілу компоненти Y за $X = x_i$

Y	y_1	y_2	$\cdots$	y_n
$p\left(Y \mid x_i\right)$	$p\left(y_1 \mid x_i\right)$	$p\left(y_2 \mid x_i\right)$	$\cdots$	$p\left(y_n \mid x_i\right)$

$$p\left(y_j \mid x_i\right)=\frac{p_{ij}}{p_i} \quad (i=1, 2, \ldots, m,\ j=1, 2, \ldots, n). \tag{7.5}$$

Приклад . Дана дискретна випадкова величина (X, Y) таблицею розподілу

Y	X			
	1	2	3	$\sum$
4	0,1	0,15	0,2	0,45
5	0,15	0,25	0,15	0,55
$\sum$	0,25	0,4	0,35	1

Знайти:

1) умовний закон розподілу випадкової величини X при $Y=4$;

2) умовний закон розподілу випадкової величини Y при $X=2$;

Розв'язання. 1) Знаходимо

$$p\left(x_1 \mid y_1\right)=\frac{p_{11}}{p_1}=\frac{0{,}10}{0{,}45}=\frac{2}{9}; \quad p\left(x_2 \mid y_1\right)=\frac{p_{21}}{p_1}=\frac{0{,}15}{0{,}45}=\frac{3}{9};$$

$$p\left(x_3 \mid y_1\right)=\frac{p_{31}}{p_1}=\frac{0{,}20}{0{,}45}=\frac{4}{9}.$$

Отже,

X	1	2	3
$p\left(X \mid y_1\right)$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{4}{9}$

Аналогічно обчислимо для випадкової величини Y при $X=2$.

$$p\left(y_1 \mid x_2\right)=\frac{p_{21}}{q_2}=\frac{0{,}15}{0{,}4}=\frac{3}{8}; \quad p\left(y_2 \mid x_2\right)=\frac{p_{22}}{q_2}=\frac{0{,}25}{0{,}4}=\frac{5}{8}.$$

Отже,

Y	4	5
$p\left(Y \mid x_2\right)$	$\frac{3}{8}$	$\frac{5}{8}$

Приклад. Авіакомпанія виконує два рейси на добу. Ймовірність затримки першого рейсу через технічні причини дорівнює 0,1, а другого – 0,05. Потрібно:

1) Скласти закон розподілу системи (X,Y), де X – кількість затримок першого рейсу, Y – сумарна кількість затримок двох рейсів.

2) Записати одномірні закони розподілів кожного компонента;

3) Записати основні числові характеристики системи випадкових величин (X,Y).

4) Записати умовні закони розподілів $\{X \mid Y=1\}, \{Y \mid X=1\}$.

5) Обчислити відповідні умовні математичні сподівання для компонент системи (X,Y).

Розв'язання. 1) Випадкова величина X набуває можливі значення 0 і 1, а Y – 0, 1 та 2. Знайдемо ймовірність p_{ij} спільної появи всіх можливих значень x_i та y_j. Позначимо p_1 – ймовірність затримки першого рейсу, p_2 – другого. Тоді ймовірності протилежних подій (своєчасного відправлення першого і другого рейсів) будуть відповідно q_1 та q_2. За умовою задачі $p_1 = 0{,}1;\ p_2 = 0{,}05$ і відповідно $q_1 = 1 - 0{,}1 = 0{,}9;\ q_2 = 0{,}95$.

Знаходимо ймовірності подій:

$$p_{11} = p\{X=0, Y=0\} = q_1 q_2 = 0{,}855;$$

$$p_{12} = p\{X=0, Y=1\} = q_1 p_2 = 0{,}045;$$

$$p_{13} = p\{X=0, Y=2\} = 0 \text{ – неможлива подія;}$$

$$p_{21} = p\{X=1, Y=0\} = 0 \text{ – неможлива подія;}$$

$$p_{22} = p\{X=1, Y=1\} = p_1 q_2 = 0{,}095;$$

$$p_{23} = p\{X=1, Y=2\} = p_1 p_2 = 0{,}005.$$

Складаємо таблицю розподілу системи:

Y	X	
	0	1
0	0,855	0
1	0,045	0,095
2	0	0,005

Умова нормування системи виконана: 0,855+0,045+0,095+0,005=1.

2) Ряди розподілу компонент X та Y знайдемо як суму ймовірностей відповідно за рядками й стовпчиками:

X	0	1
p	0,9	0,1

Y	0	1	2
p	0,855	0,140	0,005

3) Обчислюємо основні числові характеристики випадкових величин X та Y, користуючись рядами розподілу цих величин:

$$M\left(X\right)=0\cdot 0,9+1\cdot 0,1=0,1;\quad M\left(Y\right)=0\cdot 0,855+1\cdot 0,14+2\cdot 0,005=0,15,$$

$$D\left(X\right)=1\cdot 0,1-(0,1)^2=0,09;\quad D\left(Y\right)=1\cdot 0,14+4\cdot 0,005-(0,15)^2=0,1375;$$

$$\sigma_x=\sqrt{0,09}=0,3;\quad \sigma_y=\sqrt{0,1375}\approx 0,37.$$

Обчислюємо кореляційний момент випадкових величин X та Y за формулою (7.2) $K(X,Y)=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}x_i y_j p_{ij}-M\left(X\right)\cdot M\left(Y\right)$:

$$K\left(X,Y\right)=0\cdot 0\cdot 0,855+0\cdot 1\cdot 0,045+0\cdot 2\cdot 0+1\cdot 0\cdot 0+1\cdot 1\cdot 0,95+1\cdot 2\cdot 0,005-0,1\cdot 0,15=$$
$$=0,095+0,01-0,015=0,09.$$

Тоді коефіцієнт кореляції $\rho(X,Y)$ знаходимо за формулою (7.3)

$$\rho(X,Y)=\frac{K(X,Y)}{\sigma_x\cdot\sigma_y}=\frac{0,09}{0,3\cdot 0,37}\approx 0,81.$$

Одержаний результат свідчить про дуже щільну, яка наближається до лінійної, додатну кореляційну залежність між випадковими величинами X та Y.

4) Знайдемо умовні закони розподілів. У пункті 2 прикладу були обчислені ймовірності значень $X=1,\quad Y=1$:

$$P\left(X=1\right)=0,14;\quad P\left(Y=1\right)=0,1.$$

Отже, відповідні умовні ймовірності знаходимо так:

$$P\left\{X=0\,|\,Y=1\right\}=\frac{P\left\{X=0,Y=1\right\}}{P\left\{Y=1\right\}}=\frac{0,045}{0,14}=0,321;$$

$$P\left\{X=1|Y=1\right\}=\frac{P\left\{X=1,Y=1\right\}}{P\left\{Y=1\right\}}=\frac{0,095}{0,14}=0,679.$$

Отримані результати записуємо у вигляді таблиці:

X	0	1
$P\left\{X\mid Y=1\right\}$	0,321	0,679

Аналогічно обчислюємо умовні ймовірності можливих значень $y_1=0,\ y_2=1,\ y_3=2$ компоненти Y:

$$P\left\{Y=0\mid X=1\right\}=\frac{P\left\{X=1,Y=0\right\}}{P\left\{X=1\right\}}=\frac{0}{0,1}=0;$$

$$P\left\{Y=1\mid X=1\right\}=\frac{P\left\{X=1,Y=1\right\}}{P\left\{X=1\right\}}=\frac{0,095}{0,1}=0,95;$$

$$P\left\{Y=2\mid X=1\right\}=\frac{P\left\{X=1,Y=2\right\}}{P\left\{X=1\right\}}=\frac{0,005}{0,1}=0,05.$$

Отримані результати записуємо у вигляді таблиці:

Y	0	1	2
$P\left\{Y\mid X=1\right\}$	0	0,95	0,05

5) Обчислюємо умовні математичні сподівання, користуючись одержаними рядами розподілу:

$$M\left\{X\mid Y=1\right\}=\sum_{i=1}^{n}x_iP\left\{X=x_i\mid Y=1\right\}=0\cdot 0,321+1\cdot 0,679=0,679;$$

$$M\left\{Y\mid X=1\right\}=\sum_{j=1}^{n}y_jP\left\{Y=y_j\mid X=1\right\}=0\cdot 0+1\cdot 0,95+2\cdot 0,05=1,05.$$

Виконання лабораторної роботи

Завдання до теми

Завдання1.

Фірма планує відкрити ще одне кафе. З метою оптимального планування можливої кількості відвідувачів кафе протягом деякого часу аналітики фірми провели статистичне дослідження кількості відвідувачів і прибутку, який був отриманий. У таблиці представлений закон розподілу системи двох дискретних випадкових величин: X – кількість відвідувачів кафе за деякий час і Y – прибуток, який отримала фірма в ум. од. Вважати, що k – номер варіанту.

Y	X			
	k	$k+2$	$k+4$	$k+6$
k	$0{,}002(35-k)$	$0{,}002(k+40)$	$0{,}002(70-k)$	$0{,}002(k+20)$
$k+5$	$0{,}002(k+30)$	$0{,}002(55-k)$	$0{,}002(k+25)$	$0{,}002(40-k)$
$k+10$	$0{,}002(50-k)$	$0{,}002(k+5)$	$0{,}002(80-k)$	$0{,}002(k+50)$

Виконати наступні завдання:

1) Скласти закони розподілу одномірних випадкових величин X та Y;

2) Знайти математичне сподівання, дисперсію та середнє квадратичне відхилення випадкових величин X та Y;

3) Обчислити кореляційний момент $K(X,Y)$ і коефіцієнт кореляції $\rho(X,Y)$;

4) Побудувати умовний закон розподілу випадкової величини X за умови, що випадкова величина Y набуває значення $k+10$ та умовний закон розподілу випадкової величини Y за умови, що випадкова величина X набуває значення $k+2$;

5) Знайти умовні математичні сподівання $M\left(Y \mid X=k+2\right)$, $M\left(X \mid Y=k+10\right)$.

Завдання 2. Дати відповіді на подані нижче теоретичні запитання (для звіту дати письмово відповіді на 2 питання та обов'язково на 9-е запитання).

Теоретичні запитання до теми

1. Означити і пояснити закон розподілу дискретної двовимірної випадкової величини.

2. Пояснити на прикладі задання законів розподілу компонент розподілу дискретної двовимірної випадкової величини та обчислення їх числових характеристик.

3. Записати формули обчислення коваріації, її застосування.

4. Обчислення коефіцієнта кореляції, його основні властивості.

5. Навести приклад обчислення коефіцієнта кореляції, пояснити зв'язок між досліджуваними величинами.

6. Записати рівняння регресій, пояснити їх побудову.

7. Означити умовний закон розподілу.

8. Навести приклади обчислення числових характеристик для умовного закону розподілу.

9. Скласти закон розподілу дискретної двовимірної випадкової величини (навести приклад самостійно), описати всі його характеристики, зробити креслення ліній регресії. Побудувати для даного закону умовний закон розподілу однієї компоненти при зафіксованому значенні іншої та обчислити можливі числові характеристики.

Оформлення звіту та порядок захисту

Лабораторна робота виконується на аркушах А4, в ній стисло відображаються формули теоретичної частини, розв’язання завдань та отримані результати. При захисті студент повинен розуміти зміст роботи, порівняти отримані результати проведених обчислень, знати відповіді на теоретичні запитання. Подається електронний варіант завдання 1 та п.9 завдання 2.

Лабораторна робота № 8

Вибірка. Точкові оцінки числових характеристик випадкових величин

Мета роботи: навчитися розраховувати точкові оцінки числових характеристик випадкових величин, будувати гістограми та функції розподілу, використовуючи теоретичні знання та можливості застосування табличного процесора Microsoft Excel.

Теоретичні відомості

8.1. Вибірка, її характеристики

Дані у статистиці, отримані за допомогою спеціальних досліджень або із звичайних робочих записів у бізнесі, надходять до дослідника у вигляді неорганізованої маси, незалежно від того, чи є вони даними із вибіркової сукупності, чи даними з генеральної сукупності. Тому постає питання обробки і впорядкування даних.

Значення x_i вибірки називають *варіантами*. Послідовність варіант, розміщених в порядку зростання, називають *варіаційним рядом*. Якщо при цьому x_i повторюється n_i разів ($i=1, 2, \dots, k, n_1+n_2+\dots+n_k=n$), то число n_i називають *абсолютною частотою* варіанти x_i, а n_i/n – *відносною частотою* варіанти x_i.

Статистичним розподілом вибірки називають перелік варіант та відповідних частот або відносних частот. Статистичний закон розподілу зручно задавати таблицею, що встановлює зв′язок між значеннями випадкової величини та їх частотами:

x_i	x_1	x_2	$\dots$	x_k
n_i	n_1	n_2	$\dots$	n_k

Означення. ***Емпірична функція розподілу*** має такий вигляд

$$F_n^*(x)=\begin{cases}0, & x<x_1;\\ \sum\limits_{k=1}^{n} n_k/n, & x_k \le x < x_{k+1}\ (k=1,2,\dots,n-1);\\ 1, & x \ge x_n.\end{cases} \tag{8.1}$$

Означення. ***Полігоном частот*** вибірки називають ламану з вершинами в точках $(x_i, n_i), (i=1,2,\dots,k)$. ***Полігоном відносних частот*** вибірки називають ламану з вершинами в точках $\left(x_i, n_i/n\right), (i=1,2,\dots,k)$.

Якщо X – неперервна випадкова величина, то її статистичний розподіл також подають у вигляді таблиць. Для оцінки $f(x)$ за вибіркою $x_1, x_2, \cdots, x_n$ розбивають інтервал на частинні інтервали довжини h, в таблицю записують середину *i*-того інтервалу, n_i– кількість елементів вибірки *i*-того інтервалу. Прямокутники з основами h і висотами $\frac{n_i}{nh}$ у прямокутній системі координат утворюють фігуру, яку називають ***гістограмою*** вибірки.

Незміщеною сильно слушною оцінкою математичного сподівання випадкової величини X є ***вибіркове середнє*** $\overline{X}=\frac{1}{n}\sum\limits_{i=1}^{n} X_i$ і реалізацію цієї оцінки позначають

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n} x_i\,. \tag{8.2}$$

Вибіркове середнє для *дискретного статистичного ряду* обчислюють за формулою $\bar{x}=\frac{1}{n}\sum\limits_{i=1}^{k} n_i x_i$, де $\sum\limits_{i=1}^{k} n_i = n.$

Вибіркове середнє для *інтервального статистичного ряду* обчислюють за формулою $\bar{x}=\frac{1}{n}\sum\limits_{i=1}^{k} n_i z_i$, де z_i– середина i –того інтервалу, $\sum\limits_{i=1}^{k} n_i = n.$

Оцінкою дисперсії σ^2 випадкової величини X є ***вибіркова дисперсія***

$$\bar{s}^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2, \tag{8.3}$$

яка є ***зміщеною*** оцінкою для σ^2.

Величина

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2, \; s^2 = \frac{n}{n-1}\overline{s}^2 \qquad (8.4)$$

є ***незміщеною*** оцінкою дисперсії σ^2 випадкової величини X.

Якщо математичне сподівання *a* відоме, то *незміщеною сильно слушною оцінкою дисперсії* σ^2 випадкової величини X є оцінка $s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - a\right)^2$.

Вибіркову дисперсію для *дискретного статистичного ряду* обчислюють за формулою

$$\overline{s}^2 = \frac{1}{n}\sum_{i=1}^{k}\left(n_i x_i - \overline{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^{\,2} - \overline{x}^2, \qquad (8.5)$$

відповідно

$$s^2 = \frac{n}{n-1}\overline{s}^2; \; s^2 = \frac{1}{n-1}\sum_{i=1}^{k}\left(n_i x_i - \overline{x}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i x_i^{\,2} - \overline{x}^2. \qquad (8.6)$$

Вибіркову дисперсію для *інтервального статистичного ряду* обчислюють за формулою

$$\overline{s}^2 = \frac{1}{n}\sum_{i=1}^{k}\left(n_i z_i - \overline{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{k} n_i z_i^{\,2} - \overline{x}^2, \qquad (8.5')$$

відповідно,

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k}\left(n_i z_i - \overline{x}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i z_i^{\,2} - \overline{x}^2. \qquad (8.6')$$

Величина $\overline{s} = \sqrt{\overline{s}^2}$ називається ***вибірковим середнім квадратичним відхиленням,*** $s = \sqrt{s^2}$ ***– вибірковим виправленим середнім квадратичним відхиленням*** (s^2 – вибіркова виправлена дисперсія).

Означення. ***Медіаною*** M_e називають варіанту, яка ділить варіаційний ряд на дві частини, рівні за кількістю варіант**.**

Для *дискретних статистичних рядів*

$$M_e = \begin{cases} x_m, & n = 2m-1 \\ \dfrac{x_m + x_{m+1}}{2}, & \text{при } n = 2m. \end{cases} \qquad (8.7)$$

Для *інтервальних статистичних рядів*

$$M_e = x_i + h_i \frac{\frac{n}{2} - \sum_{j=1}^{i} n_j}{n_i}, \tag{8.7'}$$

де x_i – початок медіанного інтервалу (йому відповідає перша з нагромаджених частот, що перевищує половину всіх спостережень), h_i – довжина *i*-го інтервалу, n_i – частота медіанного інтервалу.

Означення. ***Модою*** називають варіанту, яка найчастіше трапляється у вибірці.

Для ***дискретних статистичних рядів***

$$M_0 = x_j, \text{ якщо } n_j = \max_i n_i. \tag{8.8}$$

Для ***інтервальних статистичних рядів***

$$M_0 = x_i + h \frac{n_i - n_{i-1}}{2n_i - n_{i-1} - n_{i+1}}, \tag{8.8'}$$

де x_i – початок інтервалу із найбільшою частотою, n_i – частота *i*-го інтервалу.

Означення. ***Початковим емпіричним моментом порядку*** *k* називається вираз

$$M_k = \frac{1}{n} \sum_{i=1}^{n} x_i^k, \tag{8.9}$$

зокрема $M_1 = \bar{x}$.

Означення. ***Центральним емпіричним моментом порядку*** *k* називається вираз

$$m_k = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^k, \tag{8.10}$$

зокрема, $m_2 = \bar{s}^2$.

Означення. ***Коефіцієнтом асиметрії A_s*** називається відношення центрального емпіричного моменту третього порядку до куба середнього квадратичного відхилення:

$$A_s = \frac{m_3}{\sigma^3} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3}. \qquad \mathbf{(8.11)}$$

Означення. ***Ексцесом E*** називається зменшене на три одиниці відношення центрального моменту четвертого порядку до четвертого степеня середнього квадратичного відхилення:

$$E = \frac{m_4}{\sigma^4} - 3. \qquad (8.12)$$

8.2. Визначення числових характеристик із використанням табличного процесору Microsoft Excel

Більшість числових характеристик у випадку незгрупованих даних можна обчислити з використанням табличного процесору Microsoft Excel. Основні вбудовані функції Excel, що застосовуються для таких розрахунків, надано у таблиці 8.1. Щоб викликати потрібну функцію, слід обрати категорію *Статистические* та ім’я функції.

Крім того, часто корисні такі функції:

− **НАИБОЛЬШИЙ** (массив, *k*) – надає *k*-е найбільше значення в ряді даних;

− **НАИМЕНЬШИЙ** (массив, *k*) – надає *k*-е найменше значення в ряді даних.

Таблиця 8.1

Статистичні функції Excel

Числові характеристики	Назва функції
Середнє	**СРЗНАЧ** (число1, число 2, …)
Середнє геометричне	**СРГЕОМ** (число1, число 2, …)
Мода	**МОДА** (число1, число 2, …)
Медіана	**МЕДИАНА** (число1, число 2, …)
Дисперсія	**ДИСП** (число1, число 2, …) **ДИСПР** (число1, число 2, …)
Середнє квадратичне відхилення	**СТАНДОТКЛОН** (число1, число 2, …)
Мінімальне значення	**МИН** (число1, число 2, …))
Максимальне значення	**МАКС** (число1, число 2, …)
Частота	**ЧАСТОТА** (массив_данных; массив_ интервалов)

8.3. Побудова гістограми засобами Microsoft Excel

Excel надає два способи побудови гістограми.

Для побудови гістограми *першим способом* необхідно:

1) Внести в лист Excel вхідні дані та інтервали, за якими ці дані будуть групуватися.

2) Знайти частоти попадання даних в інтервали за допомогою функції **ЧАСТОТА**, для чого:

– виділити діапазон комірок (на одну більше, ніж інтервалів), в яких будуть записані частоти;

– викликати f_x – *Статистические* – **ЧАСТОТА**;

– ввести посилання на комірки, що містять вхідні дані і інтервали;

– натиснути Ctrl+Shift+Enter.

3) Викликати *Вставка — Гистограмма*, появиться діалогове вікно (див. рис. 8.1).

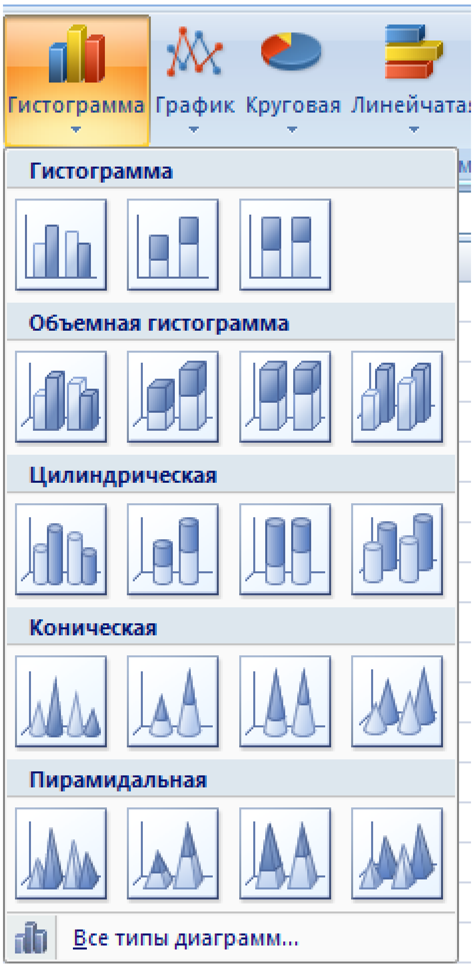


Рис. 8.1. Діалогове вікно майстра гістограм.

4) Надати необхідні для побудови гістограми параметри:

– діапазон вхідних даних, спосіб їх групування (за рядками або стовпчиками) та імена рядів даних, якщо це потрібно;

– якщо імена рядів надано, відмітити *Добавить легенду* і вказати її розміщення;

– якщо потрібно, добавити *Имена рядов*, або (та) *Имена категорій*, або (та) *Значения*;

– якщо потрібно, добавити *Заголовок*, *Линии сетки*, *Оси*, *Таблицу данных*.

Для побудови гістограми *другим способом* необхідно:

1) Внести в лист Excel вихідні дані.

2) Обрати в меню *Сервис – Анализ данных – Гистограмма*, появиться діалогове вікно (див. рис. 8.2).

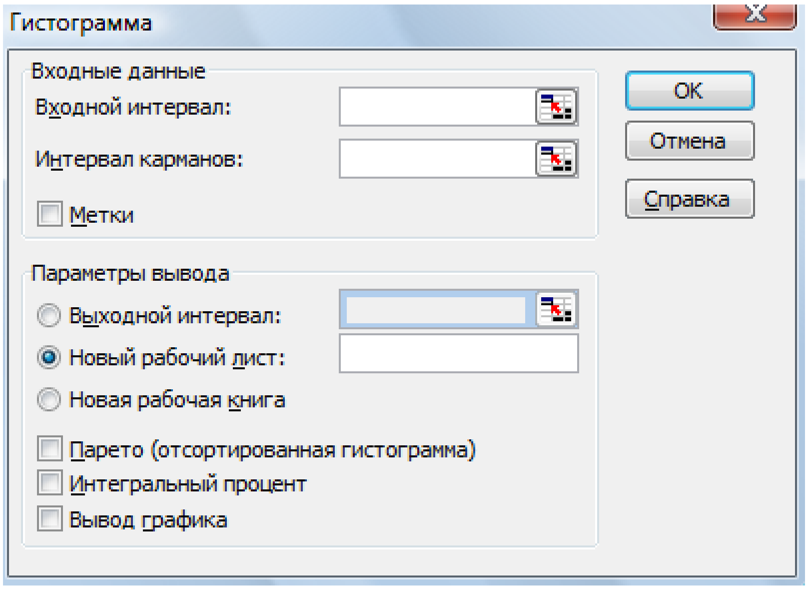


Рис. 8.2. Діалогове вікно для побудови гістограми.

3) Задати необхідні для побудови гістограми параметри:

входной диапазон – задати посилання на комірки, в яких знаходяться вхідні дані;

интервал карманов (параметр не є обов’язковим) – задати діапазон комірок і набір граничних значень у порядку зростання; якщо параметр не введений, то буде автоматично створений набір відрізків, рівномірно розподілених між мінімальним і максимальним значеннями даних;

выходной диапазон – ввести посилання на верхню ліву комірку діапазону, в який буде надано гістограму, або відмітити параметр *Новый рабочий лист* або *Новая рабочая книга*;

интегральный процент – якщо цей параметр відмічено, то будуть розраховані накопичені частоти і побудований їх графік;

вывод графика – якщо цей параметр відмічено, то буде створено автоматично діаграму, при цьому обов'язково задається значення *Новая книга*.

Приклад 1. Дана вибірка X=1, 0, –1, 0, 1, 0. Знайти точкову оцінку математичного сподівання, а також зміщену і незміщену оцінки дисперсії та середньоквадратичного відхилення.

Розв'язування. Зведемо вихідні дані до дисперсійної таблиці та виконаємо відповідні обчислення (табл. 8.2).

Таблиця 8.2. Дисперсійна таблиця до прикладу 1

n	x_i	$x_i^{\,2}$
1	1	1
2	0	0
3	–1	1
4	0	0
5	1	1
6	0	0
Σ	1	3

За даними, одержаними з табл.8.2, розрахуємо:

– математичне сподівання $\bar{x}$:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{6}\cdot 1 = \frac{1}{6};$$

– зміщену та незміщену оцінки дисперсії

$$\overline{\sigma_x^2} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2\right) - \frac{1}{n^2}\left(\sum_{i=1}^{n} x_i\right)^2 = \frac{1}{6}\cdot 3 - \frac{1}{6^2}\cdot 1^2 = \frac{17}{36};$$

$$\sigma_x^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{1}{n(n-1)}\left(\sum_{i=1}^{n} x_i\right)^2 = \frac{1}{5}\cdot 3 - \frac{1}{6\cdot 5}\cdot 1^2 = \frac{17}{30};$$

– зміщену та незміщену оцінки середньоквадратичного відхилення

$$\overline{\sigma_x} = \sqrt{\sigma_x^2} = \sqrt{\frac{17}{36}} = \frac{\sqrt{17}}{6}; \sigma_x = \sqrt{\sigma_x^2} = \sqrt{\frac{17}{30}}.$$

Приклад 2. За даними вибіркового дослідження відома заробітна платня (у грн.) 20-и службовців певної компанії (табл. 8.3). Знайти за допомогою вбудованих статистичних функцій Excel всі можливі числові характеристики за даними таблиці.

Таблиця 8.3

3560	2190	2390	3400
2180	2400	3350	2340
2900	2570	3300	3150
3680	3250	2250	3240
2180	2600	2870	3050

Розв'язання. Запишемо в лист Excel вхідні дані і числові характеристики, які можна знайти (рис. 8.3). Для знаходження характеристик введемо:

1) середнє значення (математичне сподівання) − формула **СРЗНАЧ** (A2:D6);

2) медіана − формула **МЕДИАНА** (A2:D6);

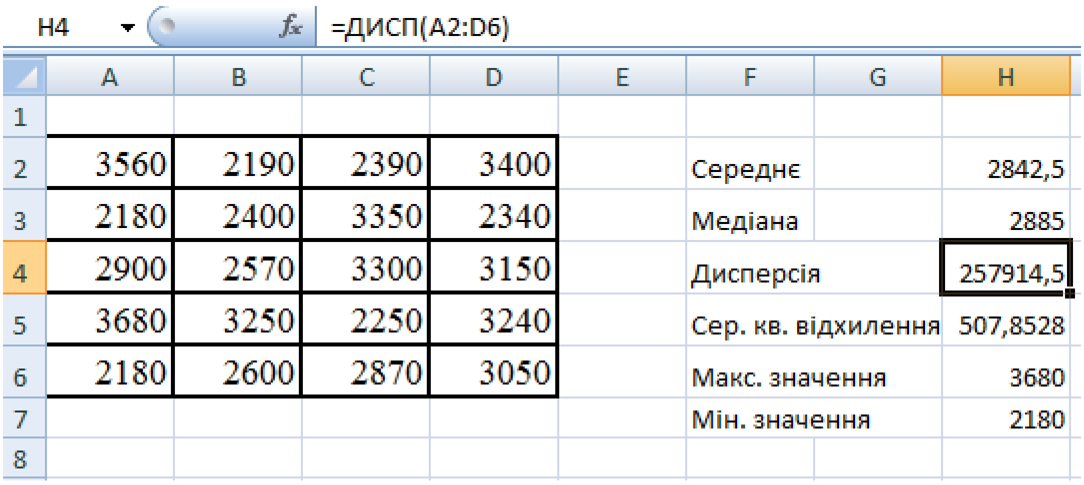
3) дисперсія (незміщена оцінка) − формула **ДИСП** (A2:D6);

4) середнє квадратичне відхилення (незміщена оцінка) − формула **СТАНДОТКЛОН** (A2:D6); також можна ввести **КОРЕНЬ** (H4), тобто обчислення за означенням (за коміркою дисперсії);

5) максимальне значення − формула **МАКС** (A2:D6);

6) мінімальне значення − формула **МИН** (A2:D6).

Зауваження. Функція **ДИСП** оцінює дисперсію за вибіркою, тобто вважається,що аргументи є вибіркою із генеральної сукупності. Якщо дані представляють всю генеральну сукупність, то слід використовувати функцію **ДИСПР.**



H4 =ДИСП(A2:D6)

	A	B	C	D	E	F	G	H
1								
2	3560	2190	2390	3400		Середнє		2842,5
3	2180	2400	3350	2340		Медіана		2885
4	2900	2570	3300	3150		Дисперсія		257914,5
5	3680	3250	2250	3240		Сер. кв. відхилення		507,8528
6	2180	2600	2870	3050		Макс. значення		3680
7						Мін. значення		2180
8								

Рис.8.3. Розрахунок числових характеристик.

Приклад 3. За даними вибіркового дослідження відома кількість родин з дітьми дошкільного віку в селах деякої області (табл. 8.4). Побудувати за допомогою Excel гістограму за даними таблиці.

Таблиця 8.4

27	36	34	46	43	28	29	37	40	43
40	33	50	37	41	32	27	43	34	32
30	41	54	42	47	35	49	49	54	36
36	51	36	24	35	25	33	38	38	36
29	51	32	36	53	30	55	44	46	38
29	44	48	30	34	46	47	36	37	36
30	58	42	46	46	29	38	44	40	30
35	35	63	47	37	29	53	41	42	41

Розв'язання. Запишемо в лист Excel вхідні дані завдання (рис. 8.4). Розрахуємо частоти попадання в інтервали (див. зміст командного рядка на рис. 8.4).

E11 {=ЧАСТОТА(B2:K9;B11:B21)}

	A	B	C	D	E	F	G	H	I	J	K
1											
2		27	36	34	46	43	28	29	37	40	43
3		40	33	50	37	41	32	27	43	34	32
4		30	41	54	42	47	35	49	49	54	36
5		36	51	36	24	35	25	33	38	38	36
6		29	51	32	36	53	30	55	44	46	38
7		29	44	48	30	34	46	47	36	37	36
8		30	58	42	46	46	29	38	44	40	30
9		35	35	63	47	37	29	53	41	42	41
10											
11		24		частоти	1						
12		28			4						
13		32			13						
14		36			17						
15		40			11						
16		44			13						
17		48			9						
18		52			5						
19		56			5						
20		60			1						
21		64			1						
22					0						

Рис. 8.4. Вхідні дані для побудови гістограми.

