

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

**ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ СЕМЕНА КУЗНЕЦЯ**

***Е. Ю. Железнякова
Л. О. Норік***

**ТЕОРІЯ ЙМОВІРНОСТЕЙ
ТА МАТЕМАТИЧНА СТАТИСТИКА**

Навчальний посібник

**Харків
ХНЕУ ім. С. Кузнеця
2025**

УДК 519.2(075.034)

Ж51

Авторський колектив: канд. фіз.-мат. наук, доцент Е. Ю. Железнякова – вступ, підрозд. 1 – 7; канд. екон. наук, доцент Л. О. Норік – підрозд. 8 – 14.

Рецензенти: професор кафедри комп'ютерної математики і аналізу даних Національного технічного університету "Харківський політехнічний інститут", д-р фіз.-мат. наук О. А. Галуза; доцент кафедри вищої математики та інформатики факультету математики і інформатики Харківського національного університету імені В. Н. Каразіна, канд. фіз.-мат. наук О. О. Аршава.

Рекомендовано до видання рішенням ученої ради Харківського національного економічного університету імені Семена Кузнеця.

Протокол № 2 від 21.01.2025 р.

Самостійне електронне текстове мережеве видання

Железнякова Е. Ю.

Ж51 Теорія ймовірностей та математична статистика [Електронний ресурс] : навчальний посібник / Е. Ю. Железнякова, Л. О. Норік. – Харків : ХНЕУ ім. С. Кузнеця, 2025. – 338 с.
ISBN 978-966-676-890-5

Розглянуто основні питання програми курсу теорії ймовірностей і математичної статистики. Згідно з темами, матеріал розподілено на два розділи. До кожного підрозділу подано основні теоретичні відомості, докладні приклади розв'язування типових задач, запропоновано вправи для самостійної роботи, тестові завдання та запитання для самоперевірки, які повністю відповідають програмі навчальної дисципліни.

Рекомендовано для здобувачів вищої освіти всіх спеціальностей як допоміжний матеріал під час самостійного вивчення навчальної дисципліни, а також для використання викладачами в процесі проведення практичних занять та організації самостійної роботи здобувачів вищої освіти.

УДК 519.2(075.034)

© Железнякова Е. Ю., Норік Л. О., 2025

© Харківський національний економічний університет імені Семена Кузнеця, 2025

ISBN 978-966-676-890-5

Зміст

Вступ.....	7
Розділ 1. Теорія ймовірностей.....	9
1. Емпіричні та логічні основи теорії ймовірностей	9
1.1. Мета та компетентності	9
1.2. Основні теоретичні відомості. Елементи комбінаторики ..	9
1.3. Приклади розв'язування задач	13
1.4. Вправи для самостійної роботи.....	20
1.5. Тестові завдання	24
1.6. Запитання для самоперевірки	26
1.7. Висновки за підрозділом 1	27
2. Основні теореми теорії ймовірностей, їхня економічна інтерпретація.....	28
2.1. Мета та компетентності	28
2.2. Основні теоретичні відомості	28
2.3. Приклади розв'язування задач	30
2.4. Вправи для самостійної роботи.....	36
2.5. Тестові завдання	41
2.6. Запитання для самоперевірки	43
2.7. Висновки за підрозділом 2	44
3. Схема незалежних випробувань	45
3.1. Мета та компетентності	45
3.2. Основні теоретичні відомості	45
3.3. Приклади розв'язування задач	49
3.4. Вправи для самостійної роботи.....	54
3.5. Тестові завдання	60
3.6. Запитання для самоперевірки	62
3.7. Висновки за підрозділом 3	63
4. Випадкові величини та їхня економічна інтерпретація	64
4.1. Мета та компетентності	64
4.2. Основні теоретичні відомості	64
4.3. Приклади розв'язування задач	70
4.4. Вправи для самостійної роботи.....	79
4.5. Тестові завдання	84

4.6. Запитання для самоперевірки	86
4.7. Висновки за підрозділом 4	86
5. Закони розподілу та числові характеристики дискретної випадкової величини.....	87
5.1. Мета та компетентності	87
5.2. Основні теоретичні відомості	87
5.3. Приклади розв'язування задач	90
5.4. Вправи для самостійної роботи.....	94
5.5. Тестові завдання	97
5.6. Запитання для самоперевірки	99
5.7. Висновки за підрозділом 5	99
6. Основні закони розподілу неперервної випадкової величини ..	100
6.1. Мета та компетентності	100
6.2. Основні теоретичні відомості	100
6.3. Приклади розв'язування задач	109
6.4. Вправи для самостійної роботи.....	121
6.5. Тестові завдання	129
6.6. Запитання для самоперевірки	132
6.7. Висновки за підрозділом 6	133
7. Багатовимірні випадкові величини	134
7.1. Мета та компетентності	134
7.2. Основні теоретичні відомості	134
7.3. Приклади розв'язування задач	149
7.4. Вправи для самостійної роботи.....	166
7.5. Тестові завдання	174
7.6. Запитання для самоперевірки	179
7.7. Висновки за підрозділом 7	180
Розділ 2. Математична статистика	181
8. Граничні теореми теорії ймовірностей.....	181
8.1. Мета та компетентності	181
8.2. Основні теоретичні відомості	181
8.3. Приклади розв'язування задач	184
8.4. Вправи для самостійної роботи.....	189
8.5. Тестові завдання	193
8.6. Запитання для самоперевірки	194
8.7. Висновки за підрозділом 8	195

9. Первинне опрацювання статистичних даних	196
9.1. Мета та компетентності	196
9.2. Основні теоретичні відомості	196
9.3. Приклади розв'язування задач	201
9.4. Вправи для самостійної роботи.....	207
9.5. Тестові завдання	210
9.6. Запитання для самоперевірки	212
9.7. Висновки за підрозділом 9	212
10. Статистичне оцінювання параметрів розподілу	213
10.1. Мета та компетентності	213
10.2. Основні теоретичні відомості.....	213
10.3. Приклади розв'язування задач	217
10.4. Вправи для самостійної роботи.....	221
10.5. Тестові завдання	225
10.6. Запитання для самоперевірки	228
10.7. Висновки за підрозділом 10	228
11. Перевірка статистичних гіпотез	229
11.1. Мета та компетентності	229
11.2. Основні теоретичні відомості.....	229
11.3. Приклади розв'язування задач	240
11.4. Вправи для самостійної роботи.....	248
11.5. Тестові завдання	254
11.6. Запитання для самоперевірки	256
11.7. Висновки за підрозділом 11	257
12. Елементи теорії кореляції.....	258
12.1. Мета та компетентності	258
12.2. Основні теоретичні відомості.....	258
12.3. Приклади розв'язування задач	267
12.4. Вправи для самостійної роботи.....	277
12.5. Тестові завдання	281
12.6. Запитання для самоперевірки	283
12.7. Висновки за підрозділом 12	283
13. Елементи дисперсійного аналізу.....	284
13.1. Мета та компетентності	284
13.2. Основні теоретичні відомості.....	284

13.3. Приклади розв'язування задач	291
13.4. Вправи для самостійної роботи.....	296
13.5. Тестові завдання	300
13.6. Запитання для самоперевірки	302
13.7. Висновки за підрозділом 13	302
14. Елементи теорії регресії	303
14.1. Мета та компетентності	303
14.2. Основні теоретичні відомості.....	303
14.3. Приклади розв'язування задач	308
14.4. Вправи для самостійної роботи.....	319
14.5. Тестові завдання	321
14.6. Запитання для самоперевірки	323
14.7. Висновки за підрозділом 14	324
Рекомендована література.....	325
Додатки.....	327
Предметний покажчик.....	335

Вступ

У сучасному світі, де величезний обсяг даних і невизначеність є невід'ємною частиною багатьох сфер діяльності, зростає потреба в спеціалістах, які володіють навичками з аналізу й інтерпретації статистичної інформації. Теорія ймовірностей та математична статистика є основою для розуміння і розв'язання проблем, пов'язаних із випадковими явищами та статистичними закономірностями.

Навчальна дисципліна "Теорія ймовірностей та математична статистика" відіграє важливу роль у науці, технологіях, економіці та багатьох інших галузях, де ухвалення рішень залежить від точного аналізу даних. Вона охоплює фундаментальні аспекти теорії ймовірностей, серед них дослідження випадкових подій, визначення їхньої ймовірності, аналіз законів розподілу дискретних і неперервних випадкових величин, їхніх числових характеристик, а також основи математичної статистики – від формування вибірових сукупностей до аналізу результатів статистичних спостережень, теорії кореляції та регресійного аналізу. Метою вивчення цієї навчальної дисципліни є формування в здобувачів вищої освіти ключових компетентностей, потрібних для застосування математичних методів у процесі дослідження економічних процесів і явищ, де математика є важливим інструментом для аналізу й ухвалення рішень.

Цей навчальний посібник створено для того, щоб допомогти здобувачам вищої освіти засвоїти основні поняття й методи теорії ймовірностей та математичної статистики через виконання практичних завдань. Його орієнтовано на здобувачів вищої освіти, які вивчають ці навчальні дисципліни в межах різних спеціальностей, серед них математика, економіка, інженерія та природничі науки.

Основна мета цього посібника – сприяти формуванню навичок з аналізу випадкових явищ, розуміння принципів побудови статистичних моделей та інтерпретації результатів статистичного аналізу. Завдяки цьому навчальному посібнику, здобувачі вищої освіти зможуть не лише засвоїти теоретичні знання, але й навчитися застосовувати їх на практиці, що є конче потрібним для формування компетентностей у сучасному світі.

Особливу увагу в цьому посібнику приділено практичній частині, що містить численні приклади економічних задач. Це дозволить здобувачам вищої освіти не лише краще зрозуміти абстрактні математичні концепції, але й побачити їхнє застосування в конкретних галузях, як-от економіка, фінанси, бізнес-аналіз та ін. Отже, посібник охоплює широкий спектр прикладних питань, що робить його корисним для здобувачів вищої освіти різних спеціальностей, зокрема тих, хто готується до кар'єри в економічній чи бізнесовій сфері.

Матеріал посібника розподілено на підрозділи. Кожний підрозділ має 7 частин: мета та компетентності, основні теоретичні відомості, приклади розв'язування задач, вправи для самостійної роботи, тестові завдання, запитання для самоперевірки та висновки за підрозділом. Виклад теоретичного матеріалу за кожним підрозділом супроводжено розв'язанням прикладів, які ілюструють його сутність. Кожний підрозділ завершує серія завдань різного рівня складності, що дозволяє здобувачам вищої освіти поступово переходити від базових до більш складних питань, а також сприяє розвитку аналітичних навичок і критичного мислення. Весь матеріал побудовано таким способом, щоб активно залучити здобувачів вищої освіти до освітнього процесу. У додатках розміщено основні статистичні таблиці, потрібні для розв'язання задач.

Цей навчальний посібник стане незамінним помічником для здобувачів вищої освіти в засвоєнні теорії ймовірностей та математичної статистики, буде сприяти поглибленню знань, розвитку вмінь застосовувати їх у практичних ситуаціях, а також підготовці до подальшого навчання та професійної діяльності.

Посібник відповідає потребам сучасної освіти та забезпечує здобувачів вищої освіти потрібними інструментами для розв'язання прикладних задач у різних сферах діяльності.

Розділ 1

Теорія ймовірностей

1. Емпіричні та логічні основи теорії ймовірностей

1.1. Мета та компетентності

Метою вивчення підрозділу є ознайомлення здобувачів вищої освіти із предметом і завданнями навчальної дисципліни "Теорія ймовірностей та математична статистика", що є теоретичною базою математичного моделювання економічних процесів і явищ; з основними означеннями теорії ймовірностей та формування системи теоретичних знань, практичних умінь і навичок із визначення ймовірностей випадкових подій та інтерпретації здобутих результатів за допомогою діаграм Венна – Ейлера.

Компетентності, що формують під час вивчення підрозділу:

знання загальних означень теорії ймовірностей;

уміння визначати умови застосування аксіом, на яких побудовано теорію ймовірностей;

уміння обчислювати ймовірності випадкових подій за класичним і геометричним означеннями ймовірності;

уміння обчислювати ймовірності випадкових подій за допомогою формул комбінаторики;

надання геометричної інтерпретації випадкових подій і дій над ними за допомогою діаграм Венна – Ейлера.

1.2. Основні теоретичні відомості. Елементи комбінаторики

Перестановками називають комбінації, що складаються з тих самих n різних елементів і відрізняються тільки порядком їхнього розташування.

Кількість усіх можливих перестановок дорівнює:

$$P_n = n!, \text{ де } n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n.$$

Розміщеннями називають комбінації, складені з n різних елементів по m елементів, що відрізняються або складом елементів, або їхнім порядком.

Кількість усіх можливих розміщень дорівнює:

$$A_n^m = n(n-1) \cdot \dots \cdot (n-m+1) \text{ або } A_n^m = \frac{n!}{(n-m)!}.$$

Сполученнями називають комбінації, складені з n різних елементів по m елементів, що відрізняються хоча б одним елементом.

Кількість сполучень обчислюють за такою формулою:

$$C_n^m = \frac{n(n-1) \cdot \dots \cdot (n-m+1)}{1 \cdot 2 \cdot 3 \cdot \dots \cdot m} \text{ або } C_n^m = \frac{n!}{m!(n-m)!}.$$

Під час розв'язання задач комбінаторики використовують такі правила:

1) *правило сум*. Якщо об'єкт A може бути вибраним із сукупності об'єктів m способами, а другий об'єкт B n способами, то вибрати або A , або B можна $m+n$ способами;

2) *правило добутку*. Якщо об'єкт A можна вибрати із сукупності об'єктів m способами й після кожного такого вибору об'єкт B можна вибрати n способами, то пару об'єктів (A, B) у наведеному порядку може бути вибрано $m \cdot n$ способами.

Класичне визначення ймовірності події

Нехай n – загальна кількість усіх рівноможливих, несумісних і тих, що утворюють повну групу наслідків випробувань, а m – кількість наслідків, що сприяють появі події A . Тоді ймовірність $P(A)$ дорівнює

$$\text{відношенню } m \text{ до } n, \text{ тобто } P(A) = \frac{m}{n}.$$

Згідно з означенням, $0 \leq P(A) \leq 1$.

Геометричне тлумачення ймовірності.

Геометрична ймовірність

Геометричне тлумачення ймовірності широко застосовують для більш наочного подання умови задачі в тому разі, якщо простір подій є дискретним, тобто є можливість застосовувати класичне означення

ймовірності. Але воно стає незамінним у тому разі, якщо простір подій є неперервним. Слід розглянути діаграму Венна – Ейлера (рис. 1.1).

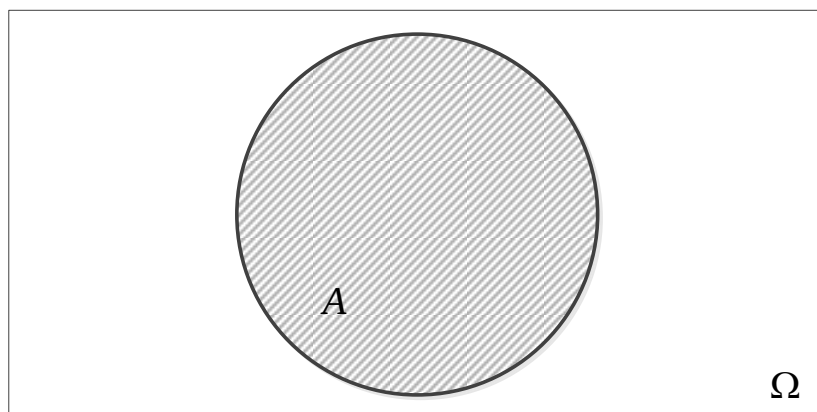


Рис. 1.1. Діаграма Венна – Ейлера для визначення ймовірності випадкової події

Простір елементарних подій позначають Ω . Це сукупність усіх можливих несумісних наслідків випробування, і на діаграмі Венна – Ейлера їй відповідає вся площа прямокутника.

Якщо A – випадкова подія і сукупності елементарних подій, що сприяють появі цієї випадкової події, відповідає площі однойменного кола, то ймовірність події A вимірюють як відношення площі кола до площі прямокутника, тобто $P(A) = \frac{S_A}{S_\Omega}$. Як і за класичного визначення ймовір-

ності випадкової події, з означення випливає, що $0 \leq P(A) \leq 1$.

Хоча найбільш поширеним є двовимірне зображення множини подій (тобто діаграми Венна – Ейлера), але геометричне означення ймовірності передбачає використання і геометричних фігур іншої вимірності, наприклад, простору елементарних подій може відповідати відрізок прямої, а випадковій події – частина цього відрізка.

За допомогою діаграм Венна – Ейлера зручно ілюструвати теореми множення та додавання ймовірностей випадкових подій. Так, на рис. 1.2 показано приклад двох сумісних випадкових подій A та B .

Отже, переріз цих подій $A \cap B$ є непорожньою множиною, якій відповідає спільна частина обох кіл, виділена на рис. 1.2 подвійним штрихуванням. Відношення площі цієї фігури до загальної площі прямокутника і є ймовірністю події добутку подій $A \cdot B$.

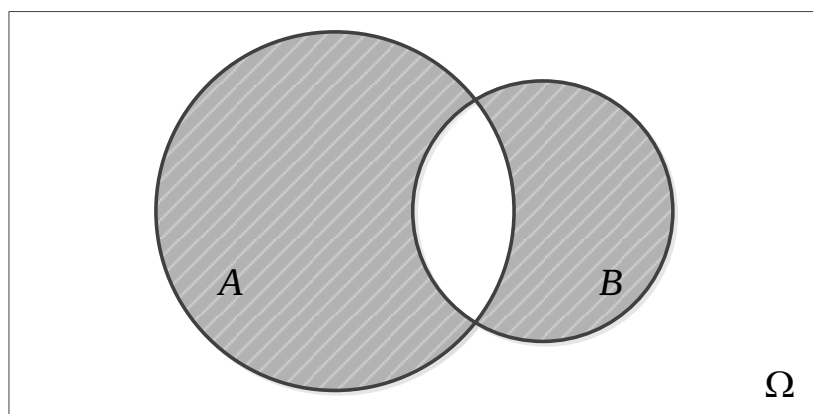


Рис. 1.2. Діаграма Венна – Ейлера, що визначає сумісні випадкові події A та B

Сполученню випадкових подій $A \cup B$, що є сумою вихідних випадкових подій, відповідає вся площа, виділена штрихуванням. Оскільки площа, позначена подвійною штриховкою, належить одночасно до обох кіл, то ймовірність суми випадкових подій A та B визначають таким співвідношенням:

$$P(A + B) = P(A) + P(B) - P(A \cdot B).$$

Слід розглянути поняття "різниця множин". Наприклад, різниці множин $A - B$, що означає $A \setminus B$, відповідає площа кола A за винятком тієї його частини, яка належить одночасно і колу A , і колу B . Отже, різницю множин $A - B$ можна також подати як переріз $A \cap \bar{B}$, де $\bar{B}$ – випадкова подія, яка полягає в тому, що подія B не відбулася. На рис. 1.3 різницю множин $A - B$ позначено подвійною штриховкою.

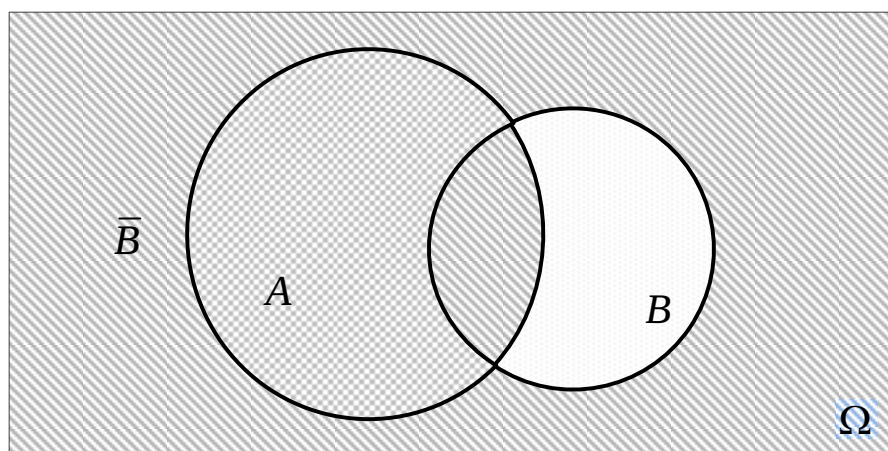


Рис. 1.3. Діаграма Венна – Ейлера, що ілюструє різницю випадкових подій A та B

Ймовірність події $A \cap \bar{B}$ можна записати в символах умовної ймовірності таким способом: $P(A | \bar{B})$, тобто ймовірність події A за умов, що подія B не відбулася (мала місце подія $\bar{B}$). Тоді ймовірність суми подій записують у такому вигляді:

$$P(A + B) = P(A | \bar{B}) + P(B | \bar{A}) + P(A \cdot B).$$

1.3. Приклади розв'язування задач

Приклад 1.1. У конкурсі беруть участь 5 фахівців. Порядок їхнього виступу визначають жеребкуванням.