Викличемо *Вставка – Гистограмма*, задамо діапазон даних.

Для зручності читання діаграми добавимо *Заголовок* та *Значения*.

Приберемо відмітку *Легенда*, оскільки імена рядів не було надано – розглядається тільки один тип даних.

Після побудови діаграми можна у разі необхідності змінити шрифти, ширину стовпчиків гістограми, їх колір та фон. Для внесення змін потрібно двічі натиснути правою кнопкою миші на відповідне поле гістограми.

Зауважимо, що на горизонтальній осі надаються не границі інтервалів, а їх порядковий номер.

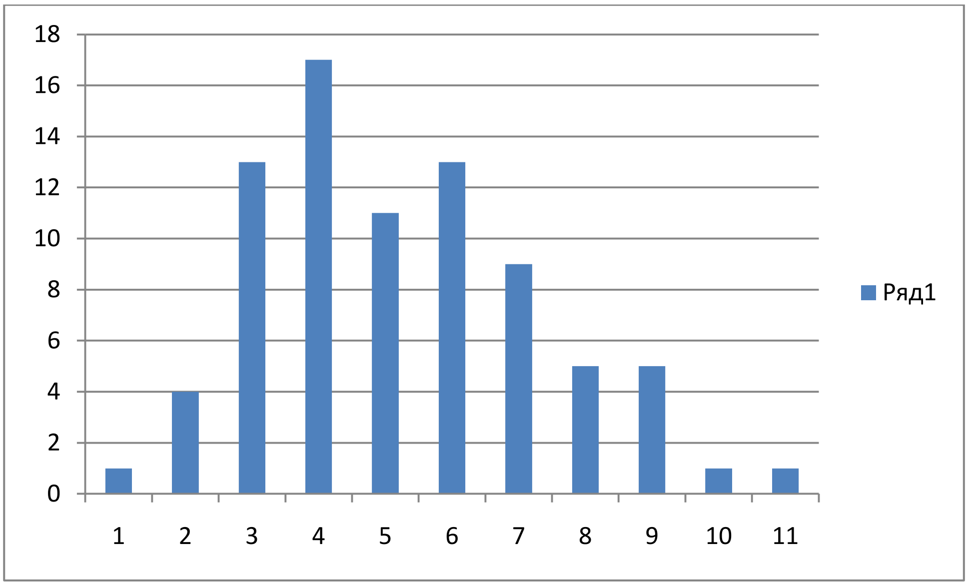


Рис. 8.5. Лист Excel із гістограмою.

Приклад 4. Обчислити вибіркове середнє та дисперсію, медіану та моду для вибірки

інтервал	$[2,4)$	$[4,6)$	$[6,8)$	$[8,10)$	$[10,12)$
n_i	2	8	35	40	15

та побудувати емпіричну функцію розподілу.

Розв'язання. Обсяг вибірки $n = 2+8+35+40+15 = 100$. Тоді

z_i	3	5	7	9	11
n_i	2	8	35	40	15

Обчислюємо числові характеристики:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i z_i = \frac{1}{100}(2\cdot 3 + 8\cdot 5 + 35\cdot 7 + 40\cdot 9 + 15\cdot 11) = 8{,}16;$$

$$\overline{s}^2 = \frac{1}{n}\sum_{i=1}^{k} n_i z_i^2 - \overline{x}^2 = \frac{1}{100}(2\cdot 9 + 8\cdot 25 + 35\cdot 49 + 40\cdot 81 + 15\cdot 121) - 8{,}16^2 =$$

$$= 69{,}88 - 66{,}5856 = 3{,}2944;$$

$$s^2 = \frac{n}{n-1}\overline{s}^2 = \frac{100}{99}\cdot 3{,}29 = 3{,}33;$$

$$M_e = 8 + 2\frac{50-45}{40} = 8{,}25; \quad M_0 = 8 + 2\frac{40-35}{80-35-15} = 8{,}33.$$

При побудові емпіричної функції розподілу (кумуляти) для інтервального статистичного розподілу вибірки за основу береться припущення, що ознака на кожному частинному інтервалі має рівномірну щільність імовірностей. Тому кумулята матиме вигляд ламаної лінії, яка зростає на кожному частковому інтервалі та наближається до 1. Отже, емпірична функція розподілу

$$F^*(x) = \begin{cases} 0, \; x \le 2, \\ 0{,}02, \; 2 < x \le 4, \\ 0{,}1, \; 4 < x \le 6, \\ 0{,}45, \; 6 < x \le 8, \\ 0{,}85, \; 8 < x \le 10, \\ 1, 10 < x \le 12, \end{cases}.$$

Графік кумуляти показано на рис. 8.6.

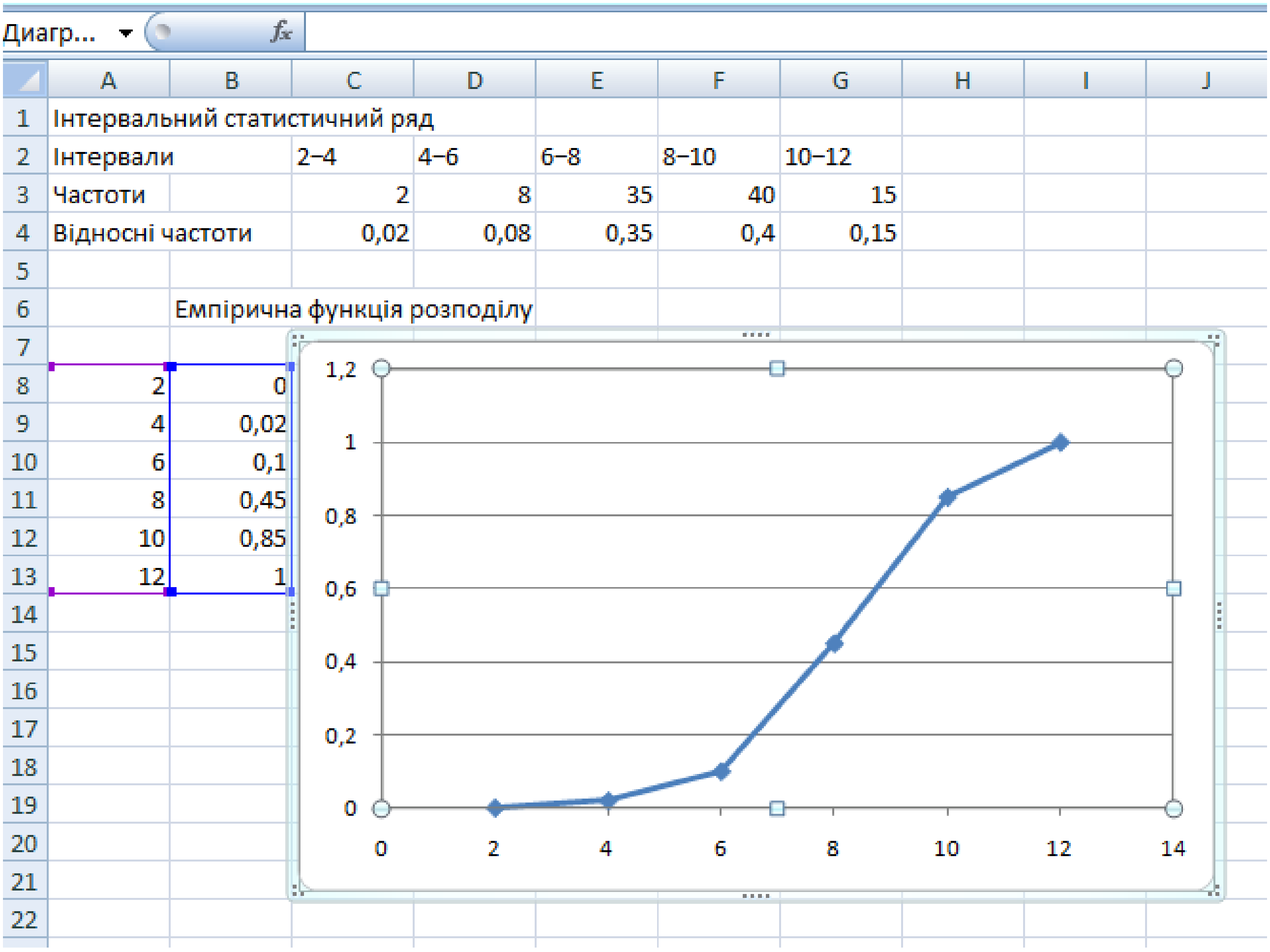


Рис. 8.6. Вигляд емпіричної функції розподілу.

Зауваження. З рисунка також визначається модальний інтервал (інтервал, що має найбільшу частоту появи), який дорівнює 8−10. Цей інтервал також є і медіанним, оскільки $F^*(8) < 0,5$, але $F^*(10) > 0,5$, то, беручи до уваги неперервність досліджуваної ознаки та властивість функції $F^*(x)$, яка є неспадною функцією, всередині інтервалу існує таке значення $X = M_e$, для якого $F^*(M_e) = 0,5.$

Відповідь. $\bar{x} = 8{,}16;\ s^2 = 3{,}33;\ M_e = 8{,}25;\ M_0 = 8{,}33.$

Виконання лабораторної роботи

Завдання до теми

Завдання 1.

Розв’язати наступні задачі за номером варіанта.

Номер варіанту	Номери завдань	Номер варіанту	Номери завдань
1	1.1, 1.2	9	1.6, 1.3
2	1.4, 1.3	10	1.9, 1.8
3	1.6, 1.5	11	1.2, 1.4
4	1.9, 1.7	12	1.4, 1.8
5	1.10, 1.8	13	1.6, 1.7
6	1.10, 1.2	14	1.2, 1.8
7	1.1, 1.5	15	1.3, 1.9
8	1.4, 1.7		

1.1. Відомі дані про безаварійну роботу автоматизованого поліграфічного комплексу (в місяцях) (табл. 8.5). Побудувати статистичний ряд за даними вибірки, визначити середній час безаварійної роботи, дисперсію і середнє квадратичне відхилення часу. Побудувати полігон частот і гістограму.

Таблиця 8.5

0,000	0,000	0,002	0,006	0,023	0,084	0,382	0,810	0,003	0,864
1,033	0,912	0,093	0,324	0,194	0,522	2,336	0,057	0,654	0,250
0,877	0,276	0,037	0,537	0,183	1,306	0,752	0,198	1,623	0,875
0,185	0,274	0,613	0,356	0,645	0,676	1,079	0,500	0,902	0,191
0,250	0,348	0,320	0,182	0,458	0,936	1,204	0,576	0,303	0,522

1.2. Інтервал між потягами у метро складає 3 хв. В таблиці 8.6 надано час очікування пасажирами потягу. Скласти інтервальний статистичний ряд, знайти середній час очікування, медіану, дисперсію і середнє квадратичне відхилення часу. Побудувати функцію розподілу величини X – часу очікування.

Таблиця 8.6

0,787	1,004	0,941	0,612	1,200	1,692	1,354	0,908	1,245	1,292
0,617	0,828	1,413	1,030	1,459	2,483	2,769	1,563	2,661	1,635
1,654	0,838	1,143	0,618	2,317	1,853	1,555	0,653	1,922	1,653
1,747	2,677	0,341	2,952	0,545	1,297	1,981	0,214	2,452	2,087
0,001	0,007	0,025	0,312	1,068	2,604	0,014	0,045	2,340	2,001

1.3. Відомі дані про інтервал часу між появою покупців у касовому залі деякого магазину (табл.8.7). Побудувати інтервальний статистичний ряд, знайти всі можливі числові характеристики. Побудувати графік функції розподілу випадкової величини X – інтервалу часу.

Таблиця 8.7

0,002	1,004	0,007	0,612	0,091	1,692	1,527	0,908	2,590	1,292
4,134	3,647	0,374	2,150	0,778	5,223	3,344	2,001	3,492	4,011
3,507	0,838	0,148	0,618	0,704	1,853	3,007	0,653	3,600	1,653
0,738	1,069	2,453	1,447	2,614	3,742	4,314	1,211	1,949	5,001
1,000	0,007	1,272	0,312	1,832	2,604	2,267	0,045	4,450	2,001

1.4. В таблиці 8.8 наведено значення прибутку 50 фірм, що належать одній корпорації (в 1000 у. од.). Знайти середнє значення прибутку, дисперсію і середнє квадратичне відхилення. Побудувати полігон частот і гістограму.

Таблиця 8.8

4,744	9,127	7,201	8,650	11,534	9,013	10,390	9,268	7,354	10,255
6,232	6,739	6,088	8,671	15,103	9,124	11,902	10,216	10,954	11,470
7,351	9,832	7,126	9,715	10,744	10,687	10,582	12,271	11,047	13,190

5,536	8,917	9,823	8,383	14,212	15,031	13,001	11,089	12,091	10,321
9,766	5,854	2,917	6,379	6,748	7,024	11,587	11,101	10,954	10,387

1.5. Відомі дані про місячний обсяг виробництва (тис. книг) підприємств поліграфічного виробництва (табл. 8.9). Побудувати інтервальний статистичний ряд, полігон частот і гістограму. Знайти всі можливі числові характеристики. Побудувати графіки функції розподілу.

Таблиця 8.9

11,240	18,545	17,750	22,560	18,355	20,424	20,650	10,780	15,590
13,720	28,505	23,170	20,360	22,450	21,590	14,565	24,295	25,140
27,655	17,786	27,045	28,650	18,670	31,445	18,540	15,598	19,720
15,230	21,240	19,535	12,934	18,195	19,074	17,037	19,610	20,970
22,075	15,090	20,754	10,195	13,580	21,490	13,987	22,645	21,218

1.6. В таблиці 8.10 наведено середні значення прибутку (в грн.) опитаних 50 сімей в поточному році. Знайти середнє значення всього прибутку, дисперсію і середнє квадратичне відхилення. Побудувати полігон частот і гістограму.

Таблиця 8.10

4844	9327	7210	8650	11540	9010	10390	9260	7350	10255
6530	6730	6088	10610	15100	9524	11900	10210	10954	11570
7350	9832	7120	9715	10745	10680	10580	12270	11047	13290
5546	8947	9923	8383	14252	15031	13011	11289	12090	10325
9760	5855	2920	6400	6748	7824	11980	11100	10954	10380

1.7. Відомі дані про інтервал часу між телефонними повідомленнями в залі телефонної станції (табл.8.11). Побудувати інтервальний статистичний ряд, знайти всі можливі числові характеристики. Побудувати графік функції розподілу випадкової величини X – інтервалу часу.

Таблиця 8.11

0,0025	1,0042	0,0072	0,6121	0,0912	1,6922	1,5274	0,9081	2,5991	1,2925
4,1342	3,6471	0,3742	2,1504	0,7789	5,2239	3,3447	2,0016	3,4927	4,0116
3,5078	0,8389	0,1486	0,6184	0,7041	1,8535	3,2271	0,6531	3,6564	1,6536
1,7382	1,0694	0,4536	1,4479	1,6146	2,7421	1,3141	1,2112	1,9492	1,0014

1.8. Наведено результати дослідження річного обсягу споживання риби та рибної продукції в кг на душу населення (табл.8.12). Побудувати інтервальний статистичний ряд, знайти всі можливі числові характеристики. Побудувати графік функції розподілу.

Таблиця 8.12

14	10,5	9,8	12,5	11,5	13,5	12,4	10,4	11	12
10,2	11,5	9,8	13,6	12,5	9,5	10,2	11,9	11,4	12,8
11,6	12,3	10,8	9,9	11,1	10,7	11,4	10,9	12,9	11,1
13,8	14	13,1	10,7	12,9	13,9	10,1	12,2	11,2	13,1

1.9. В таблиці 8.13 наведено середні значення ваги 50 шоколадок (в грамах), яка визначалась на ексериментальній лінії запуску. Знайти середнє значення всього прибутку, дисперсію і середнє квадратичне відхилення. Побудувати полігон частот і гістограму.

Таблиця 8.13

100,01	101,02	100,32	100,25	101,60	99,98	99,79	100,04	97,99	102,55
100,30	99,73	99,88	100,10	101,00	97,24	100,90	102,10	99,54	101,70
99,50	98,32	100,20	97,15	100,45	101,80	101,80	102,70	100,47	99,90
99,46	99,47	99,23	93,83	102,52	100,31	100,11	100,89	100,90	98,25
97,69	98,55	99,20	99,90	97,48	100,24	100,80	101,00	100,54	99,89

1.10. Відомі дані про можливості збою проявлення пластини в автоматичному проявному процесорі (в секундах) (табл. 8.14). Побудувати статистичний ряд за даними вибірки, визначити середній час безаварійної роботи, дисперсію і середнє квадратичне відхилення часу. Побудувати полігон частот і гістограму.

Таблиця 8.14

0,00	0,00	0,02	0,06	0,23	0,08	0,32	0,80	0,00	0,84
1,04	0,91	0,03	0,34	0,94	0,52	2,36	0,57	0,65	0,50
0,67	0,76	0,03	0,57	0,83	1,30	0,72	0,98	1,62	0,85
0,85	0,74	0,61	0,56	0,45	0,67	1,09	0,00	0,90	0,91
0,25	0,48	0,32	0,82	0,58	0,93	1,24	0,76	0,30	0,22

Завдання 2. Обчислити вибіркове середнє та дисперсію, медіану та моду для вибірки та побудувати емпіричну функцію розподілу. Всі обчислення бажано виконати за допомогою Excel.

2.1.

інтервал	$[-2,0)$	$[0,2)$	$[2,4)$	$[4,6)$	$[6,8)$
n_i	4	6	25	40	25

2.2.

інтервал	$[2,3)$	$[3,4)$	$[4,5)$	$[5,6)$	$[6,7)$
n_i	12	8	35	40	15

2.3.

інтервал	$[-12,-10)$	$[-10,-8)$	$[-8,-6)$	$[-6,-4)$	$[-4,-2)$
n_i	7	13	32	38	10

2.4.

інтервал	$[1,2)$	$[2,3)$	$[3,4)$	$[4,5)$	$[5,6)$
n_i	20	8	32	22	18

2.5.

інтервал	$[1,2;1,4)$	$[1,4;1,6)$	$[1,6;1,8)$	$[1,8;1,0)$	$[1,0;1,2)$
n_i	23	7	15	40	15

2.6.

інтервал	$[0,1)$	$[1,2)$	$[2,3)$	$[3,4)$	$[4,5)$
n_i	11	45	5	4	35

2.7.

інтервал	$[-1,2)$	$[2,5)$	$[5,8)$	$[8,11)$	$[11,13)$
n_i	7	8	15	40	30

2.8.

інтервал	$[0,2;0,4)$	$[0,4;0,6)$	$[0,6;0,8)$	$[0,8;1)$	$[1;1,2)$
n_i	21	2	8	49	20

2.9.

інтервал	$[3,4)$	$[4,5)$	$[5,6)$	$[6,7)$	$[7,8)$
n_i	25	15	25	20	15

2.10.

інтервал	$[4,5)$	$[5,6)$	$[6,7)$	$[7,8)$	$[8,9)$
n_i	2	18	25	30	25

2.11.

інтервал	$[7,8)$	$[8,9)$	$[9,10)$	$[10,11)$	$[11,12)$
n_i	9	11	5	50	25

2.12.

інтервал	$[0;0,4)$	$[0,4;0,8)$	$[0,8;1,2)$	$[1,2;1,6)$	$[1,6;2)$
n_i	30	36	5	14	15

2.13.

інтервал	$[-1,2;-0,8)$	$[-0,8;-0,4)$	$[-0,4;0)$	$[0;0,4)$	$[0,4;0,8)$
n_i	23	17	15	40	5

2.14.

інтервал	$[-5,-4)$	$[-4,-3)$	$[-3,-2)$	$[-2,-1)$	$[-1,0)$
n_i	30	18	12	20	20

2.15.

інтервал	$[-3,1)$	$[1,5)$	$[5,9)$	$[9,13)$	$[13,17)$
n_i	7	18	25	10	40

Завдання 3. Знати відповіді на подані нижче теоретичні запитання.

Теоретичні запитання до теми

1. Пояснити відмінність між статистичним рядом вибірки та інтервальним статистичним рядом.
2. Записати основні числові характеристики вибірки, пояснити їх обчислення за допомогою вбудованих статистичних функцій Excel.
3. Зміщена та незміщена оцінка дисперсії, формули обчислення.
4. Пояснити побудову гістограм.
5. Пояснити побудову емпіричної функції розподілу.

Оформлення звіту та порядок захисту

Лабораторна робота виконується на аркушах А4, в ній стисло відображаються формули теоретичної частини, розв’язання завдань самостійної роботи та отримані результати (студент здає електронний варіант проведених обчислень практичних задач.

Лабораторна робота № 9

Побудова надійних інтервалів. Обчислення коефіцієнта кореляції та перевірка його статистичної значущості

Мета роботи: навчитися будувати надійні інтервали для математичного сподівання у випадку відомої та невідомої дисперсії, обчислювати коефіцієнт кореляції, використовуючи теоретичні знання та можливості застосування табличного процесора Microsoft Excel.

Теоретичні відомості

9.1. Інтервальні оцінки параметрів розподілу

Означення. ***Інтервальною оцінкою*** параметрів розподілу генеральної сукупності називають оцінку, яка визначається двома числами – кінцями інтервалу.

Нехай $X_1, X_2, ..., X_n$ – вибірка з генеральної сукупності, θ – невідомий параметр генеральної сукупності X. При інтервальному оцінюванні невідомого параметра θ шукають такі дві функції $h_1 = h_1(X_1, X_2, \ldots, X_n)$ і $h_2 = h_2(X_1, X_2, \ldots, X_n)$, для яких завжди $h_1 < h_2$ та при заданому $\gamma \in (0,1)$ виконується умова

$$P(h_1 < \theta < h_2) \geq \gamma. \tag{9.1}$$

Тоді інтервал (h_1, h_2) називають *надійним інтервалом*, γ – *рівнем надійності*, а числа h_1 і h_2 – *нижньою та верхньою межами надійності*.

Рівень надійності задається наперед і найчастіше беруть 0,95; 0,99.

9.2. Побудова надійних інтервалів

Для знаходження надійного інтервалу математичного сподівання у ***випадку відомої дисперсії*** виконаємо наступні дії:

а) розрахуємо за допомогою дисперсійної таблиці значення математичного сподівання $\overline{x}$ за формулою $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$;

б) установимо надійну ймовірність або рівень надійності (рівень значущості);

в) за таблицею нормального закону розподілу запишемо надійний інтервал: $\left(\overline{x} - \varepsilon_{\alpha} \cdot \frac{\sigma}{\sqrt{n}}, \overline{x} + \varepsilon_{\alpha} \cdot \frac{\sigma}{\sqrt{n}} \right)$ або $x \pm \varepsilon_{\alpha} \cdot \frac{\sigma}{\sqrt{n}}$.

Побудований надійний інтервал

$$\left(\overline{x} - \frac{t\sigma}{\sqrt{n}},\ \overline{x} + \frac{t\sigma}{\sqrt{n}} \right), \tag{9.2}$$

де $t = \frac{\varepsilon\sqrt{n}}{\sigma}$, містить в собі невідомий параметр a (математичне сподівання) з ймовірністю γ. Число t при заданому значенні γ знаходимо із таблиці значень функції Лапласа.

Висновки:

1) при збільшенні обсягу n вибірки число $\varepsilon = \frac{\sigma t}{\sqrt{n}}$ зменшується, тобто точність оцінки підвищується;

2) зростання надійності $\gamma = 2\Phi(t)$ веде до збільшення t, отже, до зростання ε, або до зниження точності.

Для побудови надійного інтервалу математичного сподівання у ***випадку невідомої дисперсії*** виконаємо наступні дії:

а) розрахуємо точкові оцінки основних характеристик випадкових величин за допомогою дисперсійної таблиці за відповідними формулами (лабораторна робота № 8);

б) задамо $\gamma -$ надійну ймовірність або рівень значущості;

в) для ймовірності γ та $(n-1)$ степенів вільності за таблицею розподілу Стьюдента знайдемо $t_{\gamma,n-1}$. Перший стовпчик таблиці (додаток) відповідає кількості степенів вільності $(n-1)$, перший рядок – ймовірності γ. На перетині $(n-1)$ рядка та γ стовпчика знаходимо шукане значення $t_{\gamma,n-1}$;

г) запишемо надійний інтервал

$$\bar{x} \mp t_{\gamma,n-1} \cdot \frac{s}{\sqrt{n-1}}.$$

Зауваження. У статистиці кількістю степенів вільності певної величини називають різницю між кількістю випробувань і кількістю величин, обчислених завдяки цим випробуванням.

9.3. Визначення надійних інтервалів із використанням табличного процессору Microsoft Excel

Більшість числових характеристик у випадку незгрупованих даних можна обчислити (лаб. № 8) із використанням табличного процесору Microsoft Excel.

Ширину надійного інтервалу для генерального середнього можна знайти за допомогою вбудованої статистичної функції Excel **ДОВЕРИТ** (*альфа, станд_откл, размер*). Параметр *альфа* – це так званий рівень значущості, $\alpha = 1 - \gamma$; параметр *станд_откл* – це вибіркове середнє квадратичне відхилення *S*; параметр *размер* – це обсяг вибірки.

Приклад 1. Нехай $\sigma = 2,\ n = 25,\ \gamma = 0{,}95$. Знайти надійний інтервал для a, якщо $\bar{x} = 5$.

Розв'язання. Для знаходження t використаємо рівняння

$$2\Phi(t) = \gamma \Rightarrow 2\Phi(t) = 0{,}95 \Rightarrow \Phi(t) = 0{,}475.$$

Із таблиці значень функції Лапласа знаходимо $t = 1{,}96.$

Отже, $\varepsilon = \dfrac{2 \cdot 1{,}96}{\sqrt{25}} = 0{,}784.$

Аналогічний результат маємо при використанні обчислень, як на рис. (9.1).

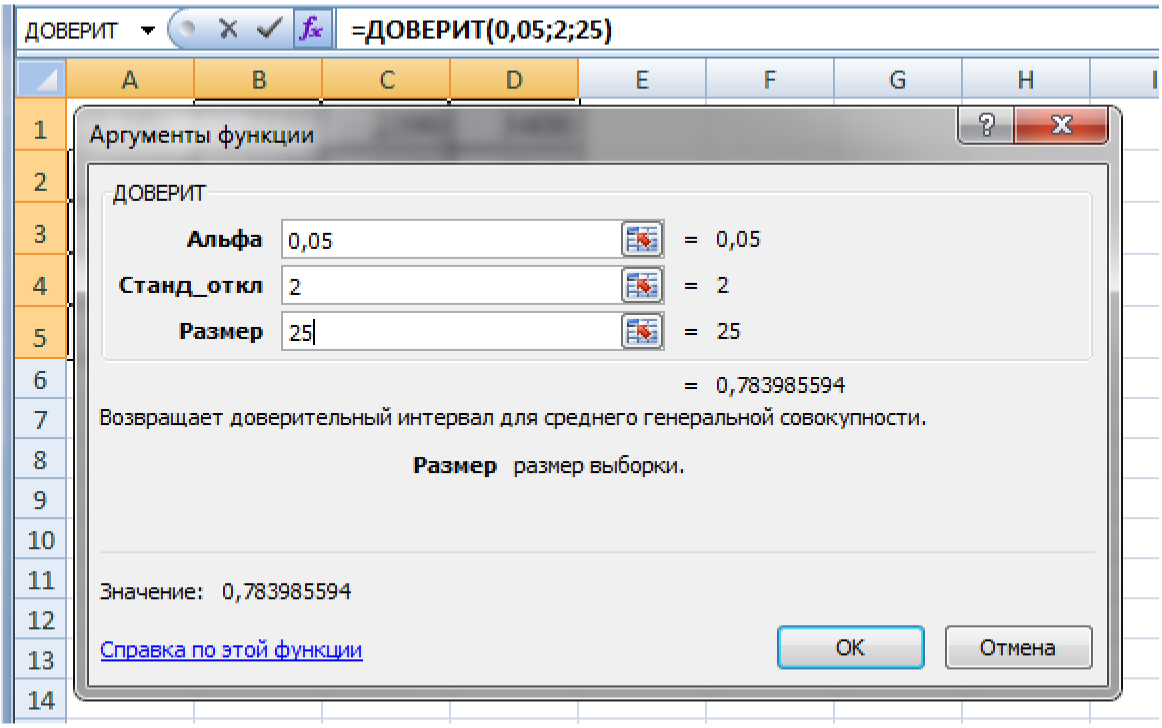


Рис.9.1. Обчислення надійного інтервала.

За формулою (9.14) отримаємо інтервал $\left(\bar{x}-0{,}784;\ \bar{x}+0{,}784\right)$ або $\left(4{,}216;\ 5{,}784\right)$.

Цей результат можна трактувати так: якщо зроблена достатньо велика кількість вибірок, то в 95% випадків значення *a* належить знайденому інтервалі, а в 5% це значення *a* може вийти за межі інтервалу.

Приклад 2. Нехай *X* – нормально розподілена випадкова величина; яка має такі параметри: $n=25$, $\bar{x}=20$, $s=0{,}4$. Знайти надійний інтервал для оцінки математичного сподівання, якщо $\gamma=0{,}95$.

Розв'язання. Знаходимо t_γ з таблиці: $t(0{,}95;\ 24)=2{,}064$. Тоді

$$t_\gamma \frac{s}{\sqrt{n}} = 2{,}064 \cdot \frac{0{,}4}{5} \approx 0{,}165.$$

Отже, $(\bar{x}-0{,}165;\ \ \bar{x}+0{,}165)$ – надійний інтервал або $(19{,}835;\ \ 20{,}165)$.

Зауваження. Значення t_γ обчислюють також засобами программного забезпечення, як на рис. 9.2, вказавши ймовірність $P=1-p=1-0{,}95=0{,}05$.

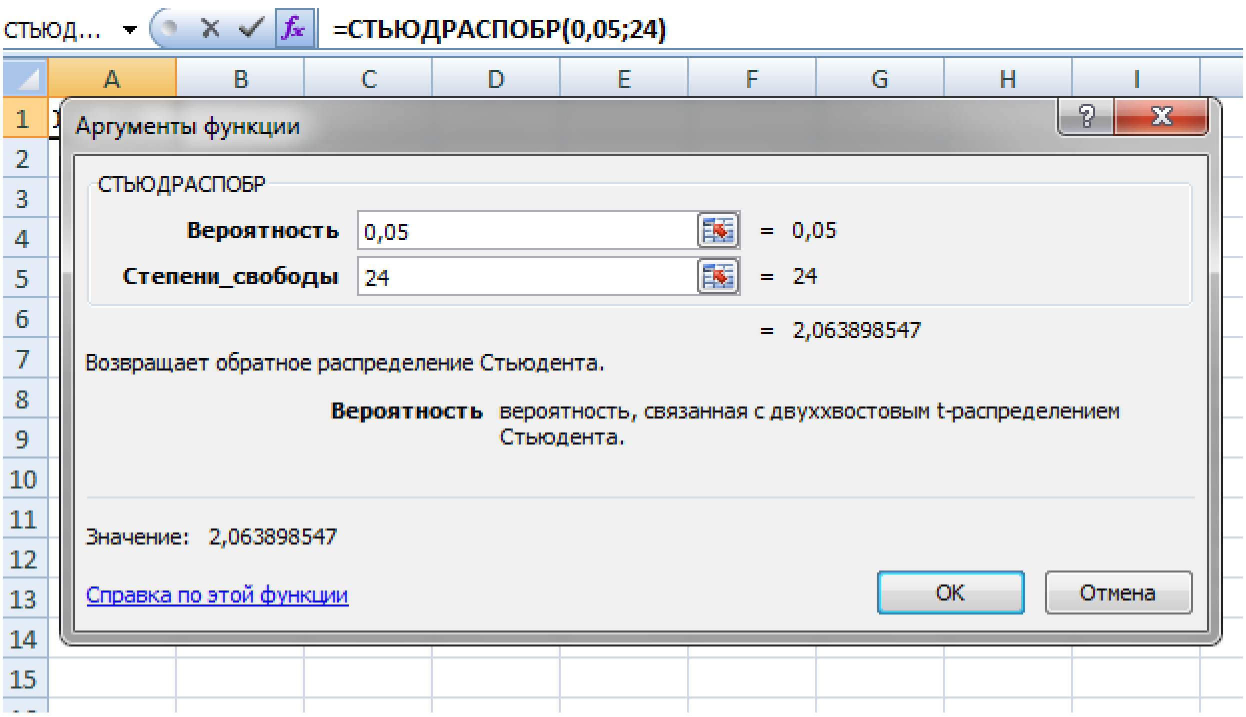


Рис. 9.2. Обчислення t_{γ} -характеристики.

Також далі проводять обчислення надійного інтервала , рис.9.3, задавши відповідно до формули (9.2) дії.

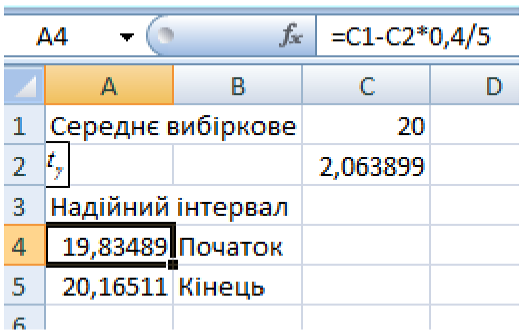


Рис. 9.3. Обчислення надійного інтервала.

Приклад 2. За даними вибіркового дослідження відома заробітна платня (у грн.) 20-и службовців певної компанії (табл. 9.1). Знайти за допомогою вбудованих статистичних функцій Excel всі можливі числові характеристики за даними таблиці (приклад із лабораторної роботи №8).

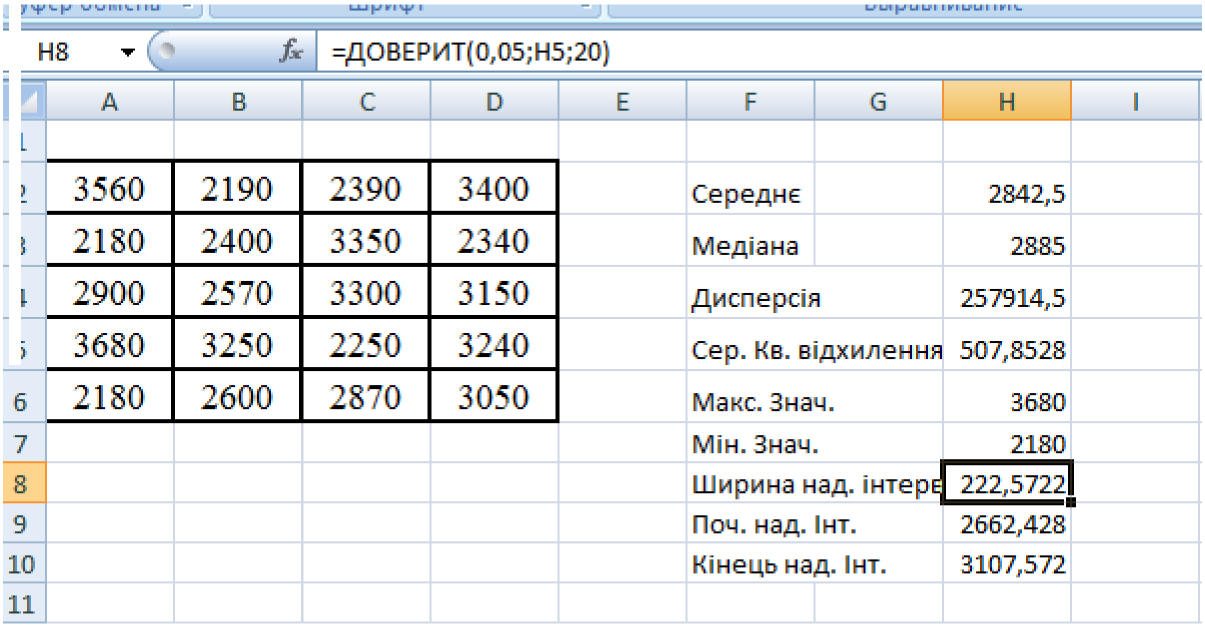
Таблиця 9.1

3560	2190	2390	3400
2180	2400	3350	2340
2900	2570	3300	3150
3680	3250	2250	3240
2180	2600	2870	3050

Розв'язання. Запишемо в лист Excel вхідні дані і числові характеристики, які можна знайти (рис. 9.4).

Для знаходження надійного інтервалу для генерального середнього знайдемо за допомогою функції **ДОВЕРИТ** його ширину (див. Рис. 5.3, командний рядок). Параметрами візьмемо $\alpha = 1 - \gamma = 1 - 0,95 = 0,05$; замість другого параметру надамо посилання на комірку, що містить розраховане значення середнього квадратичного відхилення; *размер* – це обсяг вибірки, що дорівнює 20.

Для знаходження початку інтервалу запишемо в комірку H9 формулу «=H3-H8»; для знаходження кінця – формулу «=H3+H8» в комірку H10.


H8 =ДОВЕРИТ(0,05;H5;20)

	A	B	C	D	E	F	G	H	I
1									
2	3560	2190	2390	3400		Середнє		2842,5	
3	2180	2400	3350	2340		Медіана		2885	
4	2900	2570	3300	3150		Дисперсія		257914,5	
5	3680	3250	2250	3240		Сер. Кв. відхилення		507,8528	
6	2180	2600	2870	3050		Макс. Знач.		3680	
7						Мін. Знач.		2180	
8						Ширина над. інтерв		222,5722	
9						Поч. над. Інт.		2662,428	
10						Кінець над. Інт.		3107,572	
11									

Рис. 9.4. Розрахунок числових характеристик та надійного інтервала.

Приклад 3. Побудувати надійний інтервал із ймовірністю 95% при $\sigma = 1$ і надійний інтервал у випадку невідомої дисперсії із ймовірністю 95%. Порівняти побудовані надійні інтервали.

Розв'язування. Зведемо вихідні дані до дисперсійної таблиці та виконаємо відповідні обчислення (табл. 9.2).

Таблиця 9.2. Дисперсійна таблиця до прикладу 3

n	x_i	$x_i^{\ 2}$
1	1	1

2	0	0
3	–1	1
4	0	0
5	1	1
6	0	0
Σ	1	3

За даними, одержаними з табл. 9.2, розрахуємо:

– математичне сподівання $\bar{x}$:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{6}\cdot 1 = \frac{1}{6};$$

– зміщену та незміщену оцінки дисперсії

$$\overline{\sigma_x^2} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2\right) - \frac{1}{n^2}\left(\sum_{i=1}^{n} x_i\right)^2 = \frac{1}{6}\cdot 3 - \frac{1}{6^2}\cdot 1^2 = \frac{17}{36};$$

$$\sigma_x^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{1}{n(n-1)}\left(\sum_{i=1}^{n} x_i\right)^2 = \frac{1}{5}\cdot 3 - \frac{1}{6\cdot 5}\cdot 1^2 = \frac{17}{30};$$

– зміщену та незміщену оцінки середньоквадратичного відхилення

$$\overline{\sigma_x} = \sqrt{\sigma_x^2} = \sqrt{\frac{17}{36}} = \frac{\sqrt{17}}{6};$$

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{\frac{17}{30}}.$$

Знайдемо надійний інтервал у випадку відомої дисперсії $\sigma = 1$ з надійністю 95%, тобто $\gamma = 0{,}95$. За таблицями нормального розподілу маємо число 1,96. Тоді надійний інтервал матиме вигляд

$$\left(\frac{1}{6} - 1{,}96\frac{1}{\sqrt{6}};\frac{1}{6} + 1{,}96\frac{1}{\sqrt{6}}\right);\ (–0{,}63;\ 0{,}97).$$

Знайдемо надійний інтервал для математичного сподівання у випадку невідомої дисперсії з надійністю 95%.

Для рівня зі значущістю 5% та п’ятьма степенями вільності за таблицею розподілу Стьюдента знаходимо $t_{табл(0.95,5)}$. Перший стовпчик таблиці показує кількість степенів вільності, перший рядок – рівень надійності. На перетині знаходимо табличне значення $t_{табл(0.95,5)}$. Тоді надійний інтервал матиме вигляд

$$\left(\frac{1}{6}-2,57\frac{0,566}{\sqrt{6-1}};\frac{1}{6}+2,57\frac{0,566}{\sqrt{6-1}}\right); (–0,48; 0,82).$$

9.4. Поняття кореляційного зв’язку між досліджуваними величинами

В багатьох прикладних задачах необхідно виявити залежність між двома властивостями (ознаками) X і Y одного і того ж економічного об'єкту, або між певними ознаками різних об’єктів. Якщо вказані ознаки допускають кількісне вимірювання, і, з погляду економічної теорії, виходячи з економічної характеристики об'єкту, ознака Y залежить від ознаки X, тоді X можна назвати незалежною змінною, або факторною ознакою, або просто фактором, а Y – залежною змінною або результативною ознакою.

Якщо кожному значенню факторної ознаки X відповідає одне і тільки одне значення результативної ознаки Y, то говорять, що між цими ознаками існує функціональний зв’язок: $Y=f(X)$.Якщо кожному значенню факторної ознаки X відповідає безліч значень результативної ознаки Y, то між цими ознаками існує статистичний зв’язок.

Вивчення статистичного зв'язку дуже складний і трудомісткий процес, у якому потрібно аналізувати багатомірні таблиці даних. Тому зазвичай вивчається не статистичний, а кореляційний зв'язок між X та Y. Якщо кожному значенню факторної ознаки X відповідає певне середнє значення результативної ознаки Y, то між цими ознаками існує кореляційний зв’язок, тобто кореляційною є функціональна залежність між значеннями X і середніми значеннями Y: $\overline{Y}=f(X)$.

Наприклад, відомо, що з однакових за площею ділянок землі при рівних кількостях внесеного добриву отримують різний урожай. Тому, якщо Y – урожайність зерна, а X – кількість внесеного добрива, то функціонального зв’язку між X та Y немає. Це пояснюється впливом таких випадкових факторів, як температура повітря, кількість опадів і т. ін. Однак досвід показує, що середній урожай є функцією від кількості добрива, тобто між X та Y існує кореляційний зв'язок.

Основними задачами кореляційного аналізу є:

− вивчення сили зв'язку між двома і більше ознаками досліджуваного об'єкту;

− встановлення факторів, що найбільш суттєво впливають на результативну ознаку;

− виявлення невідомих причинно-наслідкових зв'язків між ознаками об'єкту.

Групування даних для кореляційного аналізу

Вибіркові дані для вивчення кореляційного зв'язку між ознаками X та Y мають вигляд пар їх значень: $(x_1; y_1)$, $(x_2; y_2)$, …, $(x_n; y_n)$, x_i – значення величини X, y_i – значення Y, n – кількість пар значень, $i = \overline{1, n}$.

Якщо кількість пар значень достатньо велика (принаймні, $n > 20$), то для зручності розрахунків дані групуються.

Для групування даних необхідно:

1) Розбити множини значень X та Y на інтервали, їх кількість для X та Y може бути різною (позначення: k – кількість інтервалів для X; m – кількість інтервалів для Y).

2) Зобразити дані графічно: побудувати на площині точки з координатами $(x_i; y_j)$. В результаті отримується площина, розбита на прямокутники, в кожному з яких може бути множина точок (рис. 9.5). Вказане графічне зображення вибіркових даних називається полем кореляції.

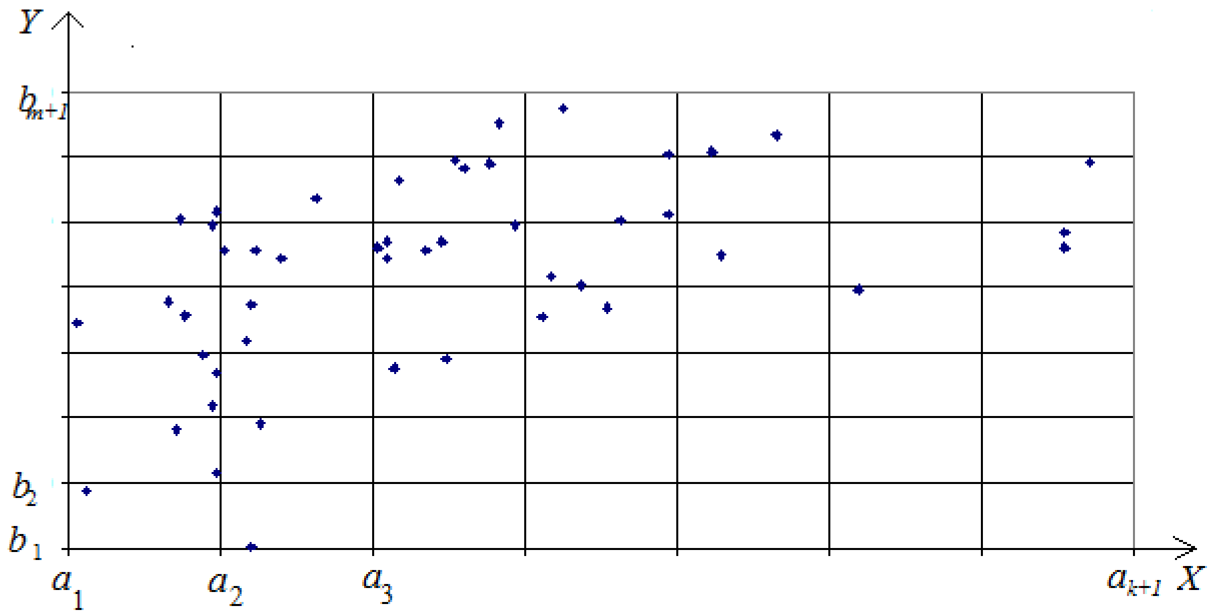


Рис. 9.5. Поле кореляції.

3) Побудувати кореляційну таблицю (табл.9.3). В першому рядку, розбитому на дві частини, записуються інтервали $[a_i;a_{i+1})$ для X та їх середини x_i. У першому стовпчику, розбитому на дві частини, записуються інтервали $[b_j;b_{j+1})$ для Y та їх середини y_j. В центральній частині таблиці записуються частоти n_{ij} – кількість точок, що потратили в прямокутник, обмежений по осі X інтервалом $[a_i;a_{i+1})$ і по осі Y інтервалом $[b_j;b_{j+1})$. В останньому рядку таблиці записуються частоти n_i для X – кількість точок, що потратили в прямокутники, які відповідають інтервалу $[a_i;a_{i+1})$, тобто $n_i = \sum_{j=1}^{m} n_{ij}$ – сума частот n_{ij} в стовпчики з номером i. В останньому стовпчику таблиці записуються частоти n_j для Y – кількість точок, що потратили в прямокутники, які відповідають інтервалу $[b_j;b_{j+1})$, тобто $n_j = \sum_{i=1}^{k} n_{ij}$ – сума частот n_{ij} в рядку з номером j.

Кореляційну таблицю можна розглядати як своєрідний подвійний статистичний ряд.

Таблиця 9.3

X (інтервали і їх середини) / Y (інтервали і їх середини)		$[a_1;a_2)$	$[a_2;a_3)$	$\ldots$	$[a_k;a_{k+1})$	$n_j=\sum_{i=1}^{k} n_{ij}$
		x_1	x_2	$\ldots$	x_k	
$[b_1;b_2)$	y_1	n_{11}	n_{21}	$\ldots$	n_{k1}	n_1
$[b_2;b_3)$	y_2	n_{12}	n_{22}	$\ldots$	n_{k2}	n_2
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$[b_m;b_{m+1})$	y_m	n_{1m}	n_{2m}	$\ldots$	n_{km}	n_m
$n_i=\sum_{j=1}^{m} n_{ij}$		n_1	n_2	$\ldots$	n_k	

4) За даними кореляційної таблиці будується ряд, що відображає залежність середнього значення *Y* від *X* (табл.9.4). В першому рядку таблиці записуються середини інтервалів x_i, в другому – відповідні середні значення $\overline{y}_{x_i}$, що знаходяться за формулами: $\overline{y}_{x_1}=\dfrac{y_1n_{11}+y_2n_{12}+\ldots+y_mn_{1m}}{n_1}$; $\overline{y}_{x_2}=\dfrac{y_1n_{21}+y_2n_{22}+\ldots+y_mn_{2m}}{n_2}$; … $\overline{y}_{x_k}=\dfrac{y_1n_{k1}+y_2n_{k2}+\ldots+y_mn_{km}}{n_k}$.

Таблиця 9.4

x_i	x_1	x_2	$\ldots$	x_k
$\overline{y}_{x_i}$	$\overline{y}_{x_1}$	$\overline{y}_{x_2}$	$\ldots$	$\overline{y}_{x_k}$
n_i	n_1	n_2	$\ldots$	n_k

В результаті отримується статистичний ряд, що містить значення *X*, відповідні середні значення *Y* та частоти. За даними такого ряду проводиться кореляційний аналіз.

9.5. Коефіцієнт кореляції Пірсона

Для оцінки тісноти (або сили) зв’язку між *X* та *Y* слугує коефіцієнт кореляції. У випадку, коли між *X* та *Y* існує *лінійний зв’язок та вибіркові дані розподілені за нормальним законом*, використовується коефіцієнт кореляції Пірсона, який зветься ще *параметричним коефіцієнтом кореляції*.

Коефіцієнт кореляції Пірсона розраховується за формулою:

$$r=\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{s_x\cdot s_y}, \qquad (9.3)$$

де $\overline{x}$ – вибіркове середнє величини *X*; $\overline{y}$ – вибіркове середнє величини *Y*; $\overline{xy}$ – вибіркове середнє величини *XY*; s_x – вибіркове середнє квадратичне відхилення величини *X*; s_y – вибіркове середнє квадратичне відхилення величини *Y*.

Враховуючи формули для знаходження вибіркових середніх і середніх квадратичних відхилень, а саме:

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{k}x_i n_i; \quad \overline{y}=\frac{1}{n}\sum_{j=1}^{m}y_j n_j; \quad \overline{xy}=\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{m}x_i y_j n_{ij};$$

$$s_x=\sqrt{\frac{1}{n}\sum_{i=1}^{k}x_i^2 n_i-\left(\frac{1}{n}\sum_{i=1}^{k}x_i n_i\right)^2}; \quad s_y=\sqrt{\frac{1}{n}\sum_{j=1}^{m}y_j^2 n_j-\left(\frac{1}{n}\sum_{j=1}^{m}y_j n_j\right)^2};$$

отримують більш зручну для розрахунків формулу:

$$r=\frac{n\sum_{i=1}^{k}\sum_{j=1}^{m}x_i y_j n_{ij}-\left(\sum_{i=1}^{k}x_i n_i\right)\left(\sum_{j=1}^{m}y_j n_j\right)}{\sqrt{n\sum_{i=1}^{k}x_i^2 n_i-\left(\sum_{i=1}^{k}x_i n_i\right)^2}\sqrt{n\sum_{j=1}^{m}y_j^2 n_j-\left(\sum_{j=1}^{m}y_j n_j\right)^2}}. \qquad (9.4)$$

У випадку незгрупованих даних розрахункова формула суттєво спрощується:

$$r=\frac{n\sum_{i=1}^{n}x_i y_i-\left(\sum_{i=1}^{n}x_i\right)\left(\sum_{i=1}^{n}y_i\right)}{\sqrt{n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2}\sqrt{n\sum_{i=1}^{n}y_i^2-\left(\sum_{i=1}^{n}y_i\right)^2}}. \qquad (9.5)$$

Властивості коефіцієнта кореляції Пірсона

1) Коефіцієнт кореляції Пірсона приймає значення на проміжку $[-1;1]$, тобто $-1\le r\le 1$.

2) Якщо $|r|\le 0{,}5$, то зв’язок вважається слабким; якщо $0{,}5<|r|\le 0{,}7$, то зв’язок вважається середнім; $|r|>0{,}7$, то зв’язок вважається сильним.

3) Якщо $r>0$, то зв’язок називається додатнім, тобто зі збільшенням значень *X* значення *Y* також збільшуються. Якщо $r<0$, то зв’язок називається від’ємним, тобто зі збільшенням значень *X* значення *Y* зменшуються.

Зауваження. Слід пам'ятати, що коефіцієнт кореляції Пірсона показує силу лінійного зв'язку. Якщо між X та Y існує сильний нелінійний зв'язок, коефіцієнт кореляції Пірсона може дорівнювати нулю.

Оскільки сила зв'язку між X та Y оцінюється за вибірковими даними, то необхідна перевірка її статистичної значущості, тобто оцінка можливості розповсюдити отримані результати на всю генеральну сукупність.

Перевірка статистичної значущості коефіцієнта кореляції Пірсона здійснюється за допомогою так званої t-статистики, яка розраховується за формулою

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}. \qquad (9.6)$$

Розраховане значення t-статистики порівнюється з критичним значенням $t_{крит}$ – табличне значення розподілу Стьюдента (додаток , таблиця 4), яке також можна знайти за допомогою вбудованої статистичної функції Excel **СТЬЮДРАСПОБР** (α; l), де α – обраний дослідником рівень значущості, l – степені вільності, $l = n-2$.

Якщо розраховане значення t-статистики більше критичного $|t| > t_{крит}$, то коефіцієнт кореляції вважається значимим на обраному рівні α.

Приклад 4. За наявними даними про рівень механізації праці X (%) і продуктивності праці Y (од. продукції/год.) для 14 однотипних підприємств (табл. 9.5) оцінити тісноту зв'язку між X і Y. Визначити можливість розповсюдження результатів розрахунків на всі підприємства такого типу.

Таблиця 9.5

X	32	30	36	40	41	47	56	54	60	55	61	67	69	76
Y	20	24	28	30	31	33	34	37	38	40	41	43	45	48

Розв'язання. Дані таблиці 9.5 є вибіркою значень X і відповідних значень Y. Оскільки кількість даних невелика ($n = 14$), то їх можна не групувати. Для оцінки тісноти зв'язку між X і Y розрахуємо коефіцієнт кореляції Пірсона за формулою (9.5) для незгрупованих даних. Розрахунки для зручності оформимо у вигляді таблиці (табл. 9.6).

Таблиця 9.6

x_i	y_i	x_i^2	y_i^2	$x_i\ y_i$
32	20	1024	400	640
30	24	900	576	720
36	28	1296	784	1008
40	30	1600	900	1200
41	31	1681	961	1271
47	33	2209	1089	1551
56	34	3136	1156	1904
54	37	2916	1369	1998
60	38	3600	1444	2280
55	40	3025	1600	2200
61	41	3721	1681	2501
67	43	4489	1849	2881
69	45	4791	2025	3105
76	48	5779	2304	3848
Суми				
724	492	40134	18138	26907

Отже,

$$r=\frac{n\sum_{i=1}^{n}x_iy_i-\left(\sum_{i=1}^{n}x_i\right)\left(\sum_{j=1}^{n}y_j\right)}{\sqrt{n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2}\sqrt{n\sum_{j=1}^{n}y_j^2-\left(\sum_{j=1}^{n}y_j\right)^2}}=\frac{14\cdot 26907-724\cdot 492}{\sqrt{14\cdot 40134-724^2}\sqrt{14\cdot 18138-492^2}}=$$

$$=\frac{20490}{\sqrt{37700}\sqrt{11868}}\approx 0{,}969.$$

Таке ж значення отримується за допомогою вбудованої функції **КОРРЕЛ**(массив1, массив2) (рис. 9.6, 9.7).

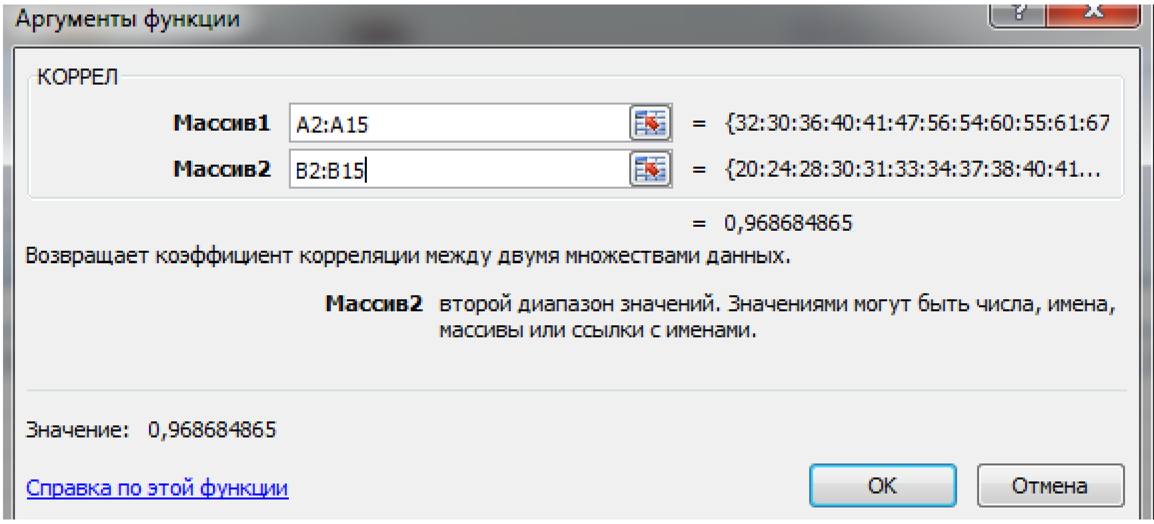
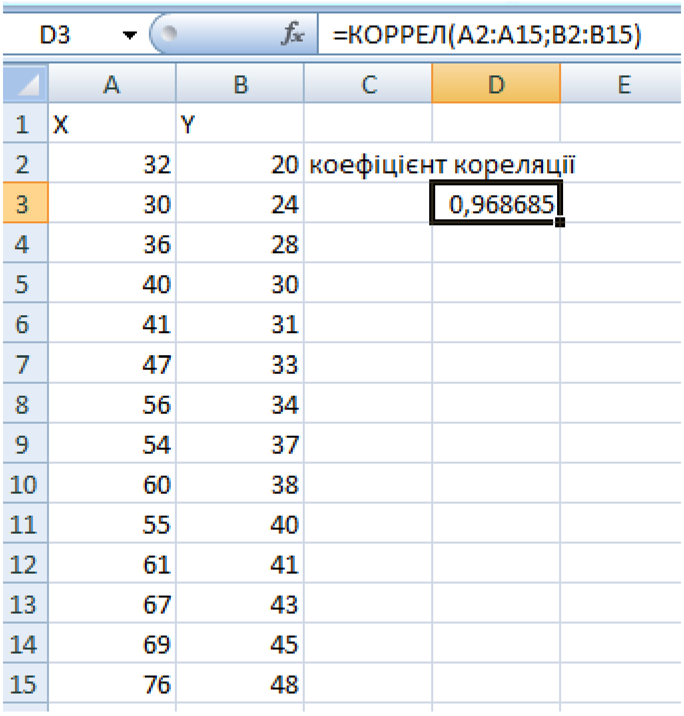


Рис. 9.6. Структура вбудованої функції **КОРРЕЛ.**



	A	B	C	D	E
1	X	Y			
2	32	20	коефіцієнт кореляції		
3	30	24		0,968685	
4	36	28			
5	40	30			
6	41	31			
7	47	33			
8	56	34			
9	54	37			
10	60	38			
11	55	40			
12	61	41			
13	67	43			
14	69	45			
15	76	48			

Рис. 9.7. Обчислення коефіцієнт кореляції.

За значенням коефіцієнта кореляції можна зробити висновок, що між X і Y існує сильний додатній зв’язок.

Перевіримо статистичну значущість знайденого коефіцієнта кореляції Пірсона. Розрахуємо t-статистику за формулою (9.6):

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}969\sqrt{14-2}}{\sqrt{1-0{,}969^2}} \approx 13{,}59.$$

Знайдемо $t_{крит}$, враховуючи, що $l = n - 2 = 14 - 2 = 12$. Оберемо рівень значущості α =0,01. Тоді $t_{крит}$=**СТЬЮДРАСПОБР** (0,01; 12)=3,055.

Оскільки розраховане значення t-статистики більше критичного $13,59 > 3,055$, то коефіцієнт кореляції можна вважати значимим на обраному рівні α =0,01.

Висновок. Між рівнем механізації праці та її продуктивністю на підприємствах, що досліджувалися, існує сильний додатній зв'язок: чим більше рівень механізації праці, тим вище її продуктивність. Висновок дійсний для всіх підприємств такого типу.

Виконання лабораторної роботи

Завдання до теми

Завдання 1.

За даною вибіркою X знайти точкову оцінку математичного сподівання, а також зміщену і незміщену оцінки дисперсії та середньоквадратичного відхилення. Побудувати надійний інтервал із ймовірністю 95%, тобто $\gamma = 0,95$ для вказаного σ і надійний інтервал у випадку невідомої дисперсії для рівня із значущістю 5%. Порівняти побудовані надійні інтервали.

1.1. $X = (-2,-1,-2,0,1,0,2,2,4,1),\ \sigma = 4.$

1.2. $X = (-2,1,2,0,3,4,2,5,0,1),\ \sigma = 1.$

1.3. $X = (2,-1,0,-3,0,2,1,4,0,1),\ \sigma = 1.$

1.4. $X = (4,2,0,3,0,1,2,3,0,1),\ \sigma = 2.$

1.5. $X = (2,-1,2,0,3,0,-2,-2,0,-1),\ \sigma = 1.$

1.6. $X = (2,-1,2,0,3,0,-2,-2,4,-1),\ \sigma = 1.$

1.7. $X = (-2,1,-2,0,-3,2,-2,4,0,1),\ \sigma = 2.$

1.8. $X = (3,1,2,0,3,1,2,3,0,1),\ \sigma = 2.$

1.9. $X = (2,-1,-1,0,0,-2,2,-4,0,1),\ \sigma = 1.$

1.10. $X = (2,1,4,0,4,0,2,4,0,1),\ \sigma = 2.$

1.11. $X = (3,1,3,3,0,2,2,3,0,1),\ \sigma = 2.$

1.12. $X = (4,1,4,0,3,4,-2,4,-2,1),\ \sigma = 2.$

1.13. $X = (-1,1,-1,0,-1,0,2,1,0,1),\ \sigma = 1.$

1.14. $X = (1,1,0,1,0,2,2,1,0,1),\ \sigma = 1.$

1.15. $X = (2,1,2,3,3,3,2,4,4,1),\ \sigma = 2.$

Завдання 2.

Користуючись теоретичними знаннями та засобами програмного забезпечення Excel, за даними варіанту обчислити середні значення, середні квадратичні відхилення та коефіцієнт кореляції.