Скільки є варіантів порядку їхнього виступу?

Розв'язання. Кожний варіант жеребкування відрізняється від інших тільки порядком виступу конкурсантів, а тому він є перестановкою з 5 елементів.

Кількість варіантів дорівнює $P_5 = 5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$.

Приклад 1.2. Розклад занять одного дня здобувача вищої освіти складається із 3 пар.

Визначмо кількість варіантів розкладу за вибору з 10 навчальних дисциплін.

Розв'язання. Кожен варіант розкладу є набором 3 навчальних дисциплін із 10, що відрізняє його від інших варіантів або складом навчальних дисциплін, або порядком їхнього розташування, тобто він є розміщенням із 10 елементів по 3 елементи.

Кількість варіантів розкладу занять дорівнює $A_{10}^3 = 10 \cdot 9 \cdot 8 = 720$.

Приклад 1.3. У шаховому турнірі беруть участь 20 юнаків.

Скільки партій вони зіграють, якщо між будь-якими двома учасниками має бути зіграно одну партію?

Розв'язання. Кожну партію відрізняє тільки склад пар учасників, тобто слід використати сполучення із 20 елементів по 2 елементи.

Отже, кількість партій дорівнює $C_{20}^2 = \frac{20!}{2!18!} = \frac{20 \cdot 19}{1 \cdot 2} = 190$.

Приклад 1.4. Із 50 стандартних і 5 нестандартних деталей для контролю навмання взято 10, що виявилися стандартними.

Визначмо ймовірність того, що наступна навмання взята деталь буде стандартною.

Розв'язання. Стандартних деталей залишилося $50 - 10 = 40$, а нестандартних 5. Усього залишилося $40 + 5 = 45$ деталей.

$$\text{Шукана ймовірність дорівнює } P = \frac{40}{45} = \frac{8}{9}.$$

Приклад 1.5. Монету кинуть 2 рази.

Визначмо ймовірність того, що хоча б один раз з'явиться "герб".

Розв'язання. За двох кидань монети можливі чотири елементарні результати: "герб" – "герб", "герб" – "цифра", "цифра" – "герб", "цифра" – "цифра". Із них сприятливими є три перші, тобто $m = 3$, а $n = 4$.

$$\text{Шукана ймовірність } P = \frac{3}{4}.$$

Приклад 1.6. На 5 картках написано букви "е", "о", "ч", "н", "в". Після ретельного перемішування беруть по 1 картці та кладуть послідовно поруч.

Яка ймовірність того, що в результаті буде складено слово "човен"?

Розв'язання. Кількість усіх можливих наслідків випробувань у цьому разі дорівнює кількості перестановок із п'яти букв, тобто:

$$P_5 = 5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120.$$

Із них лише одна утворить слово "човен".

$$\text{Тому } m = 1, \text{ а ймовірність шуканої події } P = \frac{1}{120}.$$

Приклад 1.7. Серед 17 здобувачів вищої освіти, із яких 8 дівчат, розігрують 7 квитків.

Яка ймовірність, що серед власників квитків буде 4 дівчини?

Розв'язання. Кількість можливих способів розподілити 7 квитків серед 17 здобувачів вищої освіти, дорівнює кількості сполучень із 17 елементів по 7, тобто C_{17}^7 . Кількість відбору чотирьох дівчат із восьми дорівнює C_8^4 . Кожну таку четвірку можна сполучити з кожною трійкою з 9 юнаків. Кількість таких трійок дорівнює C_9^3 .

Отже, кількість результатів такого розподілу 7 квитків, коли 4 з них дівчини та 3 юнаки, за правилом множення дорівнює $C_8^4 \cdot C_9^3$.

$$\text{Тоді шукана ймовірність дорівнює } P = \frac{C_8^4 \cdot C_9^3}{C_{17}^7} = \frac{735}{2431} \approx 0,302.$$

Приклад 1.8. Набираючи номер телефону, абонент забув останні три цифри та пам'ятає лише, що ці цифри різні. Він набрав їх навмання.

Визначмо ймовірність того, що набрані цифри є правильними.

Розв'язання. Нехай подія A – набрані три потрібні цифри. Усього можна набрати стільки різних цифр, скільки може бути складено розміщень із десяти цифр по три, тобто $A_{10}^3 = 10 \cdot 9 \cdot 8 = 720$. Це загальна кількість елементарних подій.

Сприятливим події A тільки один результат – три правильні цифри.

$$\text{Тому шукана ймовірність дорівнює } P = \frac{1}{720}.$$

Приклад 1.9. Велике коло радіусом 50 см містить у центрі маленьке коло радіусом 5 см.

Визначмо ймовірність того, що куля влучить у маленьке коло, якщо відомо, що вона влучила у велике коло.

Розв'язання. У цьому прикладі неможливо скористатися класичним означенням імовірності. Застосуємо її геометричне означення. Оскільки куля влучила у велике коло, то можна вважати, що його площа відповідає множині Ω (повній групі елементарних подій).

Тоді ймовірність випадкової події, а саме, улучання кулі в маленьке коло, визначають як відношення площі малого кола до площі великого кола, тобто:

$$P(A) = \frac{\pi r^2}{\pi R^2} = \frac{25}{2500} = 0,01.$$

Приклад 1.10. Квадратну коробку розділено на секції, три з яких є однаковими за площею квадратами, а інші три – прямокутниками, до того ж площа кожного із прямокутників є удвічі більшою за площу квадратної секції. До коробки кидають кульку.

Яка ймовірність того, що вона потрапить до квадратної секції?

Розв'язання. Імовірність улучання кульки до однієї із квадратних секцій визначають відношенням загальної площі цих трьох секцій до площі всієї квадратної коробки. Нехай один бік коробки має розмір a .

Отже, площа всієї коробки дорівнює a^2 . Загальна площа кожної квадратної секції є вдвічі більшою за площу прямокутної, звідси загальна площа всіх трьох квадратних секцій становить $\frac{a^2}{3}$.

Тоді ймовірність улучання кульки до квадратної секції дорівнює

$$P(A) = \frac{\frac{a^2}{3}}{a^2} = \frac{1}{3}.$$

Приклад 1.11. Серед 100 туристів 60 добре володіють англійською, 40 – німецькою та 20 – французькою.

Визначмо ймовірність того, що навмання вибраний турист не володіє іноземною мовою.

Розв'язання. Хоча в прикладі розглядають дискретні множини подій, умову задачі зручно подати за допомогою діаграми Венна – Ейлера (рис. 1.4).

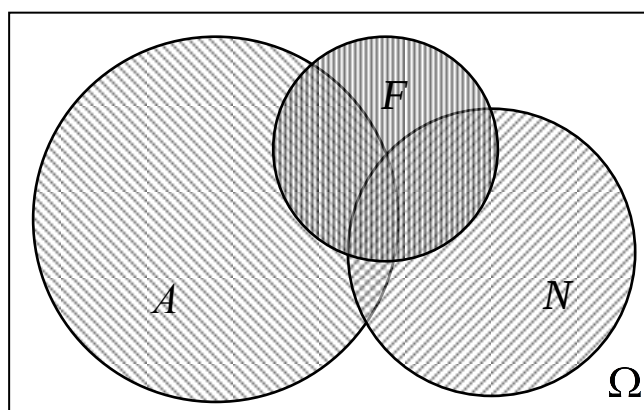


Рис. 1.4. Діаграма Венна – Ейлера до прикл. 1.11

На діаграмі множини туристів, які володіють певною іноземною мовою, позначено початковими літерами. Площина поза колом відповідає множині туристів, які не володіють цією мовою, тобто протилежна подія.

Отже, множині туристів, які не володіють жодною іноземною мовою, відповідає частина прямокутника повної групи елементарних подій, вільній від штриховки. Нехай це буде випадкова подія V , імовірність якої слід визначити.

Випадкову подію V , яка полягає в тому, що навмання вибраний турист не володіє жодною іноземною мовою, визначають як переріз множин $\bar{A}$, $\bar{N}$ та $\bar{F}$. Ці події є сумісними, але незалежними, тому для визначення ймовірності є таке співвідношення:

$$P(V) = P(\bar{A} \cap \bar{N} \cap \bar{F}) = P(\bar{A}) \cdot P(\bar{N}) \cdot P(\bar{F}) = \\ = (1 - P(A)) \cdot (1 - P(N)) \cdot (1 - P(F)) = (1 - 0,6) \cdot (1 - 0,4) \cdot (1 - 0,2) = 0,192.$$

Визначено відповідь на запитання задачі, але для перевірки розгляньмо інший шлях розв'язання.

Визначмо ймовірність випадкової події, яка полягає в тому, що навмання вибраний турист добре володіє якою-небудь іноземною мовою (оскільки ця подія є протилежною тій, імовірність якої досліджують, то її позначають $\bar{V}$). Такий турист може знати будь-яку, але тільки одну мову, дві мови або навіть три мови. Випадковій події $\bar{V}$ відповідає вся заштрихована площа на рис. 1.4.

Події A , N та F є незалежними, але сумісними, отже:

$$P(\bar{V}) = P(A) + P(N) + P(F) - P(A \cdot N) - P(A \cdot F) - P(N \cdot F) + \\ + P(A \cdot N \cdot F) = P(A) + P(N) + P(F) - P(A) \cdot P(N) - P(A) \cdot P(F) - \\ - P(N) \cdot P(F) + P(A) \cdot P(N) \cdot P(F) = 0,6 + 0,4 + 0,2 - 0,6 \cdot 0,4 - \\ - 0,6 \cdot 0,2 - 0,4 \cdot 0,2 + 0,6 \cdot 0,4 \cdot 0,2 = 0,808.$$

Звідси визначають, що $P(V) = 1 - P(\bar{V}) = 1 - 0,808 = 0,192$.

Зрозуміло, що обидва розв'язання дали той самий результат.

Приклад 1.12. На числовій прямій узято відрізок OA , початок якого міститься в початку координат, а точці A відповідає одиниця.

Відомо, що точка B , яка є внутрішньою точкою відрізка, ділить його у відношенні $1:2$. На цей відрізок довільно нанесено точку, яку позначено C .

Визначмо ймовірність того, що відстань між точками B та C буде меншою, ніж відстань від точки C до будь-якого з кінців відрізка.

Розв'язання. Множина випадкових подій, про яку йдеться мова в прикладі, є неперервною, отже, для обчислення ймовірності слід скористатися її геометричним означенням. Спочатку зробимо такий рисунок (рис. 1.5):

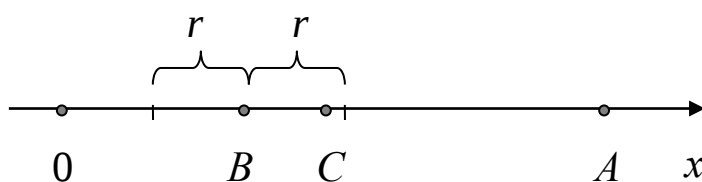


Рис. 1.5. Ілюстрація до прикл. 1.12

Оскільки точка B ділить відрізок OA у відношенні $1:2$, то це означає, що вона міститься від найближчого до неї кінця відрізка на відстані $\frac{1}{3}$

(наприклад, має координату $\frac{1}{3}$, як показано на рис. 1.5). Те, що відстань

між точками B та C є меншою за відстань від точки C до будь-якого з кінців відрізка, передбачає, що точка C має влучити до околу точки B з радіусом $r = 0,5 \cdot OB$. Імовірність такої випадкової події визначають як відношення діаметра околу до довжини відрізка OA . Оскільки точка B може мати координату або $\frac{1}{3}$, або $\frac{2}{3}$, то результат треба ще помножити на 2.

$$\text{Отже, } P(C) = \frac{2 \cdot r}{OA} \cdot 2 = \frac{2 \cdot 0,5 \cdot \frac{1}{3}}{1} \cdot 2 = \frac{2}{3}.$$

Приклад 1.13. У податковій адміністрації одного з районів міста зареєстровано 160 приватних і 4 державні підприємства.

Визначмо ймовірність того, що серед навмання вибраних трьох підприємств приватними будуть: а) три; б) два; в) не більше ніж одне.

Розв'язання.

Оскільки не ставлять умову впорядкованості підмножини з вибраних трьох підприємств, то потрібно використати комбінації.

$$\text{Тоді } n = C_{164}^3 = \frac{164!}{3! \cdot 161!} = \frac{161! \cdot 162 \cdot 163 \cdot 164}{1 \cdot 2 \cdot 3 \cdot 161!} = 721\,746.$$

$$\text{Для пункту а) } m = C_{160}^3 = \frac{160!}{3! \cdot 157!} = \frac{157! \cdot 158 \cdot 159 \cdot 160}{1 \cdot 2 \cdot 3 \cdot 157!} = 669\,920,$$

$$\text{тоді шукана ймовірність дорівнює } P = \frac{m}{n} = \frac{669\,920}{721\,746} \approx 0,93.$$

Пункт б відрізняється від пункту а тим, що вибрана група із трьох підприємств містить два приватні й одне державне.

Тому за правилом множення маємо:

$$m = C_{160}^2 \cdot C_4^1 = \frac{160!}{2! \cdot 158!} \cdot 4 = \frac{158! \cdot 159 \cdot 160}{1 \cdot 2 \cdot 158!} \cdot 4 = 50\,880.$$

$$\text{Тоді шукана ймовірність дорівнює } P = \frac{m}{n} = \frac{50\,880}{721\,746} \approx 0,08.$$

Для пункту в зміст умови "не більше ніж одне" означає "одне або жодного", тобто слід використати правило суми. Визначення одного приватного підприємства із трьох вибраних передбачає, що інші два є державними, тобто слід використати правило множення, а умова, що жодного приватного підприємства не буде визначено, передбачає, що всі три вибрані підприємства є державними.

Тоді,

$$m = C_{160}^1 \cdot C_4^2 + C_4^3 = 160 \cdot 6 + 4 = 964, \quad P = \frac{964}{721\,746} \approx 0,001.$$

1.4. Вправи для самостійної роботи

1.1. Обчисліть:

а) $15! \cdot \frac{1}{30 \cdot P_{12}}$; б) $\frac{P_7 + P_6}{P_5}$; в) $\frac{A_5^4 + A_5^3}{A_5^2}$; г) $\frac{A_6^2 \cdot P_8}{P_{10}} + \frac{A_5^2}{A_4^2}$; д) $\frac{C_{100}^{97} \cdot P_5}{66 \cdot A_{50}^2}$.

1.2. Спростіть вираз $\frac{A_n^7 - A_n^5}{A_n^5}$ і обчисліть за $n = 15$.

1.3. Спростіть вираз $\frac{C_n^2 + C_n^{n-2}}{C_{n+1}^3 + C_{n+1}^{n-2}}$ і обчисліть за $n = 11$.

1.4. Розв'яжіть рівняння $C_{x+3}^3 = 3(x+1)(x+2)$.

1.5. Є 5 прапорців різного кольору.

Скільки є різних сигналів, які можна подавати, піднімаючи по 3 прапорці в будь-якому порядку?

1.6. У кондитерській є 7 видів тістечок. Черговий покупець вибив чек на 4 тістечка.

Уважаючи, що будь-який набір тістечок, які замовляють, є рівноймовірним, обчисліть імовірність того, що покупець замовив тістечка різних видів.

1.7. Опишіть простір елементарних подій для ситуації, якщо компанія із продажу автомобілів рекламує дві нові моделі авто в соціальних мережах і на телебаченні. Компанію цікавить ефективність реклами, зокрема оцінювання того, що навімання вибрана людина має уявлення хоча б про одну із двох рекламованих моделей. Подія A – навімання вибрана особа чула рекламу в соціальних мережах, подія B – навімання вибрана особа знає про нові моделі автомобілів із реклами телебачення.

1.8. У конкурсі беруть участь 9 осіб.

Скільки є варіантів списку призерів, якщо кожне із 3 призових місць може посісти тільки одна людина?

1.9. Бухгалтери трьох компаній подають звіти в податкову інспекцію.

Запишіть події, які означають, що звіт подано: 1) лише бухгалтером однієї компанії; 2) лише бухгалтером першої компанії; 3) бухгалтерами не більше ніж двох компаній; 4) бухгалтерами принаймні двох компаній.

1.10. На участь у шкільному конкурсі "Пісня року" подали заявки 12 осіб. Порядок їхнього виступу визначали жеребкуванням.

Скільки є способів скласти програму концерту?

1.11. Для підбиття підсумків конкурсу потрібно із 30 присутніх обрати незалежне журі у складі голови, секретаря та спостерігача.

Скільки є способів такого вибору?

1.12. Учні вивчають 8 різних предметів. Розклад уроків має бути складено так, щоб на день було 3 уроки та предмети не повторювалися.

Скільки є способів скласти розклад на один день?

1.13. Скільки є способів розмістити 6 осіб за одним столом?

1.14. Слово зашифровано 6 позначками, які не повторюються.

Скільки є варіантів розшифрування, якщо позначка може відповідати одній із 30 букв алфавіту?

1.15. Десять туристів, серед яких пан Михайло та пан Петро, розміщують у готелі у двох тримісних та одному чотиримісному номерах.

Обчисліть ймовірність події, що полягає в тому, що пан Михайло та пан Петро опиняться в чотиримісному номері?

1.16. У шаховому турнірі беруть участь 12 осіб. Кожен з учасників має зіграти з кожним з інших по 2 партії.

Скільки партій мають зіграти учасники турніру?

1.17. У шухляді лежать 12 олівців однакової форми й розмірів, із яких 4 олівці є кольоровими, а інші простими.

Яка ймовірність, що, відкривши шухляду, буде взято простий олівець?

1.18. Із цифр від 1 до 9 включно навмання вибирають одну.

Визначте ймовірність того, що вибране число буде: 1) парним; 2) непарним; 3) простим; 4) більшим за 7.

1.19. У шухляді у 5 разів більше червоних куль, ніж чорних. Навмання виймають 1 кулю.

Яка ймовірність того, що вона буде червоною?

1.20. У цеху з виготовлення м'ячів для гольфа в одній коробці було 77 м'ячів правильної форми, а в іншій – 23 м'ячі з дефектами. М'ячі висипали в одну коробку.

Яка ймовірність того, що навмання витягнутий м'яч буде бракованим?

1.21. Кинуто дві гральні кістки.

Визначте імовірності таких подій:

а) сума чисел на гранях, що випали, дорівнює 7;

б) сума чисел дорівнює 8, а різниця – 4;

в) сума чисел дорівнює 5, а добуток – 4;

г) сума чисел більша за 7;

ґ) сума чисел більша за 3 і не більша за 8.

1.22. Монету кинуто два рази.

Яка ймовірність того, що обидва рази випаде "герб"?

1.23. Монету підкидають тричі.

Чому дорівнює ймовірність того, що "герб" випаде хоча б один раз?

1.24. На кожній із 5 однакових карток надруковано одну з букв: "а", "м", "р", "т", "ю". Картки ретельно перемішують.

Яка ймовірність того, що із 4 послідовно взятих карток складуть слово "юрта"?

1.25. У лотереї 1 000 квитків. Із них 500 є виграшними. Куплено 2 квитки.

Яка ймовірність того, що вони обидва є виграшними?

1.26. У шухляді 15 деталей, серед яких 10 пофарбованих. Складальник навмання виймає 3 деталі.

Яка ймовірність того, що вони виявляться пофарбованими?

1.27. У шухляді 100 деталей, із них 10 бракованих. Навмання вибрано 4 деталі.

Визначте ймовірність того, що серед них дві є бракованими.

1.28. Набираючи номер телефону, абонент забув останні дві цифри, і, пам'ятаючи лише, що ці цифри є різними, набрав їх навмання.

Визначте ймовірність того, що набрані цифри є правильними.

1.29. У партії з 10 деталей 7 стандартних.

Визначте ймовірність того, що серед 6 взятих навмання деталей 4 є стандартними.

1.30. У цеху працюють 6 чоловіків і 4 жінки. Навмання відібрано 7 осіб.

Яка ймовірність того, що серед них виявляться 3 жінки?

1.31. У групі 12 здобувачів вищої освіти, із них 8 відмінників. За списком навмання відібрано 9 здобувачів вищої освіти.

Яка ймовірність того, що серед них 5 відмінників?

1.32. На кафедрі вищої математики працює 10 викладачів.

Скількома способами з них може бути складено екзаменаційну комісію для проведення іспиту, якщо до її складу має увійти 3 викладачі?

1.33. У маркетинговому відділі компанії працює 15 співробітників, 8 із них – чоловіки.

Визначте ймовірність того, що серед 6 навмання відібраних для участі в тренінгу співробітників є хоча б один чоловік?

1.34. На столі лежать 32 екзаменаційні білети.

Яка ймовірність того, що номер навмання взятого білета буде кратним 7?

1.35. Яка ймовірність того, що в лютому навмання вибраного високосного року буде 5 субот?

1.36. Із розрізної абетки, що налічує 30 карток із різними буквами алфавіту, виймають 5 карток.

Яка ймовірність того, що із 5 букв, розташованих у порядку появи, складуть слово "спорт"?

1.37. У розробленні проекту економічного розвитку певної галузі виробництва беруть участь 4 організації.

Скількома способами може бути розподілено роботу: а) над усіма розділами; б) над першим, другим та третім розділами, якщо проєкт становить 5 розділів?

1.38. Для з'ясування ставлення споживачів до продукції фірми з виробництва харчових продуктів вибирають 4 із 12 покупців магазину. Їх потрібно прикріпити до 4 інтерв'юерів.

Скільки є різних способів складання таких груп?

1.39. У кондитерському відділі супермаркету є 10 видів шоколадних цукерок. Покупець робить замовлення на 500 г цукерок.

Визначте ймовірність того, що покупець замовив: а) по 100 г цукерок різного виду; б) 200 г цукерок одного виду та 300 г іншого; в) усі цукерки одного виду.

1.40. Множину A становлять прості додатні числа, що не перевищують 20; множину B – цілі додатні числа, які не перевищують 50

і в записі яких є цифра 7. Серед множини чисел $A + B$ навмання вибирають одне.

Визначте ймовірність того, що воно буде належати множині $A \setminus B$. Поясніть розрахунок діаграмою Венна – Ейлера.

1.41. Множину A становлять парні тризначні числа; множину B – тризначні числа, які є паліндромами, тобто мають однакове значення, якщо їх читати справа наліво або зліва направо.

Визначте ймовірність того, що вибране навмання будь-яке тризначне число буде належати обом множинам одночасно. Поясніть розв'язання діаграмою Венна – Ейлера.

1.42. Два партнери домовилися зустрітися між 18-ю та 19-ю годинами. Той, хто прийде першим, буде чекати другого чверть години, а потім піде.

Визначте ймовірність того, що зустріч відбудеться, якщо кожний із партнерів може прийти в довільний момент часу з визначеного проміжку.

1.43. На перехресті встановлено світлофор, у якому одну хвилину горить зелене світло та півхвилини – червоне, потім знову одну хвилину – зелене та півхвилини – червоне тощо. У випадковий момент часу до перехрестя під'їжджає автомобіль.

Яка ймовірність того, що він проїде перехрестя без зупинки?

1.44. Два постачальники мають привезти товари в супермаркет. Час прибуття кожного з постачальників є незалежним і однаково можливим протягом дня.

Визначте ймовірність того, що один із постачальників буде чекати на іншого до 1 год, якщо час отримання товару від першого постачальника становить 40 хв.

1.5. Тестові завдання

1.1. Кількість способів вибору k елементів із n , незважаючи на порядок відбору, дорівнює

А	Б	В	Г
$A_n^k = \frac{n!}{k!}$	$C_n^k = \frac{n!}{k!(n-k)!}$	$C_n^k = \frac{n!}{(n-k)!}$	$A_n^k = \frac{n!}{(n-k)!}$

1.2. Кількість способів вибору k елементів із n , з огляду на порядок відбору, дорівнює

А	Б	В	Г
$A_n^k = \frac{n!}{k!}$	$C_n^k = \frac{n!}{k!(n-k)!}$	$C_n^k = \frac{n!}{(n-k)!}$	$A_n^k = \frac{n!}{(n-k)!}$

1.3. Подія – це

А ймовірнісна закономірність

Б наслідок експерименту

В комплекс умов експерименту

Г зчисленна множина подій

Д інша відповідь

1.4. Простір елементарних подій – це

А множина всіх можливих наслідків експерименту

Б зчисленна множина подій

В множина сумісних подій

Г множина подій, які неможливо перелічити

Д інша відповідь

1.5. Імовірнісний простір елементарних подій позначають

А	Б	В	Г	Д
Ω	Σ	α	ω_i	інше

1.6. Експеримент – це

А простір елементарних подій

Б множина подій, що виключають появу інших

В випробування за певних умов

Г множина подій, які неможливо перелічити

1.7. Сумісні події – це

- А** поява однієї події виключає появу інших
- Б** подія, яка розкладається на елементарні події
- В** поява однієї не виключає можливості появи інших
- Г** множина подій, які неможливо перелічити
- Д** інша відповідь

1.8. Імовірність події – це

- А** показник випадковості
- Б** характеристика настання події
- В** числова міра, що характеризує ступінь об'єктивної можливості настання події
- Г** властивість настання випадкової події
- Д** інша відповідь

1.9. Для довільної випадкової події

А	Б	В	Г	Д
$P(A) < 1$	$0 < P(A) < 1$	$P(A) \geq 0$	$0 \leq P(A) \leq 1$	інше

1.10. Геометричну ймовірність обчислюють за такою формулою

А	Б	В	Г	Д
$P(A) = \frac{m(A)}{m(\Omega)}$	$m(\Omega) \cdot m(A)$	$P(A) = \frac{m(\Omega)}{m(A)}$	$m(A)$	інше

1.6. Запитання для самоперевірки

- 1.1. Дайте означення таких основних понять теорії ймовірностей: "випробування", "випадкова подія", "імовірність випадкової події".
- 1.2. Назвіть аксіоми, на яких ґрунтується теорія ймовірностей.
- 1.3. Сформулюйте властивості, притаманні ймовірності події.
- 1.4. Дайте визначення вірогідної, неможливої та випадкової подій.
- 1.5. Назвіть критерій, за яким визначають малоймовірні події.
- 1.6. Сформулюйте критерій, за яким розрізняють сумісні й несумісні події. Наведіть приклад їхнього зображення за допомогою діаграми Венна – Ейлера.

- 1.7. Дайте означення повної групи несумісних випадкових подій.
- 1.8. Поясніть, чи відрізняється повна група несумісних випадкових подій від простору елементарних подій.
- 1.9. Дайте означення суми випадкових подій.
- 1.10. Наведіть приклад зображення суми несумісних подій за допомогою діаграми Венна – Ейлера.
- 1.11. Наведіть приклад зображення суми сумісних подій за допомогою діаграми Венна – Ейлера.
- 1.12. Дайте означення добутку випадкових подій. Наведіть приклад його зображення за допомогою діаграми Венна – Ейлера.
- 1.13. Дайте означення різниці двох випадкових подій. Наведіть приклад її зображення за допомогою діаграми Венна – Ейлера.
- 1.14. Поясніть, чому є кілька означень імовірності випадкової події.
- 1.15. Дайте означення статистичної ймовірності.
- 1.16. Дайте означення класичної ймовірності. Поясніть це тлумачення за допомогою діаграми Венна – Ейлера.
- 1.17. Дайте означення геометричної ймовірності. Поясніть це означення за допомогою геометричної побудови.
- 1.18. Визначте межі використання класичного означення ймовірності.
- 1.19. Поясніть, яких умов потрібно дотримуватися під час побудови діаграм Венна – Ейлера для того, щоб їх можна було застосовувати для безпосереднього обчислення ймовірності.

1.7. Висновки за підрозділом 1

Після вивчення цього підрозділу здобувач вищої освіти має:

- вільно володіти формулами комбінаторики та розуміти, у яких випадках треба використовувати той чи той вид сполучень;
- уміти розв'язувати задачі, використовуючи класичне та геометричне визначення ймовірності;
- надавати геометричну інтерпретацію випадкових подій і дій над ними за допомогою діаграм Венна – Ейлера.

Література: [2; 3; 6; 7; 13].

2. Основні теореми теорії ймовірностей, їхня економічна інтерпретація

2.1. Мета та компетентності

Метою вивчення підрозділу є формування системи теоретичних знань, практичних умінь і навичок з основних теорем теорії ймовірностей та умов їхнього використання для обчислення ймовірності випадкових подій.

Компетентності, що формують під час вивчення підрозділу:

- знання сутності теорем множення та додавання ймовірностей;
- знання означення умовної ймовірності;
- знання означення повної ймовірності події й апостеріорної ймовірності гіпотез за формулою Баєса;
- уміння застосовувати основні формули комбінаторики та теореми теорії ймовірностей для обчислення ймовірності випадкової події;
- уміння здійснювати класифікацію умов застосування теорем і аксіом теорії ймовірностей та ухвалювати рішення щодо застосування певного підходу до обчислення ймовірності.

2.2. Основні теоретичні відомості

Під час побудови будь-якої теорії після того, як названо об'єкти дослідження, їхні основні відношення, а також аксіоми, яким підлягають ці об'єкти та їхні відношення, можна переходити до розгляду теорем.

Слід розглянути основні теореми теорії ймовірностей.

Теореми множення ймовірностей

Ймовірність одночасної появи двох подій (ймовірність добутку подій) дорівнює добутку ймовірності появи однієї з подій на умовну ймовірність іншої події в припущенні, що перша подія вже відбулася:

$$P(AB) = P(A) \cdot P(B|A),$$

де $P(A)$ – ймовірність події A (безумовна ймовірність події A);

$P(B|A)$ – ймовірність появи події B за умови, що подія A вже відбулася (умовна ймовірність події B).

$$\text{Або} \quad P(AB) = P(B) \cdot P(A|B),$$

де $P(B)$ – імовірність події B (безумовна ймовірність події B);

$P(A|B)$ – імовірність появи події A за умови, що подія B вже відбулася (умовна ймовірність події A).

Якщо ймовірність появи події B не залежить від того, відбулася чи ні подія A , то такі події є взаємно незалежними, і для кожної випадкової події ймовірність її появи за умови, що перша з подій уже відбулася, дорівнює безумовній імовірності: $P(A|B) = P(A)$ та $P(B|A) = P(B)$ і теорема множення ймовірностей набирає такої форми:

$$P(AB) = P(A) \cdot P(B).$$

Наслідком теореми множення ймовірностей є теорема, що дозволяє визначити ймовірність появи хоча б однієї з випадкових подій.

Теорема. Імовірність появи хоча б однієї з випадкових подій $A_1, A_2, \dots, A_n$, які є незалежними в сукупності, дорівнює різниці між одиницею та добутком ймовірностей протилежних подій $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_n$.

Якщо $P(A_i) = p_i$ ($i = \overline{1, n}$), то, відповідно, ймовірність появи протилежної події позначають $P(\bar{A}_i) = 1 - p_i = q_i$. Згідно із цими позначеннями, теорема про ймовірність появи хоча б однієї з випадкових подій набирає такої форми:

$$P(A) = 1 - q_1 \cdot q_2 \cdot \dots \cdot q_n.$$

Окремим випадком теореми про визначення ймовірності появи хоча б однієї події є формула для обчислення ймовірності появи хоча б однієї події в разі, якщо для всіх подій, що розглядають, ймовірність їхньої появи є однаковою.

Нехай для кожної з випадкових подій $A_1, A_2, \dots, A_n$ ймовірність появи становить $P(A_i) = p$.

Тоді для протилежної події $P(\bar{A}) = 1 - p = q$ і $P(A) = 1 - q^n$.

Теореми додавання ймовірностей

Ймовірність суми несумісних випадкових подій A та B дорівнює сумі їхніх безумовних ймовірностей:

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

Ймовірність появи однієї із двох сумісних подій дорівнює сумі їхніх ймовірностей, за винятком ймовірності їхнього добутку:

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

Формула повної ймовірності

Нехай подія A може відбуватися лише за умови появи однієї з несумісних випадкових подій (гіпотез) $B_1, B_2, \dots, B_n$, що утворюють повну групу подій, і ймовірність кожної з подій цієї групи $P(B_i)$ відомо, а також є відомими умовні ймовірності події A в разі, якщо реалізують певну гіпотезу, тобто $P(A|B_i)$, де $i = \overline{1, n}$.

Тоді ймовірність події A визначають за такою формулою:

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \dots + P(B_n)P(A|B_n).$$

Формула Бесса

Нехай подія A настає лише за умови появи однієї з несумісних випадкових подій (гіпотез) $B_1, B_2, \dots, B_n$, що утворюють повну групу випадкових подій. Якщо припустити, що подія A вже відбулася, і потрібно визначити ймовірність того, що подія A відбулася саме завдяки реалізації гіпотези B_i , тоді:

$$P(B_i|A) = \frac{P(B_i) \cdot P(A|B_i)}{\sum_{i=1}^n P(B_i) \cdot P(A|B_i)}, \quad i = \overline{1, n}.$$

2.3. Приклади розв'язування задач

Приклад 2.1. Ймовірність складання іспиту здобувачем вищої освіти на "відмінно" дорівнює 0,3; на "добре" – 0,45; на "незадовільно" – 0,1; що він не з'явиться на іспит – 0,05.

Яка ймовірність того, що здобувач вищої освіти складе іспит?

Розв'язання. Очевидно, щоб скласти іспит, здобувач вищої освіти має скласти іспит на "відмінно", "добре" або на "задовільно". Імовірність отримати "задовільно" дорівнює

$$P = 1 - (0,3 + 0,45 + 0,1 + 0,05) = 0,1.$$

За теоремою додавання ймовірностей, імовірність скласти іспит дорівнює $P = 0,3 + 0,45 + 0,1 = 0,85$.

Приклад 2.2. Імовірність улучання в мішень за одного пострілу першим стрільцем дорівнює 0,8, а другим – 0,9.

Визначмо ймовірність того що: а) обидва стрільці влучать у мішень; б) тільки перший стрілець улучить у мішень; в) тільки другий стрілець улучить у мішень; г) один стрілець улучить у мішень; ґ) жоден стрілець не влучить у мішень; д) хоча б один стрілець улучить у мішень.

Розв'язання. Нехай подія A – це ціль, у яку влучив перший стрілець, а подія B – ціль, у яку влучив другий стрілець.

За умовою $P(A) = 0,8$; $P(B) = 0,9$, тоді ймовірності протилежних подій дорівнюють $P(\bar{A}) = 0,2$; $P(\bar{B}) = 0,1$.

Отже:

а) $P(A \text{ та } B) = P(AB) = P(A) \cdot P(B) = 0,8 \cdot 0,9 = 0,72$;

б) $P(A \text{ та } \bar{B}) = P(A\bar{B}) = P(A) \cdot P(\bar{B}) = 0,8 \cdot 0,1 = 0,08$;

в) $P(\bar{A} \text{ та } B) = P(\bar{A}B) = P(\bar{A}) \cdot P(B) = 0,2 \cdot 0,9 = 0,18$;

г) $P(\bar{A} \cdot B \text{ або } A \cdot \bar{B}) = P(\bar{A} \cdot B) + P(A \cdot \bar{B}) = 0,18 + 0,08 = 0,26$;

ґ) $P(\bar{A} \text{ та } \bar{B}) = P(\bar{A} \cdot \bar{B}) = P(\bar{A}) \cdot P(\bar{B}) = 0,2 \cdot 0,1 = 0,02$;

д) $P(\text{хоча б один}) = 1 - P(\bar{A} \cdot \bar{B}) = 1 - 0,02 = 0,98$.

Приклад 2.3. Для деякої місцевості середня кількість ясних днів у липні дорівнює 25.

Яка ймовірність того, що перші два дні липня будуть ясними?

Розв'язання. Імовірність того, що першого липня буде ясний день (подія A), дорівнює $P(A) = \frac{25}{31}$.

Імовірність того, що другого липня буде ясний день (подія B) за умови, що першого липня був ясний день, тобто умовна ймовірність події B дорівнює $P(B|A) = \frac{24}{30} = \frac{4}{5}$.

Тоді ймовірність того, що перші два дні липня будуть ясними, за теоремою множення ймовірностей залежних подій дорівнює

$$P(AB) = P(A) \cdot P(B|A) = \frac{25}{31} \cdot \frac{4}{5} = \frac{20}{31}.$$

Приклад 2.4. Імовірність того, що подія настане хоча б один раз у трьох незалежних випробуваннях, дорівнює 0,936.

Визначмо ймовірність появи події в одному випробуванні.

Розв'язання. Імовірність появи події хоча б один раз у трьох незалежних випробуваннях дорівнює $P(A) = 1 - q^3$, де $q = 1 - p$.

За умовою $P(A) = 0,936$.

Тоді $0,936 = 1 - q^3$, звідси $q^3 = 0,064$, а $q = 0,4$ і шукана ймовірність $p = 1 - 0,4 = 0,6$.

Приклад 2.5. Імовірність того, що деталь визнають нестандартною, дорівнює 0,1.

Визначмо, яку кількість деталей потрібно буде взяти, щоб з ймовірністю 0,998 можна було стверджувати, що хоча б одна з них буде стандартною.

Розв'язання. Нехай n – кількість деталей. Тоді ймовірність того, що всі деталі є нестандартними, дорівнює $(0,1)^n$. Імовірність визначення хоча б однієї стандартної деталі дорівнює $1 - (0,1)^n$, за умовою 0,998, звідси $1 - (0,1)^n = 0,998 \Rightarrow (0,1)^n = 0,002 \Rightarrow n = 2,7 \approx 3$ (n – ціле число).

Приклад 2.6. На склад надходять деталі із трьох верстатів. Перший верстат дає 20 %, другий – 30 %, третій – 50 % усіх деталей,

що надходять на складання. Перший верстат дає 0,2 % браку, другий – 0,3 %, третій – 0,1 %.

Визначмо ймовірність того, що до складу надійшла бракована деталь.

Розв'язання. Позначмо через B_1 подію "деталь виготовлено на першому верстаті", B_2 – "деталь виготовлено на другому верстаті", B_3 – "деталь виготовлено на третьому верстаті".

Тоді, за умовою задачі, маємо:

$$P(B_1) = 0,2, P(B_2) = 0,3, P(B_3) = 0,5.$$

Перевірка: $0,2 + 0,3 + 0,5 = 1$.

Нехай подія A – "на склад надійшла бракована деталь".

Тоді $P(A|B_1) = 0,002$; $P(A|B_2) = 0,003$; $P(A|B_3) = 0,001$.

За формулою повної ймовірності, буде:

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3).$$

$$P(A) = 0,2 \cdot 0,002 + 0,3 \cdot 0,003 + 0,5 \cdot 0,001 = 0,0018.$$

Приклад 2.7. У першій шухляді 5 білих і 10 чорних куль, у другій – 3 білі й 7 чорних. Із другої шухляди до першої переклали 1 кулю, а потім із першої шухляди навмання вийняли 1 кулю.

Яка ймовірність того, що вона є білою?

Розв'язання. Можливі дві гіпотези: B_1 – "перекладено білу кулю"; B_2 – "перекладено чорну кулю".

$$\text{Тоді } P(B_1) = \frac{3}{10}, \quad P(B_2) = \frac{7}{10}.$$

$$\text{Перевірка: } \frac{3}{10} + \frac{7}{10} = 1.$$

Нехай подія A – "вийнято білу кулю з першої шухляди" після перекладання до неї кулі із другої шухляди.

Очевидно, що:

$$P(A|B_1) = \frac{6}{16}, \quad P(A|B_2) = \frac{5}{16}.$$

За формулою повної ймовірності, буде:

$$P(A) = \frac{3}{10} \cdot \frac{6}{16} + \frac{7}{10} \cdot \frac{5}{16} = \frac{53}{160}.$$

Приклад 2.8. У першій шухляді 12 свердел, із них 8 використаних, а в другій – 8 свердел, із них 4 використано. Із першої шухляди навмання перекладають до другої 2 свердла.

Визначмо ймовірність того, що вийняте після цього із другої шухляди свердло ще жодного разу не використовували.

Розв'язання. Можливі три гіпотези: B_1 – "із першої шухляди до другої перекладено 2 невикористані свердла"; B_2 – "із першої шухляди до другої перекладено 1 використане й 1 невикористане свердло"; B_3 – "із першої шухляди до другої перекладено 2 використані свердла".

Ймовірності цих гіпотез дорівнюють:

$$P(B_1) = \frac{4}{12} \cdot \frac{3}{11} = \frac{1}{11},$$

$$P(B_2) = \frac{4}{12} \cdot \frac{8}{11} + \frac{8}{12} \cdot \frac{4}{11} = \frac{16}{33},$$

$$P(B_3) = \frac{8}{12} \cdot \frac{7}{11} = \frac{14}{33}.$$

$$\text{Перевірка: } P(B_1) + P(B_2) + P(B_3) = \frac{1}{11} + \frac{16}{33} + \frac{14}{33} = \frac{3+16+14}{33} = 1.$$

Нехай подія A – "свердло, що вийняли із другої шухляди після перекладання, буде жодного разу не використаним".