Варіант 1		
n	X_i	Y_i
1	1,8	36,1
2	2,4	38,3
3	2,5	30,6
4	2,3	32,1
5	2,3	37,6
6	2,5	34,8
7	2,4	34,2
8	2,5	34,2
9	2,1	32,5

Варіант 2		
n	X_i	Y_i
1	2	38,2
2	2,1	36,9
3	2,3	39,7
4	2,6	37,2
5	2,8	31,7
6	2,8	30,1
7	2,2	39,9
8	2,3	38,8
9	2,6	38,4

Варіант 3		
n	X_i	Y_i
1	3	29,4
2	2,6	35,4
3	2,3	39,7
4	2,5	37,1
5	2,2	35,7
6	2,4	40,2
7	2,2	39,4
8	2,6	43,7
9	2,6	38,4

Варіант 4		
n	X_i	Y_i
1	2,8	14
2	2,4	17,1
3	2,3	18,2
4	2,5	17,4
5	2,7	16,1
6	2,4	18,8
7	2,3	32,2
8	1,9	31
9	2,3	32,4

Варіант 5		
n	X_i	Y_i
1	2	35
2	2,3	43,7
3	2,7	31,9
4	2,2	37,3
5	2,4	40,9
6	2,3	38,8
7	2,3	35,7
8	2,6	43,2
9	2,7	30,5

Варіант 6		
n	X_i	Y_i
1	2,4	40,2
2	2,2	39,4
3	2,6	43,7
4	2,6	38,4
5	2,3	38,8
6	2,2	39,9
7	2,8	30,1
8	2,8	31,7
9	2,8	37,2

Варіант 7		
n	X_i	Y_i
1	2,3	32,1
2	1,9	31
3	2,3	32,4
4	2,5	33,2
5	2,6	31,2
6	2	34,8
7	1,9	35,4
8	2,4	33
9	2,2	34,8

Варіант 8		
n	X_i	Y_i
1	2,8	13,8
2	2,7	14,8
3	2,4	16,9
4	2,3	16,8
5	2,5	14,8
6	2,5	17,9
7	2,5	17,6
8	2,4	15,7
9	2,3	15,2

Варіант 9		
n	X_i	Y_i
1	2,7	14,9
2	2,5	16,1
3	2,1	19,7
4	2,8	14
5	2,4	17,1
6	2,3	18,2
7	2,5	17,4
8	2,7	16,1
9	2,4	18

Варіант 10		
n	X_i	Y_i
1	30,2	5000
2	32	5200
3	32	5350
4	37	5880
5	30	5430
6	30	5430
7	30	5350
8	29	5740
9	33	5570

Варіант 11		
n	X_i	Y_i
1	29	5350
2	33	2740
3	31	5570
4	30	5530
5	34	6020
6	38	7010
7	31	6420
8	39	7150
9	39,5	7190

Варіант 12		
n	X_i	Y_i
1	5,1	15,7
2	7,2	17
3	2	5,2
4	5,1	17,5
5	10	22,1
6	12	25,8
7	9,4	23
8	6,9	17
9	3,4	9,1

Варіант 13		
n	X_i	Y_i
1	5,4	3,7
2	7,6	8
3	2,3	3,2
4	5,9	2,5
5	11	4,9
6	12,6	4,8
7	10,4	5,1
8	4,9	2,2
9	2,4	1,1

Варіант 14		
n	X_i	Y_i
1	1,4	12,7
2	2,6	18
3	2,3	16,2
4	5,1	25,5
5	6	24,1
6	5,6	24,8
7	10,4	55
8	4,9	33
9	2,4	18,1

Варіант 15		
n	X_i	Y_i
1	0,4	13,7
2	0,6	18
3	0,3	6,2
4	0,9	15,5
5	1	24,1
6	1,6	24,8
7	1,4	25
8	0,6	13
9	0,3	8,1

Завдання 3.

За експериментальними даними вимірювань, де ΔE – колірні відмінності виміряного кольору від еталонного значення, D – оптична густина фарби, 80, 40 – розтискування растрової точки (80+k, вказано k) при стабільній подачі фарби та зволожувального розчину та при коригуванні подачі фарби автоматизованою системою контролю параметрів відбитка визначити силу зв'язку між обраними факторами та порівняти їх.

Дані вимірювань при стабільній подачі фарби та зволожувального розчину

Відбиток	3 ділянка				4 ділянка				5 ділянка				6 ділянка			
	ΔE	D	80	40	ΔE	D	80	40	ΔE	D	80	40	ΔE	D	80	40
5	16,24	0,89	4	5	16,99	0,88	5	5	14,15	0,96	4	6	10,83	1,06	5	6
5	11,75	0,96	4	5	13,38	0,95	4	5	12,51	0,99	4	5	9,78	1,08	5	6
10	16,11	0,89	5	7	16,86	0,87	5	6	14,45	0,93	5	6	11,38	1,03	5	7
10	12,84	0,96	4	6	13,84	0,92	4	5	12,57	0,96	4	6	9,84	1,06	5	7
15	16,31	0,87	5	7	17,10	0,86	5	7	14,09	0,94	5	7	11,26	1,03	6	7
15	12,73	0,95	5	6	14,17	0,91	5	6	13,34	0,96	4	6	10,22	1,05	5	7
20	15,37	0,89	4	6	15,99	0,88	5	6	14,11	0,93	5	6	11,20	1,02	5	6
20	12,76	0,98	5	5	13,64	0,93	4	6	12,79	0,99	4	6	10,43	1,06	5	6
25	12,96	0,95	4	7	12,90	0,96	5	6	11,37	1,01	5	7	9,06	1,09	5	7
25	10,75	1,02	5	7	12,09	0,99	5	7	11,15	1,02	5	7	8,90	1,09	6	8
30	13,93	0,93	5	7	14,47	0,92	5	6	12,88	0,97	5	7	9,99	1,06	5	7
30	11,53	0,99	5	6	12,96	0,96	4	6	11,97	1,01	4	6	9,63	1,08	5	7
35	14,46	0,91	5	7	15,79	0,89	5	6	13,64	0,94	5	7	10,36	1,05	5	7
35	12,05	0,97	5	6	14,02	0,93	5	6	12,96	0,96	4	6	10,13	1,06	5	7
40	14,59	0,91	5	7	16,20	0,86	5	6	13,90	0,93	5	7	11,15	1,02	5	7
40	12,10	0,97	5	7	13,40	0,94	4	6	12,96	0,96	4	6	10,49	1,04	5	7

Відбиток	7 ділянка				8 ділянка				9 ділянка				10 ділянка			
	ΔE	D	80	40	ΔE	D	80	40	ΔE	D	80	40	ΔE	D	80	40
5	8,70	1,13	5	7	5,68	1,26	5	7	3,28	1,36	6	8	2,34	1,44	6	9
5	6,76	1,18	5	6	4,37	1,29	6	7	2,93	1,40	6	8	2,97	1,53	7	10
10	8,62	1,12	5	7	5,84	1,22	5	8	3,32	1,34	6	8	1,59	1,45	7	10
10	7,10	1,16	5	7	4,79	1,28	6	8	2,79	1,37	7	9	2,62	1,52	8	10
15	8,46	1,11	6	9	5,69	1,23	6	9	3,01	1,34	7	9	1,73	1,46	7	10
15	7,60	1,17	5	8	5,00	1,26	6	9	3,14	1,40	7	9	3,14	1,54	8	11
20	8,43	1,12	6	7	5,43	1,23	5	8	3,03	1,35	6	8	1,85	1,43	7	10
20	7,48	1,16	5	7	4,61	1,28	6	8	2,80	1,40	6	9	2,94	1,54	8	11
25	6,97	1,16	6	9	4,92	1,26	6	9	2,87	1,37	6	10	2,34	1,46	8	11
25	6,32	1,19	6	9	4,73	1,29	6	9	3,06	1,41	8	11	3,18	1,55	9	13
30	7,81	1,14	5	7	4,97	1,26	6	8	2,99	1,37	6	9	2,48	1,48	7	10
30	7,35	1,16	5	8	4,76	1,31	6	9	2,77	1,43	7	11	2,66	1,52	9	12

35	8,03	1,14	5	7	5,42	1,25	6	9	3,20	1,38	6	10	2,78	1,49	7	11
35	7,66	1,17	5	7	4,65	1,29	6	9	3,07	1,42	7	9	3,51	1,59	9	12
40	8,71	1,11	6	7	5,74	1,23	6	9	3,37	1,36	6	9	2,75	1,48	7	10
40	7,58	1,15	5	7	5,02	1,27	6	8	2,97	1,43	7	10	3,31	1,58	8	11

Відбиток	11 ділянка				12 ділянка				13 ділянка				14 ділянка			
	ΔE	D	80	40	ΔE	D	80	40	ΔE	D	80	40	ΔE	D	80	40
5	2,68	1,52	7	10	3,31	1,58	8	10	3,64	1,55	9	10	3,54	1,57	8	10
5	4,03	1,65	9	11	4,76	1,69	9	11	4,03	1,69	9	12	4,56	1,68	10	12
10	2,11	1,52	7	10	3,19	1,53	9	11	3,24	1,55	9	11	2,94	1,55	9	11
10	3,66	1,64	9	12	4,58	1,70	10	12	4,58	1,65	9	12	4,48	1,66	10	12
15	2,38	1,53	8	11	3,18	1,56	9	12	2,97	1,55	9	12	2,41	1,52	9	11
15	3,88	1,65	9	12	4,88	1,68	10	13	4,67	1,64	10	13	4,52	1,64	9	12
20	2,47	1,53	7	11	3,41	1,58	9	11	3,19	1,55	9	11	2,59	1,55	9	11
20	4,17	1,66	9	12	4,88	1,73	10	12	4,97	1,70	9	13	4,42	1,67	10	14
25	3,02	1,54	8	11	3,54	1,57	9	12	3,28	1,56	9	12	2,55	1,50	9	12
25	4,29	1,65	10	13	4,80	1,68	10	14	4,35	1,62	10	14	3,87	1,61	10	12
30	3,34	1,56	8	11	4,04	1,60	9	11	3,66	1,57	9	11	2,99	1,54	8	11
30	4,40	1,68	9	13	5,08	1,69	10	13	4,33	1,63	10	12	3,93	1,65	10	13
35	3,59	1,55	8	12	4,18	1,60	9	12	3,93	1,57	9	11	3,06	1,53	8	10
35	4,65	1,67	10	14	5,66	1,74	11	15	4,76	1,68	9	13	4,35	1,66	10	12
40	3,48	1,53	8	13	3,64	1,58	10	12	3,89	1,56	9	12	3,13	1,53	8	12
40	4,76	1,67	10	14	5,39	1,72	10	14	4,90	1,66	10	13	4,16	1,65	10	12

Відбиток	15 ділянка				16 ділянка				17 ділянка			
	ΔE	D	80	40	ΔE	D	80	40	ΔE	D	80	40
5	3,46	1,55	8	9	3,18	1,51	8	9	3,12	1,48	7	9
5	4,40	1,68	9	10	3,59	1,65	8	10	3,10	1,59	8	9
10	2,60	1,52	8	10	2,13	1,48	8	9	2,33	1,44	7	9
10	3,57	1,62	10	11	3,15	1,57	8	10	2,60	1,52	8	9
15	1,87	1,49	7	11	1,84	1,46	7	10	1,92	1,41	7	10
15	3,53	1,62	9	12	2,93	1,56	8	9	2,57	1,51	8	9
20	2,04	1,49	7	10	1,75	1,44	7	9	2,09	1,39	7	8
20	3,54	1,59	9	12	2,63	1,55	8	11	2,28	1,51	8	9
25	2,11	1,45	7	11	2,47	1,41	7	9	3,37	1,33	7	10
25	3,15	1,55	10	12	2,31	1,50	8	10	2,78	1,42	7	9
30	2,25	1,48	7	9	2,67	1,38	7	9	3,62	1,33	6	9
30	3,03	1,57	9	11	2,06	1,51	7	10	2,76	1,41	7	9
35	2,43	1,47	6	10	2,92	1,37	7	8	4,21	1,30	7	8
35	3,21	1,56	8	11	2,32	1,47	7	10	3,45	1,40	7	9
40	2,55	1,44	7	10	2,97	1,36	7	9	4,31	1,29	6	10
40	3,00	1,54	8	11	2,23	1,44	7	9	3,40	1,36	7	9

Дані вимірювань при коригуванні подачі фарби
автоматизованою системою контролю параметрів відбитка

Відбиток	3 ділянка				4 ділянка				5 ділянка				6 ділянка			
	ΔE	D	80	40	ΔE	D	80	40	ΔE	D	80	40	ΔE	D	80	40
1	11,97	1,02	6	9	13,19	0,98	7	9	11,67	1,03	7	9	9,36	1,11	7	9
1	9,90	1,08	6	8	11,35	1,04	6	8	11,10	1,07	6	8	8,70	1,14	6	9
101	11,25	1,05	6	10	12,45	1,01	6	9	10,96	1,07	7	10	8,58	1,15	7	10
101	9,45	1,10	6	8	11,42	1,06	6	9	10,74	1,09	6	9	8,50	1,16	7	10
201	11,96	1,03	6	10	12,37	1,02	6	9	10,97	1,06	7	10	8,69	1,14	7	10
201	9,54	1,09	6	8	11,04	1,06	6	9	10,39	1,10	6	9	8,09	1,16	7	10
301	9,23	1,11	7	10	9,96	1,10	6	9	8,20	1,15	7	10	6,42	1,23	7	10
301	7,38	1,18	7	9	8,25	1,15	6	9	7,82	1,20	6	11	5,54	1,25	8	10
401	8,34	1,15	7	11	7,77	1,16	8	13	6,81	1,21	7	11	4,98	1,28	8	13
401	5,41	1,25	8	10	6,76	1,23	7	10	5,70	1,26	7	10	4,23	1,33	8	12

Відбиток	7 ділянка				8 ділянка				9 ділянка				10 ділянка			
	ΔE	D	80	40	ΔE	D	80	40	ΔE	D	80	40	ΔE	D	80	40
1	7,08	1,19	7	10	4,60	1,29	7	10	1,83	1,41	8	12	1,02	1,53	9	15
1	6,43	1,22	7	10	3,80	1,33	7	11	0,76	1,49	9	13	1,78	1,61	10	14
101	6,35	1,23	7	12	3,77	1,33	8	12	1,74	1,43	8	14	1,38	1,56	10	16
101	6,10	1,22	7	11	3,23	1,34	8	12	0,65	1,46	9	17	1,63	1,63	11	16
201	6,70	1,21	7	11	3,98	1,33	8	12	1,60	1,43	9	14	1,31	1,55	11	15
201	5,72	1,27	7	11	3,35	1,36	8	13	0,67	1,51	10	13	1,98	1,64	11	16
301	4,82	1,29	7	11	3,36	1,35	8	14	1,32	1,45	8	14	1,04	1,51	10	16
301	4,33	1,33	8	11	2,64	1,41	8	12	0,59	1,52	9	13	1,82	1,63	11	15
401	4,03	1,31	8	12	2,49	1,38	9	14	1,34	1,44	9	14	1,12	1,51	11	17
401	3,36	1,37	8	11	1,99	1,42	8	13	0,51	1,52	10	14	1,52	1,59	11	16

Відбиток	11 ділянка				12 ділянка				13 ділянка				14 ділянка			
	ΔE	D	80	40	ΔE	D	80	40	ΔE	D	80	40	ΔE	D	80	40
1	3,01	1,64	10	15	3,44	1,61	12	16	2,34	1,57	11	15	0,99	1,53	10	14
1	3,57	1,69	12	15	4,44	1,75	12	16	3,54	1,67	11	15	2,02	1,62	11	14
101	3,42	1,63	11	18	4,01	1,68	12	18	2,27	1,59	12	17	0,62	1,50	11	14
101	4,04	1,76	13	17	4,33	1,75	13	18	2,63	1,66	11	16	1,00	1,58	11	15
201	3,06	1,64	11	18	4,27	1,70	12	17	2,48	1,60	11	16	0,67	1,53	10	15
201	4,40	1,77	13	17	4,79	1,81	13	17	3,07	1,69	11	16	1,17	1,60	11	15
301	2,28	1,60	10	17	2,23	1,59	11	16	0,95	1,54	11	17	1,50	1,46	9	13
301	3,24	1,70	12	17	3,26	1,72	13	17	1,81	1,63	10	16	0,52	1,55	10	14
401	2,05	1,58	11	19	1,91	1,57	11	16	0,58	1,49	10	17	1,75	1,43	10	15
401	2,32	1,67	11	17	2,58	1,65	11	16	1,17	1,56	10	15	0,97	1,49	9	15

Відбиток	15 ділянка				16 ділянка				17 ділянка			
	ΔE	D	80	40	ΔE	D	80	40	ΔE	D	80	40
1	1,46	1,42	9	12	3,65	1,33	8	12	4,96	1,28	7	11
1	0,54	1,50	9	13	1,81	1,42	9	12	3,58	1,34	7	11
101	2,74	1,38	8	13	5,88	1,24	8	12	8,11	1,17	7	11
101	2,16	1,42	9	13	4,90	1,33	8	12	6,86	1,21	7	10
201	2,78	1,38	8	12	5,81	1,26	7	11	7,98	1,17	7	11
201	2,21	1,43	9	13	4,57	1,32	8	11	6,75	1,23	8	11
301	4,05	1,31	8	13	6,13	1,24	7	11	7,86	1,17	7	11
301	2,73	1,39	9	12	5,39	1,31	8	11	6,88	1,22	7	11
401	4,19	1,32	8	13	5,54	1,26	8	12	6,27	1,22	8	12
401	2,62	1,40	9	13	4,29	1,32	8	11	5,25	1,26	8	11

Номер варіанта відповідає номеру ділянки, варіанти 1 та 2 є, відповідно, варіантами 16 та 17.

Завдання 4. Дати відповіді на подані нижче теоретичні запитання, одне із них оформити письмово.

Теоретичні запитання до теми

1. Дати пояснення інтервальної оцінки парметрів розподілу.
2. Записати покрокове знаходження надійного інтервалу математичного сподівання у випадку відомої дисперсії.
3. Описати знаходження надійного інтервалу математичного сподівання у випадку невідомої дисперсії.
4. Пояснити характеристики вбудованої статистичної функції Excel **ДОВЕРИТ** та функції **СТЬДРАСПОБР,** їх практичне використання.

Додаткові завдання

1. При формуванні портфелю поставок для фірми було обрано 100 поставщиків, які працювали із фірмою у минулому році. Знайти надійний інтервал для долі поставщиків, що несвоєчасно здійснюють поставки на рівні надійності 0,999, якщо у вибірці таких 45.

Таблиця 9.7

11,240	18,545	17,750	22,560	18,355	20,424	20,650	10,780	15,590
13,720	28,505	23,170	20,360	22,450	21,590	14,565	24,295	25,140
27,655	17,786	27,045	28,650	18,670	31,445	18,540	15,598	19,720
15,230	21,240	19,535	12,934	18,195	19,074	17,037	19,610	20,970
22,075	15,090	20,754	10,195	13,580	21,490	13,987	22,645	21,218

2. Відомі дані про розмір вкладів в банку (табл. 9.8). Знайти з надійністю 0,96 надійний інтервал для середнього розміру вкладу.

Таблиця 9.8

Розмір вкладу (тис. грн.)	10-30	31-50	51-70	71-90	91-110	111-130
Кількість вкладчиків	1	3	10	30	60	7

Оформлення звіту та порядок захисту

Лабораторна робота виконується на аркушах А4, в ній стисло відображаються формули теоретичної частини, хід роботи та отримані результати. При захисті студент повинен розуміти зміст роботи, порівняти отримані результати проведених обчислень, а також знати відповіді на теоретичні запитання. Також студент здає електронний варіант проведених обчислень практичних задач. Обов’язковими є проведення обчислень коефіцієнта кореляції із записом, як в таблиці 9.6.

Лабораторна робота № 10

Кореляційний зв'язок між досліджуваними величинами. Побудова ліній регресії

Мета роботи − навчитися обчислювати і визначати статистичну значущість кореляції, обчислювати та аналізувати множинні та частинні коефіцієнти кореляції, використовуючи теоретичні знання та можливості табличного процесора Microsoft Excel; будувати лінії регресії та аналізувати їх взаємне розміщення.

Теоретичні відомості

10.1. Множинний та частинний коефіцієнти кореляції

У випадку, коли досліджуваний об’єкт або явище характеризується більш ніж двома ознаками X_1, X_2, … , X_k, необхідно вивчати множинні залежності. Для оцінки сили зв’язку між певною ознакою X_i та усіма іншими ознаками слугує ***множинний коефіцієнт кореляції***, який позначається R_i.

Для розрахунку ***множинного коефіцієнта кореляції*** необхідно:

1) Побудувати матрицю парних коефіцієнтів кореляції $r_{ij},\ i=\overline{1,k}$ між ознаками X_i та X_j:

$$A=\begin{pmatrix} 1 & r_{12} & \dots & r_{1k} \\ r_{21} & 1 & \dots & r_{2k} \\ \dots & \dots & \dots & \dots \\ r_{k1} & r_{k2} & \dots & r_{kk} \end{pmatrix}. \tag{10.1}$$

2) Знайти визначник $|A|$ матриці A та алгебраїчне доповнення A_{ii} елемента r_{ii} цієї матриці.

3) Розрахувати множинний коефіцієнт кореляції за формулою:

$$R_i=\sqrt{1-\frac{|A|}{A_{ii}}}\ . \tag{10.2}$$

Перевірка статистичної значущості множинного коефіцієнта кореляції здійснюється за допомогою t-статистики, яка розраховується за формулою:

$$t = \frac{R^2(n-k)}{(1-R^2)(k-1)}, \tag{10.3}$$

де n – кількість взаємопов'язаних значень ознак $X_i,\, i=\overline{1,k}$.

Розраховане значення t-статистики порівнюється з критичним значенням $F_{крит}$. $F_{крит}$ – табличне значення розподілу Фішера (додаток, таблиця 5), яке також можна знайти за допомогою вбудованої статистичної функції Excel **FРАСПОБР** (α; l_1; l_2), де α – обраний дослідником рівень значущості, l_1; l_2 – степені вільності, $l_1 = k-1$; $l_2 = n-k$.

Якщо розраховане значення t-статистики більше критичного $|t| > F_{крит}$, то множинний коефіцієнт кореляції вважається значимим на обраному рівні значущості α.

У випадку, коли необхідно дослідити кореляційний зв'язок між ознаками X_i та X_j, $i=\overline{1,k}$, $j=\overline{1,k}$, із множини ознак $X_1, X_2, ..., X_k$ досліджуваного об'єкту або явища, вільний від впливу всіх інших ознак, розраховується *частинний коефіцієнт* кореляції, який позначається R_{ij}.

Для розрахунку ***частинного коефіцієнта кореляції*** необхідно:

1) Побудувати матрицю парних коефіцієнтів кореляції A.

2) Знайти алгебраїчні доповнення A_{ii}, A_{jj}, A_{ij} елементів r_{ii}, r_{jj}, r_{ij} відповідно.

3) Розрахувати частинний коефіцієнт кореляції за формулою:

$$R_{ij} = \frac{-A_{ij}}{\sqrt{A_{ii}A_{jj}}}. \tag{10.4}$$

Перевірка статистичної значущості частинного коефіцієнта кореляції здійснюється за допомогою t-статистики, яка розраховується за формулою:

$$t = \frac{R_{ij}\sqrt{n-k+2}}{\sqrt{1-R_{ij}^{\ 2}}}, \tag{10.5}$$

де n – кількість взаємопов'язаних значень ознак X_i, $i=\overline{1,k}$.

Розраховане значення t-статистики порівнюється з критичним значенням $t_{крит}$. $t_{крит}$ – табличне значення розподілу Стьюдента, яке також можна знайти за допомогою вбудованої статистичної функції Excel **СТЬЮДРАСПОБР** (α; l), де α– обраний дослідником рівень значущості, l – степені вільності, $l = n - k + 2$.

Якщо розраховане значення t-статистики більше критичного $|t| > t_{крит}$, то частинний коефіцієнт кореляції вважається значимим на обраному рівні значущості α.

Зауваження. 1. Вважається, що для коректного використання множинного і частинного коефіцієнтів кореляції необхідно, щоб вибіркові дані мали сумісний нормальний розподіл, однак перевірка цієї умови на практиці зазвичай не виконується, оскільки пов'язана зі значними труднощами у розрахунках.

2. Замість парного коефіцієнта кореляції Пірсона можна використовувати також парний коефіцієнт кореляції Спірмена.

3. Кореляційна матриця завжди симетрична відносно головної діагоналі, оскільки $r_{ij} = r_{ji}$, $i=\overline{1,k}$, $j=\overline{1,k}$. Елементи головної діагоналі завжди дорівнюють 1, оскільки вони є коефіцієнтами кореляції X_i та X_i.

Приклад 1. Для вивчення залежності урожайності зернових культур *Z* (ц/га) від якості пашні *X* (бали) і кількості внесеного добрива *Y* (кг/га) було проведено дослідження 6 фермерських хазяйств, результати якого надано у таблиці 10.1. Визначити силу зв'язку між *Z* та *X* та *Y*, використовуючи множинний коефіцієнт кореляції. Порівняти силу зв'язку між *Z* та *X* та між *Z* та *Y* за частинними коефіцієнтами кореляції.

Таблиця 10.1

X	26	35	36	40	41	45
Y	2,1	2,3	2,4	2,6	2,9	3
Z	18	21	22,1	25,3	28	28,5

Розв'язання. За умов задачі необхідно для об'єкту, що характеризується трьома ознаками X, Y та Z ($k=3$), розрахувати множинний коефіцієнт кореляції R_Z і частинні коефіцієнти кореляції R_{XZ} та R_{YZ} на основі 6 взаємопов'язаних трійок вибіркових даних $(x_i, y_i, z_i), i=\overline{1,n}, n=6$.

Побудуємо матрицю парних коефіцієнтів кореляції, які обчислимо за формулою

$$r=\frac{n\sum_{i=1}^{n}x_iy_i-\left(\sum_{i=1}^{n}x_i\right)\left(\sum_{i=1}^{n}y_i\right)}{\sqrt{n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2}\sqrt{n\sum_{i=1}^{n}y_i^2-\left(\sum_{i=1}^{n}y_i\right)^2}}$$

(випадок незгрупованих даних).

Розрахунки для зручності оформимо у вигляді таблиці (табл. 10.2).

Таблиця 10.2

Розрахункова таблиця							Суми
x_i	26	35	36	40	41	45	223
y_i	2,1	2,3	2,4	2,6	2,9	3	15,3
z_i	18	21	22,1	25,3	28	28,5	142,9
x_i^2	676	1225	1296	1600	1681	2025	8503
y_i^2	4,41	5,29	5,76	6,76	8,41	9	39,63
z_i^2	324	441	488,41	640,09	784	812,25	3489,75
$x_i\,y_i$	54,6	80,5	86,4	104	118,9	135	579,4
$x_i\,z_i$	468	735	795,6	1012	1148	1282,5	5441,1
$y_i\,z_i$	37,8	48,3	53,04	65,78	81,2	85,5	371

Отже, за формулою маємо:

$$r_{XY}=r_{YX}=\frac{n\sum_{i=1}^{n}x_iy_i-\left(\sum_{i=1}^{n}x_i\right)\left(\sum_{i=1}^{n}y_i\right)}{\sqrt{n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2}\sqrt{n\sum_{i=1}^{n}y_i^2-\left(\sum_{i=1}^{n}y_i\right)^2}}=\frac{6\cdot 579,4-223\cdot 15,3}{\sqrt{6\cdot 8503-223^2}\sqrt{6\cdot 39,63-15,3^2}}\approx$$

$$\approx 0,935;$$

$$r_{XZ}=r_{ZX}=\frac{n\sum_{i=1}^{n}x_iz_i-\left(\sum_{i=1}^{n}x_i\right)\left(\sum_{i=1}^{n}z_i\right)}{\sqrt{n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2}\sqrt{n\sum_{i=1}^{n}z_i^2-\left(\sum_{i=1}^{n}z_i\right)^2}}=\frac{6\cdot 5441{,}1-223\cdot 142{,}9}{\sqrt{6\cdot 8503-223^2}\sqrt{6\cdot 3489{,}75-142{,}9^2}}\approx$$

$$\approx 0{,}954;$$

$$r_{YZ}=r_{ZY}=\frac{n\sum_{i=1}^{n}y_iz_i-\left(\sum_{i=1}^{n}y_i\right)\left(\sum_{i=1}^{n}z_i\right)}{\sqrt{n\sum_{i=1}^{n}y_i^2-\left(\sum_{i=1}^{n}y_i\right)^2}\sqrt{n\sum_{i=1}^{n}z_i^2-\left(\sum_{i=1}^{n}z_i\right)^2}}=\frac{6\cdot 371{,}62-15{,}3\cdot 142{,}9}{\sqrt{6\cdot 39{,}63-15{,}3^2}\sqrt{6\cdot 3489{,}75-142{,}9^2}}\approx$$

$$\approx 0{,}991;$$

Таким чином, кореляційна матриця має вигляд:

$$A=\begin{pmatrix}1 & 0{,}935 & 0{,}954\\ 0{,}935 & 1 & 0{,}991\\ 0{,}954 & 0{,}991 & 1\end{pmatrix}.$$

Знайдемо визначник $|A|$ матриці A та алгебраїчне доповнення $A_{ZZ}=A_{33}$:

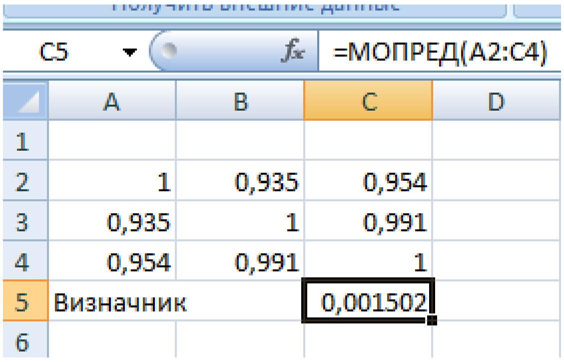
$$|A|=\begin{vmatrix}1 & 0{,}935 & 0{,}954\\ 0{,}935 & 1 & 0{,}991\\ 0{,}954 & 0{,}991 & 1\end{vmatrix}=1+2\cdot 0{,}935\cdot 0{,}991\cdot 0{,}954-0{,}954^2-0{,}991^2-$$

$$-0{,}935^2\approx 0{,}0015;$$

$$A_{ZZ}=A_{33}=(-1)^{3+3}\begin{vmatrix}1 & 0{,}935\\ 0{,}935 & 1\end{vmatrix}=1-0{,}935^2\approx 0{,}1258;$$

тоді $R_Z=R_3=\sqrt{1-\frac{|A|}{A_{33}}}=\sqrt{1-\frac{0{,}0015}{0{,}1258}}\approx 0{,}994$. Значення множинного коефіцієнта кореляції R_z показує, що величина Z сильно пов’язана з X та Y.

Зауваження. Можна значно спростити розрахунки, використовуючи вбудовану математичну функцію **МОПРЕД**, яка дозволяє знайти визначник заданої матриці (рис. 10.1), а також задавати формули для обчислення множинного коефіцієнта кореляції.



C5 f_x =МОПРЕД(A2:C4)

	A	B	C	D
1				
2	1	0,935	0,954	
3	0,935	1	0,991	
4	0,954	0,991	1	
5	Визначник		0,001502	
6				

Рис. 10.1. Обчислення визначника матриці.

Перевіримо статистичну значущість множинного коефіцієнта кореляції R_Z. Знайдемо t-статистику за формулою (10.3):

$$t=\frac{R^2(n-k)}{(1-R^2)(k-1)}=\frac{0,994^2(6-3)}{(1-0,994^2)(3-1)}\approx 124,09.$$

Знайдемо $F_{крит}$, враховуючи, що $l_1=k-1=3-1=2$, $l_2=n-k=6-3=3$. Оберемо рівень значущості α=0,01. Тоді $F_{крит}$=**FРАСПОБР** (0,01; 2; 3)=30,82. Оскільки t> $F_{крит}$, то множинний коефіцієнт кореляції R_Z є статистично значимим на рівні значущості α=0,01.