Тоді умовні ймовірності події A за умови, що відбулися події B_1 , B_2 , B_3 , відповідно, дорівнюють:

$$P(A|B_1) = \frac{6}{10};$$

$$P(A|B_2) = \frac{5}{10};$$

$$P(A|B_3) = \frac{4}{10}.$$

За формулою повної ймовірності, буде:

$$P(A) = \frac{1}{11} \cdot \frac{6}{10} + \frac{16}{33} \cdot \frac{5}{10} + \frac{14}{33} \cdot \frac{4}{10} = \frac{77}{165}.$$

Приклад 2.9. У першій шухляді 20 білих куль, у другій – 10 білих і 10 чорних куль, у третій – 20 чорних куль. Із вибраної навмання шухляди вийняли кулю. Вона є білою.

Яка ймовірність того, що ця куля з першої шухляди?

Розв'язання. Нехай гіпотеза B_1 – "кулю вийнято з першої шухляди", B_2 – "кулю вийнято із другої шухляди", B_3 – "кулю вийнято із третьої шухляди".

Тоді, за умовою, що ці події є рівноймовірними, буде:

$$P(B_1) = P(B_2) = P(B_3) = \frac{1}{3}.$$

Подія A – "вийнято білу кулю", тоді умовні ймовірності дорівнюють:

$$P(A|B_1) = 1, P(A|B_2) = \frac{10}{20} = \frac{1}{2}, P(A|B_3) = 0.$$

За формулою повної ймовірності, буде:

$$P(A) = \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot 0 = \frac{1}{2}.$$

Шукану ймовірність $P(B_1|A)$ слід визначати за формулою Баєса:

$$P(B_1|A) = \frac{P(B)P(A|B_1)}{P(A)} = \frac{\frac{1}{3} \cdot 1}{\frac{1}{2}} = \frac{2}{3}.$$

Як видно, після переоцінювання, тобто після того, як було враховано факт появи події A , ймовірність гіпотези B_1 зросла удвічі.

2.4. Вправи для самостійної роботи

2.1. У шухляді 10 білих, 15 червоних, 20 синіх і 25 чорних куль. Вийняли одну кулю.

Яка ймовірність, що вийнята куля мала такий колір: а) білий або синій; б) синій або червоний; в) білий, чорний або синій?

2.2. У шухляді 6 білих і 8 чорних куль. Узяли 2 кулі.

Яка ймовірність того, що вони обидві є білими?

2.3. Здобувач вищої освіти вивчив 20 питань із 25.

Визначте ймовірність, що він відповість на три питання.

2.4. Серед виробів, виготовлених на окремому заводі, 98 % відповідають стандарту. Із них 60 % – першого ґатунку і 40 % – другого. Узято один виріб.

Яка ймовірність того, що взятий виріб виявиться виробом першого ґатунку?

2.5. На 10 картках надруковано цифри 0, 1, 2, ..., 9.

Визначте ймовірність того, що 3 навмання взяті й поставлені в ряд картки становлять число 125.

2.6. У першій шухляді 2 білі й 10 чорних куль, у другій – 8 білих і 2 чорні. Із кожної шухляди взяли по кулі.

Яка ймовірність того, що: а) обидві кулі є білими; б) одна куля є білою, а друга – чорною; в) кулі одного кольору?

2.7. На вході в супермаркет встановлено відеокамеру для підрахунку кількості відвідувачів. Коли два відвідувачі входять до магазину разом і один іде попереду другого, то перший із них буде врахованим електронним пристроєм з імовірністю 0,97, другий – із імовірністю 0,93, а обидва – з імовірністю 0,92.

Визначте ймовірність того, що пристрій відсканує принаймні одного із двох відвідувачів, які входять разом.

2.8. Імовірність виграшу одного квитка лотереї дорівнює $1/20$.

Яка ймовірність того, що маючи 5 квитків, можна виграти: а) за всіма квитками; б) за жодним із квитків; в) хоча б за одним квитком?

2.9. Імовірність того, що споживач побачить рекламу певного продукту по телебаченню, дорівнює 0,07. Імовірність того, що споживач побачить рекламу того самого продукту на рекламному стенді, дорівнює 0,09.

Уважаючи, що обидві події є незалежними, визначте ймовірність того, що споживач побачить: а) обидві реклами; б) хоча б одну рекламу.

2.10. Туристична компанія продає тури для відпочинку за кордоном. Компанія створила список клієнтів у трьох регіонах і розсилає їм поштою каталог літніх турів. Менеджер компанії вважає, що ймовірність того, що компанія не дістане відгуку на розіслані пропозиції з кожного регіону, дорівнює 0,3.

Обчисліть ймовірність того, що компанія дістане відповідь хоча б з одного регіону.

2.11. Білет містить три питання. Ймовірність того, що здобувач вищої освіти відповість на перше і друге питання дорівнює 0,9; на третє – 0,8.

Визначте ймовірність того, що здобувач вищої освіти складе іспит, якщо для цього потрібно відповісти: а) на всі питання; б) хоча б на два питання.

2.12. Ймовірність улучання в мішень за одного пострілу дорівнює 0,2. Спортсмен на змаганні здійснив 10 пострілів.

Визначте ймовірність улучання в ціль.

2.13. У майстерні стоять 4 токарські верстати. Ймовірність того, що в заданий момент кожен із верстатів працює, дорівнює 0,7.

Визначте ймовірність того, що в заданий момент працює хоча б один верстат.

2.14. Ймовірність улучання в ціль за одного пострілу двох спортсменів дорівнює 0,38.

Визначте ймовірність улучання в ціль за одного пострілу першого спортсмена, якщо для другого спортсмена ця ймовірність дорівнює 0,8.

2.15. Вироби перевіряє експерт на нестандартність. Ймовірність того, що виріб є стандартним, дорівнює 0,9.

Визначте ймовірність того, що із двох перевірених виробів тільки один є стандартним.

2.16. У шухляді 10 куль, серед яких 6 білих. Навмання витягають 4 кулі.

Яка ймовірність того, що всі кулі виявляться білими?

2.17. Ймовірність хоча б одного влучання стрільцем за трьох пострілів дорівнює 0,875.

Визначте ймовірність улучання за одного пострілу.

2.18. Імовірність улучання в ціль для окремого стрільця дорівнює 60 %. Зроблено 5 пострілів.

Яка ймовірність того, що: а) усі 5 пострілів виявляться вдалими; б) стрілець жодного разу не влучить у ціль; в) стрілець улучить у ціль хоча б один раз; г) скільки пострілів достатньо зробити, щоб імовірність улучання в ціль досягла 95 %?

2.19. Є 4 шухляди. У першій шухляді є 1 біла і 1 чорна куля, у другій – 2 білі та 3 чорні кулі, у третій – 3 білі й 5 чорних, у четвертій – 4 білі та 7 чорних куль. Вибирають навмання шухляду та виймають із неї 1 кулю.

Яка ймовірність того, що вийнята куля є білою?

2.20. Дві однакові на вигляд коробки містять: одна – 3 червоні та 2 чорні олівці, друга – 2 червоні та 4 чорні олівці. Із навмання взятої коробки виймають 1 олівець.

Яка ймовірність того, що він виявиться чорним?

2.21. На заводі болти виготовляють на 3 верстатах, що виробляють, відповідно, 25, 30 та 45 % усієї кількості болтів. У продукції верстатів брак становить 4, 3 і 2 %, відповідно.

Яка ймовірність того, що болт, випадково взятий з усієї продукції, що надійшла, виявиться бракованим?

2.22. Для компанії, що будує термінали для аеропортів, можливість укласти угоду у країні A дорівнює 0,6; можливість укласти його в країні B дорівнює 0,4. Імовірність того, що угоди будуть укладеними і в країні A , і в країні B , дорівнює 0,18.

Визначте ймовірність того, що компанія укладе угоду: а) хоча б в одній країні; б) лише в одній країні.

2.23. На склад надходять деталі із трьох автоматичних ліній. Перша лінія дає 0,3 % браку, друга – 0,2 % браку, третя – 0,4 % браку.

Визначте ймовірність потрапляння на склад бракованої деталі, якщо з першої лінії надходить 1 000 деталей, із другої – 2 000 деталей, із третьої – 2 500 деталей.

2.24. Із коробки, що містить 4 білі та 2 чорні кулі, перекладено 1 кулю в коробку, що містить 5 білих і 8 чорних куль. Після цього із другої коробки було взято 1 кулю.

Яка ймовірність того, що вона буде чорною?

2.25. Під час набору спеціалістів в агенції розглядають спеціалістів трьох рівнів: провідних, старших та молодших – у відношенні 1 : 3 : 6. У процесі реалізації проекту провідний спеціаліст демонструє креативні ідеї з імовірністю 0,9; старший – 0,5; молодший – 0,2.

Яка ймовірність того, що проект буде складатися із креативних розробок?

2.26. У відділі *A* інституту працюють 5 інженерів і 3 старші інженери, а у відділі *B* – 8 інженерів і 2 старші інженери. Із відділу *A* у відділ *B* перевели 1 співробітника.

Визначте ймовірність того, що 3 співробітники, випадково вибрані з нового складу відділу *A*, є старшими інженерами.

2.27. У першій шухляді 10 куль, із них 2 є чорними; у другій – 20 куль, із них 16 є чорними. Із кожної шухляди витягли по 1 кулі, а потім із цих 2 куль навмання взяли одну.

Яка ймовірність того, що вона є білою?

2.28. У першій бригаді 5 робітників мають I розряд і 4 – II розряд. У другій бригаді 6 робітників мають I розряд і 7 – II розряд. Із першої бригади до другої переведено 2 робітники.

Визначте ймовірність того, що 2 робітники, навмання узяті з нового складу другої бригади, мають I розряд.

2.29. Компанія планує укласти угоду на постачання сільськогосподарського обладнання в одну із країн, що розвиваються. Якщо основний конкурент компанії не буде одночасно претендувати на укладання угоди, то можливість її укладання оцінюють як 0,55; в іншому випадку – як 0,35. За оцінками експертів компанії, ймовірність того, що конкурент висуне свої пропозиції щодо укладання угоди, дорівнює 0,40.

Визначте ймовірність укладання угоди цією компанією.

2.30. У партії 600 лампочок: 200 виготовлено на першому заводі, 250 – на другому, 150 – на третьому. Ймовірність того, що лампочка є стандартною, для першого заводу дорівнює 0,97; для другого – 0,91; для третього – 0,93. Навмання взято 1 лампочку, що є стандартною.

Яка ймовірність того, що її виготовлено на першому заводі?

2.31. Серед деталей, що надходять до складу, із першого верстата є 0,1 % бракованих, із другого – 0,2 %, із третього – 0,25 %, із четвертого – 0,5 %. Продуктивність їх відноситься як 4 : 3 : 2 : 1, відповідно. Узята навмання деталь виявилася бракованою.

Яка ймовірність того, що її виготовлено: а) на першому верстаті; б) на другому верстаті; в) на третьому верстаті; г) на четвертому верстаті? Як перевірити правильність обчислень?

2.32. Є 5 коробок: дві з них містять по 2 білі та 3 чорні кулі, дві – по 1 білій і 3 чорні кулі й одна – 4 білі та 1 чорну кулю. Із навмання взятої коробки вийняли кулю. Вона є білою.

Яка ймовірність того, що кулю вийняли з коробки із 4 білими й 1 чорною кулею?

2.33. У рекламній агенції працює три групи дизайнерів: перша група обслуговує 30 компаній, друга – 20, третя – 50. Протягом одного місяця кошти, витрачені на рекламу дизайнерами першої групи, повертають до 45 % компаній, другої – до 50 %, третьої – до 30 %. Упродовж місяця компанія окупила кошти, що було витрачено на рекламу.

Визначте ймовірність того, що цю компанію обслуговує друга група дизайнерів.

2.34. Групу спортсменів становлять 2 відмінні, 2 гарні й 4 посередні стрільці. Імовірність улучання в мішень (за одного пострілу) для відмінного, гарного та посереднього стрільців дорівнює, відповідно, 0,9; 0,8 і 0,7. Навмання вибраний стрілець вистрілив двічі; було зазначено 1 влучення й 1 промах.

Яким, найімовірніше, був цей стрілець: відмінним, гарним чи посереднім?

2.35. На молокозаводі вершкове масло фасують на двох технологічних лініях. Імовірність виходу продукції, що відповідає ДСТУ з першої лінії дорівнює 0,93, а із другої – 0,95. Навмання взятий пакет масла відповідає ДСТУ.

Установіть лінію, на якій, найімовірніше, було фасовано цей пакет.

2.36. Дослідженнями психологів встановлено, що чоловіки й жінки по-різному реагують на певні життєві ситуації.

Результати досліджень показали, що 75 % жінок позитивно реагують на ці ситуації, тим часом як 45 % чоловіків реагують на них негативно. 25 жінок і 15 чоловіків заповнили анкету, у якій відобразили своє ставлення до пропонованих ситуацій. Випадково витягнута анкета містить негативну реакцію.

Визначте ймовірність того, що її заповнював чоловік.

2.37. До міської лікарні надходить у середньому 80 % хворих із різними формами інфекції, 15 % – із захворюваннями серцево-судинної системи та 5 % – із захворюваннями нервової системи. Імовірність повного одужання хворого на інфекційні захворювання дорівнює 0,9; на захворювання серцево-судинної системи – 0,5 та на захворювання нервової системи – 0,2. Хворий, який надійшов до лікарні, одужав.

Визначте ймовірність того, що цей хворий мав: а) інфекційне захворювання; б) захворювання серцево-судинної системи; в) захворювання нервової системи. Зробіть перевірку визначених результатів.

2.5. Тестові завдання

2.1. Формула додавання ймовірностей несумісних подій має такий вигляд

А $P(A + B) = P(A) + P(B)$

Б $P(A + B) = P(A) + P(B) - P(AB)$

В $P(A \cdot B) = P(A) \cdot P(B)$

Г інша відповідь

2.2. Формула додавання ймовірностей сумісних подій має такий вигляд

А $P(A + B) = P(A) + P(B)$

Б $P(A + B) = P(A) + P(B) - P(AB)$

В інша відповідь

2.3. Дві події називають протилежними, якщо

А вони є несумісними

Б вони є сумісними та становлять повну групу

В вони є несумісними та становлять повну групу

2.4. Дві події називають незалежними, якщо

А імовірність сумісної появи цих подій дорівнює добутку ймовірності однієї з них та умовної ймовірності другої

Б імовірність появи однієї з таких подій дорівнює сумі їхніх імовірностей

В імовірність однієї з них не залежить від появи другої

Г імовірність однієї з них змінюється під час появи другої

2.5. Імовірність сумісної появи двох залежних подій дорівнює

А	Б	В
$P(AB) = P(A) \cdot P(B A)$	$P(AB) = P(A) + P(B)$	$P(AB) = P(A) \cdot P(B)$

2.6. Імовірність появи хоча б однієї з випадкових подій обчислюють за такою формулою

А $P(A) = 1 - q_1 \cdot q_2 \cdot \dots \cdot q_n$

Б $P(A) = 1 + q_1 \cdot q_2 \cdot \dots \cdot q_n$

В $P(A) = 1 - q_1 \cdot q_2$

2.7. Формулу повної ймовірності записують у такому вигляді

А $P(A) = P(B_i) \cdot P(A|B_i), i = \overline{1, n}$

Б $P(A) = \sum_{i=1}^n P(B_i) \cdot P(A|B_i), i = \overline{1, n}$

В $P(A) = \sum_{i=1}^n P(B_i) \cdot P(B_i|A), i = \overline{1, n}$

Г інша відповідь

2.8. Формула Баєса має такий вигляд

А $P(B_i|A) = \frac{\sum_{i=1}^n P(B_i) \cdot P(A|B_i)}{P(B_i) \cdot P(A|B_i)}, i = \overline{1, n}$

Б $P(A) = \sum_{i=1}^n P(B_i) \cdot P(A|B_i), i = \overline{1, n}$

$$\text{В } P(B_i|A) = \frac{P(B_i) \cdot P(A|B_i)}{\sum_{i=1}^n P(B_i) \cdot P(A|B_i)}, i = \overline{1, n}$$

Г інша відповідь

2.9. Формули Баєса дозволяють

А переоцінити ймовірність гіпотези за умови, що випадкова подія A здійснилася

Б переоцінити ймовірність гіпотези за умови, що випадкова подія A ще не здійснилася

В інше

2.10. Після переоцінювання всіх гіпотез $\sum_{i=1}^n P(B_i|A)$ дорівнює

А	Б	В	Г
0	1	2	інша відповідь

2.6. Запитання для самоперевірки

2.1. Які події називають взаємно незалежними?

2.2. Дайте означення умовної ймовірності та поясніть, як її визначають.

2.3. Поясніть, чим відрізняється умовна ймовірність від безумовної.

2.4. Сформулюйте теорему множення ймовірностей.

2.5. Наведіть формулу для визначення ймовірності добутку двох сумісних подій.

2.6. Сформулюйте теорему додавання ймовірностей.

2.7. Визначте, чи можна формулу для додавання ймовірностей несумісних випадкових подій уважати теоремою, та обґрунтуйте висновки.

2.8. Наведіть формулу для визначення суми трьох сумісних подій і розкрийте зміст кожної події, що входить до складу цієї суми.

2.9. Наведіть основні положення про ймовірність появи хоча б однієї з випадкових подій.

2.10. За формулою визначення ймовірності появи хоча б однієї події поясніть, на які події слід було б зважати як на сприятливі.

2.11. Поясніть, які події під час визначення повної ймовірності вважають гіпотезами. Яким вимогам відповідають гіпотези?

2.12. Наведіть формулу повної ймовірності.

2.13. Назвіть теоретичні положення, на основі яких виводять формулу повної ймовірності.

2.14. Поясніть різницю між апіорною та апостеріорною ймовірностями. Як ви вважаєте, чи завжди їхні значення мають відрізнятися одне від одного?

2.15. Сформулюйте теорему гіпотез (для визначення апостеріорної ймовірності).

2.16. Доведіть, що сума апостеріорних імовірностей має дорівнювати одиниці.

2.17. Поясніть, за яким принципом будують дерево станів, яку інформацію воно відображає і як за його допомогою обчислюють розв'язок задач щодо визначення повної ймовірності події.

2.7. Висновки за підрозділом 2

Після вивчення цього підрозділу здобувач вищої освіти має:

знати означення взаємно незалежних подій;

знати умови використання теорем додавання ймовірностей;

знати умови використання теорем множення ймовірностей;

уміти розв'язувати задачі, використовуючи теореми додавання та множення ймовірностей;

чітко розрізняти випадки, коли треба застосовувати формулу повної ймовірності та формули Баєса;

уміти переоцінити ймовірність появи події після того, як стає відомим результат випробування;

уміти зробити перевірку результатів розрахунків;

уміти обчислювати ймовірність того, що випадкова подія відбудеться хоча б один раз.

Література: [2; 3; 6; 7; 13].

3. Схема незалежних випробувань

3.1. Мета та компетентності

Метою вивчення підрозділу є надання уявлення про повторні випробування за схемою Бернуллі та набуття навичок з обчислення ймовірності появи події в цих випробуваннях певну кількість разів.

Компетентності, що формують під час вивчення підрозділу:

уміння виконувати обчислення за формулою Бернуллі та застосовувати довідкові таблиці значень функцій Гаусса та Лапласа для визначення ймовірностей за асимптотичними теоремами;

застосовування формули Бернуллі, локальної й інтегральної теорем Муавра – Лапласа, теореми Пуассона для обчислення ймовірності появи випадкової події певну кількість разів під час випробувань за схемою Бернуллі;

здатність здійснювати аналіз умов, у яких проводять повторні випробування, та рекомендувати використання формули Бернуллі й асимптотичних теорем для обчислення ймовірності того, що випадкова подія відбудеться певну кількість разів у процесі незалежних випробувань.

3.2. Основні теоретичні відомості

Формула Бернуллі

Якщо ймовірність появи події A в кожному з n незалежних випробувань є постійною і дорівнює p , то ймовірність $P_n(m)$ того, що подія настане саме m разів, обчислюють за формулою Бернуллі:

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

де $q = 1 - p$.

Ймовірність настання події A принаймні один раз під час n незалежних випробувань за схемою Бернуллі, дорівнює:

$$P_n(m \geq 1) = 1 - q^n,$$

де $q = 1 - p$.

Імовірність того, що подія A під час проведення n незалежних випробувань за схемою Бернуллі настане не менше ніж m_1 і не більше ніж m_2 разів, визначають за такою формулою:

$$P_n(m_1 \leq m \leq m_2) = \sum_{m_1}^{m_2} P_n(m).$$

Число m_0 появи події A в n незалежних випробуваннях називають *найімовірнішим*, якщо ймовірність появи події A є найбільшою. Це число визначають із такої нерівності:

$$np - q \leq m_0 \leq np + p.$$

Якщо ймовірність $P(A)$ появи події A в кожному з n незалежних випробувань є постійною і дорівнює p , то можна визначити ймовірність $P_n(m)$ того, що подія A настане саме m разів.