Для обчислення частинних коефіцієнтів кореляції $R_{XZ}=R_{13}$ та $R_{YZ}=R_{23}$ знайдемо алгебраїчні доповнення:

$$A_{13}=(-1)^{1+3}\begin{vmatrix}0,935 & 1\\ 0,954 & 0,991\end{vmatrix}=0,935\cdot 0,991-0,954\approx -0,027;$$

$$A_{23}=(-1)^{2+3}\begin{vmatrix}1 & 0,935\\ 0,954 & 0,991\end{vmatrix}=(-1)(0,991-0,935\cdot 0,954)\approx -0,099;$$

$$A_{11}=(-1)^{1+1}\begin{vmatrix}1 & 0,991\\ 0,991 & 1\end{vmatrix}=(1-0,991^2)\approx 0,018;$$

$$A_{22}=(-1)^{2+2}\begin{vmatrix}1 & 0,954\\ 0,954 & 1\end{vmatrix}=(1-0,954^2)\approx 0,09.$$

Тоді за формулою (10.4) маємо:

$$R_{13}=\frac{-A_{13}}{\sqrt{A_{11}A_{33}}}=\frac{-(-0{,}027)}{\sqrt{0{,}018\cdot 0{,}126}}\approx 0{,}577; \qquad R_{23}=\frac{-A_{23}}{\sqrt{A_{22}A_{33}}}=\frac{-(-0{,}099)}{\sqrt{0{,}09\cdot 0{,}126}}\approx 0{,}929.$$

Значення частинних коефіцієнтів кореляції показують, що величина *Z* пов'язана з величиною *Y* сильніше, ніж з величиною *X*.

Перевіримо статистичну значущість частинного коефіцієнта кореляції R_{13}. Знайдемо t-статистику за формулою (10.5):

$$t=\frac{R_{ij}\sqrt{n-k+2}}{\sqrt{1-R_{ij}^{\,2}}}=\frac{0{,}577\sqrt{6-3+2}}{\sqrt{1-0{,}577^2}}\approx 1{,}581.$$

Знайдемо критичне значення $t_{крит}$, враховуючи, що $l=n-k+2=5$. Оберемо рівень значущості α =0,01.

Тоді $t_{крит}$=**СТЬЮДРАСПОБР**(0,01;5)=4,032 (рис. 10.2). Оскільки розраховане значення t-статистики менше критичного $|t|<t_{крит}$, то частинний коефіцієнт кореляції R_{13} не є значимим на рівні значущості α =0,01.

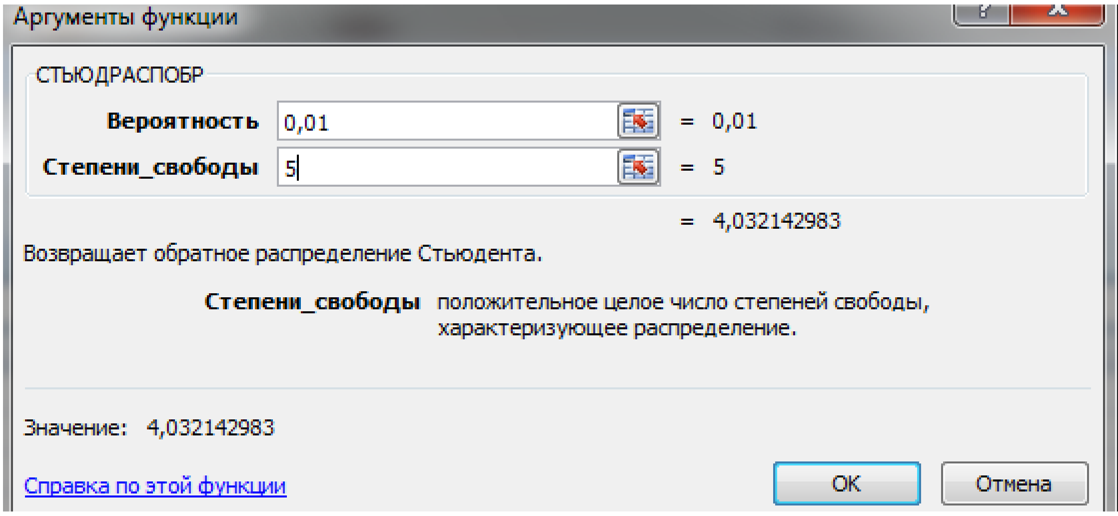


Рис. 10.2. Структура вбудованої функції Excel **СТЬЮДРАСПОБР.**

Якщо $|t|>t_{кр}$, то вважають коефіцієнт кореляції значущим, якщо ж $|t|<t_{кр}$ – то незначущим.

Відповідно до рівня значущості розглядається часткова класифікація кореляційних зв'язків:

1) висока значима кореляція – для *r*, який відповідає рівню статистичної значущості $\alpha \le$0,01;

2) значима кореляція – для r, який відповідає рівню статистичної значущості $\alpha \le 0{,}05$;

3) тенденція достовірного зв'язку – для r, який відповідає рівню статистичної значущості $\alpha \le 0{,}1$;

4) незначима кореляція − при r, який не досягає рівня статистичної значущості.

Перевіримо статистичну значущість частинного коефіцієнта кореляції R_{23}. Знайдемо t-статистику:

$$t = \frac{R_{ij}\sqrt{n-k+2}}{\sqrt{1-R_{ij}^{\;2}}} = \frac{0,929\sqrt{6-3+2}}{\sqrt{1-0,929^2}} \approx 5,614\,.$$

Оскільки розраховане значення t-статистики більше критичного $|t| > t_{крит}$, то частинний коефіцієнт кореляції R_{23} є значимим на рівні значущості α =0,01.

Висновок. Урожайність зернових культур сильно пов’язана з якістю пашні і кількістю внесеного добриву. При цьому урожайність значно сильніше залежіть від кількості добрива, ніж від якості пашні. Сила зв’язку між урожайністю та якістю пашні середня та не є статистично значимою.

Зауваження. Вбудовані сервісні функції Microsoft Excel дозволяють розраховувати коефіцієнти кореляції Пірсона.

Приклад і результати розрахунків парних коефіцієнтів кореляції надано на рис. 10.3.

Клітини матриці, що розташовані вище головної діагоналі звичайно надаються незаповненими, оскільки матриця симетрична відносно головної діагоналі.

	A	B	C	D	E	F	G	H	I
1									
2		Кореляційний аналіз							
3	Вхідні дані								
4		Значення					Стовп. 1	Стовп. 2	Стовп. 3
5	1	1	328	0,054		Рядок 1	1		
6	2	2	329	0,101		Рядок 2	0,8914	1	
7	3	3	329	0,099		Рядок 3	0,563423	0,692214	1
8	4	4	345	0,019					
9	5	5	352	0,065					
10	6	6	370	0,053					
11	7	7	377	0,178					
12	8	8	385	0,174					
13	9	9	396	0,289					
14	10	10	399	0,195					
15	11	11	390	0,102					
16	12	12	373	0,138					
17									

Рис. 10.3. Результати розрахунку коефіцієнтів кореляції.

10.2. Побудова прямих регресії

Нехай X і Y – випадкові величини, зв'язок між якими треба вивчити. В результаті n випробувань отримали n точок $(x_1, y_1), (x_2, y_2), \dots,$ (x_n, y_n). Вибіркові середні та дисперсії позначимо $\bar{x}, \bar{y}, \bar{s}_x^{\,2}, \bar{s}_y^{\,2}$. За оцінку коефіцієнта кореляції $\rho = \rho(X, Y)$ можна взяти

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\bar{s}_x \bar{s}_y} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{n\bar{s}_x \bar{s}_y} \tag{10.6}$$

за невеликої кількості випробувань.

Аналогічно визначаються рівняння вибіркових прямих регресії: прямої регресії Y на X

$$y - \bar{y} = r\frac{\bar{s}_y}{\bar{s}_x}(x - \bar{x}) \tag{10.7}$$

і прямої регресії X на Y

$$x - \bar{x} = r\frac{\bar{s}_x}{\bar{s}_y}(y - \bar{y})\,. \tag{10.8}$$

Прямі регресії доцільно шукати у тому випадку, коли точки (x_i, y_i) $(i = 1, 2, \ldots, n)$ групуються навколо деякої прямої.

Якщо $y = \alpha_1 x + \alpha_0$ – рівняння вибіркової прямої регресії *Y* на *X*, то згідно з означенням повинно бути

$$\Phi(\alpha_1, \alpha_0) = \sum_{i=1}^{n} \left(y_i - \alpha_1 x_i - \alpha_0\right)^2 \to \min .$$

Диференціюючи за α_1 і α_0 функцію $\Phi(\alpha_1, \alpha_0)$ отримуємо систему рівнянь для знаходження α_1 і α_0, з якої

$$\begin{cases} \alpha_0 = \bar{y} - \alpha_1 \bar{x}, \\ \alpha_1 = \dfrac{\sum\limits_{i=1}^{n} x_i y_i - n\overline{x}\overline{y}}{\sum\limits_{i=}^{n} x_i^2 - n\bar{x}^2} = \dfrac{\sum\limits_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2}. \end{cases}$$

Якщо значення x_i відомі без похибок, а значення y_i незалежні та точні, то оцінка дисперсії (похибка вимірювань) величин y_i визначаються за формулою $\sigma^2 = \dfrac{\Phi_{\min}}{n-2}$, де $\Phi_{\min} = \sum\limits_{i=1}^{n} \left(y_i - \alpha_1 x_i - \alpha_0\right)^2.$

Оцінки дисперсій коефіцієнтів α_1 і α_0 визначаються за формулами

$$\sigma_{\alpha_0}^2 = \frac{\sum\limits_{i=1}^{n} x_i^2}{n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2} \cdot \frac{\Phi_{\min}}{n-2}, \quad \sigma_{\alpha_1}^2 = \frac{n}{n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2} \cdot \frac{\Phi_{\min}}{n-2}.$$

Якщо величини y_i мають нормальний розподіл, то для коефіцієнтів α_1 і α_0 справджуються такі надійні інтервали:

$$\left(\alpha_k - t(n-2, \gamma)\sigma_{\alpha_k}, \alpha_k + t(n-2, \gamma)\sigma_{\alpha_k}\right), k = 0, 1,$$

де α_k – оцінки, отримані методом найменших квадратів, а число $t(n-2, \gamma)$ знаходиться за таблицею Стьюдента при кількості степенів вільності $k = n - 2$ і $\gamma = P\left(|t| < t(n-2, \gamma)\right)$.

Приклад 2. Для торгових агентів компанії зареєстровані дві ознаки: X – витрати на представництво, Y – обсяг продажів товарів за певний час (в тис. грн). Дані наведені в таблиці 10.3

Таблиця 10.3

x_i	6,5	6,5	6,2	6,7	6,9	6,5	6,1	6,7
y_i	105	125	110	120	140	135	95	130

Знайти вибірковий коефіцієнт кореляції, рівняння прямої регресії Y на X, похибку вимірювань, надійні інтервали для коефіцієнтів прямої регресії при $\gamma = 0{,}95$.

Розв'язання. Усі результати обчислень заносимо в таблицю 10.4 $(n = 8)$.

Таблиця 10.4

№	x_i	y_i	$x_i y_i$	x_i^2	y_i^2	$\widehat{y}_i$	$\left(y_i - \widehat{y}_i\right)^2$
1	6,5	105	682,5	42,25	11025	119,4120	207,7057
2	6,5	125	812,5	42,25	15625	119,4120	31,2257
3	6,2	110	682,0	38,44	12100	105,2981	22,1079
4	6,7	120	804,0	44,89	14400	128,8212	77,8136
5	6,9	140	966,0	47,61	19600	138,2305	3,1311
6	6,5	135	877,5	42,25	18225	119,4110	242,9857
7	6,1	95	579,5	37,21	9025	100,5934	31,2861
8	6,7	130	871,0	44,89	16900	128,8212	1,3896
$\sum$	52,1	960	6275	339,79	116900		617,6454

$$\bar{x} = \frac{1}{8} \cdot 52{,}1 = 6{,}5125; \quad \bar{y} = \frac{1}{8} \cdot 960 = 120;$$

$$\bar{s}_x^2 = \frac{1}{8} \cdot 339{,}79 - 6{,}5125^2 = 0{,}0611; \quad \bar{s}_x = 0{,}2472;$$

$$\bar{s}_y^2 = \frac{1}{8} \cdot 116900 - 120^2 = 212{,}5; \quad \bar{s}_y = 14{,}5774; 23$$

$$r = \frac{6275 - 8 \cdot 6{,}5125 \cdot 120}{8 \cdot 0{,}2472 \cdot 14{,}5774} = 0{,}7978.$$

Записуємо пряму регресії Y на X за формулою (6.11) $y-\bar{y}=r\frac{\bar{s}_y}{\bar{s}_x}(x-\bar{x})$, тому

$$\alpha_1=r\frac{\bar{s}_y}{\bar{s}_x}=0{,}7978\cdot\frac{14{,}5774}{0{,}2472}=47{,}0463;$$

$$\alpha_0=\bar{y}-\alpha_1\bar{x}=120-47{,}0463\cdot 6{,}5125=-186{,}3890;$$

$$\hat{y}=47{,}0463x-186{,}3890.$$

Обчислимо:

$$\hat{y}_1=47{,}0463\cdot 6{,}5-186{,}389=119{,}4120;$$

$$\hat{y}_2=47{,}0463\cdot 6{,}5-186{,}389=119{,}4120;$$

$$\hat{y}_3=47{,}0463\cdot 6{,}2-186{,}389=105{,}2981;$$

$$\hat{y}_4=47{,}0463\cdot 6{,}7-186{,}389=128{,}8212;$$

$$\hat{y}_5=47{,}0463\cdot 6{,}9-186{,}389=138{,}2305;$$

$$\hat{y}_6=47{,}0463\cdot 6{,}5-186{,}389=119{,}4120;$$

$$\hat{y}_7=47{,}0463\cdot 6{,}1-186{,}389=100{,}5934;$$

$$\hat{y}_8=47{,}0463\cdot 6{,}7-186{,}389=128{,}8212.$$

$$\Phi_{\min}=617{,}6454.$$

Знаходимо

$$\Delta=n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2=8\cdot 339{,}79-52{,}1^2=3{,}91.$$

Тоді $\sigma^2=\frac{617{,}6454}{8-2}=102{,}94044;$

$$\sigma_{\alpha_0}^2=\frac{339{,}79}{3{,}91}\cdot 102{,}9409=8945{,}8537;\quad \sigma_{\alpha_0}=94{,}5825;$$

$$\sigma_{\alpha_1}^2=\frac{8}{3{,}91}\cdot 102{,}9409=210{,}6208;\quad \sigma_{\alpha_1}=14{,}1278.$$

З таблиці розподілу Стьюдента знаходимо $t(6;0{,}95)=2{,}45.$ Аналогічно, за допомогою $t_{крит}$=**СТЬЮДРАСПОБР**(0,05;6)=2,4469. Тоді з ймовірністю 0,95

$$-186{,}389-2{,}45\cdot 94{,}5825<\alpha_0<-186{,}389+2{,}45\cdot 94{,}5825,$$

тобто $-418{,}1161 < \alpha_0 < 45{,}3381.$

$$47{,}0463 - 2{,}45 \cdot 14{,}1278 < \alpha_1 < 47{,}0463 + 2{,}45 \cdot 14{,}1278,$$

тобто $12{,}4332 < \alpha_1 < 81{,}6594.$

Приклад 3.

За даними таблиці 10.5 (опитування однієї з груп економічної спеціальності) проаналізувати зв'язок між вибраними означеними факторами за частинними і множинним коефіцієнтами кореляції., побудувати лінії регресії.

Таблиця 10.5

D	E	F	G	H	I	J
Дані студентів		Зріст X	Вага Y	ЗНО Z	Вишка U	макек V
1		165	52	176	70	80
2		165	64	175	97	97
3		175	76	186,5	85	62
4		172	54	171	90	80
5		168	49	176	80	90
6		165	46	173	88	90
7		158	49	175	80	80
8		180	90	180	68	70
9		165	43	176	97	80
10		194	90	177	85	80
11		166	55	174	80	80
12		169	50	170	78	70
13		154	43	192	90	80
14		158	45	192	90	90
15		170	57	189,5	97	90
16		160	47	176	80	80
17		168	52	185	80	80

Розв'язання. За даними таблиці обчислення та вигляд ліній регресій для U, V (U − середня оцінка з вищої математики за два семестри в балах, V − оцінка з макроекономіки за семестр в балах) показано на рис. 10.4 та 10.4.

L	M	N	O
U	V		
84,41176	81,11765	середнє	
74,88235	74,48529	дисперсія	
8,653459	8,630486	кв. відхил.	
0,506451		кореляція	
0,507799	коефіцієн	0,505106	
41,19143		42,63689	
43,22034		38,48075	
u=0,508v+43,22		v=0,505u+38,48	

Рис.10.4. Обчислення всіх значень для ліній регресій.

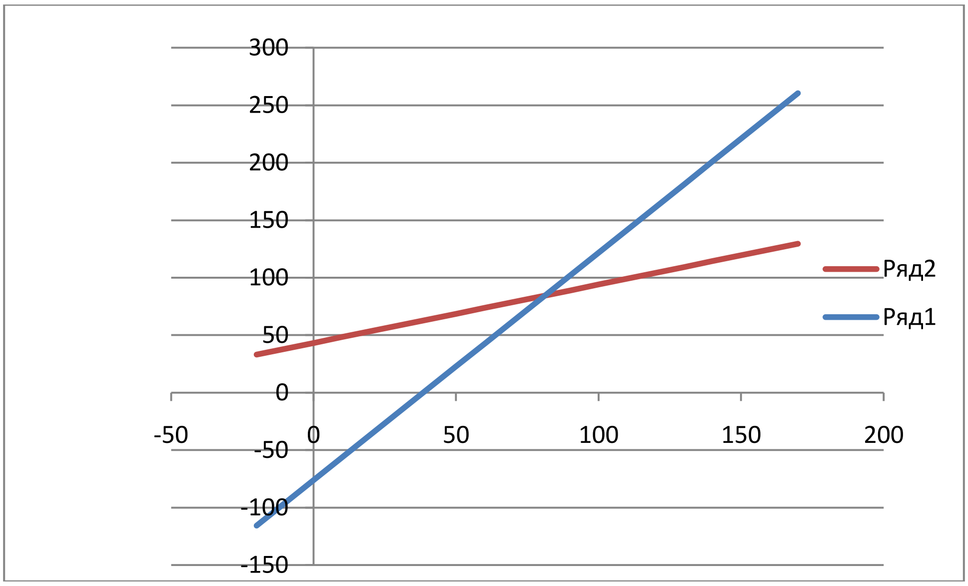


Рис. 10.5. Вигляд ліній регресій.

Виконання лабораторної роботи

Завдання до теми

Завдання 1. Розв'язати наступні задачі:

– непарні варіанти: 1.1, 1.3;

– парні варіанти: 1.2, 1.4.

1.1. В таблиці 10.6 наведено дані про роздрібний товарообіг *Z* (млрд. грн.), середню кількість населення *X* (млн. осіб) та середній дохід *Y* (млн. грн.). Проаналізувати зв'язок між *Z* та *X* і *Y* за частинними і множинним коефіцієнтами кореляції.

Таблиця 10.6

Z	1,2	1,3	2,5	1,4	1,2	0,2	2,4	4,1	1,1
X	1,4	1,4	2,5	1,5	1,3	0,3	2,6	4,2	1,1
Y	1,3	1,3	1,4	1,8	1,5	1,6	1,8	1,9	1,6

1.2. В таблиці 10.7 наведені дані про щомісячний прибуток *Z* (тис. у. од.), витрати на рекламу *X* (тис. у. од.) та вкладення капіталу в цінні папери *Y* (тис. у. од.). Проаналізувати зв'язок між *Z* та *X* і *Y* за частинними і множинним коефіцієнтами кореляції.

Таблиця 10.7

Z	10	12	12	14	16	17	18
X	0,2	0,5	0,3	0,5	0,5	0,6	0,8
Y	0,8	0,2	1	1,2	0,9	1	1,1

1.3. Для дослідження впливу капіталовкладень *X* (млн. грн.) на отриманий річний прибуток *Y* (млн. грн.) було зібрано статистичні дані за 20 великими підприємствами (табл. 10.8). Визначити силу зв'язку між означеними факторами.

Таблиця 10.8

Y \ *X*	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
1,5 – 2,5	1	-	-	-	-
2,5 – 3,5	2	5	2	-	-
3,5 – 4,5	-	3	3	2	-
4,5 – 5,5	-	-	-	2	-

1.4. В таблиці 10.9 наведено дані про рівень витрат *X* (%) та річний дохід *Y* (млн. грн.), які було зібрано за 50 супермаркетами. Визначити силу зв'язку між означеними факторами.

Таблиця 10.9

X / *Y*	4 – 6	6 – 8	8 – 10	10 – 12	12 – 14
0,5 – 2,0	-	-	2	3	1
2,0 – 3,5	-	4	5	1	-
3,5 – 5,0	-	8	5	5	-

5,0 – 6,5	3	8	2	-	-
6,5 – 8,0	2	1	-	-	-

Завдання 2. Зібрати дані своєї академічної групи, як у наведеному прикладі 3, проаналізувати зв’язок між вибраними означеними факторами за частинними і множинним коефіцієнтами кореляції, побудувати лінії регресії.

Завдання 3. Дати відповіді на подані нижче теоретичні запитання, одне із запитань 3 або 4 оформити письмово.

Теоретичні запитання до теми

1. Пояснити побудову множинного коефіцієнта кореляції.
2. Побудова частинного коефіцієнтів кореляції.
3. Пояснити доцільність використання вбудованої статистичної функції Excel **FРАСПОБР.**
4. Показати випадки використання функції **СТЬЮДРАСПОБР.**
5. Побудова прямих регресії, їх особливість.
6. Пояснити застосування функцій програмного забезпечення Excel для розв’язування практичних задач кореляційного аналізу.

Оформлення звіту та порядок захисту

Лабораторна робота виконується на аркушах А4, в ній стисло відображаються формули теоретичної частини, хід роботи та отримані результати. Особливу увагу звернути на повне оформлення завдання 2 (усі таблиці, графіки, обчислення та до них пояснення). Також студент здає електронний варіант проведених обчислень практичних задач.

Лабораторна робота № 11

Перевірка законів розподілу випадкових величин: розподіл із рівномірною щільністю та розподіл Пуассона

Мета роботи: навчитись перевіряти гіпотези про закони розподілу величини *X,* використовуючи теоретичні знання та всі набуті навики обчислень за допомогою застосування табличного процесора Microsoft Excel.

Теоретичні відомості

11.1. Поняття про статистичні гіпотези

На основі статистичних даних при розв'язуванні практичних задач необхідно зробити припущення про вигляд закону розподілу випадкової величини X. При цьому для остаточного вирішення питання про вигляд закону розподілу доцільно перевірити, наскільки зроблене припущення узгоджується з дослідними даними. Через обмежену кількість спостережень емпіричний закон розподілу, зазвичай, в деякій мірі, відрізняється від передбачуваного, навіть, якщо припущення про вигляд закону розподілу виявилося правильним. В зв'язку з цим виникає наступна задача: чи розбіжність між емпіричним і передбачуваним (теоретичним) законом розподілу є наслідком обмеженості числа спостережень, а чи вона є істотною і пов'язана з тим, що істинний закон розподілу випадкової величини суттєво відрізняється від передбачуваного. Для розв'язування цієї задачі служать так звані «критерії згоди».

«*Критерієм згоди»* називають критерій перевірки гіпотези про передбачуваний вигляд закону розподілу, наприклад, критерії згоди: Колмогорова, Пірсона (критерій χ^2), Романовського та інші.

При застосуванні певних статистичних методів обробки даних вибірки часто ставляться вимоги до розподілу даних або до числових характеристик.

Означення. ***Статистичною гіпотезою*** називається будь-яке припущення про властивості досліджуваної величини, висунуте на основі статистичних даних.

За змістом статистичні гіпотези можна віднести до таких типів:

1) Гіпотези про вид закону розподілу досліджуваної величини.

2) Гіпотези про числові характеристики досліджуваної величини.

3) Гіпотези про рівність числових характеристик досліджуваних величин.

4) Гіпотези про належність досліджуваних величин до однієї генеральної сукупності.

5) Гіпотези про вид моделі, що описує взаємозв'язок між досліджуваними величинами.

6) Гіпотези про належність досліджуваних величин до одного класу.

Прийняття основної або однієї з альтернативних гіпотез здійснюється на основі дослідження статистичних даних. Дослідження проводиться за певним ***критерієм***, який обирається відповідно до змісту гіпотези і виду наявних статистичних даних.

Якщо сформульовані гіпотези H_0 – основна та H_1, як альтернативна (конкуруюча), і обраний критерій перевірки справедливості основної гіпотези, то прийняття H_0 означає відкидання H_1, а відкидання H_0 означає справедливість H_1.

Оскільки прийняття гіпотези здійснюється на основі статистичних даних, то завжди існує ймовірність помилки.

Означення. Імовірність відкидання гіпотези H_0, якщо вона справедлива, називається ймовірністю ***помилки першого роду*** або ***рівнем значущості***, який позначається α. Величина $1-\alpha$ є ймовірністю прийняття справедливої гіпотези і називається ***рівнем надійності***.

Ймовірність прийняття гіпотези H_0, якщо вона не вірна, називається ймовірністю ***помилки другого роду*** і позначається β. Величина $1-\beta$ є ймовірністю відкидання невірної гіпотези і називається ***потужністю критерію***.

Чим менше рівень значущості, тим менше ймовірність відкинути вірну гіпотезу. Зазвичай рівень значущості обирається дослідником рівним 0,1; 0,05;

0,01 або 0,001. Якщо, наприклад, обраний рівень значущості α=0,01, то ризик відкинути вірну гіпотезу виникає в одному випадку із ста.

Зауваження. Перевірка статистичної гіпотези не надає точного висновку щодо її вірності або невірності. *Прийняття гіпотези означає, що на прийнятому рівні значущості вона не протирічить статистичним даним.*

Перевірка статистичних гіпотез зазвичай здійснюється за такими ***етапами:***

1) Висунення припущень про вид розподілу досліджуваної величини (величин) або про її числові характеристики.

2) Формулювання статистичних гіпотез.

3) Вибір критерію перевірки відповідно до змісту гіпотез і статистичних даних.

4) Вибір рівня значущості залежно від вимог до точності результатів дослідження.

5) Розрахунок значення обраного критерію за статистичними даними.

6) Порівняння розрахованого значення критерію з його критичним значенням і прийняття або відкидання основної гіпотези.

11.2. Перевірка гіпотези про вид закону розподілу досліджуваної величини

Перевірка гіпотези про вид закону розподілу досліджуваної величини має велике значення для прикладних досліджень. Необхідність такої перевірки виникає при виборі критерію, оскільки для багатьох з них висувається вимога нормального розподілу статистичних даних. Означені гіпотези перевіряються при проектуванні систем масового обслуговування, перевірки якості продукції або праці і т. ін.

Припустимо, що з деякої генеральної сукупності X, яка розглядається як випадкова величина, обрана вибірка $\{x_1, x_2, \ldots x_n\}$. За даними вибірки побудовано статистичний ряд (табл. 11.1), що містить варіанти x_i та відповідні частоти n_i, $i=\overline{1,k}$, k – кількість варіант у випадку дискретного ряду. У випадку інтервального ряду x_i – середини інтервалів, k – кількість інтервалів.

Таблиця 11.1

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k

Отриманий на основі вибіркових даних статистичний ряд називається ***емпіричним законом розподілу*** величини *X*.

За даними статистичного ряду можна знайти числові характеристики, які є вибірковими параметрами закону розподілу *X*. Вид закону розподілу визначається відповідно до умов здобуття вибірки або залежно від виду графіка емпіричної щільності розподілу (гістограми) у випадку неперервної випадкової величини *X* і полігону частот, якщо величина *X* дискретна. Параметри обраного закону розподілу заміняються відповідними вибірковими параметрами.

Означення. Закон розподілу випадкової величини *X*, параметрами якого є відповідні вибіркові числові характеристики, називається ***теоретичним законом розподілу***.

При здійсненні такої заміни немає впевненості, що закон розподілу обраний правильно. Тому розроблено процедуру, яка дозволяє оцінити ступінь відповідності обраного закону даним вибірки. Критерії, в яких формулюється лише одна гіпотеза H_0, і необхідно перевірити, чи узгоджуються статистичні дані з цією гіпотезою, чи ні, називають ***критеріями згоди***. Найбільш відомим з яких є **критерій Пірсона** χ^2 (хі-квадрат).

Критерій Пірсона χ^2 обчислюється за формулою:

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(n_i - n_i'\right)^2}{n_i'}, \tag{11.1}$$

де n_i' – частоти, отримані за теоретичним законом розподілу (теоретичні частоти).

З формули (11.1) видно, що у випадку, коли відповідні теоретичні та емпіричні частоти співпадають, $\chi^2 = 0$. Отже, чим ближче χ^2 до нуля, тим краще узгоджуються вибіркові дані та обраний теоретичний закон розподілу.

Розраховане значення критерію χ^2 порівнюється з його критичним значенням $\chi^2_{\alpha,l}$, яке знаходиться за статистичними таблицями (додаток, таблиця 3) або за допомогою вбудованої статистичної функції Excel **ХИ2ОБР**(α ,*l*). Параметри функції **ХИ2ОБР**: α – рівень значущості; *l* – степені вільності, $l = k - r - 1$, де k – кількість груп емпіричного розподілу, r – кількість параметрів теоретичного розподілу (наприклад, для нормального розподілу $r = 2$, оскільки параметрів два – a і σ). Якщо $\chi^2 < \chi^2_{\alpha,l}$, то гіпотеза про закон розподілу приймається. У противному випадку гіпотеза відкидається.

За критерієм Романовського розглядають величину

$$a = \frac{\left|\chi^2 - l\right|}{\sqrt{2l}},$$

де l – степені вільності. Якщо $a \geq 3,$ то розбіжність між теоретичними та дослідними даними слід вважати не випадковою. Якщо $a < 3$, то таку розбіжність можна вважати випадковою, тобто емпіричні дані узгоджуються з обраним теоретичним розподілом.

Зауваження. У деяких статистичних таблицях критичне значення χ^2 надається залежно від рівня надійності γ , $\gamma = 1 - \alpha$.

Перевірка гіпотези про закон розподілу величини X здійснюється за такими пунктами:

1) З генеральної сукупності *X* здобувається вибірка і будується статистичний ряд.

2) Висувається гіпотеза про закон розподілу випадкової величини *X*.

3) Знаходяться вибіркові параметри обраного закону розподілу.

4) Розраховуються теоретичні частоти.

5) Розраховується критерій χ^2 за формулою (11.1).

6) Обирається рівень значущості α (або рівень надійності γ) і знаходиться критичне значення $\chi^2_{\alpha,l}$ (або $\chi^2_{\gamma,l}$).

7) Порівнюються розраховане і критичне значення критерію χ^2 і робиться висновок про справедливість висунутої гіпотези.

11.3. Приклади перевірки законів розподілу випадкової величини

11.3.1. Закон розподілу Пуассона

Приклад 1. На одній з міських АТС фіксувалася кількість телефонних дзвінків в годину. Спостереження велися протягом 100 годин, їх результати представлені в таблиці 11.2. Чи можна вважати навантаження на АТС стандартним?

Таблиця 11.2

Кількість викликів в годину	0	1	2	3	4	5	6	7
Кількість спостережень	6	27	26	20	10	5	5	1

Розв'язання. Навантаження на АТС можна вважати стандартним, якщо випадкова величина X – кількість телефонних дзвінків, що поступили, підкоряється закону розподілу Пуассона. Тому необхідно перевірити гіпотезу про закон розподілу випадкової величини. Сформулюємо гіпотези:

H_0 – випадкова величина X підкоряється закону розподілу Пуассона;

H_1 – випадкова величина X не підкоряється закону розподілу Пуассона.

Означення. ***Закон Пуассона*** має вигляд:

$$p_k = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0,1,\ldots; \lambda > 0, \qquad (11.2)$$

де $\lambda -$ параметр розподілу. Крім того, відомо, що $\bar{x} = \lambda; \quad s^2 = \lambda$. Отже, для встановлення параметра λ достатньо знайти $\bar{x}$, обчислити s^2 та порівняти їх.

Випадкова величина X – кількість викликів в годину; тоді кількість спостережень – це відповідні значенням X частоти n_i, а таблиця 11.2 є статистичним рядом і емпіричним законом розподілу величини X. Знайдемо $\overline{X}$ і s^2. Для зручності обчислення оформимо у вигляді таблиці (табл. 11.3).

Таблиця 11.3

x_i	0	1	2	3	4	5	6	7	Суми
n_i	6	27	26	20	10	5	5	1	100
$x_i\ n_i$	0	27	52	60	40	25	30	7	241

$\left(x_i - \overline{x}\right)^2 n_i$	34,85	53,68	4,37	6,96	25,28	33,54	64,44	21,07	244,19

Отже,

$$\overline{x} = \frac{1}{n}\sum_{i=1} x_i n_i = \frac{1}{100}\cdot 241 = 2,41; \quad s^2 = \frac{1}{n}\sum_{i=1}(x_i - \overline{x})^2 n_i = \frac{1}{100}\cdot 244,19 \approx 2,44.$$

Оскільки повинна виконуватися рівність $\overline{x} = \lambda; s^2 = \lambda,$ то як параметр можна вибрати або $\overline{x}$, або s^2, або їх середнє арифметичне. Виберемо $\lambda = \frac{\overline{x} + s^2}{2} \approx 2,426$. Таким чином, гіпотеза H_0 – це припущення, що величина X розподілена згідно із законом Пуассона за формулою (11.2).

Перевіримо правильність гіпотези за допомогою критерію Пірсона. Знайдемо теоретичні частоти, використовуючи формулу (11.2):

$$p_k = \frac{e^{-2,426} 2,426^k}{k!}, \quad k = 0,1,\ldots,$$

де за k візьмемо значення X, тобто x_i.

Відмітимо, що p_i – ймовірність того, що X прийме значення x_i, тобто статистично вони є відносними частотами, теоретичні частоти знаходитимемо за формулою: $n_i' = np_i$.

Для зручності при обчисленні теоретичних частот продовжимо таблицю, складену на основі статистичного ряду (табл. 11.2).

Отже, за результатами розрахунків $\chi^2 = \sum_{i=0}^{7} \frac{\left(n_i - n_i'\right)^2}{n_i'} = 5,739$.

Для даного завдання $l = k - r - 1 = 8 - 1 - 1 = 6$. Виберемо рівень значущості α=0,01 і знайдемо за допомогою таблиць або функції **ХИ2ОБР** табличного процесора Excel значення $\chi^2_{\alpha,l}$: $\chi^2_{0,01;6}$=16,812. Оскільки для такого рівня надійності, тобто $\gamma = 0,99$: $\chi^2 < \chi^2_{\alpha,l}$, то гіпотезу H_0 про розподіл Пуассона можна прийняти.

Таблиця 11.4

x_i	0	1	2	3	4	5	6	7	Суми
n_i	6	27	26	20	10	5	5	1	100
$p_i = \dfrac{e^{-2,426} 2,426^{x_i}}{x_i!}$	0,09	0,21	0,26	0,21	0,13	0,06	0,03	0,01	
$n_i' = np_i$	8,84	21,44	26,01	21,03	12,76	6,19	2,50	0,87	
$\dfrac{(n_i - n_i')^2}{n_i'}$	0,91	1,44	4,65E-06	0,05	0,59	0,23	2,49	0,02	5,739

Зауваження. В таблицю внесено для ймовірності p_i наближені значення, а на рис. 11.1 точніші, обчислені за допомогою статистичної функції **ПУАССОН** табличного процесора Excel . Також показано використання вбудованої функції **ХИ2ОБР.**

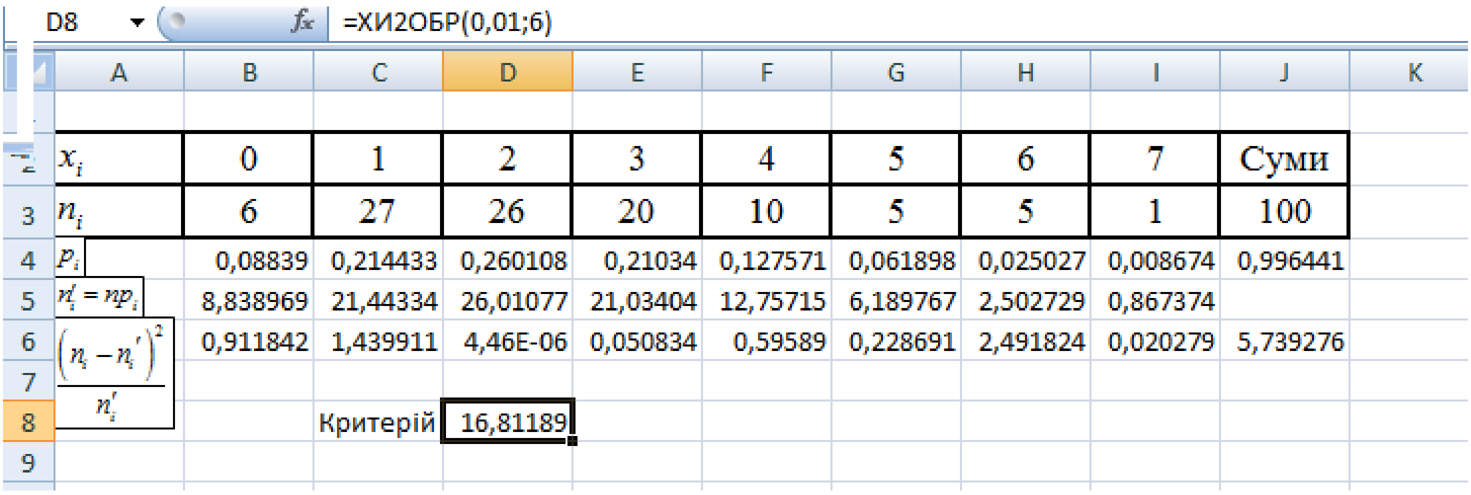


Рис.11.1. Використання функції **ХИ2ОБР.**

Висновок. Навантаження на АТС можна вважати стандартним.

11.3.2. Рівномірний закон розподілу

Приклад 2. З метою впорядкування роботи міського суспільного транспорту фіксувався час очікування в хвилинах пасажирами тролейбусів на декількох маршрутах. Було проведено 200 вимірювань, їх результати представлені в таблиці 11.5. Чи можна вважати, що перевезення по перевірених маршрутах забезпечені раціонально?

Таблиця 11.5

Час очікування	1 – 3	3 – 5	5 – 7	7 – 9	9 – 11	11 – 13
Кількість спостережень	25	30	48	35	42	20

Розв'язання. Можна вважати, що перевезення по перевірених маршрутах забезпечені раціонально, якщо випадкова величина X – час очікування пасажирами транспорту підкоряється рівномірному закону розподілу. Отже, задача зводиться до перевірки гіпотези про рівномірний закон розподілу випадкової величини. Сформулюємо гіпотези:

H_0 – випадкова величина X розподілена рівномірно;

H_1 – випадкова величина X не розподілена рівномірно.

Означення. Щільність розподілу випадкової величини, яка є ***рівномірно розподіленою***, має вигляд

$$f(x) = \begin{cases} \dfrac{1}{b-a}, a \le x \le b, \\ 0, \qquad x < a, x > b, \end{cases} \qquad (11.3)$$

де $a; b$ − параметри розподілу.

Відомо, що $\bar{x} = \dfrac{a+b}{2}$; $s^2 = \dfrac{(b-a)^2}{12}$, тобто для встановлення параметрів a і b потрібно знайти $\bar{x}$ і s^2, після чого розв'язати систему

$$\begin{cases} \bar{x} = \dfrac{a+b}{2}, \\ s^2 = \dfrac{(b-a)^2}{12}. \end{cases} \qquad (11.4)$$

Оскільки за умов задачі випадкова X – час очікування транспорту, то кількість спостережень – це відповідні значенням X частоти n_i, а таблиця 7.3 – це інтервальний статистичний ряд і емпіричний закон розподілу X. Знайдемо $\bar{x}$ і s^2. За x_i візьмемо середини відповідних інтервалів. Для зручності обчислення оформимо у вигляді таблиці (табл.11.6).

Таблиця11.6

$[a_i;a_{i+1})$	1 – 3	3 – 5	5 – 7	7 – 9	9 – 11	11 – 13	Суми
x_i	2	4	6	8	10	12	
n_i	25	30	48	35	42	20	200
$x_i\ n_i$	50	120	288	280	420	240	1398
$\left(x_i-\overline{x}\right)^2 n_i$	622,5	268,2	47,045	35,704	380,52	502	1856

Отже, $\overline{x}=\frac{1}{n}\sum_{i=1}^{6}x_i n_i=\frac{1}{200}\cdot 1398=6,99$; $s^2=\frac{1}{n}\sum_{i=1}^{6}(x_i-\overline{x})^2 n_i=\frac{1}{200}\cdot 1856\approx 9,2799$.

Складемо систему (11.4) для визначення параметрів рівномірного розподілу і розв'яжемо її:

$$\begin{cases} 6,99=\dfrac{a+b}{2} \\ 9,2799=\dfrac{\left(b-a\right)^2}{12} \end{cases} \Rightarrow \begin{cases} a+b=13,98 \\ \left(b-a\right)^2=111,3588 \end{cases} \Rightarrow \begin{cases} a+b=13,98 \\ b-a\approx 10,55 \end{cases} \Rightarrow \begin{cases} a\approx 1,715 \\ b\approx 12,265 \end{cases}.$$

Таким чином, гіпотеза H_0 – це припущення, що X розподілена за рівномірним законом із щільністю розподілу:

$$f(x)=\begin{cases} \dfrac{1}{12,265-1,715}=\dfrac{1}{10,55}, & 1,715\le x\le 12,265 \\ 0, & x<1,715, \quad x>12,265 \end{cases}.$$

Перевіримо справедливість гіпотези H_0 за допомогою критерію Пірсона. Для знаходження теоретичних частот використаємо формулу $n_i'=np_i$, а ймовірності попадання в інтервали p_i знайдемо за формулою

$$P\left(\alpha<X<\beta\right)=\frac{\beta-\alpha}{b-a}.$$

Для зручності обчислень теоретичних частот складемо таблицю (табл.11.7). Врахуємо, що $\beta-\alpha=2$; $b-a=12,265-1,715=10,55$ для всіх інтервалів, окрім першого і останнього.

Для першого інтервалу $\beta-\alpha=\beta-a=3-1,715=1,285$; для останнього інтервалу $\beta-\alpha=b-\alpha=12,265-11=1,265$.

Таблиця 11.7

$[a_i; a_{i+1})$	1 – 3	3 – 5	5 – 7	7 – 9	9 – 11	11 – 13	Суми
n_i	25	30	48	35	42	20	200
$\beta - \alpha$	1,285	2	2	2	2	1,265	
$p_i = \dfrac{\beta - \alpha}{10,55}$	0,122	0,19	0,19	0,19	0,19	0,12	
$n_i' = np_i$	24,360	37,915	37,915	37,915	37,915	23,981	
$\dfrac{\left(n_i - n_i'\right)^2}{n_i'}$	0,017	1,6522	2,6827	0,2241	0,4402	0,6609	5,677

Отже, $\chi^2 = \sum_{i=0}^{7} \dfrac{\left(n_i - n_i^{'}\right)^2}{n_i^{'}} = 5,677$. Таку ж таблицю формується в Excel набагато швидше, задаючи вихідні дані та описуючи відповідні функції для розрахунків, одну з яких видно на рис. 11.2, а перевірка критерію на рис. 11.3.

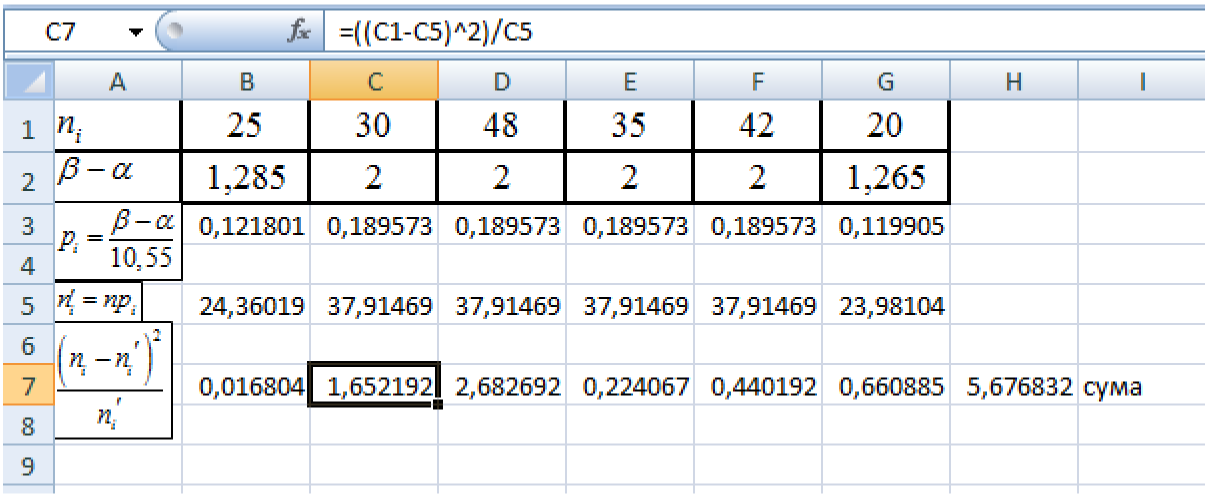


Рис.11.2. Задання функції для обчислення.

Для даного завдання $l = k - r - 1 = 6 - 2 - 1 = 3$. Виберемо рівень значущості α=0,01 і знайдемо за допомогою таблиць або функції **ХИ2ОБР** табличного процесора Excel значення $\chi^2_{\alpha,l}$: $\chi^2_{0,01;3}$=11,34 (рис. Оскільки для такого рівня надійності (0,99) $\chi^2 < \chi^2_{\alpha,l}$, то гіпотезу про рівномірний розподіл приймаємо.

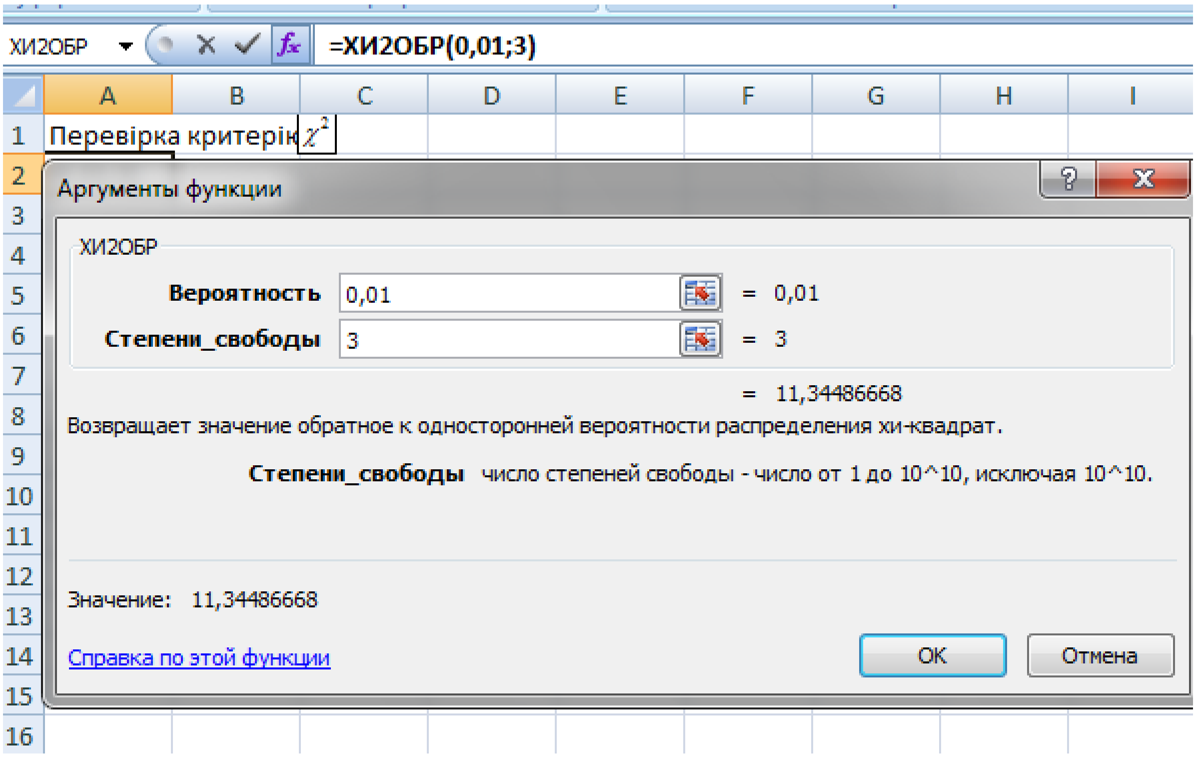


Рис. 11.3. Перевірка критерію.

Висновок. Перевезення по перевірених маршрутах організовані раціонально.

Виконання лабораторної роботи

Завдання до теми

Завдання 1.

За даними побудувати гістограму розподілу, перевірити правильність гіпотези H_0 – *випадкова величина X розподілена згідно із законом Пуассона* за допомогою критерію Пірсона та Романовського.

№	*X*	0	1	2	3	4	5	6	7	8	9	10
1	*n*	37	37	18	6	2	0	0	0	0	0	0
2		29	36	22	9	3	1	0	0	0	0	0
3		22	33	25	13	5	2	0	0	0	0	0
4		17	30	27	16	7	2	1	0	0	0	0
5		14	27	27	18	9	4	1	0	0	0	0
6		11	24	27	20	11	5	2	0	0	0	0
7		8	21	26	21	13	7	3	1	0	0	0
8		6	18	24	22	15	8	4	2	1	0	0
9		5	15	22	22	17	10	5	2	1	0	0
10		4	13	20	22	18	12	6	3	1	1	0
11		0	4	10	20	28	34	32	26	18	12	8

12		0	4	12	22	32	34	32	26	16	10	6
13		2	6	14	26	34	34	30	22	16	8	4
14		2	6	16	28	36	36	30	20	14	8	4
15		2	8	20	30	36	34	28	20	12	6	4

Завдання 2.

За поданими нижче даними на одному із поліграфічних виробництв (дані вимірювань при стабільній подачі фарби та зволожувального розчину) для одного із показників (вказана величина в таблиці значень є усереднене значення п’яти вимірювань, тому всіх даних, відповідно, не 16, а 80):

1) побудувати таблицю статистичного розподілу;

2) побудувати гістограму;

3) вирівняти дослідні дані за допомогою закону із рівномірною щільністю;

4) перевірити узгодженість між емпіричними та теоретичними даними за допомогою критерія Пірсона.

Дані вимірювань при стабільній подачі фарби та зволожувального розчину

Відбиток	3 ділянка				4 ділянка				5 ділянка				6 ділянка			
	ΔE	D	80	40	ΔE	D	80	40	ΔE	D	80	40	ΔE	D	80	40
5	16,24	0,89	4	5	16,99	0,88	5	5	14,15	0,96	4	6	10,83	1,06	5	6
5	11,75	0,96	4	5	13,38	0,95	4	5	12,51	0,99	4	5	9,78	1,08	5	6
10	16,11	0,89	5	7	16,86	0,87	5	6	14,45	0,93	5	6	11,38	1,03	5	7
10	12,84	0,96	4	6	13,84	0,92	4	5	12,57	0,96	4	6	9,84	1,06	5	7
15	16,31	0,87	5	7	17,10	0,86	5	7	14,09	0,94	5	7	11,26	1,03	6	7
15	12,73	0,95	5	6	14,17	0,91	5	6	13,34	0,96	4	6	10,22	1,05	5	7
20	15,37	0,89	4	6	15,99	0,88	5	6	14,11	0,93	5	6	11,20	1,02	5	6
20	12,76	0,98	5	5	13,64	0,93	4	6	12,79	0,99	4	6	10,43	1,06	5	6
25	12,96	0,95	4	7	12,90	0,96	5	6	11,37	1,01	5	7	9,06	1,09	5	7
25	10,75	1,02	5	7	12,09	0,99	5	7	11,15	1,02	5	7	8,90	1,09	6	8
30	13,93	0,93	5	7	14,47	0,92	5	6	12,88	0,97	5	7	9,99	1,06	5	7
30	11,53	0,99	5	6	12,96	0,96	4	6	11,97	1,01	4	6	9,63	1,08	5	7
35	14,46	0,91	5	7	15,79	0,89	5	6	13,64	0,94	5	7	10,36	1,05	5	7
35	12,05	0,97	5	6	14,02	0,93	5	6	12,96	0,96	4	6	10,13	1,06	5	7
40	14,59	0,91	5	7	16,20	0,86	5	6	13,90	0,93	5	7	11,15	1,02	5	7
40	12,10	0,97	5	7	13,40	0,94	4	6	12,96	0,96	4	6	10,49	1,04	5	7

Відбиток	7 ділянка				8 ділянка				9 ділянка				10 ділянка			
	ΔE	D	80	40	ΔE	D	80	40	ΔE	D	80	40	ΔE	D	80	40
5	8,70	1,13	5	7	5,68	1,26	5	7	3,28	1,36	6	8	2,34	1,44	6	9
5	6,76	1,18	5	6	4,37	1,29	6	7	2,93	1,40	6	8	2,97	1,53	7	10
10	8,62	1,12	5	7	5,84	1,22	5	8	3,32	1,34	6	8	1,59	1,45	7	10
10	7,10	1,16	5	7	4,79	1,28	6	8	2,79	1,37	7	9	2,62	1,52	8	10
15	8,46	1,11	6	9	5,69	1,23	6	9	3,01	1,34	7	9	1,73	1,46	7	10
15	7,60	1,17	5	8	5,00	1,26	6	9	3,14	1,40	7	9	3,14	1,54	8	11
20	8,43	1,12	6	7	5,43	1,23	5	8	3,03	1,35	6	8	1,85	1,43	7	10
20	7,48	1,16	5	7	4,61	1,28	6	8	2,80	1,40	6	9	2,94	1,54	8	11
25	6,97	1,16	6	9	4,92	1,26	6	9	2,87	1,37	6	10	2,34	1,46	8	11
25	6,32	1,19	6	9	4,73	1,29	6	9	3,06	1,41	8	11	3,18	1,55	9	13
30	7,81	1,14	5	7	4,97	1,26	6	8	2,99	1,37	6	9	2,48	1,48	7	10
30	7,35	1,16	5	8	4,76	1,31	6	9	2,77	1,43	7	11	2,66	1,52	9	12
35	8,03	1,14	5	7	5,42	1,25	6	9	3,20	1,38	6	10	2,78	1,49	7	11
35	7,66	1,17	5	7	4,65	1,29	6	9	3,07	1,42	7	9	3,51	1,59	9	12
40	8,71	1,11	6	7	5,74	1,23	6	9	3,37	1,36	6	9	2,75	1,48	7	10
40	7,58	1,15	5	7	5,02	1,27	6	8	2,97	1,43	7	10	3,31	1,58	8	11

Відбиток	11 ділянка				12 ділянка				13 ділянка				14 ділянка			
	ΔE	D	80	40	ΔE	D	80	40	ΔE	D	80	40	ΔE	D	80	40
5	2,68	1,52	7	10	3,31	1,58	8	10	3,64	1,55	9	10	3,54	1,57	8	10
5	4,03	1,65	9	11	4,76	1,69	9	11	4,03	1,69	9	12	4,56	1,68	10	12
10	2,11	1,52	7	10	3,19	1,53	9	11	3,24	1,55	9	11	2,94	1,55	9	11
10	3,66	1,64	9	12	4,58	1,70	10	12	4,58	1,65	9	12	4,48	1,66	10	12
15	2,38	1,53	8	11	3,18	1,56	9	12	2,97	1,55	9	12	2,41	1,52	9	11
15	3,88	1,65	9	12	4,88	1,68	10	13	4,67	1,64	10	13	4,52	1,64	9	12
20	2,47	1,53	7	11	3,41	1,58	9	11	3,19	1,55	9	11	2,59	1,55	9	11
20	4,17	1,66	9	12	4,88	1,73	10	12	4,97	1,70	9	13	4,42	1,67	10	14
25	3,02	1,54	8	11	3,54	1,57	9	12	3,28	1,56	9	12	2,55	1,50	9	12
25	4,29	1,65	10	13	4,80	1,68	10	14	4,35	1,62	10	14	3,87	1,61	10	12
30	3,34	1,56	8	11	4,04	1,60	9	11	3,66	1,57	9	11	2,99	1,54	8	11
30	4,40	1,68	9	13	5,08	1,69	10	13	4,33	1,63	10	12	3,93	1,65	10	13
35	3,59	1,55	8	12	4,18	1,60	9	12	3,93	1,57	9	11	3,06	1,53	8	10
35	4,65	1,67	10	14	5,66	1,74	11	15	4,76	1,68	9	13	4,35	1,66	10	12
40	3,48	1,53	8	13	3,64	1,58	10	12	3,89	1,56	9	12	3,13	1,53	8	12
40	4,76	1,67	10	14	5,39	1,72	10	14	4,90	1,66	10	13	4,16	1,65	10	12

Відбиток	15 ділянка				16 ділянка				17 ділянка			
	ΔE	D	80	40	ΔE	D	80	40	ΔE	D	80	40
5	3,46	1,55	8	9	3,18	1,51	8	9	3,12	1,48	7	9
5	4,40	1,68	9	10	3,59	1,65	8	10	3,10	1,59	8	9
10	2,60	1,52	8	10	2,13	1,48	8	9	2,33	1,44	7	9
10	3,57	1,62	10	11	3,15	1,57	8	10	2,60	1,52	8	9

15	1,87	1,49	7	11	1,84	1,46	7	10	1,92	1,41	7	10
15	3,53	1,62	9	12	2,93	1,56	8	9	2,57	1,51	8	9
20	2,04	1,49	7	10	1,75	1,44	7	9	2,09	1,39	7	8
20	3,54	1,59	9	12	2,63	1,55	8	11	2,28	1,51	8	9
25	2,11	1,45	7	11	2,47	1,41	7	9	3,37	1,33	7	10
25	3,15	1,55	10	12	2,31	1,50	8	10	2,78	1,42	7	9
30	2,25	1,48	7	9	2,67	1,38	7	9	3,62	1,33	6	9
30	3,03	1,57	9	11	2,06	1,51	7	10	2,76	1,41	7	9
35	2,43	1,47	6	10	2,92	1,37	7	8	4,21	1,30	7	8
35	3,21	1,56	8	11	2,32	1,47	7	10	3,45	1,40	7	9
40	2,55	1,44	7	10	2,97	1,36	7	9	4,31	1,29	6	10
40	3,00	1,54	8	11	2,23	1,44	7	9	3,40	1,36	7	9

Зауваження. Для виконання завдань 1), 2) використовувати набуті навики лабораторної роботи № 8). Номер варіанта відповідає номеру ділянки, варіанти 1 та 2 є, відповідно, варіантами 16 та 17.

Завдання 3. Дати відповіді на подані нижче теоретичні запитання, одне із них оформити письмово.

Теоретичні запитання до теми

1. Гіпотези, їх різновиди.
2. Помилки першого і другого роду, їх визначення.
3. Пояснити етапи перевірки статистичної гіпотези.
4. Описати критерії згоди, їх застосування та використання.
5. Описати рівномірний закон розподілу, його числові характеристики.
6. Пояснити вигляд гістограми для розподілу Пуассона, записавши вигляд його диференціальної функції розподілу та вказавши параметр λ.

Оформлення звіту та порядок захисту

Лабораторна робота виконується на аркушах А4, в ній стисло відображаються формули теоретичної частини, які відповідають кожному законові розподілу, перевірці гіпотези. Студент здає електронний варіант проведених обчислень практичних задач, причому із повним описанням всіх використаних можливостей табличного процесора Microsoft Excel.

Лабораторна робота № 12

Статистична перевірка гіпотези про нормальний закон розподілу

Мета роботи: навчитись перевіряти гіпотезу про нормальний закон розподілу величини *X,* використовуючи теоретичні знання та всі набуті навики обчислень за допомогою застосування табличного процесора Microsoft Excel.

Теоретичні відомості

12.1. Центральна гранична теорема

Центральна гранична теорема – теорема теорії ймовірності про збіжність розподілу суми незалежних однаково розподілених випадкових величин до нормального розподілу. Вона і підкреслює особливість нормального розподілу.

Теорема Ляпунова. *Якщо для послідовності попарно незалежних випадкових величин* $X_1, X_2, \ldots, X_n, \ldots$ *можна знайти таке число* $\delta > 0$*, що*

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} M\left|X_k - M(X_k)\right|^{2+\delta}}{\left(\sqrt{\sum_{k=1}^{n} D(X_k)}\right)^{2+\delta}} = 0,$$

то

$$\lim_{n\to\infty} P\left\{ \frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} M(X_k)}{\sqrt{\sum_{k=1}^{n} D(X_k)}} < x \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$$

Зауваження. На практиці зазвичай найлегше перевірити умову Ляпунова для $\delta = 1$. Якщо послідовність випадкових величин задовольняє умову Ляпунова, то вона задовольняє також умову Лінденберга. Обернене твердження неправильне.

Теорема Лінденберга-Леві. *Якщо попарно незалежні випадкові величини* $X_1, X_2, \ldots, X_n, \ldots$*однаково розподілені і мають математичне сподівання а і дисперсію* σ^2*, то*

$$\lim_{n\to\infty} P\left\{\frac{\sum_{k=1}^{n} X_k - na}{\sigma\sqrt{n}} < x\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt .$$

За теоремою Ляпунова вплив кожного окремого доданка на суму при великих n дуже малий, і у разі необмеженого збільшення кількості доданків закон розподілу їх суми необмежено наближається до нормального з математичним сподіванням і дисперсією, які дорівнюють сумам відповідних числових характеристик доданків.

Центральна гранична теорема пояснює велике поширення нормального закону розподілу і є теоретичною основою застосування нормального розподілу для багатьох практичних задач: *за широких припущень сума великого (але скінченного) числа незалежних випадкових величин розподілена згідно із законом, близьким до нормального.* Наприклад, на відлагодженому виробництві якість продукції змінюється за нормальним законом внаслідок того, що виробнича похибка є результатом сумарної дії великого числа випадкових величин. Окремим випадком центральної граничної теореми є інтегральна формула Лапласа.

Центральна гранична теорема пояснює достатньо широке застосування нормального закону розподілу: *якщо випадкова величина формується під впливом багатьох незалежних факторів, кожен з яких здійснює на неї незначний вплив, то розподіл цієї величини близький до нормального.* Отже, важливими є вміння перевірки відповідності вибіркових даних нормальному розподілу.

Порядок дій при перевірці статистичних гіпотез

Для перевірки правильності основної статистичної гіпотези H_0 необхідно:

1) визначити гіпотезу H_1, альтернативну до гіпотези H_0;

2) обрати статистичну характеристику перевірки;

3) визначити допустиму ймовірність похибки першого роду, тобто рівень значущості α;

4) знайти за відповідною таблицею критичну область (критичну точку) для обраної статистичної характеристики.

До критичної області належать такі значення статистичної характеристики, при яких гіпотеза H_0 відхиляється на користь альтернативної гіпотези H_1.

12.2. Критерій згоди χ^2 про вигляд нормального розподілу

Критерій згоди χ^2 можна використовувати для будь-яких розподілів. Розглянемо його для перевірки гіпотези H_0: генеральна сукупність має нормальний закон розподілу $N(a, \sigma^2)$, де $a = \bar{x}, \sigma^2 = s^2$.