Формула Пуассона

У разі, якщо $n \rightarrow \infty$, а $p \rightarrow 0$ для обчислення $P_n(m)$ застосовують асимптотичну формулу Пуассона:

$$P_n(m) \approx \frac{\lambda^m}{m!} e^{-\lambda},$$

де $\lambda = np$.

Значення функції $P_n(m) \approx \frac{\lambda^m}{m!} e^{-\lambda}$ для різних значень $\lambda = np$ наведено в додатку А.

Рекомендують застосувати формулу Пуассона, якщо $\lambda \leq 10$.

Локальна теорема Лапласа

Якщо ймовірність p появи події A в кожному з n незалежних випробувань є постійною, а кількість випробувань n достатньо великою,

то ймовірність $P_n(m)$ того, що подія A настане m разів у цих випробуваннях, обчислюють за такою формулою:

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

$$\text{де } x = \frac{m - np}{\sqrt{npq}};$$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} - \text{функція Гаусса.}$$

Таблицю значень функції $\varphi(x)$ наведено в додатку Б, причому $\varphi(-x) = \varphi(x)$. Треба вважати, що для значень $x > 4$ $\varphi(x) \approx 0$.

За допомогою *інтегральної теореми Лапласа* визначають наближено ймовірність того, що в n незалежних випробуваннях подія A настане від m_1 до m_2 разів (включно):

$$P_n(m_1 \leq m \leq m_2) \approx \Phi(x_2) - \Phi(x_1),$$

$$\text{де } x_1 = \frac{m_1 - np}{\sqrt{npq}};$$

$$x_2 = \frac{m_2 - np}{\sqrt{npq}};$$

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

Значення функції Лапласа $\Phi(x)$ наведено в додатку В, $\Phi(-x) = -\Phi(x)$. Треба вважати, що для $x > 5$ $\Phi(x) \approx 0,5$.

Інтегральна теорема Лапласа дає незначну похибку, якщо $npq \geq 20$. Інакше треба застосовувати формулу Бернуллі та теорему додавання ймовірностей, а якщо p або q наближається до нуля, то застосовують наближену формулу Пуассона.

Наслідком інтегральної формули Лапласа є формула ймовірності відхилення відносної частоти $\frac{m}{n}$ появи події A в n незалежних випробуваннях від її ймовірності p на величину $\varepsilon > 0$ (за абсолютною величиною):

$$P\left(\left|\frac{m}{n} - p\right| \leq \varepsilon\right) \approx 2\Phi(t),$$

де $t = \varepsilon \sqrt{\frac{n}{pq}}$.

Нерівність $\left|\frac{m}{n} - p\right| \leq \varepsilon$ є рівносильною такій нерівності:

$$p - \varepsilon \geq \frac{m}{n} \geq p + \varepsilon \text{ або } n(p - \varepsilon) < \frac{m}{n} \leq n(p + \varepsilon).$$

За допомогою формули $P\left(\left|\frac{m}{n} - p\right| \leq \varepsilon\right) \approx 2\Phi(t)$ можна також визначити кількість випробувань n , за яких із заданою ймовірністю p відхилення відносної частоти $\frac{m}{n}$ події A від імовірності p її появи за абсолютною величиною не буде перевищувати задане додатне число ε :

$$n = \frac{t^2 pq}{\varepsilon^2}.$$

Із тієї самої формули можна за заданої ймовірності p і кількості випробувань n обчислити похибку ε :

$$\varepsilon = t \sqrt{\frac{pq}{n}}.$$

Формула $p - \varepsilon \geq \frac{m}{n} \geq p + \varepsilon$ означає, що відносна частота $\frac{m}{n}$ міститься в симетричному інтервалі.

Якщо частота $\frac{m}{n}$ події A міститься в межах від α до β (включно),

то:

$$P_n(\alpha \leq \frac{m}{n} \leq \beta) \approx \Phi(x_2) - \Phi(x_1),$$

$$\text{де } x_1 = \frac{(\alpha - p)\sqrt{n}}{\sqrt{pq}};$$

$$x_2 = \frac{(\beta - p)\sqrt{n}}{\sqrt{pq}}.$$

3.3. Приклади розв'язування задач

Приклад 3.1. Схожість насіння заданої рослини становить 90 %.

Визначмо ймовірність того, що із 4 посіяних насінин зійдуть: а) 3;
б) не менше від 3.

Розв'язання:

а) за умовою $n = 4$; $m = 3$; $p = 0,9$.

За формулою Бернуллі, шукана ймовірність дорівнює:

$$P_4(3) = C_4^3 \cdot 0,9^3 \cdot 0,1 \approx 0,2916;$$

б) слід визначити ймовірність події, яка полягає в тому, що із 4 насінин зійде 3 або 4.

Шукана ймовірність буде дорівнювати:

$$P_4(3) + P_4(4) \approx 0,2916 + 0,9^4 \approx 0,9477.$$

Приклад 3.2. Імовірність виготовлення нестандартної деталі дорівнює 0,05.

Визначмо ймовірність того, що із 5 взятих навмання деталей 4 є стандартними.

Розв'язання. За умовою $n = 5$; $m = 4$; $p = 0,95$; $q = 0,05$.

Імовірність того, що із 5 взятих навмання деталей 4 є стандартними можна обчислити за формулою Бернуллі:

$$P_5(4) = C_5^4 \cdot 0,95^4 \cdot 0,05 \approx 0,2036.$$

Приклад 3.3. Імовірність виробництва стандартної деталі дорівнює 0,98.

Визначмо найімовірніше число стандартних деталей серед 625.

Розв'язання. Якщо $n = 625$; $p = 0,98$; $q = 0,02$, то

$$612,48 \leq m_0 \leq 613,48,$$

звідси $m_0 = 613$.

Приклад 3.4. Імовірність улучання в мішень за одного пострілу дорівнює 0,2.

Визначмо найімовірніше число влучань у ціль за 14 пострілів.

Розв'язання. Якщо $n = 14$; $p = 0,2$; $q = 0,8$, тоді найімовірніше число влучень у мішень за 14 пострілів обчислюють за допомогою такої нерівності:

$$14 \cdot 0,2 - 0,8 \leq m_0 \leq 14 \cdot 0,2 + 0,2;$$

$$2 \leq m_0 \leq 3.$$

Отже, визначено два найімовірніші числа $m_0 = 2$ і $m_0 = 3$.

Приклад 3.5. Імовірність того, що банкомат спрацює правильно під час подавання в нього купюри, дорівнює 0,97.

Скільки потрібно подати купюр у банкомат, щоб найімовірніше число випадків правильної роботи банкомата становило 100?

Розв'язання. За умовою $p = 0,97$; $q = 0,03$; $m_0 = 100$.

Число n визначають із такої подвійної нерівності:

$$n \cdot 0,97 - 0,03 \leq 100 \leq n \cdot 0,97 + 0,97;$$

$$\text{або } \begin{cases} n \cdot 0,97 \leq 100,03; \\ n \cdot 0,97 \geq 99,03; \end{cases} \quad \text{тоді } \begin{cases} n \leq 103,12; \\ n \geq 102,09. \end{cases}$$

Звідси $n = 103$.

Приклад 3.6. Імовірність народження хлопчика дорівнює 0,515.

Визначмо ймовірність того, що із 200 народжених дітей хлопчиків і дівчат буде порівно.

Розв'язання. За умовою $n = 200$; $m = 100$; $p = 0,515$; $q = 0,485$.

За локальною теоремою Лапласа:

$$P_{200} \approx \frac{1}{\sqrt{200 \cdot 0,515 \cdot 0,485}} \varphi(x) = \frac{1}{7,068} \varphi(x);$$

$$x = \frac{100 - 200 \cdot 0,515}{7,068} = -0,42; \quad \varphi(-0,42) = \varphi(0,42) = 0,3653.$$

$$\text{І тоді } P_{200}(100) \approx \frac{1}{7,068} \cdot \varphi(-0,42) \approx \frac{0,3653}{7,068} \approx 0,052.$$

Приклад 3.7. Імовірність того, що буде виявлено помилку на сторінці книги, дорівнює 0,002. Перевірено 500 сторінок.

Визначмо ймовірність того, що з помилкою будуть: а) 3 сторінки; б) від 3 до 5 сторінок.

Розв'язання:

а) за умовою $p = 0,002$; $n = 500$.

Звідси $\lambda = n \cdot p = 0,002 \cdot 500 = 1$.

Тоді ймовірність $P(3) \approx \frac{1^3}{3!} e^{-1} = 0,0613$;

б) для обчислення ймовірності $P(3 \leq m \leq 5)$ слід використати теорему додавання:

$$P(3 \leq m \leq 5) = P(3) + P(4) + P(5) \approx 0,0613 + \frac{1^4}{4!} e^{-1} + \frac{1^5}{5!} e^{-1} = 0,0613 + 0,0153 + 0,0031 = 0,0797.$$

Приклад 3.8. Визначмо ймовірність того, що серед 1 000 новонароджених хлопчиків буде: а) від 465 до 550 включно; б) не менше за 550; в) не більше за 465, якщо ймовірність народження хлопчика дорівнює 0,5.

Розв'язання:

а) за умовою $n = 1\,000$; $p = 0,5$; $q = 0,5$; $m_1 = 465$; $m_2 = 550$.

Тоді ймовірність $P(465 \leq m \leq 550) \approx \Phi(x_2) - \Phi(x_1)$.

Обчислімо:

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

За таблицею значень функції Лапласа (див. додаток В) визначмо $\Phi(x_1)$ і $\Phi(x_2)$ та обчислімо шукану ймовірність:

$$P(465 \leq m \leq 550) \approx \Phi(3,16) - \Phi(-2,21) = \Phi(3,16) + \Phi(2,21) = 0,9856;$$

б) за умовою $m_1 = 550$; $m_2 = 1000$.

$$P(550 \leq m \leq 1000) \approx \Phi(x_2) - \Phi(x_1);$$

Обчислімо:

$$x_1 = \frac{550 - 1000 \cdot 0,5}{\sqrt{1000 \cdot 0,5 \cdot 0,5}} = 3,16; \quad x_2 = \frac{1000 - 1000 \cdot 0,5}{\sqrt{1000 \cdot 0,5 \cdot 0,5}} \approx 32.$$

За таблицею значень функції Лапласа (див. додаток В) визначмо $\Phi(3,16)$ та $\Phi(32)$.

Отже,

$$P(550 \leq m \leq 1000) \approx \Phi(32) - \Phi(3,16) = 0,5 - 0,4992 = 0,0008;$$

в) за умовою $m_1 = 0$; $m_2 = 464$.

Очевидно, що

$$P(0 \leq m \leq 464) = 1 - P(465 \leq m \leq 1000) \approx 1 - (0,9856 + 0,0008) = 0,0136.$$

Приклад 3.9. Імовірність улучання в мішень за одного пострілу дорівнює 0,7.

Яка ймовірність того, що за 500 пострілів частота влучання в мішень відхилиться від імовірності p не більше ніж на 0,04 (за абсолютною величиною)?

Розв'язання. За умовою $n = 500$; $p = 0,7$; $q = 0,3$; $\varepsilon = 0,04$.

Тоді шукану ймовірність обчислюють за такою формулою:

$$P\left(\left|\frac{m}{500} - 0,7\right| \leq 0,04\right) \approx 2\Phi\left(0,04\sqrt{\frac{500}{0,3 \cdot 0,7}}\right) = 2\Phi(1,95) = 0,9488.$$

Отже, те, що частота влучання в мішень буде міститися в межах від 0,66 до 0,74 ($0,7 \pm 0,04$), можна стверджувати з імовірністю 0,948 8.

Приклад 3.10. Імовірність улучання в мішень за одного пострілу дорівнює 0,3.

Скільки потрібно зробити пострілів, щоб з імовірністю 0,996 відхилення частоти влучань від імовірності $p = 0,3$ не перевищувало 0,04 (за абсолютною величиною)?

Розв'язання. Для розв'язання задачі потрібно використати таку формулу:

$$P\left(\left|\frac{m}{n} - p\right| \leq \varepsilon\right) = 2\Phi(t),$$

де $t = \varepsilon \sqrt{\frac{n}{pq}}$.

За умовою $2\Phi(t) = 0,996$, тоді $\Phi(t) = 0,498$. Значення t слід визначити за таблицею значень функції $\Phi(x)$ (див. додаток В): $t = 2,88$.

Звідси $\varepsilon \sqrt{\frac{n}{pq}} = 2,88$ або $0,04 \sqrt{\frac{n}{0,3 \cdot 0,7}} = 2,88$;

тоді n буде дорівнювати: $n = 0,21 \cdot 72^2 \approx 1\,089$.

Отже, слід зробити 1 089 пострілів.

Приклад 3.11. Імовірність улучання в мішень за кожного з 800 пострілів дорівнює 0,3.

У яких межах буде міститися частота влучань, щоб імовірність невиходу за межі становила 0,962 4?

Розв'язання. За умовою $2\Phi(t) = 0,962\,4$.

За таблицею значень функції $\Phi(x)$ (див. додаток В) t набуває значення: $t = 2,08$.

Як і в попередньому прикладі для обчислення ε слід розв'язати таке рівняння:

$$\varepsilon \sqrt{\frac{n}{pq}} = 2,08.$$

Звідси за $n = 800$; $p = 0,3$; $q = 0,7$; $\varepsilon = 0,0337$.

Отже, за 800 пострілів з імовірністю 0,926 4 можна стверджувати, що частота влучання в мішень буде міститися в межах від 0,266 3 до 0,333 7 ($0,3 \pm 0,033\ 7$).

3.4. Вправи для самостійної роботи

3.1. У партії кількість великих деталей є удвічі більшою від кількості дрібних.

Яка ймовірність того, що серед узятих навмання 10 деталей виявиться 6 великих?

3.2. Імовірність улучання в мішень за кожного окремого пострілу $p = 0,8$.

Яка ймовірність того, що за 5 пострілів у мішень буде влучено?

3.3. Бавовна містить 10 % коротких волокон.

Яка ймовірність того, що в навмання взятому пучку із 5 волокон буде не більше від 2 коротких?

3.4. Схожість насіння деякої рослини оцінюють імовірністю $p = 0,8$.

Визначте ймовірність того, що з 8 посіяних насінин зійде не менше від 6.

3.5. Імовірність появи події в окремому випробуванні дорівнює 0,75.

Яка ймовірність того, що за восьмикратного повторення випробування ця подія відбудеться понад 6 разів?

3.6. Монету кидають 5 разів.

Визначте ймовірність того, що "герб" випаде хоча б 2 рази.

3.7. 30 % виробів підприємства – це продукція вищого сорту. Хтось придбав 6 виробів.

Яка ймовірність того, що 4 вироби з придбаних є виробами вищого сорту?

3.8. Імовірність своєчасного прибуття потяга на станцію дорівнює 0,8.

Яка ймовірність того, що із 4 очікуваних потягів: а) 3 прибудуть вчасно; б) не менше від 3 потягів прибудуть вчасно; в) не більше від 3 потягів прибудуть вчасно; г) принаймні 1 із потягів прибуде вчасно?

3.9. Що ймовірніше: виграти в рівносильного суперника 3 партії із 4 чи 5 із 8?

3.10. Із 10 пострілів стрілець у середньому влучає в мішень 8 разів.

Яка ймовірність того, що із 3 незалежних пострілів він точно 2 рази влучить у ціль?

3.11. Як, використовуючи формулу Бернуллі, обчислити ймовірність того, що в n випробуваннях Бернуллі деяка подія не настане жодного разу? Чи можна обчислити цю ймовірність, не звертаючись до формули Бернуллі?

3.12. Як, використовуючи формулу Бернуллі, обчислити ймовірність того, що в випробуваннях Бернуллі "успіх" відбудеться хоча б один раз? Чи можна обчислити цю ймовірність, не звертаючись до формули Бернуллі?

3.13. Туристична компанія "Мандрівник" кожного літа відправляє туристів у країни Європи. Досліджено, що ймовірність замовлення туру в Італію становить 0,7.

Визначте ймовірність того, що з 10 клієнтів компанії 6 замовлять тур в Італію.

3.14. У результаті багаторічних спостережень установлено, що ймовірність випадання дощу 1 жовтня в заданому місті дорівнює $\frac{1}{7}$.

Визначте найімовірніше число дощових днів 1 жовтня в заданому місті за 40 років.

3.15. Дані тривалої перевірки якості стандартних деталей показали, що брак у середньому становить 7,5 %.

Визначте найімовірніше число цілком справних деталей у партії із 39 штук.

3.16. За якої кількості пострілів найімовірніше число влучань дорівнює 16, якщо ймовірність улучання за одного пострілу є 0,7?

3.17. Чому дорівнює ймовірність p появи події в кожному із 39 незалежних випробувань, якщо найімовірніше число появи події в цих випробуваннях дорівнює 25?

3.18. Маркетологами підприємства було оцінено, що ймовірність реалізації кожної із пляшок води "Миргородська" буде становити 0,7.

Скільки пляшок води "Миргородська" потрібно підготувати для продажу, щоб найімовірніше число реалізованих пляшок становило 750?

3.19. Зроблено 6 пострілів по об'єкту. Імовірність улучань в об'єкт за одного пострілу дорівнює 0,8.

Визначте: а) найімовірніше число влучань; б) імовірність найімовірнішого числа влучань; в) імовірність того, що об'єкт буде зруйновано, якщо для цього досить хоча б двох улучань.

3.20. У компанії працює 600 співробітників. Імовірність звільнення за власним бажанням кожного з них за певний період становить 0,25.

Визначте ймовірність того, що за певний період звільняться за власним бажанням 30 співробітників.

3.21. Імовірність улучання за одного пострілу дорівнює 0,6.

Визначте: а) імовірність того, що за 5 400 пострілів у мішень буде влучено 3 240 разів; б) найімовірніше число влучань за 5 400 пострілів.

3.22. Верстат-автомат дає 75 % виробів I сорту.

Визначте найімовірніше число виробів I сорту з партії із 320 штук і ймовірність цього числа.

3.23. У шухляді є 3 чорні та 4 білі кулі. Кулі виймають таким способом, що кожну вийняту кулю повертають у шухляду.

Визначте ймовірність того, що за 250 спроб білу кулю буде вийнято 100 разів.

3.24. Визначте ймовірність одночасної зупинки 30 машин зі 100 тих, що працюють, якщо ймовірність зупинки однієї машини дорівнює 0,2.

3.25. Імовірність бракованого виробу дорівнює 0,2.

Визначте ймовірність того, що в партії із 400 виробів буде 104 бракованих.

3.26. Банк "Пумб" проводить опитування щодо якості обслуговування клієнтів. Практика показує, що до 65 % клієнтів можна додзвонитися з першого разу.

Визначте ймовірність того, що із 500 клієнтів до 330 оператор додзвониться з першого разу?

3.27. На обліку служби зайнятості району стоїть 600 безробітних. Відомо, що перша пропозиція працевлаштування задовольняє тільки 40 %.

Визначте ймовірність того, що працевлаштованими з першого разу буде: 1) не більше ніж 30 осіб; 2) від 300 до 400 осіб.

3.28. Підручник видано накладом 1 000 прим. Імовірність виготовлення бракованого примірника підручника дорівнює 0,004.

Яка ймовірність того, що наклад має 5 бракованих книг?

3.29. На одній із технологічних ліній підприємства масло розфасовують у пачки. Відомо, що на кожні 1 000 пачок три не відповідають стандарту.

Яка ймовірність того, що із 500 відправлених на реалізацію пачок не будуть відповідати стандарту: 1) три; 2) хоча б три.

3.30. Частка браку за деякого технічного процесу становить 0,2 %.

Визначте ймовірність того, що в партії з 1 000 виробів бракованих буде 25.

3.31. За результатами перевірок фіскальними органами виявлено, що малі підприємства регіону допускають порушення фінансової дисципліни з ймовірністю 0,001.

Визначте ймовірність того, що із 5 000 зареєстрованих у регіоні малих підприємств допускають порушення фінансової дисципліни більше ніж 5 підприємств.

3.32. Завод відправив на базу 5 000 якісних виробів. Імовірність того, що в дорозі виріб буде зіпсованим, дорівнює 0,000 2.

Визначте ймовірність того, що: а) на базу надійдуть тільки 3 зіпсовані вироби; б) на базу надійдуть від 2 до 3 зіпсованих виробів.

3.33. Робітниця обслуговує 800 веретен. Імовірність обриву пряжі на кожному з веретен протягом деякого проміжку часу дорівнює 0,005.

Визначте: а) найімовірніше число обривів; б) ймовірність того, що протягом проміжку часу буде не більше ніж 10 обривів.

3.34. Імовірність появи події в кожному із 2 100 незалежних випробувань дорівнює 0,7.

Визначте ймовірність того, що подія настане: а) від 1 470 до 1 500 разів; б) не менше ніж 1 470 разів; в) не більше ніж 1 469 разів.

3.35. Імовірність того, що деталь не пройшла перевірку ВТК, дорівнює 0,3.

Визначте ймовірність того, що серед 100 деталей пройдуть перевірку ВТК не більше ніж 79.

3.36. Стрілець улучає в ціль із ймовірністю $p = 0,75$.

Яка ймовірність того, що за 100 пострілів кількість улучань буде:
а) не меншою за 71 і не більшою за 80; б) не меншою за 70 і не більшою за 80; в) не меншою за 81; г) не більшою за 70.

3.37. На кожні 20 штампованих виробів із пластмаси припадає в середньому 3 дефектні.

Визначте ймовірність того, що із 50 взятих навмання виробів понад 42 будуть без дефекту.

3.38. Стрілець уражає мішень з ймовірністю $p = 0,75$.

Яка ймовірність того, що за 400 пострілів відносна частота влучень відхилиться від величини p менше ніж на 0,035?

3.39. Стрілець уражає ціль з ймовірністю $p = 0,75$.

У яких межах має бути кількість m улучень за 400 пострілів, щоб ймовірність невиходу за ці межі дорівнювала 0,97?

3.40. Стрілець уражає мішень з ймовірністю $p = 0,75$.

Скільки пострілів має зробити стрілець, щоб ймовірність того, що відхилення відносної частоти влучання від величини p буде меншим ніж 0,035; дорівнювала 0,95?

3.41. Ймовірність появи події в кожному з 10 000 незалежних випробувань дорівнює 0,75.

Визначте ймовірність того, що відносна частота появи події відхилиться від ймовірності за абсолютною величиною не більше ніж на 0,000 1.

3.42. У коробці є білі й чорні кулі у відношенні 4 : 1. Після виймання кулі реєструють її колір, а кулю повертають до коробки.

Чому дорівнює кількість виймання, за якої з ймовірністю 0,972 2 можна чекати, що абсолютна величина відхилення відносної частоти появи білої кулі від її ймовірності буде не більшою ніж 0,01?

3.43. Ймовірність появи події в кожному з 900 незалежних випробувань дорівнює 0,5.

Визначте таке додатне число ε , щоб з ймовірністю 0,769 8 абсолютна величина відхилення відносної частоти появи події від її ймовірності p не перевищувала ε .

3.44. Імовірність появи події в кожному з 10 000 незалежних випробувань дорівнює 0,75.

Визначте таке додатне число ε , щоб з імовірністю 0,979 абсолютна величина відхилення відносної частоти появи події від її ймовірності p не перевищувала ε .

3.45. Імовірність улучання в мішень за одного пострілу дорівнює 0,6.

Визначте:

а) імовірність того, що за 12 пострілів ціль буде вражено 7 разів;

б) імовірність того, що за 200 пострілів ціль буде вражено:
1) не менше ніж 111 і не більше ніж 130 разів; 2) не менше ніж 110 і не більше ніж 130 разів; 3) не більше ніж 110 разів; 4) не менше ніж 115 разів;

в) імовірність того, що за 600 пострілів частота влучання в мішень відхилиться від імовірності 0,6 за абсолютною величиною не більше ніж на 0,03;

г) межі влучань у мішень за 600 пострілів, щоб імовірність невиходу за ці межі дорівнювала 0,993;

г') кількість пострілів по мішені, за якою з імовірністю 0,993 можна чекати, що відхилення частоти влучань від імовірності 0,6 за абсолютною величиною не перевищать 0,03.

3.46. Гральну кістку кидають 80 разів.

Визначте з імовірністю 0,9973 межі, у яких буде наставати кількість m появи цифри "6".

3.47. Імовірність того, що клієнт в кафе замовить першу страву дорівнює 0,75.

Визначте ймовірність того, що в кафе зі 100 клієнтів першу страву замовлять: 1) не менше ніж 80 і не більше ніж 90 клієнтів; 2) не менше ніж 80 клієнтів; 3) не більше ніж 79 клієнтів.

3.48. Місто щодня відвідує 1 000 туристів, які вдень ідуть обідати. Кожен із них вибирає для обіду один із двох міських ресторанів з однаковими ймовірностями та незалежно один від одного. Власник одного з ресторанів бажає, щоб із ймовірністю приблизно 0,98 усі туристи, що прийшли до його ресторану, могли там одночасно пообідати.

Скільки місць має бути для цього в його ресторані?

3.49. Контролер якості продукції перевіряє партію кетчупу, фасованого в пляшки. Імовірність того, що пляшку кетчупу буде забраковано дорівнює 0,05.

Яку кількість пляшок треба перевірити, щоб із ймовірністю 0,98 за абсолютною величиною відносна частота появи бракованих пляшок відхилялася від 0,98 не більше ніж на 0,02?

3.50. Під час надходження грошей у банк потрапляє в середньому 0,1 % фальшивих купюр різним номіналом.

Визначте ймовірність того, що з 10 000 купюр, які надійшли в банк, відхилення від указанного відсотка фальшивих за абсолютною величиною не перевищує 0,02.

3.5. Тестові завдання

3.1. Формула Бернуллі має такий вигляд

А	Б	В	Г
$P_n(k) \approx \frac{\lambda^k}{k!} e^{-\lambda}$	$P_n(k) = p^k q^{n-k}$	$P_n(k) = C_n^k p^k q^{n-k}$	$P_n(k) \approx \frac{\lambda^k}{k!}$

3.2. Формула Пуассона має такий вигляд

А	Б	В	Г
$P_n(k) \approx \frac{\lambda^k}{k!} e^{-\lambda}$	$P_n(k) = p^k q^{n-k}$	$P_n(k) = C_n^k p^k q^{n-k}$	$P_n(k) \approx \frac{\lambda^k}{k!}$

3.3. Імовірність того, що в n незалежних випробуваннях подія настане від k_1 до k_2 разів, обчислюють за допомогою

А локальної теореми Муавра – Лапласа

Б інтегральної теореми Лапласа

В формули Бернуллі

Г формули Пуассона

Д формули Баєса

3.4. Подію вважають малоїмовірною, якщо

А	Б	В	Г
$n \rightarrow 0, p \rightarrow 0$	$n \rightarrow \infty, p \rightarrow 0$	$n \rightarrow 0, p \rightarrow 1$	$n \rightarrow \infty, p > 0,1$

3.5. У випадках, якщо $n \leq 30$, а $p > 0,1$, слід використовувати

- А** формулу Пуассона
- Б** формулу Бернуллі
- В** формулу повної ймовірності
- Г** інтегральну формулу Лапласа
- Д** локальну формулу Муавра – Лапласа

3.6. Які із зазначених властивостей функції $\varphi(x)$ є правильними

- А** $\varphi(-x) = \varphi(x)$ та якщо $x > 4$, то $\varphi(x) \approx 0$
- Б** $\varphi(-x) = \varphi(x)$ та якщо $x > 0$, то $\varphi(x) \approx 1$
- В** $\varphi(-x) = \varphi(x)$ та якщо $x > 0$, то $\varphi(x) \approx 0$
- Г** $\varphi(-x) = -\varphi(x)$ та якщо $x > 4$, то $\varphi(x) \approx 0$

3.7. Які із зазначених властивостей функції $\Phi(x)$ є правильними

- А** $\Phi(-x) = \Phi(x)$ та якщо $x > 5$, то $\Phi(x) \approx 0,5$
- Б** $\Phi(-x) = -\Phi(x)$ та якщо $x > 0$, то $\Phi(x) \approx 0,5$
- В** $\Phi(-x) = -\Phi(x)$ та якщо $x > 5$, то $\Phi(x) \approx 0,5$
- Г** $\Phi(-x) = -\Phi(x)$ та якщо $x > 5$, то $\Phi(x) \approx 0$

3.8. Кількість випробувань n , за яких із заданою ймовірністю p

відхилення відносної частоти $\frac{m}{n}$ події A від імовірності p її появи

за абсолютною величиною не буде перевищувати задане додатне число ε , дорівнює

А	Б	В	Г
$n = \frac{\varepsilon^2 pq}{t^2}$	$n = \frac{t^2 pq}{\varepsilon^2}$	$n = \frac{tp^2 q}{\varepsilon^2}$	$n = \frac{t^2 p^2 q}{\varepsilon^2}$

3.9. За локальною теоремою Муавра – Лапласа ймовірність того, що в n незалежних випробуваннях подія настане k разів, приблизно дорівнює

А $P_n(k) \approx \frac{\varphi(x)}{\sqrt{npq}}$, де $x = \frac{k - np}{\sqrt{npq}}$

Б $P_n(k) \approx \frac{\varphi(x)}{\sqrt{npq}}$, де $x = \frac{k}{\sqrt{npq}}$

В $P_n(k) \approx \varphi(x)$, де $x = \frac{k - np}{\sqrt{npq}}$

Г $P_n(k) \approx \varphi(x)$, де $x = k - np$

3.10. Найімовірніше число появи події A в n незалежних випробуваннях m_0 міститься в інтервалі

А $np - q \leq m_0 \leq np + p$

Б $np - p \leq m_0 \leq np + q$

В $0 \leq m_0 \leq 1$

Г $np + q \leq m_0 \leq np - p$

3.6. Запитання для самоперевірки

3.1. Що таке "повторні випробування"? Наведіть приклади.

3.2. Яким умовам відповідає схема Бернуллі?

3.3. За яких умов слід застосовувати формулу Бернуллі?

3.4. Чи має обмеження формула Бернуллі за кількістю випробувань?

3.5. Дайте означення поняття "довірча ймовірність".

3.6. Наведіть формулу для визначення кількості випробувань, що із заданою надійністю забезпечує появу певної випадкової події хоча б один раз.

3.7. Чому формули Муавра – Лапласа та Пуассона називають асимптотичними?

3.8. Сформулюйте локальну теорему Муавра – Лапласа.

3.9. Як впливає ймовірність появи події в окремому випробуванні на точність обчислень за локальною теоремою Муавра – Лапласа?

3.10. Сформулюйте інтегральну теорему Муавра – Лапласа.

3.11. За допомогою якої формули можна визначити ймовірність відхилення частоти появи події від її ймовірності на величину, що не перевищує заздалегідь обумовленого значення?

3.12. Порівняйте можливості застосування формули Пуассона та локальної формули Муавра – Лапласа.

3.13. Наведіть формулу Пуассона.

3.14. Чому закон, для характеристики якого застосовують формулу Пуассона, називають законом рідкісних подій?

3.15. Наведіть приклади повторних випробувань, якщо для визначення ймовірності появи події в серії випробувань не можна застосовувати формулу Бернуллі.

3.7. Висновки за підрозділом 3

Після вивчення цього підрозділу здобувач вищої освіти має:

знати ознаки, за якими здійснюють класифікацію схеми повторних випробувань та особливості застосування асимптотичних теорем;

знати умови застосування формули Бернуллі, формули Пуассона, локальної та інтегральної формул Лапласа;

уміти виконувати обчислення за формулами Бернуллі та Пуассона;

застосовувати довідкові таблиці значень функцій Гаусса та Лапласа для визначення ймовірностей за локальною та інтегральною теоремами Лапласа.

Література: [2; 3; 7; 8; 13; 16].

4. Випадкові величини та їхня економічна інтерпретація

4.1. Мета та компетентності

Метою вивчення підрозділу є надання уявлення про типи випадкових величин і способи визначення їхніх законів розподілу; про функцію розподілу, її властивості та особливості побудови.

Компетентності, що формують під час вивчення підрозділу:

знання означення випадкової величини;

уміння будувати функцію розподілу випадкової величини;

знання основних числових характеристик випадкової величини та їхніх властивостей;

уміння визначати початкові й центральні теоретичні моменти.

4.2. Основні теоретичні відомості

Випадковою величиною називають величину, яка в результаті випробування може набути того чи того значення, але якого саме – невідомо. Випадкові величини позначають великими літерами, а їхні можливі значення – відповідними малими літерами латинського алфавіту, наприклад, $X(x_1, x_2, \dots, x_n)$.

Дискретною (перервною) випадковою величиною називають випадкову величину, яка набуває окремих ізольованих можливих значень із певними ймовірностями. Кількість можливих значень дискретної випадкової величини може бути скінченною або зліченною.

Неперервною називають випадкову величину, яка може набувати всіх значень із деякого скінченного чи нескінченного проміжку.

Законом розподілу (рядом розподілу) дискретної випадкової величини називають відповідність між можливими значеннями та їхніми ймовірностями. Його можна задати у вигляді таблиці (табл. 4.1).

Таблиця 4.1

Закон розподілу дискретної випадкової величини

X	x_1	x_2	x_3	...	x_n
P	p_1	p_2	p_3	...	p_n

У першому рядку табл. 4.1 розміщено можливі значення x_i , а в другому – відповідні ймовірності p_i , де $\sum_{i=1}^n p_i = 1$.

Закон розподілу дискретної випадкової величини можна задати аналітично (у вигляді формули): $P(X = x_i) = \varphi(x_i)$, а також графічно.

Для цього в прямокутній системі координат будують точки $M_1(x_1, p_1)$, $M_2(x_2, p_2)$, ..., $M_n(x_n, p_n)$ (x_i – можливі значення X , p_i – відповідні ймовірності) і з'єднують їх відрізками прямих.

Побудовану фігуру називають **многокутником розподілу** (рис. 4.1).

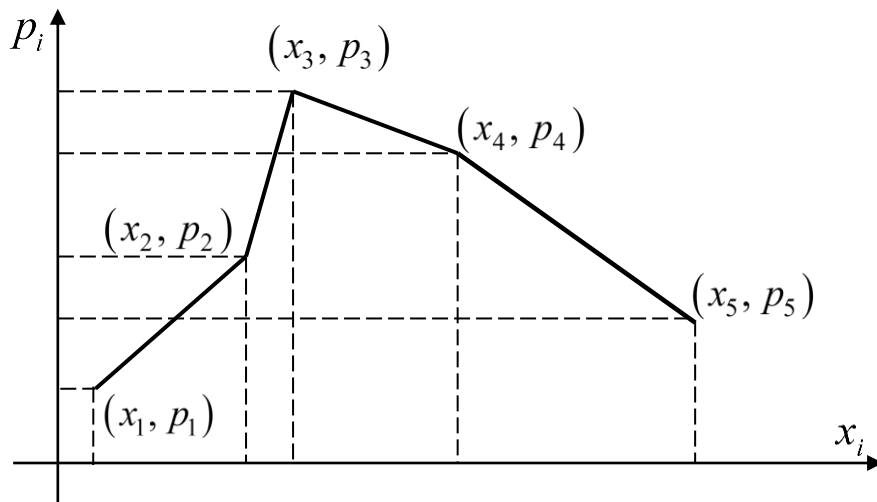


Рис. 4.1. Многокутник розподілу

Дискретну випадкову величину можна задати й функцією розподілу.

Функцією розподілу ймовірностей випадкової величини називають функцію $F(x)$, яка визначає ймовірність того, що випадкова величина X у результаті випробування набуде значення, яке менше ніж x :

$$F(x) = P(X < x) = \sum_{x_i < x} p_i.$$

Тут для кожного значення x становлять ймовірності тих значень x_i , які розташовано ліворуч від точки x .

Числові характеристики дискретної випадкової величини

Характеристикою середнього значення випадкової величини є математичне сподівання.

Математичним сподіванням дискретної випадкової величини називають суму добутків усіх її можливих значень на відповідні ймовірності:

$$M(X) = x_1 p_1 + x_2 p_2 + \dots + x_n p_n = \sum_{i=1}^n x_i p_i.$$

Якщо дискретна випадкова величина X набуває скінченної множини можливих значень, то $M(X) = \sum_{i=1}^{\infty} x_i p_i$, за умови, що ряд є збіжним абсолютно.

Математичне сподівання показує, яке значення випадкової величини слід очікувати в середньому під час випробувань.

Властивості математичного сподівання:

1) математичне сподівання сталої величини дорівнює цій сталій:

$$M(C) = C;$$

2) сталий множник можна виносити за знак математичного сподівання:

$$M(CX) = CM(X);$$

3) математичне сподівання суми скінченного числа випадкових величин дорівнює сумі їхніх математичних сподівань:

$$M(X_1 + X_2 + \dots + X_n) = M(X_1) + M(X_2) + \dots + M(X_n).$$

Наслідок – математичне сподівання різниці двох випадкових величин дорівнює різниці їхніх математичних сподівань:

$$M(X - Y) = M(X) - M(Y);$$

4) математичне сподівання добутку скінченного числа взаємно незалежних випадкових величин дорівнює добутку їхніх математичних сподівань:

$$M(X_1 \cdot X_2 \cdot \dots \cdot X_n) = M(X_1) \cdot M(X_2) \cdot \dots \cdot M(X_n).$$

Іншою важливою характеристикою випадкової величини є дисперсія, яка характеризує міру розсіювання значень випадкової величини навколо її математичного сподівання.

Дисперсією випадкової величини X називають математичне сподівання квадрата відхилення цієї величини від її математичного сподівання:

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

Для розрахунку використовують ще й таку формулу:

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

Властивості дисперсії:

1) дисперсія сталої дорівнює нулю:

$$D(C) = 0;$$

2) сталий множник можна виносити за знак дисперсії, підносячи його до квадрата:

$$D(CX) = C^2 D(X);$$

3) дисперсія суми попарно незалежних випадкових величин дорівнює сумі їхніх дисперсій:

$$D(X_1 + X_2 + \dots + X_n) = D(X_1) + D(X_2) + \dots + D(X_n);$$

4) дисперсія різниці двох незалежних випадкових величин дорівнює сумі їхніх дисперсій:

$$D(X - Y) = D(X) + D(Y).$$

Дисперсія має розмірність квадратних одиниць вимірюваної величини. Щоб обчислити розмірність міри розсіювання в одиницях вимірюваної величини, розглядають квадратний корінь із дисперсії:

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

який називають **середнім квадратичним відхиленням** (стандартним відхиленням).

В економічних задачах часто середнє квадратичне відхилення σ використовують як міру ризику.

*Додаткові числові характеристики розподілу
дискретної випадкової величини*

Початковим моментом k -го порядку випадкової величини X називають математичне сподівання величини X^k :

$$\nu_k = M(X^k).$$

Зокрема, початковий момент 1-го порядку дорівнює математичному сподіванню випадкової величини X :

$$\nu_1 = M(X);$$

початковий момент 2-го порядку:

$$\nu_2 = M(X^2).$$

Центральним моментом k -го порядку випадкової величини X називають математичне сподівання величини $(X - M(X))^k$.

$$\mu_k = M((X - M(X))^k).$$

Зокрема, центральний момент 1-го порядку дорівнює нулю:

$$\mu_1 = M(X - M(X));$$

центральний момент 2-го порядку дорівнює дисперсії:

$$\mu_2 = M(X - M(X))^2.$$

Центральні моменти зручніше обчислювати за формулами, що визначають центральні моменти через початкові:

$$\mu_2 = \nu_2 - \nu_1^2;$$

$$\mu_3 = \nu_3 - 3\nu_1\nu_2 + 2\nu_1^3;$$

$$\mu_4 = \nu_4 - 4\nu_1\nu_3 + 6\nu_1^2\nu_2 - 3\nu_1^4.$$

Однією з поширених додаткових характеристик розподілу є **коефіцієнт варіації**.

Він визначає у відсотках частку середнього квадратичного відхилення, що припадає на одиницю математичного сподівання:

$$V = \frac{\sigma}{M(X)} \cdot 100 \%.$$

Коефіцієнт варіації є зручною характеристикою, за якою можна оцінювати, за яким саме законом є розподіленою випадкова величина.

Разом із математичним сподіванням показниками центральних тенденцій є мода та медіана.

Модой випадкової величини ($Mo(X)$) називають те її значення, якому відповідає найбільше значення ймовірності.

Медіаною випадкової величини $Me(X)$ називають таке її значення, яке за кількістю значень випадкової величини ділить навпіл ряд розподілу.

Додатковими числовими характеристиками розподілу є також асиметрія та ексцес випадкової величини. Формально їх можна визначати і для дискретної випадкової величини, однак переважно їх застосовують як характеристики розподілу неперервної випадкової величини. Саме для визначення цих характеристик застосовують теоретичні моменти вищих порядків. Так, *асиметрію*, або *коефіцієнт асиметрії* розподілу, визначають таким співвідношенням:

$$A_s = \frac{\mu_3}{\sigma^3},$$

де μ_3 – центральний момент 3-го порядку;

σ – середнє квадратичне відхилення випадкової величини.