Розбиваємо числову вісь на r інтервалів, що не перетинаються: $h_1, h_2, \ldots, h_r$; n_i – кількість елементів вибірки, що потрапили в інтервал h_i $(i = 1, 2, \ldots, r)$. Позначимо через $p_i = P(X \in h_i)$; тоді np_i –теоретичні частоти. За формулою обчислення ймовірності

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right) \tag{12.1}$$

для нормального розподілу одержимо при $a = \bar{x}, \sigma = s, h_i = (x_i, x_{i+1})$

$$p_i = P\left(x_i < X < x_{i+1}\right) = \Phi\left(\frac{x_{i+1} - \bar{x}}{s}\right) - \Phi\left(\frac{x_i - \bar{x}}{s}\right).$$

За критерій перевірки гіпотези H_0 беруть величину

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i}. \tag{12.2}$$

Число k степенів вільності дорівнює $k = r - l - 1$, де l – кількість параметрів розподілу. Для нормального розподілу $k = r - 3$.

За умови справедливості нульової гіпотези за даним рівнем значущості α і кількістю степенів вільності k за таблицею розподілу χ^2 знаходимо критичну точку $\chi^2_{k;\alpha}$ з рівності $P(\chi^2 > \chi^2_{k;\alpha}) = \alpha$. Якщо $\chi^2 < \chi^2_{k;\alpha}$, то гіпотеза H_0 приймається, якщо $\chi^2 \geq \chi^2_{k;\alpha}$, то гіпотеза H_0 відхиляється. У випадку, коли гіпотеза відкидається, необхідно, по-перше, збільшити об'єм вибірки (кількість іспитів) і провести нову перевірку, якщо й це не допоможе, то треба знайти інший вираз для закону розподілу.

Висновок. Для того, щоб при заданому рівні значущості α перевірити гіпотезу про нормальний розподіл генеральної сукупності, треба:

1. Обчислити вибіркову середню $\bar{x}$ і вибіркове середнє квадратичне відхилення σ.

2. Обчислити теоретичні частоти.

Якщо ознака X генеральної сукупності має неперервний розподіл імовірностей, то теоретичні частоти обчислюються за формулою

$$n_i' = np_i, \qquad (12.3)$$

де n – обсяг вибірки, а p_i – імовірність того, що випадкова величина X потрапить в i-й частковий інтервал. Вона обчислюється в загальному випадку за формулами того закону розподілу, який припускаємо на основі обробки статистичного розподілу вибірки.

Якщо є підстави для припущення, що ознака генеральної сукупності X має нормальний закон розподілу, то теоретичні частоти можна обчислювати за формулами:

$$n_i' = \frac{nh}{\sigma}\cdot\varphi(u_i) = \frac{nh}{\sigma}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{(x_i-\bar{x}_B)^2}{2\sigma^2}}, \qquad (12.4)$$

де n – обсяг вибірки; h – довжина *часткового* інтервалу; $\bar{x}_{\text{В}}$ – вибіркова середня величина; σ – вибіркове середнє квадратичне відхилення; $\varphi(u_i)$ – щільність імовірностей для загального нормального закону розподілу або

$$n_i' = n\cdot\left(\Phi\left(\frac{x_{i+1}-\bar{x}_{\text{В}}}{\sigma}\right) - \Phi\left(\frac{x_i-\bar{x}_{\text{В}}}{\sigma}\right)\right), \qquad (12.5)$$

де $\Phi\left(\frac{x_{i+1}-\bar{x}_{\text{В}}}{\sigma}\right)$, $\Phi\left(\frac{x_i-\bar{x}_{\text{В}}}{\sigma}\right)$ – функції Лапласа.

3. Порівняти емпіричні і теоретичні частоти за допомогою критерію Пірсона. Для цього обчислити значення критерію за формулою (8.2).

4. За таблицею критичних точок розподілу χ^2 за заданим рівнем значущості α та числу степенів вільності знаходять критичну точку $\chi^2_{k;\alpha}$ та роблять відповідні висновки стосовно зробленої гіпотези.

12.3. Приклади перевірки статистичних гіпотез

Приклад 1. За заданим інтервальним статистичним розподілом випадкової величини X − маса новонароджених дітей (табл. 12.1),

Таблиця 12.1

h=0,5	1−1,5	1,5−2	2−2,5	2,5−3	3−3,5	3,5−4	4−4,5
n_i	10	20	50	35	28	15	12

при рівні значущості $\alpha = 0,01$ перевірити правильність H_0 про нормальний закон розподілу ознаки X − маси новонароджених дітей.

Розв'язання. Для визначення теоретичних частот $n_i' = np_i$ необхідно обчислити $\overline{x}_B$, σ.

Дискретний статистичний розподіл буде таким:

x_i	1,25	1,75	2,25	2,75	3,25	3,75	4,25
n_i	10	20	50	35	28	15	12

$$n = \sum n_i = 170.$$

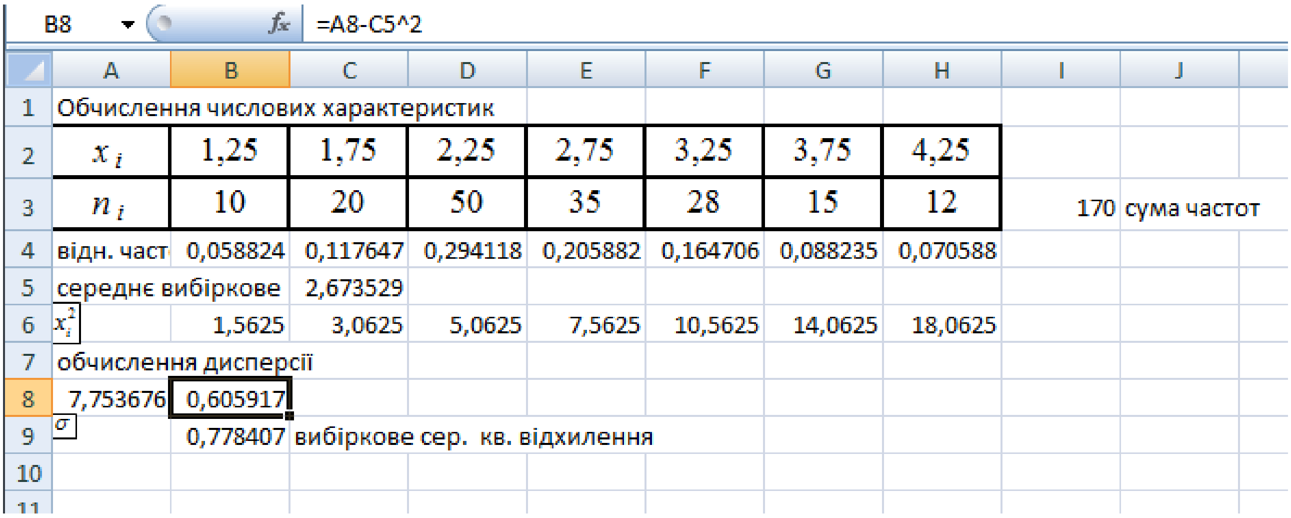
$$\overline{x}_{\text{В}} = \frac{\sum x_i n_i}{n} = = \frac{12,5 + 35 + 112,5 + 96,25 + 91 + 56,25 + 51,0}{170} = \frac{454,5}{170} = 2,67;$$

$$\frac{\sum x_i^2 n_i}{n} = \frac{1,25^2 \cdot 10 + 1,75^2 \cdot 20 + 2,25^2 \cdot 50 + 2,75^2 \cdot 35 + 3,25^2 \cdot 28 +}{170}$$

$$\frac{+3,75^2 \cdot 15 + 4,25^2 \cdot 12}{170} = \frac{1318,125}{170} = 7,75;$$

$$s^2 = \frac{\sum x_i^2 n_i}{n} - (\overline{x}_{\text{В}})^2 = 7,75 - (2,67)^2 = 7,75 - 7,1289 = 0,6211;$$

$$\sigma = \sqrt{s^2} = \sqrt{0,6211} \approx 0,79.$$



B8 =A8-C5^2

	A	B	C	D	E	F	G	H	I	J
1	Обчислення числових характеристик									
2	x_i	1,25	1,75	2,25	2,75	3,25	3,75	4,25		
3	n_i	10	20	50	35	28	15	12	170	сума частот
4	відн. част	0,058824	0,117647	0,294118	0,205882	0,164706	0,088235	0,070588		
5	середнє вибіркове		2,673529							
6	x_i^2	1,5625	3,0625	5,0625	7,5625	10,5625	14,0625	18,0625		
7	обчислення дисперсії									
8	7,753676	0,605917								
9	σ	0,778407	вибіркове сер. кв. відхилення							
10										
11										

Рис. 12.1. Числові характеристики.

Обчислення теоретичних частот за формулою (12.5) подано в таблиці (12.2).

Таблиця 12.2

x_i	x_{i+1}	n_i	$z_i = \frac{x_i - \bar{x}_B}{\sigma_B}$	$z_{i+1} = \frac{x_{i+1} - \bar{x}_B}{\sigma_B}$	$\Phi(z_i)$	$\Phi(z_{i+1})$	$n_i' = n(\Phi(z_{i+1}) - \Phi(z_i))$
1	1,5	10	– 2,11	– 1,48	– 0,4821	– 0,4306	9
1,5	2	20	– 1,48	– 0,85	– 0,4306	– 0,3023	22
2	2,5	50	– 0,85	– 0,22	– 0,3023	0,0871	37
2,5	3	35	– 0,22	0,42	– 0,0871	0,1628	43
3	3,5	28	0,42	1,05	0,1628	0,3531	32
3,5	4	15	1,05	1,68	0,3531	0,4535	17
4	4,5	12	1,68	2,32	0,4535	0,4898	6

Обчислення спостережуваного значення статистичного критерію χ^2 дається нижче на рис. 12.2, де теоретичні частоти обчислені із використанням локальної формули Лапласа та, відповідно, використанням для функції **НОРМРАСП** логічного значення 0. Також критерій задано спрощеною формулою через відносні частоти.

C10 f_x =J9*170

	A	B	C	D	E	F	G	H	I	J
1		Дискретний статистичний ряд								
2	Середина інтервалу		1,25	1,75	2,25	2,75	3,25	3,75	4,25	Сума
3	Частота		10	20	50	35	28	15	12	170
4	Відносна частота		0,058824	0,117647	0,294118	0,205882	0,164706	0,088235	0,070588	1
5	Середнє вибіркове		2,6735	2,6735	2,6735	2,6735	2,6735	2,6735	2,6735	
6	Виб. кв. відхилення		0,7784	0,7784	0,7784	0,7784	0,7784	0,7784	0,7784	
7	Лок. ф. Лапласа		0,096271	0,253547	0,442008	0,510047	0,389581	0,196967	0,065917	
8	Теорет. частоти		0,048135	0,126774	0,221004	0,255023	0,19479	0,098483	0,032958	0,977169
9	(Pi-Wi)^2/Pi		0,002373	0,000657	0,024188	0,009469	0,004646	0,001066	0,042963	0,085363
10	X^2		14,51173							
11	Хи2обр		13,2767							
12	Гіпотеза не приймається									

Рис. 12.2. Обчислення для статистичного критерію.

Отже, маємо

$$\chi^2 = \sum \frac{(n_i - np_i)^2}{np_i} = 14{,}51.$$

За таблицею знаходимо значення

$$\chi^2_{\text{кр}}(\alpha = 0{,}01;\ k = 7 - 2 - 1 = 4) = \chi^2_{\text{кр}}(0{,}01;\ 4) = 13{,}3.$$

Правобічна критична область показана на рис. 12.3.

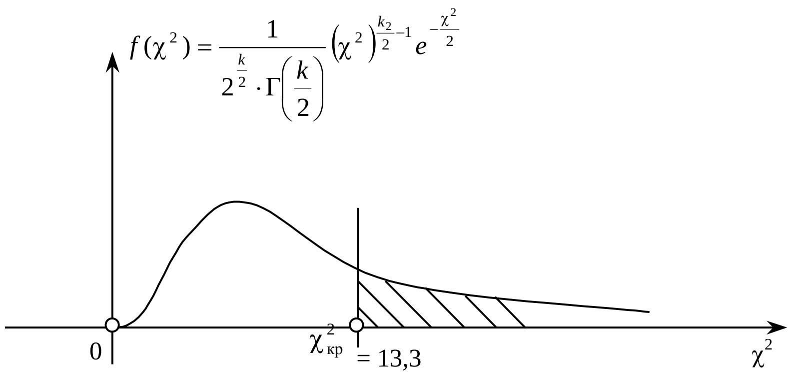


Рис. 12.3. Зображення критичної області та точки.

Висновок. Оскільки $\chi^2 \in [0;\ 13{,}3]$, то не маємо підстав для прийняття H_0 про нормальний закон розподілу ознаки генеральної сукупності X.

Приклад 2. Вимірювання зросту юнаків віком 17 років (табл. 12.4) дав такі результати:

Таблиця 12.4

$h = 4$	154–158	158–162	162–166	166–170	170–174	174–178	178–182	182–186
n_i	8	14	20	32	12	8	4	2

Визначити гіпотетично, який закон розподілу має ознака $X-$ зріст юнака. При рівні значущості $\alpha = 0{,}01$ перевірити правильність висунутої нульової гіпотези.

Розв'язання. Для заданого статистичного розподілу побудуємо гістограму частот (рис. 12.4).

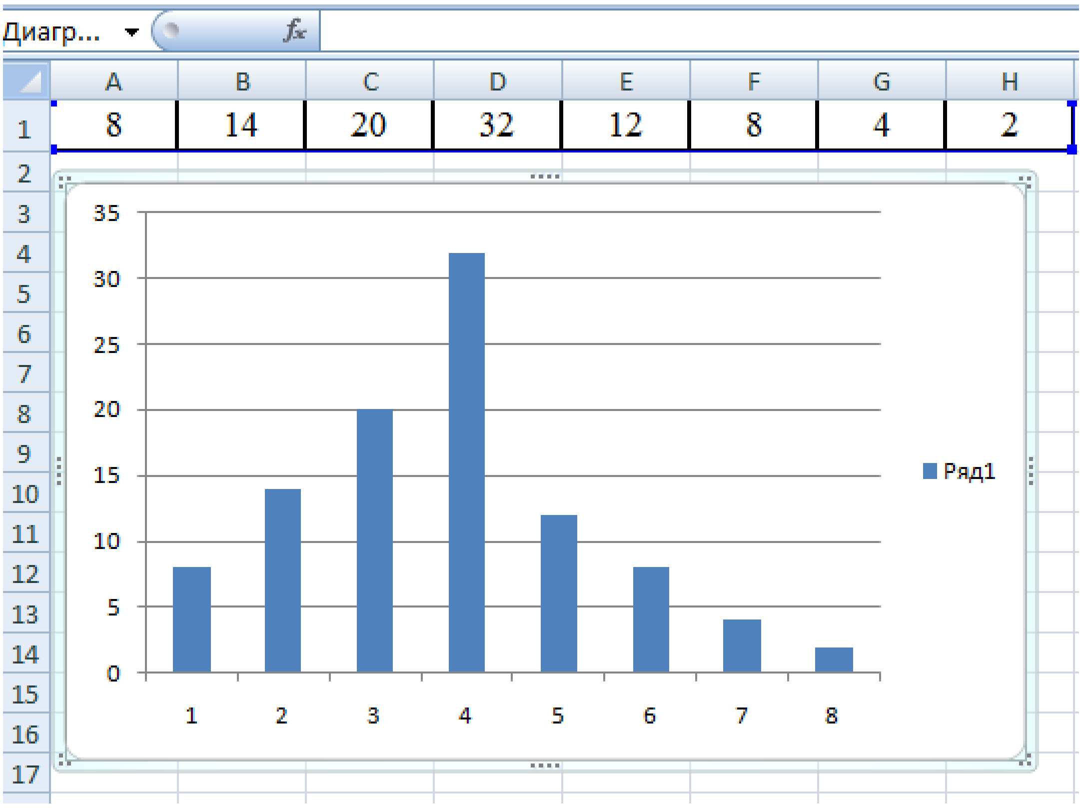


Рис. 12.4. Гістограма частот.

За формою гістограми частот можемо припустити, що ознака X має нормальний закон розподілу. Отже, висуваємо нульову гіпотезу H_0: ознака X має нормальний закон розподілу ймовірностей. Для перевірки справедливості H_0 використаємо критерій узгодженості Пірсона.

Отже, необхідно обчислити теоретичні частоти, а для цього знайдемо значення $\overline{x}_{\text{В}}$, σ, побудувавши дискретний розподіл за заданим інтервальним, а саме:

x_i	156	160	164	168	172	176	180	184
n_i	8	14	20	32	12	8	4	2

$$\overline{x}_B = \frac{\sum x_i n_i}{n} = \frac{156\cdot 8 + 160\cdot 14 + 164\cdot 20 + 168\cdot 32 +}{100}$$

$$\frac{+172\cdot 12 + 176\cdot 8 + 180\cdot 4 + 184\cdot 2}{100} = \frac{16704}{100} = 167,04;$$

$$\frac{\sum x_i^2 n_i}{n} = \frac{156^2\cdot 8 + 160^2\cdot 14 + 164^2\cdot 20 + 168^2\cdot 32 + 172^2\cdot 12 + 176^2\cdot 8 +}{100}$$

$$\frac{+180^2\cdot 4 + 184^2\cdot 2}{100} = \frac{2794304}{100} = 27943,04;$$

$$s^2 = \frac{\sum x_i^2 n_i}{n} - (\overline{x}_{\text{В}})^2 = 27943,04 - (167,04)^2 =$$

$$= 27943,04 - 27902,3616 = 40,68;$$

$$\sigma = \sqrt{s^2} = \sqrt{40,68} \approx 6,38.$$

Обчислення теоретичних частот наведено в табл. 12.5:

Таблиця 12.5

x_i	x_{i+1}	n_i	$z_i = \frac{x_i - \overline{x}_{\text{В}}}{\sigma_{\text{В}}}$	$z_{i+1} = \frac{x_{i+1} - \overline{x}_{\text{В}}}{\sigma_{\text{В}}}$	$\Phi(z_i)$	$\Phi(z_{i+1})$	$n_i' = n(\Phi(z_{i+1}) - \Phi(z_i))$
154	158	8	– 2,04	– 1,42	– 0,4793	– 0,4222	6
158	162	14	– 1,42	– 0,79	– 0,4222	– 0,2852	14
162	166	20	– 0,79	– 0,16	– 0,2852	– 0,0636	22
166	170	32	– 0,16	0,464	– 0,0636	0,1772	24
170	174	12	0,464	1,09	0,1772	0,3621	19
174	178	8	1,09	1,72	0,3621	0,4573	10
178	182	4	1,72	2,34	0,4573	0,4904	3
182	186	2	2,34	2,97	0,4904	0,4986	1

Обчислення спостережуваного значення χ^2 наведено нижче в таблиці:

n_i	np_i	$n_i - np_i$	$(n_i - np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$
8	6	2	4	0,667
14	14	0	0	0
20	22	– 2	4	0,182
32	24	8	64	2,667
12	19	– 7	49	2,579
8	10	– 2	4	0,4
4	3	1	1	0,333
2	1	1	1	1

$$\chi^2 = \sum_{i=1}^{8} \frac{(n_i - np_i)^2}{np_i} = 7{,}828.$$

За таблицею (додаток) знаходимо значення

$$\chi^2_{\alpha,k}(\alpha = 0{,}01;\ k = 8 - 2 - 1) = \chi^2_{\alpha,k}(0{,}01;\ 5) = 15{,}1.$$

Критична область даного прикладу зображена на рис. 12.3.

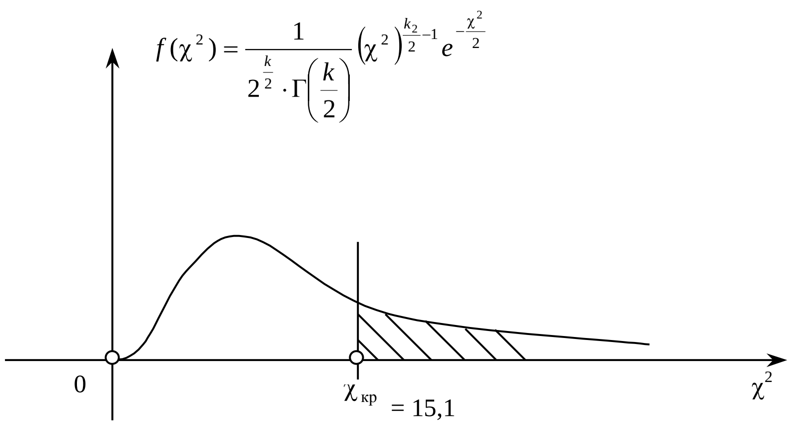


Рис. 12.5. Критична область.

Висновок. Оскільки $\chi^2 \in [0;\ 15{,}1]$, немає підстав для відхилення нульової гіпотези H_0 про нормальний закон розподілу ймовірностей ознаки X.

Приклад 3. За наданим інтервальним статистичним рядом (табл. 12.6) знайти закон розподілу випадкової величини X.

Таблиця 12.6

$[a_i; a_{i+1})$	[–2;–1,2)	[–1,2;–0,4)	[–0,4;0,4)	[0,4;1,2)	[1,2;2)
n_i	6	11	21	7	5

Розв'язання. Для визначення виду закону розподілу побудуємо гістограму за даними таблиці 12.6 (рис. 12.6). За видом гістограми висуваємо гіпотезу про нормальний закон розподілу даної випадкової величини:

H_0 – випадкова величина X розподілена за нормальним законом;

H_1 – випадкова величина X не розподілена за нормальним законом.

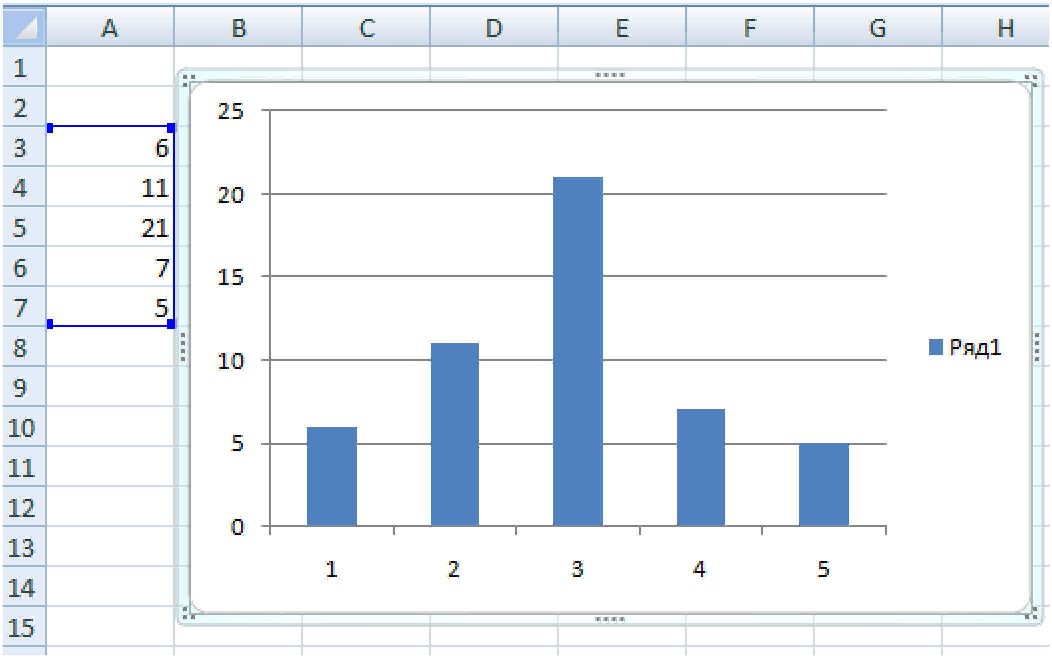


Рис. 12.6. Гістограма за даними таблиці 12.6.

Щільність розподілу випадкової величини, розподіленої за нормальним законом має вид $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$, де a і σ – параметри розподілу. Знайдемо означені параметри, враховуючи, що $\overline{x} = a$; $S^2 = \sigma^2$. Розрахунки оформимо у вигляді таблиці (табл. 12.7).

Таблиця 12.7

$[a_i;a_{i+1})$	$[-2;-1{,}2)$	$[-1{,}2;-0{,}4)$	$[-0{,}4;0{,}4)$	$[0{,}4;1{,}2)$	$[1{,}2;2)$
n_i	6	11	21	7	5
x_i	-1,6	-0,8	0	0,8	1,6
$x_i n_i$	-9,6	-8,8	0	5,6	8
$(x_i-\overline{x})^2 n_i$	13,572	5,452	0,194	5,620	14,382

Знайдемо вибіркове середнє, вибіркову дисперсію і вибіркове середнє квадратичне відхилення за відповідними формулами:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \frac{1}{50}(-1{,}6\cdot 6 - 0{,}8\cdot 11 + 0\cdot 21 + 0{,}8\cdot 7 + 1{,}6\cdot 5) = -0{,}096;$$

$$s^2 = \frac{1}{n}\sum_{i=1}(x_i - \overline{x})^2 n_i = \frac{1}{50}(13,572 + 5,452 + 0,194 + 5,620 + 14,382) = 0,7844$$

$$\sigma = \sqrt{s^2} \approx 0,886\,.$$

Отже, параметрами теоретичного закону розподілу є величини: $\overline{x} = a = -0,096; \quad s = \sigma = 0,886.$

Для знаходження значення критерію χ^2 розрахуємо теоретичні частоти n_i'. Для зручності обчислень побудуємо табл. 12.8.

Таблиця 12.8

$[a_i; a_{i+1})$	$[-2;-1,2)$	$[-1,2;-0,4)$	$[-0,4;0,4)$	$[0,4;1,2)$	$[1,2;2)$
n_i	6	11	21	7	5
x_i	-1,6	0,8	0	0,8	1,6
$x_i n_i$	-9,6	8,8	0	5,6	8
$\left(x_i - \overline{x}\right)^2 n_i$	13,572	5,452	0,194	5,620	14,382
$\frac{a_i - \overline{x}}{S}$	-2,1498	-1,2465	-0,3433	0,56	1,4633
$\Phi(\frac{a_i - \overline{x}}{S})$	-0,958	-0,785	-0,266	0,425	0,856
p_i	0,0856	0,2595	0,3455	0,2155	0,063
$n_i' = np_i$	4,325	12,975	17,275	10,775	3,15
$\frac{\left(n_i - n_i'\right)^2}{n_i'}$	0,649	0,301	0,803	1,323	1,087

За формулою критерію маємо

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(n_i - n_i'\right)^2}{n_i'} \approx 4,16\,.$$

Знайдемо критичне значення $\chi^2_{\alpha,l}$, враховуючи, що $l = k - r - 1$=5-2-1=2. Рівень значущості α оберемо рівним 0,1. За допомогою Excel знаходимо **ХИ2ОБР**(0,1; 2)=4,6.

Отже, оскільки $\chi^2 < \chi^2_{\alpha,l}$, гіпотеза H_0 про нормальний розподіл приймається, гіпотеза H_1 відкидається.

Дані вимірювань на одному із поліграфічних виробництв та їх статистична обробка

Приклад 4. Дані вимірювань з перших трьох текстових накладів, друкування яких здійснювалось в різні дні декілька разів по 100 відбитків, було використано для перевірки підлягання отриманих значень у вибірках нормальному закону розподілу за критерієм Пірсона.

Розглядається приклад перевірки нормальному розподілу вибірки з вимірювання значень, отриманих на ділянці дії фарбового ножа № 3 для тесту, де вимірювання було проведено на відбитку № 90, для якого значення можна вважати стабілізованими. В табл. 12.9 наведено дані поведених вимірювань, де $\Delta E -$колірні відмінності виміряного кольору від еталонного значення, D – оптична густина фарби, 80, 40 – розтискування растрової точки (80+k, вказано k) при стабільній подачі фарби та зволожувального розчину.

Таблиця 12.9

Результати вимірювання відбитка з тестовим зображенням

№ вимірювання	ΔE	D	ΔSвідн 80%	ΔSвідн 40%
1	*2*	*3*	*4*	*5*
1	16,24	0,89	5	6
2	16,11	0,89	5	7
3	16,21	0,86	5	7
4	16,25	0,89	4	7
5	16,19	0,93	4	7
6	16,18	0,9	5	7
7	16,31	0,91	5	7
8	16,17	0,97	5	7
9	16,2	0,91	5	7
10	16,12	0,9	5	7
11	16,22	0,92	4	7
12	16,18	0,94	5	7
13	16,23	0,88	5	7

14	16,09	0,89	5	6
15	16,25	0,93	5	7
16	16,17	0,89	4	7
17	16,21	0,87	4	7
18	16,25	0,89	5	7
19	16,19	0,92	4	7
20	16,16	0,93	5	7
21	16,2	0,89	5	7
22	16,19	0,95	5	6
23	16,28	0,93	5	7
24	16,13	0,91	4	7
25	16,17	0,9	5	7
26	16,2	0,91	5	7
27	16,28	0,93	5	7
28	16,15	0,91	5	7
29	16,17	0,9	4	7
30	16,2	0,91	5	7
31	16,12	0,9	4	7
32	16,22	0,92	5	7
33	16,18	0,94	4	7
34	16,29	0,88	5	7
35	16,09	0,89	5	7
36	16,25	0,93	5	6
37	16,11	0,89	5	7
38	16,21	0,87	5	7
39	16,09	0,89	5	7
40	16,24	0,9	5	7
41	16,11	0,89	4	7
42	16,31	0,87	5	7
43	16,25	0,89	4	7
44	16,19	0,92	5	7
45	16,22	0,88	5	6
<X>	16,2	0,9	4,73	6,89
σ	0,0596	0,0234	0,4539	0,2939

У наведеній вище таблиці одразу подано пораховані значення середнього арифметичного значення вибірки та середньоквадратичного відхилення.

Згідно з методикою розрахунку *за критерієм Пірсона*, визначаємо інтервали для кожної з вибірок (для ΔE, D, ΔS відн 80 % та ΔS відн 40 %) та кількість значень, що у них потрапляють. Розраховуємо ймовірності потрапляння випадкового значення у задані інтервали відповідно до очікузваного середнього значення та середньоквадратичного відхилення, використовуючи функцію пакету MS Excel **НОРМРАСП**(x; $x_{ср}$; σ; 1).

Після цього, коригуючи отримані значення ймовірностей (за необхідної для критерію Пірсона умови, що сума ймовірностей обов’язково повинна бути рівною 1), розраховуємо величини χ^2-статистики.

Розраховане значення порівнюємо з критичним, яке також попередньо розрахуємо за формулою в MS Excel **ХИ2ОБР**($\alpha;k$). Для кожної з вибірок кількість степенів вільності буде дещо відрізнятись, наприклад, у даному випадку для вибірки за ΔE $k = 2$, для вибірки за D $k = 9$, а для вибірок за ΔS відн 80 % за ΔS відн 40 % $k = 1$.

У всіх випадках розрахунки показують, що оскільки значення величин χ^2 є меншими за значення $\chi^2_{крит}$ (5,723<5.99146; 12,39<16,9189; 2,837<3,8414; 1,007<3,8414), то гіпотеза про відповідність вибірок нормальному закону розподілу приймається.

Зауваження. Обчислення проведено за допомогою функцій пакету MS Excel.

В табл. Б.1, Б.2, Б.3 та Б.4 показані розрахунки відповідності нормальному розподілу вибірки даних за $\Delta \boldsymbol{E}$, за D, за ΔS відн 80 % та ΔS відн 40 %, відповідно.

Зауваження. В таблицях нижче ряд даних є наближеними, тому можливі, на перший погляд, неточності.