Ексцесом, або **коефіцієнтом ексцесу** випадкової величини, називають величину, що визначають за такою формулою:

$$E_k = \frac{\mu_4}{\sigma^4} - 3,$$

де μ_4 – центральний момент 4-го порядку;

σ – середнє квадратичне відхилення випадкової величини.

4.3. Приклади розв'язування задач

Приклад 4.1. Дискретну випадкову величину X задано законом розподілу (табл. 4.2).

Таблиця 4.2

Закон розподілу випадкової величини X

X	1	5	8	10
p	0,3	0,4	0,2	0,1

Побудуємо многокутник розподілу.

Розв'язання. У прямокутній системі координат на осі абсцис слід відкласти можливі значення випадкової величини, а на осі ординат – відповідні ймовірності.

Таким способом будують точки $M_1(1; 0,3)$, $M_2(5; 0,4)$, $M_3(8; 0,2)$, $M_4(10; 0,1)$.

Якщо з'єднати ці точки відрізками прямих, то буде многокутник розподілу (рис. 4.2).

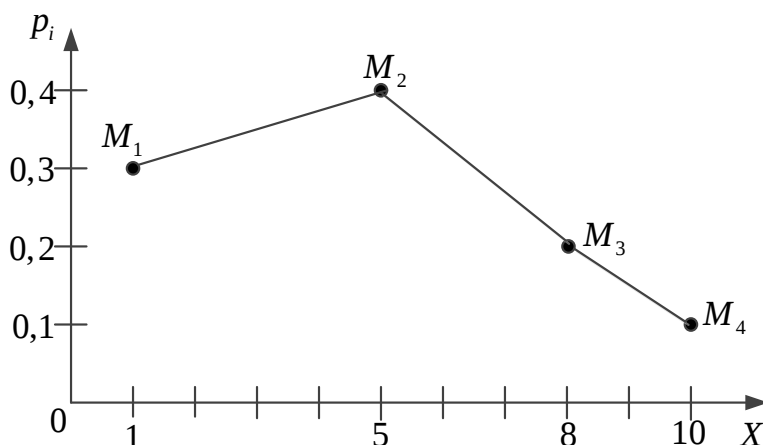


Рис. 4.2. Многокутник розподілу

Приклад 4.2. Два стрільці виконують по одному пострілу в мішень. Імовірність улучання для першого дорівнює 0,6, а для другого – 0,5.

Складімо закон розподілу кількості влучань у мішень.

Розв'язання. Дискретна випадкова величина X (кількість улучань у мішень) має такі значення:

$x_1 = 0$ (обидва стрільці не влучили в мішень),

$x_2 = 1$ (один улучив),

$x_3 = 2$ (обидва влучили).

Визначмо відповідні ймовірності.

Ймовірність улучання для першого $p_1 = 0,6$, а ймовірність промаху:
 $q_1 = 1 - 0,6 = 0,4$;

ймовірність улучення для другого $p_2 = 0,5$, а ймовірність промаху:
 $q_2 = 1 - 0,5 = 0,5$.

Тоді ймовірність того, що обидва стрільці не влучили в мішень:

$$P(X = 0) = q_1 \cdot q_2 = 0,4 \cdot 0,5 = 0,2;$$

ймовірність того, що один улучив:

$$P(X = 1) = p_1 q_2 + p_2 q_1 = 0,6 \cdot 0,5 + 0,6 \cdot 0,4 = 0,5;$$

ймовірність того, що обидва влучили:

$$P(X = 2) = p_1 \cdot p_2 = 0,6 \cdot 0,5 = 0,3.$$

Перевірка: $0,2 + 0,5 + 0,3 = 1$.

Ряд розподілу кількості влучань у мішень наведено в табл. 4.3.

Таблиця 4.3

Закон розподілу випадкової величини X

X	0	1	2
p	0,2	0,5	0,3

Приклад 4.3. Спортсмен стріляє в ціль до першого влучання, але встигає зробити не більше від 4 пострілів.

Складімо закон розподілу кількості пострілів, зроблених спортсменом, якщо ймовірність улучання в ціль за одного пострілу дорівнює 0,7.

Розв'язання. Кількість пострілів, зроблених спортсменом, може набувати таких значень: 1 (перший постріл улучає в ціль), 2 (перший постріл – промах, другий – улучання), 3 (два перші постріли – промахи, третій – улучання), 4 (перші три постріли – промахи, а четвертий – улучання або промах).

За умовою задачі, $P_4(1) = 0,7$.

За теоремою множення ймовірностей, буде:

$$P_4(2) = 0,3 \cdot 0,7 = 0,21;$$

$$P_4(3) = 0,3 \cdot 0,3 \cdot 0,7 = 0,063;$$

$$P_4(4) = 0,3 \cdot 0,3 \cdot 0,3 = 0,027$$

(тому що результат четвертого пострілу не має значення).

Закон розподілу шуканої випадкової величини X – кількість пострілів зручно надати у вигляді таблиці (табл. 4.4).

Таблиця 4.4

Закон розподілу випадкової величини X

Кількість пострілів	1	2	3	4
p_i	0,700	0,210	0,063	0,027

Перевірка: $0,700 + 0,210 + 0,063 + 0,027 = 1$.

Приклад 4.4. Визначмо функцію розподілу випадкової величини X , заданої законом розподілу (див. табл. 4.4).

Визначмо ймовірність того, що випадкова величина X набуде значення, не меншого за 1 і не більшого за 4.

Розв'язання. Випадкова величина X не набуває значень, менших за 1. Тому для $x \leq 1$ події $X < x$ є неможливими, а $F(x) = 0$.

Якщо $1 < x \leq 2$, то $F(x) = 0,700$, тому що X може набути тільки значення $x = 1$ з імовірністю $p = 0,700$.

Якщо $2 < x \leq 3$, то $F(x) = 0,700 + 0,210 = 0,910$, тому що X може набути одного зі значень $x = 1$ або $x = 2$ з імовірностями $p = 0,700$ та $p = 0,210$, що потребує застосування теореми додавання ймовірностей для обчислення шуканої ймовірності.

Аналогічно, якщо $3 < x \leq 4$, то $F(x) = 0,700 + 0,210 + 0,063 = 0,973$.

Якщо $x > 4$, то $F(x) = 1$, тому що подія $X \leq 1$ є достовірною і ймовірність її дорівнює 1.

Отже, за $x > 4$ $F(x) = 0,700 + 0,210 + 0,063 + 0,270 = 1$.

Тому функція розподілу випадкової величини X , а також її графік (рис. 4.3) мають такий вигляд:

$$F(x) = \begin{cases} 0, & \text{якщо } x \leq 1; \\ 0,700, & \text{якщо } 1 < x \leq 2; \\ 0,910, & \text{якщо } 2 < x \leq 3; \\ 0,973, & \text{якщо } 3 < x \leq 4; \\ 1, & \text{якщо } x > 4. \end{cases}$$

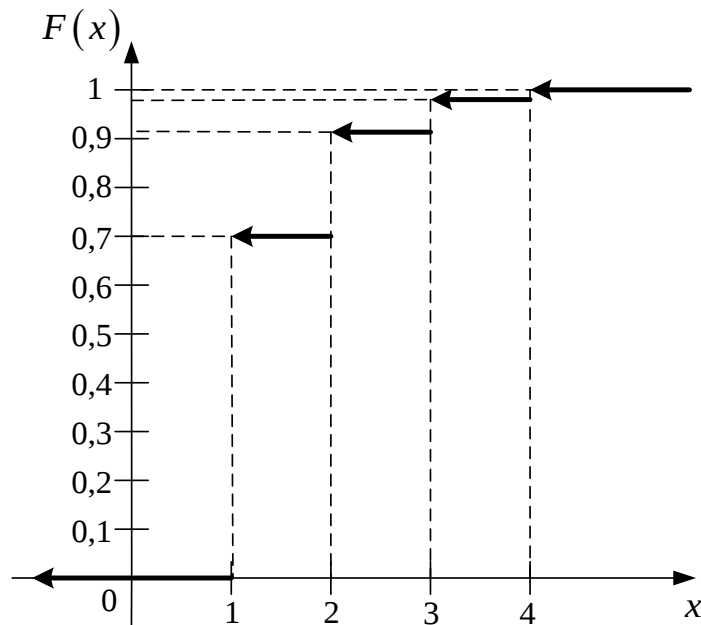


Рис. 4.3. Графік інтегральної функції розподілу

Імовірність того, що випадкова величина X набуде значення, не меншого за 1 і не більшого за 4, дорівнює:

$$P(1 \leq X \leq 4) = F(4) - F(1) = 0,973 - 0 = 0,973.$$

Приклад 4.5. Для випадкової величини X , заданої законом розподілу (табл. 4.5), обчислімо $M(X)$, $M(X^2)$, $M^2(X)$, $D(X)$.

Таблиця 4.5

Закон розподілу випадкової величини X

x_i	5	10	15
p_i	0,2	0,6	0,2

Розв'язання. $M(X) = \sum_{i=1}^n x_i p_i$.

Тоді $M(X) = 5 \cdot 0,2 + 10 \cdot 0,6 + 15 \cdot 0,2 = 10$.

Далі слід записати закон розподілу випадкової величини X^2 .
Це зручно зробити у вигляді таблиці (табл. 4.6).

Таблиця 4.6

Закон розподілу випадкової величини X^2

x_i^2	25	100	225
p_i	0,2	0,6	0,2

Тоді $M(X^2) = 25 \cdot 0,2 + 100 \cdot 0,6 + 225 \cdot 0,2 = 110$.

Дисперсію слід обчислити за такою формулою:

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

Отже, $D(X) = 110 - 10^2 = 10$.

Приклад 4.6. Випадкову величину X задано законом розподілу (табл. 4.7).

Таблиця 4.7

Закон розподілу випадкової величини X

x_i	1	2	3
p_i	p_1	p_2	p_3

Також задано $M(X) = 2,3$; $M(X^2) = 5,9$.

Визначмо p_1, p_2, p_3 .

Розв'язання. Очевидно, що

$$p_1 + p_2 + p_3 = 1;$$

$$M(X) = 1 \cdot p_1 + 2 \cdot p_2 + 3 \cdot p_3;$$

$$M(X^2) = 1 \cdot p_1 + 4 \cdot p_2 + 9 \cdot p_3.$$

Тоді слід розв'язати систему рівнянь:

$$\begin{cases} p_1 + p_2 + p_3 = 1; \\ p_1 + 2 \cdot p_2 + 3 \cdot p_3 = 2,3; \\ p_1 + 4 \cdot p_2 + 9 \cdot p_3 = 5,9. \end{cases}$$

Звідси $p_1 = 0,2$; $p_2 = 0,3$; $p_3 = 0,5$.

Приклад 4.7. Визначмо математичне сподівання і дисперсію випадкової величини X , заданої законом розподілу (табл. 4.8).

Таблиця 4.8

Закон розподілу випадкової величини X

X	2	5	7	10	12
p	0,1	0,2	0,4	0,2	0,1

Розв'язання. Математичне сподівання випадкової величини X буде дорівнювати:

$$M(X) = 2 \cdot 0,1 + 5 \cdot 0,2 + 7 \cdot 0,4 + 10 \cdot 0,2 + 12 \cdot 0,1 = 7,2.$$

Для обчислення дисперсії за означенням потрібно використати таку формулу:

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

Розрахунки зручно оформити у вигляді таблиці (табл. 4.9).

Таблиця 4.9

Розрахункова таблиця

x_i	p_i	$x_i p_i$	$x_i^2 p_i$	$x_i - 7,2$	$(x_i - 7,2)^2$	$(x_i - 7,2)^2 p_i$
2	0,1	0,2	0,4	-5,2	27,04	2,704
5	0,2	1,0	5,0	-2,2	4,84	0,968
7	0,4	2,8	19,6	-0,2	0,04	0,016
10	0,2	2,0	20,0	2,8	7,84	1,568
12	0,1	1,2	14,1	4,8	23,04	2,304
—	—	7,2	59,4	—	—	7,56

У перший стовпець таблиці запишімо значення випадкової величини, у другий – відповідні ймовірності. Для визначення математичного сподівання треба спочатку обчислити добуток $x_i p_i$ (третій стовпець).

Для цього числа першого стовпця слід помножити на відповідні числа

другого стовпця, а сума чисел $\left(\sum_{i=1}^n x_i p_i \right)$ цього стовпця і є математичним

сподіванням, а саме: $M(X) = \sum_{i=1}^n x_i p_i = 7,2$.

Для обчислення дисперсії спочатку визначмо $M(X^2) = \sum_{i=1}^n x_i^2 p_i$,

для цього заповнімо четвертий стовпець, де числа третього стовпця слід помножити на відповідні числа першого стовпця. Обчислимо суму чисел

цього стовпця і визначмо $M(X^2) = 59,4$.

Для обчислення дисперсії за означенням спочатку заповнімо п'ятий стовпець, де від кожного x_i віднімімо $M(X) = 7,2$.

Далі визначмо $(x_i - 7,2)^2$, тобто кожне число із п'ятого стовпця піднесімо до квадрата і запишімо результат у шостий стовпець. Тепер визначмо $(x_i - 7,2)^2 p_i$, для цього кожне число із шостого стовпця помножмо на відповідні числа другого стовпця і результат запишімо у сьомий стовпець. Сума чисел сьомого стовпця і є дисперсією, тобто $D(X) = 7,56$.

Тепер обчислимо дисперсію за такою формулою:

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

За табл. 4.9, $M(X^2) = \sum_{i=1}^n x_i^2 p_i = 59,4$.

Тоді $D(X) = 59,4 - (7,2)^2 = 7,56$.

Отже, за двома формулами дисперсії збіглися.

Приклад 4.8. Дискретну випадкову величину X задано законом розподілу (табл. 4.10).

Визначмо початкові моменти 1, 2, 3 і 4-го порядку.

Таблица 4.10

Закон розподілу випадкової величини X

X	2	4	5
p	0,3	0,6	0,1

Розв'язання:

початковий момент 1-го порядку $\nu_1 = M(X)$,

$$M(X) = x_1 \cdot p_1 + x_2 \cdot p_2 + x_3 \cdot p_3,$$

тому $\nu_1 = 2 \cdot 0,3 + 4 \cdot 0,6 + 5 \cdot 0,1 = 3,5$;

початковий момент 2-го порядку $\nu_2 = M(X^2)$,

$$M(X^2) = x_1^2 \cdot p_1 + x_2^2 \cdot p_2 + x_3^2 \cdot p_3,$$

тому $\nu_1 = 2^2 \cdot 0,3 + 4^2 \cdot 0,6 + 5^2 \cdot 0,1 = 13,3$;

початковий момент 3-го порядку $\nu_3 = M(X^3)$,

$$M(X^3) = x_1^3 \cdot p_1 + x_2^3 \cdot p_2 + x_3^3 \cdot p_3,$$

звідси $\nu_1 = 2^3 \cdot 0,3 + 4^3 \cdot 0,6 + 5^3 \cdot 0,1 = 53,3$;

початковий момент 4-го порядку $\nu_4 = M(X^4)$,

$$M(X^4) = x_1^4 \cdot p_1 + x_2^4 \cdot p_2 + x_3^4 \cdot p_3,$$

звідси $\nu_1 = 2^4 \cdot 0,3 + 4^4 \cdot 0,6 + 5^4 \cdot 0,1 = 220,9$.

Приклад 4.9. Дискретну випадкову величину X задано законом розподілу (табл. 4.11).

Таблиця 4.11

Закон розподілу випадкової величини X

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

Визначмо центральні моменти 1, 2, 3 та 4-го порядку.

Розв'язання. Центральний момент першого порядку дорівнює:

$$\mu_1 = M(X - M(X)).$$

Спочатку визначмо $M(X)$:

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

$$\mu_1 = (x_1 - M(X)) \cdot p_1 + (x_2 - M(X)) \cdot p_2 + (x_3 - M(X)) \cdot p_3.$$

Ураховуючи дані задачі, буде:

$$\mu_1 = (-1 - 1,4) \cdot 0,4 + (2 - 1,4) \cdot 0,3 + (4 - 1,4) \cdot 0,3 = 0,$$

що й підтвердили розрахунки.

Далі визначмо центральний момент 2-го порядку:

$$\mu_2 = M(X - M(X))^2 - \text{дисперсія};$$

$$\mu_2 = (x_1 - M(X))^2 \cdot p_1 + (x_2 - M(X))^2 \cdot p_2 + (x_3 - M(X))^2 \cdot p_3.$$

Використаймо такі дані задачі:

$$\mu_2 = (-1 - 1,4)^2 \cdot 0,4 + (2 - 1,4)^2 \cdot 0,3 + (4 - 1,4)^2 \cdot 0,3 = 4,44.$$

Тепер обчислімо центральний момент 3-го порядку:

$$\mu_3 = M(X - M(X))^3;$$

$$\mu_3 = (x_1 - M(X))^3 \cdot p_1 + (x_2 - M(X))^3 \cdot p_2 + (x_3 - M(X))^3 \cdot p_3.$$

Використовуючи умови задачі, буде:

$$\mu_3 = (-1 - 1,4)^3 \cdot 0,4 + (2 - 1,4)^3 \cdot 0,3 + (4 - 1,4)^3 \cdot 0,3 = 0,192.$$

Центральний момент 4-го порядку:

$$\mu_4 = M(X - M(X))^4,$$

$$\text{або } \mu_4 = (x_1 - M(X))^4 \cdot p_1 + (x_2 - M(X))^4 \cdot p_2 + (x_3 - M(X))^4 \cdot p_3.$$

Для цієї задачі буде:

$$\mu_4 = (-1 - 1,4)^4 \cdot 0,4 + (2 - 1,4)^4 \cdot 0,3 + (4 - 1,4)^4 \cdot 0,3 = 27,0192.$$

4.4. Вправи для самостійної роботи

4.1. У студентській групі 4 юнаки та 6 дівчат. На чергування випадково відібрано двох осіб.

Складіть закон розподілу кількості дівчат серед відібраних.

4.2. Імовірність виготовлення нестандартного виробу під час деякого технологічного процесу дорівнює 0,05. Із партії беруть виріб і перевіряють його якість. Якщо він буде нестандартним, випробування припиняють, а партію затримують. Якщо виріб буде стандартним, беруть наступний і так далі, але перевіряють не більше ніж п'ять виробів.

Складіть закон розподілу кількості перевірених виробів.

4.3. Комплекс із виробництва та фасування морозива становлять 4 лінії, що працюють незалежно. Імовірність виходу з ладу кожної лінії протягом місяця дорівнює 0,2.

Складіть закон розподілу дискретної випадкової величини X – кількості ліній, що вийшли з ладу протягом місяця.

4.4. Імовірність того, що банкомат під час введення коду спрацює правильно, становить 0,99. Клієнт має 3 спроби.

Складіть закон розподілу дискретної випадкової величини X – кількості спроб введення клієнтом коду в банкомат.

4.5. Дискретну випадкову величину задано законом розподілу (табл. 4.12).

Таблиця 4.12

Закон розподілу випадкової величини X

X	0	2	4	5
p	0,1	0,6	?	0,1

Визначте інтегральну функцію $F(x)$ і побудуйте її графік.

4.6. 3 урни, що містить 4 білі та 3 чорні кулі, навмання виймають 2 кулі. Нехай X – кількість вийнятих білих куль.

Визначте функцію розподілу випадкової величини X , побудуйте графік цієї функції.

4.7. Кожен із 3 стрільців стріляє по цілі 1 раз. Імовірності влучання в ціль за одного пострілу з 1, 2 і 3-го стрільців дорівнюють, відповідно, 0,7; 0,8; 0,9.

Визначте функцію розподілу випадкової величини X – кількості влучань у ціль.

4.8. Напишіть закон розподілу випадкової величини X – кількості появ "герба" за двох підкидань монети.

4.9. Із шухляди, що містить 2 білі та 2 чорні кулі, навмання виймають кулю. Вийняту кулю повертають у шухляду, після чого з неї виймають навмання другу кулю. Нехай X – кількість білих куль, вийнятих за обох спроб.

Складіть закон розподілу, функцію розподілу випадкової величини X ; побудуйте графік функції розподілу.

4.10. У місті є 5 комерційних банків. Для кожного з них ризик банкрутства протягом року становить 10 %.

Складіть закон розподілу кількості банків, які можуть збанкрутувати протягом наступного року й обчисліть числові характеристики цього розподілу.

4.11. Пару гральних кісток кидають 10 разів.

Яка ймовірність того, що: а) 12 очок випадуть 2 рази; б) 3 очки випадуть 2 рази; в) 7 очок випадуть 4 рази?

4.12. Задано випадкову величину X (табл. 4.13).