Таблиця Б.1

Розрахунок відповідності нормальному розподілу вибірки за ΔE

Знач. ΔE	К-ть елем. n_i	$x_{i\min}$	$x_{i\max}$	$P(x_{i\min})$	$P(x_{i\max})$	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	2	3	4	5	6	7	8	9	10
16,10	8,00	16,075	16,125	0,022	0,120	0,098	4,6	3,4	2,585
16,15	7,00	16,125	16,175	0,120	0,368	0,248	11,6	-4,6	1,820
16,20	17,00	16,175	16,225	0,368	0,692	0,324	15,1	1,9	0,228
16,25	8,00	16,225	16,275	0,692	0,910	0,218	10,2	-2,2	0,467
16,30	5,00	16,275	16,325	0,910	0,985	0,075	3,5	1,5	0,623

n= 45 Σ=0,963 χ^2= 5,723

5,9914645

n^*= 47 $\chi^2_{крит}$= 4

k= 2,00

Таблиця Б.2

Розрахунок відповідності нормальному розподілу вибірки за *D*

Знач. ΔE	К-ть елем.	$x_{i\min}$	$x_{i\max}$	$P(x_{i\min})$	$P(x_{i\max})$	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	2	3	4	5	6	7	8	9	10
0,86	1,00	0,855	0,865	0,017	0,045	0,028	1,3	-0,3	0,067
0,87	3,00	0,865	0,875	0,045	0,103	0,058	2,6	0,4	0,050
0,88	3,00	0,875	0,885	0,103	0,201	0,098	4,5	-1,5	0,494
0,89	12,00	0,885	0,895	0,201	0,340	0,139	6,4	5,6	4,936
0,9	6,00	0,895	0,905	0,340	0,506	0,166	7,6	-1,6	0,335
0,91	6,00	0,905	0,915	0,506	0,607	0,165	7,5	-1,5	0,318
0,92	4,00	0,915	0,925	0,670	0,807	0,137	6,3	-2,3	0,823
0,93	6,00	0,925	0,935	0,807	0,902	0,095	4,4	1,6	0,620
0,94	2,00	0,935	0,945	0,902	0,957	0,027	2,5	-0.5	0,111
0,95	1,00	0,945	0,955	0,957	0,984	0,011	1,2	-0,2	0,042
0,96	0,00	0,955	0,965	0,984	0,995	0,004	0,5	-0,5	0,498
0,97	1,00	0,965	0,975	0,995	0,999	0,004	0,2	0,8	4,098

n=45 Σ=0,982 χ^2= 12,390

n^*=46 $\chi^2_{крит}$= 16,91897762

k=9,00

Таблиця Б.3

Розрахунок відповідності нормальному розподілу вибірки за ΔS відн 80 %

Знач. ΔE	К-ть елем.	$x_{i\min}$	$x_{i\max}$	$P(x_{i\min})$	$P(x_{i\max})$	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	2	3	4	5	6	7	8	9	10
3,00	0,00	2,5	3,500	0,000	0,003	0,003	0,1	-0,1	0,148
4,00	12,00	3,500	4,500	0,003	0,304	0,300	13,5	-1,5	0,170
5,00	33,00	4,500	5,500	0,304	0,954	0,651	29,3	3,7	0,470
6,00	0,00	5,500	6,500	0,954	1,000	0,046	2,0	-2,0	2,049

n=45 Σ=1,000 χ^2=2,837

n*=45 $\chi^2_{\text{крит}}$=3,841459149

k=1,00

Таблиця Б.4

Розрахунок відповідності нормальному розподілу вибірки за ΔS відн 40 %

Знач. ΔE	К-ть елем.	$x_{i\min}$	$x_{i\max}$	$P(x_{i\min})$	$P(x_{i\max})$	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	2	3	4	5	6	7	8	9	10
5	0,00	4,5	5,500	0,000	0,000	0,000	0,0	0,0	0,000
6	5,00	5,500	6,500	0,000	0,093	0,093	4,2	0,8	0,161
7	40,00	6,500	7,500	0,093	0,981	0,888	40,0	0,0	0,000
8	0,00	7,500	8,500	0,981	1,000	0,019	0,8	-0,8	0,846

n=45 Σ=1,000 χ^2=1,007

n*=45 $\chi^2_{\text{крит}}$=3,841459149

k=1,00

На основі отриманих даних зроблено висновок про те, що дійсно дані експериментальних вимірювань підлягають нормальному розподілу.

Виконання лабораторної роботи

Завдання до теми

Завдання 1.

Користуючись теоретичними знаннями та засобами програмного забезпечення Excel:

1) Побудувати таблицю статистичного розподілу.

2) Побудувати гістограму.

3) Вирівняти дослідні дані за допомогою нормального закону розподілу.

4) Перевірити узгодженість між емпіричними та теоретичними даними за допомогою критерія Пірсона та Романовського.

Варіант 1

2,37	-0,94	0,58	-0,38	-0,72	0,76	1,55	-0,53	1,41	1,03
-0,06	0,29	0,06	1,40	0,32	1,65	0,61	2,72	-1,03	0,48
0,61	2,05	1,12	-0,94	0,46	1,18	0,93	0,48	0,34	0,48
-0,50	1,58	1,39	2,30	0,83	1,19	-0,48	0,93	1,07	0,84
1,34	1,14	1,48	3,08	2,73	-1,14	-0,48	1,63	1,31	0,08
-0,10	1,52	2,41	0,16	0,31	0,60	2,75	-0,01	0,33	-0,13
0,51	0,81	-0,23	1,26	1,89	0,89	1,93	1,02	2,26	0,31
1,97	2,48	1,88	1,96	1,67	0,08	0,66	0,98	1,91	-0,11
0,67	1,18	2,30	3,15	1,24	0,81	0,73	0,65	0,79	0,63
1,45	1,31	1,42	1,23	1,84	1,99	2,05	2,50	2,55	2,90

Варіант 2

1,60	0,07	0,90	0,39	0,19	0,76	1,29	0,14	1,01	0,27
1,15	0,70	0,56	1,05	0,65	1,37	0,83	1,76	-0,04	0,72
0,96	1,70	1,13	0,19	0,70	1,06	0,96	0,71	0,68	0,66
0,20	1,29	1,01	1,55	0,80	1,19	0,35	0,77	1,15	0,85
1,19	1,11	1,16	2,02	1,88	-0,02	0,50	1,49	1,20	0,54
0,31	1,42	1,54	0,74	0,59	0,83	1,91	0,38	0,56	0,37
0,77	0,80	0,28	1,23	1,26	0,76	1,40	1,15	1,53	0,62
1,36	1,62	1,36	1,40	1,43	0,64	0,84	0,78	1,47	0,47
0,93	1,01	1,54	2,15	1,07	0,81	0,82	0,78	0,85	0,97
1,06	1,16	1,04	1,24	1,35	1,44	1,66	1,69	1,97	1,79

Варіант 3

3,33	0,18	1,52	0,99	0,38	1,74	2,78	0,23	2,20	0,67
2,19	1,46	1,03	2,09	1,38	2,83	1,81	3,53	-0,47	1,25
1,61	3,36	2,41	0,21	1,27	2,29	1,73	1,01	1,49	1,21
0,22	2,67	2,28	3,19	1,62	2,28	0,98	1,88	2,38	1,54
2,18	2,41	2,41	4,29	3,58	-0,27	0,83	2,88	2,21	1,16
0,86	2,85	3,07	1,36	1,33	1,95	3,96	0,67	1,02	0,85
1,96	1,61	0,64	2,46	2,56	1,90	2,88	2,45	3,42	1,23
2,62	3,05	2,64	2,59	2,60	1,28	1,90	1,97	2,85	0,84
1,94	2,28	3,09	4,17	2,02	1,82	1,84	1,97	1,66	1,86

2.40	2,01	2,41	2,47	2,99	2,67	3,47	3,38	3,94	3,54

Варіант 4

2,57	1,10	1,82	1,42	1,14	1,94	2,41	1,07	2,07	1,27
2,25	1,54	1,52	2,16	1,66	2,48	1,89	2,77	0,75	1,71
1,78	2,63	2,01	1,24	1,61	2,19	1,79	1,68	1,65	1,61
1,24	2,40	2,04	2,55	1,82	2,00	1,33	1,84	2,05	1,78
2,23	2,23	2,22	3,18	2,80	0,86	1,42	1,64	2,32	2,55
1,38	2,39	2,70	1,59	1,52	1,77	2,87	1,28	1,57	1,33
1,89	1,78	1,26	2,13	2,35	1,99	2,32	2,06	2,51	1,72
2,40	2,65	2,34	2,26	2,29	1,69	1,84	1,93	2,37	1,31
1,97	2,06	2,69	3,23	2,10	1,76	1,83	1,82	1,91	1,83
2,01	2,13	2,03	2,06	2,49	2,37	2,55	2,65	2,89	2,78

Варіант 5

4,24	1,32	2,80	1,95	1,47	2,55	3,90	1,18	3,36	1,79
3,04	2,11	2,30	3,08	2,32	3,87	2,61	4,90	0,93	2,29
2,64	4,26	3,01	1,06	2,13	3,39	2,60	2,07	2,46	2,22
1,08	3,79	3,17	4,18	2,51	3,39	1,90	2,63	3,37	2,82
3,28	3,04	3,37	5,32	4,73	0,62	1,99	3,97	3,27	2,02
1,65	3,72	4,26	2,00	2,25	2,55	4,78	1,63	2,07	1,92
2,84	2,56	1,75	3,31	3,56	2,67	3,54	3,16	4,06	2,45
3,99	4,43	3,83	3,69	3,56	2,05	2,86	2,61	3,69	1,65
2,96	3,04	4,37	5,23	3,40	2,54	2,54	2,59	2,60	2,77
3,20	3,06	3,14	3,32	3,94	3,94	4,43	4,23	4,98	4,64

Варіант 6

3,71	2,16	2,98	2,40	2,14	2,79	3,29	2,08	3,22	2,31
3,20	2,53	2,56	3,11	2,65	3,26	2,93	3,75	1,86	2,52
2,90	3,59	3,04	2,19	2,56	3,14	2,86	2,51	2,69	2,67
2,03	3,42	3,05	3,54	2,77	3,11	2,48	2,87	3,02	2,81
3,20	3,10	3,04	4,20	3,85	1,93	2,50	3,27	3,17	2,58
2,32	3,48	3,52	2,72	2,51	2,80	3,95	2,46	2,57	2,33
2,85	2,89	2,43	3,22	3,26	2,78	3,36	3,17	3,51	2,73
3,41	3,59	3,28	3,32	3,41	2,54	2,98	2,91	3,42	2,29
2,99	3,08	3,67	4,06	3,06	2,91	2,96	2,95	2,89	3,00
3,24	3,15	3,18	3,16	3,40	3,45	3,61	3,51	3,85	3,76

Варіант 7

5,24	2,29	3,84	2,61	2,01	3,66	4,71	2,36	4,11	2,92

4,44	3,38	3,18	4,08	3,43	4,56	3,83	5,57	1,61	3,14
3,61	5,48	4,34	2,27	3,24	4,37	3,84	3,35	3,16	3,01
2,33	4,97	4,05	5,07	3,55	4,31	2,95	3,53	4,08	3,65
4,43	4,29	4,11	6,07	5,88	1,57	2,82	4,98	4,01	3,34
2,58	4,64	5,19	3,06	3,40	3,66	5,82	2,90	3,30	2,61
3,64	3,97	2,90	4,46	4,89	3,62	4,51	4,45	5,45	3,08
4,65	5,25	4,82	4,85	4,98	3,08	3,80	3,62	4,59	2,86
3,96	4,37	5,41	6,40	4,03	3,92	3,77	3,53	3,54	3,66
4,40	4,50	4,02	4,34	4,99	4,95	5,24	5,37	5,46	5,75

Варіант 8

4,52	3,18	3,86	3,33	3,24	3,82	4,31	3,22	4,22	3,38
4,17	3,71	3,51	4,09	3,71	4,33	3,82	4,93	2,93	3,71
3,81	4,62	4,24	3,05	3,53	4,16	3,86	3,65	3,57	3,63
3,01	4,40	4,10	4,54	3,95	4,03	3,48	3,77	3,86	4,04
4,25	4,05	4,08	5,15	4,89	2,90	3,28	4,28	4,20	3,61
3,45	4,42	4,55	3,63	3,64	3,78	4,86	3,37	3,64	3,40
3,91	3,78	3,36	4,06	4,42	3,93	4,47	4,19	4,50	3,59
4,48	4,64	4,46	4,41	4,36	3,55	3,99	3,78	4,26	3,45
3,86	4,24	4,72	5,11	4,07	3,78	3,99	3,84	3,82	3,99
4,24	4,22	4,17	4,23	4,33	4,34	4,61	4,60	4,81	4,94

Варіант 9

6,25	3,44	4,61	3,67	3,14	4,52	5,88	3,18	5,07	3,94
5,09	4,20	4,38	5,07	4,32	5,55	4,53	6,73	2,92	4,13
4,55	6,42	5,03	3,47	4,01	5,11	4,66	4,14	4,02	4,01
3,37	5,84	5,24	6,05	4,57	5,03	3,66	4,85	5,48	4,72
5,26	5,33	5,45	7,38	6,74	2,60	3,83	5,79	5,21	4,28
3,51	6,01	6,29	4,41	4,15	4,73	6,81	3,68	4,47	4,02
4,66	4,81	3,92	5,15	5,87	3,98	4,98	5,109	6,48	4,07
5,74	6,26	5,56	5,54	5,98	4,22	4,94	4,92	5,97	3,67
4,81	4,45	6,31	7,19	5,25	4,56	4,79	4,86	4,88	4,56
5,02	5,29	5,02	5,01	5,65	5,66	6,11	6,49	6,72	6,86

Варіант 10

5,53	4,12	4,97	4,41	4,16	4,93	5,31	4,18	5,06	4,29
5,23	4,45	4,54	5,24	4,65	5,31	4,77	5,92	3,88	4,61
4,83	5,74	5,13	4,17	4,45	5,14	4,77	4,68	4,66	4,55
4,15	5,42	5,19	5,65	4,83	5,05	4,48	4,95	5.24	4,84
5,15	5,16	5,12	6,07	5,99	3,85	4,43	5,42	5,03	4,56
4,38	5,35	5,61	4,60	4,64	4,80	5,98	4,27	4,63	4,30

4,88	4,78	4,29	5,14	5,35	4,85	5,28	5,03	5,72	4,69
5,35	5,51	5,38	5,34	5,27	4,61	4,96	4,79	5,42	4,50
4,85	5,02	5,70	6,22	5,24	4,80	4,76	4,96	4,84	4,97
5,20	5,01	5,45	5,03	5,49	5,36	5,57	5,56	5,94	5,86

Варіант 11

7,17	4,34	5,87	4,65	4,12	5,81	6,85	4,29	6,47	4,69
6,29	5,14	5,07	6,20	5,27	6,86	5,66	7,98	3,59	5,15
5,76	7.16	6,24	4,13	5,11	6,04	5,86	5,04	5,10	5,17
4,31	6,58	6,49	7,48	5,88	6,12	4,56	5,81	6,19	5,93
6,09	6,23	6,38	8,37	7,56	3,90	4,69	6,75	6,05	5,20
4,65	6,96	7,19	5,35	5,23	5,94	7,55	4,72	5,41	4,89
5,86	5,94	6,67	6,07	6,58	5,62	6,74	6,25	7,14	5,11
6,80	7,19	4,61	6,67	6,90	5,12	5,74	5,77	6,57	4,93
5,88	6,41	4,34	8,05	6,18	5,68	5,59	5,69	5,79	4,67
6,07	6,13	6,34	6,05	6,84	6,65	7,05	7,06	7,97	7,88

Варіант 12

6,62	5,14	5,78	5,40	5,13	5,98	6,49	5,19	6,03	5,38
6,02	5,74	5,63	6,20	5,71	6,33	5,91	6,91	4,78	5,59
5,94	6,55	6,22	5,11	5,74	6,10	5,82	5,54	5,64	5,57
5,20	6,48	6,03	6,66	5,93	6,03	5,32	5,82	6,21	5,81
6,02	6,10	6,18	7,23	6,76	4,83	5,49	6,39	6,17	5,68
5,35	6,31	6,51	5,68	5,74	5,82	6,83	5,31	5,66	5,46
5,81	5,97	5,27	6,01	6,29	5,91	6,47	5,68	6,37	5,60
6,33	6,62	6,35	6,33	6,47	5,61	5,84	5,87	6,37	5,31
5,84	6,12	6,65	7,17	6,04	5,75	5,95	5,87	5,84	5,87
6,09	6,13	5,75	6,21	6,42	6,29	6,75	6,59	6,98	6,80

Варіант 13

9,58	8,09	9,00	8,31	8,01	8,81	9,25	8,05	9,24	8,45
9,07	8,53	8,64	9,22	8,71	9,40	8,86	9,88	7,79	8,65
8,93	9,60	9,07	8,01	8,70	9,11	8,79	8,63	8,68	8,74
8,18	9,37	9,16	9,68	8,84	9,12	8,33	8,76	9,00	8,88
9,07	9,19	9,18	10,03	9,93	7,96	8,45	9,30	9,22	8,54
8,30	9,34	9,74	8,68	8,74	8,99	9,91	8,31	8,66	8,35
8,87	8,78	8,33	9,17	9,29	9,00	9,47	9,11	9,37	8,60
9,38	9,68	9,35	9,26	9,47	8,66	8,84	8,89	9,46	8,50
8,87	9,05	9,65	10,20	9,04	8,96	8,95	8,87	8,78	8,82
9,16	9,19	9,11	9,18	9,42	9,41	9,75	9,69	9,83	9,80

Варіант 14

11,36	8,41	9,94	8,78	8,37	9,56	10,58	8,37	10,45	8,75
10,29	9,29	9,38	10,31	9,05	10,62	9,64	11,83	7,87	9,28
9,79	11,04	10,43	8,09	9,07	10,40	9,65	9,14	9,39	9,07
8,14	10,59	10,02	11,47	9,69	10,19	8,78	9,86	10,50	9,59
10,39	10,37	10,42	12,10	11,65	7,99	8,90	10,60	10,34	9,08
8,52	10,91	11,11	9,34	9,06	9,79	11,56	8,79	9,02	8,46
9,90	9,71	8,67	10,05	10,89	9,87	10,59	10,20	11,41	9,46
10,81	11,28	11,20	10,56	10,77	9,46	9,98	10,01	10,92	8,97
9,93	10,05	10,79	12,24	10,21	9,93	9,56	9,74	9,56	9,81
10,41	10,05	10,36	10,19	10,95	10,59	8,75	11,47	11,64	11,79

Варіант 15

14,24	12,29	13,84	12,61	12,01	13,66	14,71	12,36	14,11	12,92
13,44	11,38	13,18	14,08	13,43	14,56	13,83	15,57	11,61	13,14
12,61	12,48	14,34	12,27	13,24	14,37	13,84	13,35	13,16	13,01
11,33	14,97	14,05	15,07	13,55	14,31	12,95	13,53	14,08	13,65
13,43	13,29	14,11	16,07	15,88	11,57	12,82	14,98	14,01	13,34
11,58	13,64	15,19	13,06	13,40	13,66	15,82	12,90	13,30	12,61
12,64	12,97	12,90	14,46	14,89	13,62	14,51	14,45	15,45	13,08
13,65	14,25	14,82	14,85	14,98	13,08	13,80	13,62	14,59	12,86
12,96	13,37	15,41	16,40	14,03	13,92	13,77	13,53	13,54	13,66
13,40	13,50	14,02	14,34	14,99	14,95	15,24	15,37	15,46	15,75

Завдання 2.

За експериментальними даними вимірювань – дані в лабораторній роботі 11 (завдання 2), де $\Delta E -$колірні відмінності виміряного кольору від еталонного значення, D – оптична густина фарби, 80, 40 – розтискування растрової точки (80+k, вказано k) при стабільній подачі фарби та зволожувального розчину та при коригуванні подачі фарби автоматизованою системою контролю параметрів відбитка для одного з наведених показників висунути гіпотезу про нормальний закон розподілу даної випадкової величини та перевірити її за критерієм Пірсона.

Зауваження. Для виконання завдань 1), 2) використовувати набуті навики лабораторної роботи № 8). Номер варіанта відповідає номеру ділянки, варіанти 1 та 2 є, відповідно, варіантами 16 та 17.

Завдання 3. Дати відповіді на подані нижче теоретичні запитання, одне із них оформити письмово.

Теоретичні запитання до теми

1. Обґрунтувати практичне використання центральної граничної теореми.
2. Пояснити критерій згоди χ^2 про вигляд нормального розподілу.
3. Записати порядок дій перевірки гіпотези про нормальний розподіл генеральної сукупності.
4. Виписати формули обчислення теоретичних частот для нормального розподілу та пояснити використання відповідної вбудованої функції в Excel для їх знаходження.
5. Побудова правобічної критичної області та визначення її критичної точки.

Оформлення звіту та порядок захисту

Лабораторна робота виконується на аркушах А4, в ній стисло відображаються формули теоретичної частини, хід роботи та отримані результати. При захисті студент повинен розуміти зміст роботи, порівняти отримані результати проведених обчислень, а також знати відповіді на теоретичні запитання. Також студент здає електронний варіант проведених обчислень практичних задач.

Додаткове завдання

1. Оформити звіт про виконану роботу на аркушах А4 за допомогою Microsoft Word.
2. Зібрати статистичні дані (або використати готові) та опрацювати їх (за тематикою лабораторних робіт 8-12).

Таблиця 1

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

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2035	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0941	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0271	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0159	0155	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Таблиця 2

Таблиця значень функції Лапласа $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,00000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1,5	43319	43448	43574	43699	43872	43943	44062	44179	44295	44408
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48860	48899
2,3	48928	48966	48983	49010	49036	49061	49086	49110	49134	49158
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861
3,0	0,49865		3,1	49903	3,2	49931	3,3	49952	3,4	49966
3,5	49977		3,6	49984	3,7	49989	3,8	49993	3,9	49995
4,0	499968									
4,5	499997									
5,0	49999997									

Таблиця 3

Таблиця значень χ^2 в залежності від k і рівня значущості α

Кількість степенів вільності k	α 0,95	α 0,90	α 0,50	α 0,30	α 0,20	α 0,10	α 0,05	α 0,025	α 0,01
1	0,004	0,016	0,455	1,074	1,642	2,71	3,84	5,0	6,6
2	0,103	0,211	1,386	2,41	3,22	4,60	5,99	7,4	9,2
3	0,352	0,584	2,37	3,67	4,64	6,25	7,82	9,4	11,3
4	0,711	1,064	3,36	4,88	5,99	7,78	9,49	11,1	13,3
5	1,145	1,61	4,35	6,06	7,29	9,24	11,07	12,8	15,1
6	1,635	2,20	5,35	7,23	8,56	10,64	12,59	14,4	16,8
7	2,17	2,83	6,35	8,38	9,80	12,02	14,07	16,0	18,5
8	2,73	3,49	7,34	9,52	11,03	13,36	15,51	17,5	20,1
9	3,32	4,17	8,34	10,66	12,24	14,68	16,92	19,0	21,7
10	3,94	4,86	9,34	11,78	13,44	15,99	18,31	20,5	23,2
11	4,58	5,58	10,34	12,90	14,63	17,28	19,68	21,9	24,7
12	5,23	6,30	11,34	14,01	15,84	18,55	21,0	23,3	26,2
13	5,89	7,04	12,34	15,12	16,98	19,81	22,4	24,7	27,7
14	6,57	7,79	13,34	16,22	18,15	21,1	23,7	26,1	29,1
15	7,26	8,55	14,34	17,32	19,31	22,3	25,0	27,5	30,6
16	7,96	9,31	15,34	18,42	20,5	23,5	26,3	28,8	32,0
17	8,67	10,08	16,34	19,51	21,6	24,8	27,6	30,2	33,4
18	9,39	10,86	17,34	20,6	22,8	26,0	28,9	31,5	34,8
19	10,11	11,65	18,34	21,7	23,9	27,2	30,1	32,9	36,2
20	10,85	12,44	19,34	22,8	25,0	28,4	31,4	34,2	37,6
21	11,59	13,24	20,3	23,9	26,2	29,6	32,7	35,5	38,9
22	12,34	14,04	21,3	24,9	27,3	30,8	33,9	36,8	40,3
23	13,09	14,85	22,3	26,0	28,4	32,0	35,2	38,1	41,6
24	13,85	15,66	23,3	27,0	29,6	33,2	36,4	39,4	43,0
25	14,61	16,47	24,3	28,2	30,7	34,4	37,7	40,6	44,3
26	15,38	17,29	25,3	29,2	31,8	35,6	38,9	41,9	45,6
27	16,15	18,11	26,3	30,3	32,9	36,7	40,1	43,2	47,0
28	16,93	18,94	27,3	31,4	34,0	37,9	41,3	44,5	48,3
29	17,71	19,77	28,3	32,5	35,1	39,1	42,6	45,7	49,6
30	18,49	20,60	29,3	33,5	36,2	40,3	43,8	47,0	50,9

Таблиця 4

Таблиця значень $t_\gamma = t(k,\gamma)$ (розподіл Стьюдента)

Кількість степенів вільності	$\gamma = 1-\alpha$					
k	0,9	0,95	0,98	0,99	0,998	0,999
1	6,31	12,7	31,82	63,7	318,3	637,0
2	2,92	4,30	6,97	9,92	22,33	31,6
3	2,35	3,18	4,54	5,84	10,21	12,9
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,37	3,00	3,50	4,79	5,40
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,06	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,96
18	1,74	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,51	3,79
23	1,71	2,07	2,50	2,81	3,49	3,77
24	1,71	2,06	2,49	2,80	3,47	3,75
25	1,71	2,06	2,49	2,79	3,45	3,72
26	1,71	2,06	2,48	2,78	3,44	3,71
27	1,71	2,05	2,47	2,77	3,42	3,69
28	1,70	2,05	2,47	2,76	3,41	3,67
29	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	20,2	2,42	2,70	3,31	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,29

Таблиця 5

Критичні точки розподілу *F* Фішера-Снедекора

Рівень значущості $\alpha = 0{,}01$

k_2	k_1											
	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6082	6106
2	98,49	99,01	90,17	99,25	99,33	99,30	99,34	99,36	99,36	99,40	99,41	99,42
3	34,12	30,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	15,54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	9,86	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,45

Рівень значущості $\alpha = 0{,}05$

k_2	k_1											
	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,37	19,40	19,41
3	10,12	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,20	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38

Таблиця 6

Значення $p_k = \dfrac{\lambda^k}{k!} e^{-\lambda}$

k	λ								
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066
1	0,0905	0,1637	0,2222	0,2681	0,3033	0,3293	0,3476	0,3595	0,3659
2	0,0045	0,0164	0,0333	0,0536	0,0758	0,0988	0,1217	0,1438	0,1647
3	0,0002	0,0011	0,0033	0,0072	0,0126	0,0198	0,0284	0,0383	0,0494
4		0,0001	0,0003	0,0007	0,0016	0,0030	0,0050	0,0077	0,0111
5				0,0001	0,0002	0,0004	0,0007	0,0012	0,0020
6							0,0001	0,0002	0,0003
k	λ								
	1,0	2,0	3,0	4,0	5,0	6,0	7,0	8,0	9,0
0	0,3679	0,1353	0,0498	0,0183	0,0067	0,0025	0,0009	0,0003	0,0001
1	0,3679	0,2707	0,1494	0,0733	0,0337	0,0149	0,0064	0,0027	0,0011
2	0,1839	0,2707	0,2240	0,1465	0,0842	0,0446	0,0223	0,0107	0,0050
3	0,0613	0,1804	0,2240	0,1954	0,1404	0,0892	0,0521	0,0286	0,0150
4	0,0153	0,0902	0,1680	0,1954	0,1755	0,1339	0,0912	0,0573	0,0337
5	0,0031	0,0361	0,1008	0,1563	0,1755	0,1606	0,1277	0,0916	0,0607
6	0,0005	0,0120	0,0504	0,1042	0,1462	0,1606	0,1490	0,1221	0,0911
7	0,0001	0,0034	0,0216	0,0595	0,1044	0,1377	0,1490	0,1396	0,1171
8		0,0009	0,0081	0,0298	0,0653	0,1033	0,1304	0,1396	0,1318
9		0,0002	0,0027	0,0132	0,0363	0,0688	0,1014	0,1241	0,1318
10			0,0008	0,0053	0,0181	0,0413	0,0710	0,0993	0,1186
11			0,0002	0,0019	0,0082	0,0225	0,0452	0,0722	0,0970
12			0,0001	0,0006	0,0034	0,0113	0,0264	0,0481	0,0728
13				0,0002	0,0013	0,0052	0,0142	0,0296	0,0504
14				0,0001	0,0005	0,0022	0,0071	0,0169	0,0324
15					0,0002	0,0009	0,0033	0,0090	0,0194
16						0,0003	0,0014	0,0045	0,0109
17						0,0001	0,0006	0,0021	0,0058
18							0,0002	0,0009	0,0029
19							0,0001	0,0004	0,0014
20								0,0002	0,0006
21								0,0001	0,0003
22									0,0001

Список використаної літератури

1. Барковський В. В. Теорія ймовірностей та математична статистика / В. В. Барковський, Н. В. Барковська, О. К. Лопатін − К., ЦУЛ , 2002. − 448 с.

2. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике / В. Е. Гмурман. – М.: Высш. шк., 1997. – 400 с.

3. Гуткевич С.О. Політика ефективного розвитку підприємств: управлінський аспект : монографія / С.О. Гуткевич, Л.П. Шендерівська. – Київ: НТУУ «КПІ», 2016. – 211 с.

4. Іванюта І. Д. Елементи теорії ймовірностей та математичної статистики / І. Д. Іванюта, В. І. Рибалка, І. А. Рудоміно-Дусятська. – К.: Слово, 2003.–272с.

5. Коханівський О.П. Завдання для індивідуальної роботи з математичної статистики (для студентів видавничо-поліграфічного факультету)./ О.П. Коханівський, Б.П. Орел, В.А. Шовський. − К., НТУУ «КПІ», 1999. −40с.

6. Кушлик-Дивульська О.І. Теорія ймовірностей та математична статистика: [навч. посіб.] ⁄ О.І. Кушлик-Дивульська, Н.В. Поліщук, Б.П. Орел, П.І. Штабалюк. – Київ, НТУУ «КПІ», 2010. – 136 с.

7. Кушлик-Дивульська О.І. Теорія ймовірностей та математична статистика: навч. посіб. ⁄ О. І. Кушлик-Дивульська, Н. В. Поліщук, Б. П. Орел, П.І. Штабалюк.− К.: НТУУ «КПІ», 2014. – 212 с.

8. Кушлик-Дивульська О.І. Методичні вказівки до виконання лабораторних робіт з дисципліни «Теорія ймовірностей і математична статистика» для напряму підготовки 6.030601 «Менеджмент» спеціальності 7.03060101 «Менеджмент організацій і адміністрування» для студентів видавничо-поліграфічного інституту [Електронний ресурс]/ НТУУ «КПІ»; уклад. О.І. Кушлик-Дивульська, Б.Р. Кушлик.− Електронні текстові данні (1 файл: 3,74 Мбайт).– Київ: НТУУ «КПІ», 2015.–161с. – Назва з екрана.– Доступ: http://ela.kpi.ua/handle/123456789/14333.

9. Кушлик Б.Р. Удосконалення технології друкування малотиражної продукції плоским офсетним друком: дис. … канд. техн. наук: 05.05.01/Київ, НТУУ «КПІ», 2014.−190с. – Бібліогр.: с. 134-148.

10. Шендерівська Л. П. Eконометричне моделювання прибутку поліграфічних підприємств / Л. П. Шендерівська // Економіко-математичне моделювання: зб. мат. Першої нац. наук.-метод. конф., м. Київ, 30 верес.-1 жовт. 2016 року. – Київ, 2016. – С. 386-388.

11. Шендерівська Л. П. Основні напрями розвитку підприємств поліграфічної галузі / Л. П. Шендерівська // Інтелект XXI. − 2015. − № 1. − С. 56-62.

Зміст