Таблиця 4.13

Закон розподілу випадкової величини X

x_i	5	10	15	20
p_i	0,2	0,3	0,4	0,1

Обчисліть $M(X)$ і $D(X)$.

4.13. Задано випадкову величину X (табл. 4.14).

Таблиця 4.14

Закон розподілу випадкової величини X

x_i	4	6	x_3
p_i	0,5	0,3	p_3

Визначте x_3 і p_3 , якщо $M(X) = 8$.

4.14. Задано випадкову величину X (табл. 4.15).

Таблиця 4.15

Закон розподілу випадкової величини X

x_i	-1	0	1
p_i	p_1	p_2	p_3

Відомо, що $M(X) = 0,1$; $M(X^2) = 0,9$.

Визначте p_1 , p_2 , p_3 .

4.15. Випадкова величина X може набувати лише двох значень: 1 і 2. Визначте ймовірність цих значень, якщо:

а) $M(X) = \frac{4}{3}$; б) $D(X) = \frac{1}{4}$.

4.16. Випадкова величина X може набувати тільки таких значень: $-1, 0$ і 1 .

Визначте ймовірності цих значень, якщо відомо, що:

а) $M(X) = 0, D(X) = \frac{2}{3}$; б) $M(X) = \frac{1}{6}, D(X) = \frac{17}{36}$.

4.17. Визначте математичне сподівання і дисперсію:

- а) кількості очок, що випали за одного кидання гральної кістки;
б) суми очок, що випали за кидання 2 гральних кісток.

4.18. Визначте математичне сподівання і середнє квадратичне відхилення добутку очок, що випали за кидання 2 гральних кісток.

4.19. Виконують ряд пострілів з імовірністю влучання $0,8$ за кожного пострілу. Стрільбу ведуть до першого влучання, але виконують не більше від 4 пострілів.

Визначте математичне сподівання кількості зроблених пострілів.

4.20. Визначте математичне сподівання та дисперсію кількості виробів, виготовлених на автоматичній лінії за відповідного налаштування за період між двома переналаштуваннями, якщо за нормального налаштування ймовірність виготовлення бракованого виробу дорівнює $0,06$, а переналаштування здійснюють після виготовлення 3-го бракованого виробу.

4.21. За табл. 4.16 розподілу випадкової величини X , визначте її математичне сподівання і середнє квадратичне відхилення.

Таблица 4.16

Закон розподілу випадкової величини X

x_i	0	1	2	3	4
p_i	0,1	0,3	0,3	0,2	0,1

4.22. Із коробки, що містить 3 білі та 3 чорні кулі, одночасно на-вмання витягають 3 кулі.

Визначте математичне сподівання і дисперсію кількості вийнятих білих куль.

4.23. Клієнти банку, не пов'язані один з одним, не повертають кре-дити у строк із ймовірністю $0,1$.

Складіть закон розподілу випадкової величини X – кількості повернених у строк кредитів із 5 виданих.

Визначте математичне сподівання, дисперсію та середнє квадратичне відхилення цієї випадкової величини.

4.24. Відмова у виконанні певної операції для кожної з десяти фінансових установ дорівнює 0,9.

Визначте середню кількість відмов і дисперсію відмов у цих фінансових установах.

4.25. Із колоди із 32 карт навмання витягають 2 карти.

Визначте математичне сподівання кількості вийнятих тузів.

4.26. В одній коробці лежать 4 білі та 1 чорна куля, в іншій – 2 білі та 3 чорні кулі. Із навмання вибраної коробки витягають: а) 1 кулю; б) 2 кулі.

Визначте математичне сподівання кількості вийнятих білих куль для випадків а і б.

4.27. Із першої коробки, що містить 1 білу і 4 чорні кулі, навмання перекладають 1 кулю в другу коробку, що містить 3 білі та 1 чорну кулю. Потім кулі другої коробки перемішують і навмання виймають 2 кулі.

Визначте математичне сподівання, дисперсію і середнє квадратичне відхилення кількості вийнятих білих куль.

4.28. Робітник обслуговує 4 верстати. Імовірність того, що протягом години верстат не потребує нагляду робітника, дорівнює для першого верстата 0,80; для другого – 0,85; для третього – 0,70; для четвертого – 0,75.

Визначте математичне сподівання і дисперсію кількості верстатів, що не потребують нагляду робітника протягом години.

4.29. Дискретну випадкову величину X задано законом розподілу:

а)	<table><tr><td>X</td><td>1</td><td>3</td><td>5</td></tr><tr><td>P</td><td>0,3</td><td>0,5</td><td>0,2</td></tr></table>	X	1	3	5	P	0,3	0,5	0,2	б)	<table><tr><td>X</td><td>-2</td><td>0</td><td>2</td></tr><tr><td>P</td><td>0,3</td><td>0,6</td><td>0,1</td></tr></table>	X	-2	0	2	P	0,3	0,6	0,1
X	1	3	5																
P	0,3	0,5	0,2																
X	-2	0	2																
P	0,3	0,6	0,1																

Визначте початкові моменти 1, 2, 3 та 4-го порядків.

4.30. Доведіть, що центральний момент 4-го порядку пов'язано з початковими моментами таким співвідношенням:

$$\mu_4 = \nu_4 - 4\nu_3\nu_1 + 6\nu_1^2\nu_2 - 3\nu_1^4.$$

4.5. Тестові завдання

4.1. Дискретною випадковою величиною називають випадкову величину, яка

А набуває окремих ізольованих можливих значень із певними ймовірностями

Б набуває всіх можливих значень із деякого інтервалу

В інше

4.2. Сума ймовірностей у таблиці закону розподілу дорівнює

А	Б	В	Г
0	1	2	інше

4.3. Математичне сподівання дискретної випадкової величини обчислюють за такою формулою

А	Б	В	Г
$M(X) = \sum_{i=1}^n x_i$	$M(X) = \sum_{i=1}^n x_i p_i$	$M(X) = \sum_{i=1}^n x_i^2 p_i$	$M(X) = \sum_{i=1}^n p_i$

4.4. Дисперсію дискретної випадкової величини обчислюють за такою формулою

А $D(X) = \sum_{i=1}^n x_i^2 \cdot p_i$

Б $D(X) = \sum_{i=1}^n x_i \cdot p_i - M(X)$

В $D(X) = \sum_{i=1}^n x_i^2 \cdot p_i - (M(X))^2$

Г $D(X) = \sum_{i=1}^n x_i \cdot p_i$

4.5. Функція розподілу $F(x)$ визначає ймовірність події

А	Б	В	Г
$F(x) = P(X = x)$	$F(x) = P(X > x)$	$F(x) = P(X < x)$	$F(x) = P(x_1 < X < x_2)$

4.6. Початковий момент k -го порядку обчислюють за такою формулою

А	Б	В	Г
$\nu_k = M(X^2)$	$\nu_k = M(X^k)$	$\nu_k = M(X_k)$	$\nu_k = M(X)$

4.7. Центральний момент k -го порядку обчислюють за такою формулою

А	Б	В	Г
$\mu_k = (X)^k$	$\mu_k = (X - M(X))^2$	$\mu_k = (X - M(X^k))$	$\mu_k = (X - M(X))^k$

4.8. Початковий момент 1-го порядку дорівнює

А	Б	В	Г
$D(X)$	0	$M(X)$	$\sigma(X)$

4.9. Центральний момент 2-го порядку дорівнює

А	Б	В	Г
$D(X)$	0	$M(X)$	$\sigma(X)$

4.10. Коефіцієнт варіації обчислюють за такою формулою

А	Б	В	Г
$V = \frac{M(X)}{\sigma} \cdot 100\%$	$V = \frac{\sigma}{M(X)} \cdot 100\%$	$V = \frac{\sigma}{M(X)}$	$V = M(X) \cdot \sigma$

4.6. Запитання для самоперевірки

- 4.1. Що називають законом розподілу дискретної випадкової величини та як його можна задати?
- 4.2. Як побудувати багатокутник розподілу?
- 4.3. Як перевірити правильність складеного закону розподілу випадкової величини?
- 4.4. Які висновки можна зробити за складеним законом розподілу випадкової величини?
- 4.5. Які числові характеристики випадкової величини вам відомо?
- 4.6. Що таке "математичне сподівання"?
- 4.7. Назвіть властивості математичного сподівання.
- 4.8. Дайте означення дисперсії.
- 4.9. Назвіть властивості дисперсії. Що вона характеризує?
- 4.10. Який вигляд має формула для обчислення дисперсії?
- 4.11. Яких значень може набувати дисперсія?
- 4.12. Як обчислити середнє квадратичне відхилення випадкової величини?
- 4.13. Запишіть формули для обчислення початкового і центрального моментів k -го порядку.
- 4.14. Що називають модою випадкової величини?
- 4.15. Що називають медіаною випадкової величини?
- 4.16. Що характеризує коефіцієнт варіації та як його обчислити?
- 4.17. Як обчислити коефіцієнт ексцесу?
- 4.18. Як обчислити коефіцієнт асиметрії?

4.7. Висновки за підрозділом 4

Після вивчення теми здобувач вищої освіти має вміти:

- складати закон розподілу дискретної випадкової величини та робити контрольну перевірку;
- будувати багатокутник розподілу;
- створювати функцію розподілу та будувати її графік;
- обчислювати основні числові характеристики дискретної випадкової величини;
- обчислювати додаткові числові характеристики;
- визначати початкові та центральні теоретичні моменти.

Література: [3; 7; 8; 13; 16].

5. Закони розподілу та числові характеристики дискретної випадкової величини

5.1. Мета та компетентності

Метою вивчення цього підрозділу є ознайомлення з основними законами розподілу дискретних випадкових величин, навчання складати закони розподілу дискретних випадкових величин та обчислювати їхні числові характеристики.

Компетентності, що формують під час вивчення підрозділу:

знання означення законів розподілу дискретних випадкових величин;

уміння будувати функцію розподілу випадкової величини за певним законом розподілу;

знання основних числових характеристик випадкової величини та вміння їх обчислювати;

уміння робити висновки.

5.2. Основні теоретичні відомості

Біноміальним називають закон розподілу дискретної випадкової величини X – кількості появи події в n незалежних випробуваннях, у кожному із яких імовірність появи події дорівнює p , якщо відповідні значення ймовірностей обчислюють за формулою Бернуллі:

$$P_n(X = k) = C_n^k p^k q^{n-k}, (k = 0, 1, 2, \dots, n).$$

Біноміальний закон розподілу є двопараметричним, оскільки його визначають параметрами n (кількість випробувань) та p (імовірність появи події в окремому випробуванні).

Математичне сподівання і дисперсію випадкової величини, розподіленої за біноміальним законом, визначають за такими формулами:

$$M(X) = np, D(X) = npq.$$

Випадкову величину X називають розподіленою за *пуассонівським законом*, якщо її ймовірності визначають за формулою Пуассона:

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

де $\lambda = np$.

Закон розподілу Пуассона є однопараметричним, єдиним його параметром є $\lambda = np$.

Для випадкової величини, розподіленої за законом Пуассона:

$$M(X) \approx D(X) = \lambda.$$

Ще одним прикладом закону розподілу дискретної випадкової величини, що можна задати в аналітичній формі, є *геометричний розподіл*. Нехай здійснюють незалежні й однорідні випробування, для кожного із яких імовірність p появи певної події є сталою величиною.

У цих випробуваннях спостерігають за появою певної події. Випробування тривають доти, доки з'явиться ця подія. Випадковою величиною є кількість випробувань, яка передуює появі цієї події. Випадкова величина X має **геометричний закон розподілу**, якщо відповідні значення її ймовірностей обчислюють за такою формулою:

$$P_n(X = k) = (1 - p)^k \cdot p, \quad k = 0, 1, 2, \dots, k.$$

Або

$$P_n(X = k) = q^{k-1} \cdot p, \quad k = 1, 2, 3, \dots, n.$$

Геометричний розподіл є однопараметричним і його визначають параметром p .

Для випадкової величини, розподіленої за геометричним законом, математичне сподівання і дисперсію визначають за такими формулами:

$$M(X) = \frac{q}{p}, \quad D(X) = \frac{q}{p^2} \text{ у разі, якщо } k = 0, 1, 2, 3, \dots, n,$$

$$\text{або } M(X) = \frac{1}{p}, \quad D(X) = \frac{q}{p^2} \text{ у разі, якщо } k = 1, 2, 3, \dots, n.$$

Нехай є партія з N од. товару, із яких M од. є якісними. Із цієї партії вибирають n од. товару таким способом, що кожна одиниця має однакову ймовірність потрапити до вибірки. Якщо випадковою величиною вважати кількість m од. якісного товару серед множини n , то закон розподілу цієї випадкової величини називають гіпергеометричним. Випадкова величина X має **гіпергеометричний закон розподілу**, якщо відповідні значення її ймовірностей обчислюють за такою формулою:

$$P_N(X = m) = \frac{C_M^m \cdot C_{N-M}^{n-m}}{C_N^n},$$

де C_M^m – кількість сполучень із M по m , тобто кількість можливостей вибрати m якісних елементів із їхньої множини, що містить M елементів;

C_{N-M}^{n-m} – кількість сполучень із $N - M$ по $n - m$, тобто кількість можливостей заповнити $n - m$ вільних місць (ці місця залишилися у вибірковій сукупності після того, як m місць було зайнято якісними елементами) неякісними елементами, загальна кількість яких становить $N - M$;

C_N^n – кількість сполучень із N по n , тобто загальна кількість можливостей вибрати n елементів із множини, що містить N елементів.

Відповідно, добуток $C_M^m \cdot C_{N-M}^{n-m}$ дорівнює кількості сприятливих елементарних подій на просторі елементарних подій, загальна кількість яких становить C_N^n . Гіпергеометричний закон розподілу характеризують за трьома параметрами: N , M та n .

Для випадкової величини, розподіленої за гіпергеометричним законом математичне сподівання і дисперсію визначають за такими формулами:

$$M(X) = m \cdot \frac{n}{N};$$

$$D(X) = M \cdot \frac{n}{N-1} \cdot \left(1 - \frac{n}{N}\right) \cdot \left(1 - \frac{M}{N}\right).$$

5.3. Приклади розв'язування задач

Приклад 5.1. Імовірність скласти іспит на "відмінно" для кожного із 6 здобувачів вищої освіти дорівнює 0,4.

Складімо закон розподілу кількості відмінних оцінок, отриманих здобувачами вищої освіти на іспиті.

Визначмо $M(X)$ і $D(X)$.

Розв'язання. Нехай випадкова величина X – кількість відмінних оцінок, отриманих здобувачами вищої освіти на іспиті.

Вона може набувати таких значень: 0, 1, 2, 3, 4, 5, 6.

За умовою $p = 0,4$; $q = 0,6$.

Імовірності, із якими величина X набуває своїх можливих значень, слід обчислювати за формулою Бернуллі:

$$P_6(0) = C_6^0 \cdot 0,4^0 \cdot 0,6^6 \approx 0,047, \quad (C_6^0 = 1);$$

$$P_6(1) = C_6^1 \cdot 0,4^1 \cdot 0,6^5 \approx 0,187;$$

$$P_6(2) = C_6^2 \cdot 0,4^2 \cdot 0,6^4 \approx 0,311;$$

$$P_6(3) = C_6^3 \cdot 0,4^3 \cdot 0,6^3 \approx 0,276;$$

$$P_6(4) = C_6^4 \cdot 0,4^4 \cdot 0,6^2 \approx 0,138;$$

$$P_6(5) = C_6^5 \cdot 0,4^5 \cdot 0,6^1 \approx 0,037;$$

$$P_6(6) = 0,4^6 \approx 0,004.$$

Закон розподілу цієї випадкової величини X зручно навести у вигляді таблиці (табл. 5.1).

Таблиця 5.1

Закон розподілу випадкової величини X

x_i	0	1	2	3	4	5	6
p_i	0,047	0,187	0,311	0,276	0,138	0,037	0,004

Перевірка: $0,047 + 0,187 + 0,311 + 0,276 + 0,138 + 0,037 + 0,004 = 1$.

Із табл. 5.1 видно: найбільш імовірно, що здобувачами вищої освіти буде отримано 2 оцінки "відмінно", тому що $P_6(2) = 0,311$ є найбільшою

ймовірністю. Практично неможливо отримати 6 оцінок "відмінно" тому, що $P_6(6) = 0,004$. Малоімовірно отримати 5 оцінок "відмінно" або жодної ($P_6(0) = 0,047$ і $P_6(5) = 0,037$). Але практично вірогідно отримати від 1 до 4 оцінок "відмінно" тому, що

$$P_6(1 \leq k \leq 4) = 0,187 + 0,311 + 0,276 + 0,138 = 0,912.$$

$$M(X) = n \cdot p = 6 \cdot 0,4 = 2,4; \quad D(X) = n \cdot p \cdot q = 6 \cdot 0,4 \cdot 0,6 = 1,44.$$

Приклад 5.2. Припускаючи однаковими ймовірності народження хлопчика і дівчинки, складімо закон розподілу випадкової величини X , яка обчислює кількість хлопчиків у родині, що має 5 дітей.

Розв'язання. Випадкова величина X може набувати 6 значень: 0, 1, 2, 3, 4 і 5. Імовірності того, що вона набуде кожне з них, слід визначати за формулою Бернуллі $P_n(m) = C_n^m p^m q^{n-m}$, за $n = 5$, $p = 0,5$ і $m = 0, \dots, 5$. Після цього потрібно скласти такий закон розподілу випадкової величини X (табл. 5.2).

Таблиця 5.2

Закон розподілу випадкової величини X

x_i	0	1	2	3	4	5
p_i	0,03125	0,15623	0,31250	0,31250	0,15625	0,03125

Очевидно, що $\sum_{i=1}^6 p_i = 1$.

Із табл. 5.2 видно, що найбільш імовірно, що в родині буде 2 хлопчики та 3 дівчинки або навпаки.

Приклад 5.3. Обладнання складається зі 100 елементів, що працюють незалежно один від одного. Імовірність відмови будь-якого елемента $p = 0,01$.

Складімо закон розподілу кількості відмов.

Розв'язання. Нехай випадкова величина X – кількість відмов елементів, яка може набувати таких значень 0, 1, 2, 3, ..., 100.

Імовірності цих значень обчислімо за формулою Пуассона тому, що $n=100$ – велика, а $p=0,01$ – мала ймовірність.

За умовою $\lambda = 100 \cdot 0,01 = 1$.

$$P_n(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}, \quad (e^{-1} \approx 0,3679);$$

$$P(X = 0) = \frac{e^{-1}}{0!} \approx 0,3679; \quad P(X = 1) = \frac{e^{-1}}{1!} \approx 0,3679;$$

$$P(X = 2) = \frac{e^{-1}}{2!} \approx 0,1839; \quad P(X = 3) = \frac{e^{-1}}{3!} \approx 0,0613;$$

$$P(X = 4) = \frac{e^{-1}}{4!} \approx 0,0153; \quad P(X = 5) = \frac{e^{-1}}{5!} \approx 0,0031;$$

$$P(X = 6) = \frac{e^{-1}}{6!} \approx 0,0005.$$

За $k \geq 7$ імовірності $P(X \geq 7) \approx 0$.

Отже, закон розподілу заданої випадкової величини має такий вигляд (табл. 5.3).

Таблиця 5.3

Закон розподілу випадкової величини X

x_i	0	1	2	3	4	5	6
p_i	0,3679	0,3679	0,1839	0,0613	0,0153	0,0031	0,0005

Перевірка: $0,3679 + 0,3679 + 0,1839 + 0,0613 + 0,0153 + 0,0031 + 0,0005 = 1$.

Аналіз табл. 5.3 дозволяє зробити деякі висновки. Найбільш імовірно, що відмовить 1 або жодний елемент. Практично неможливо, що відмовлять 5 або більше елементів.

Приклад 5.4. Звітність містить певні неточності, які за експрес-перевірки виявляють з імовірністю 0,7. Виконують перевірку всієї документації

до виявлення першого порушення. Випадковою величиною вважають кількість документів, у яких не було помічено порушень.

Складімо закон розподілу цієї випадкової величини.

Розв'язання. Оскільки випробування проводять до першої появи події (виявлення порушення) і ймовірність появи події в кожному випробуванні є сталою величиною, то закон розподілу випадкової величини є геометричним. Випадкова величина може набувати таких значень: 1, 2, 3, ..., отже, це нескінченна дискретна випадкова величина. Для обчислення ймовірності кожного її значення слід користуватися такою формулою:

$$P_n(X = k) = (1 - p)^k \cdot p, \quad k = 0, 1, 2, \dots, k.$$

$$\text{Наприклад, } P_1(X = 0) = (1 - 0,7)^0 \cdot 0,7 = 0,7;$$

$$P_2(X = 1) = (1 - 0,7)^1 \cdot 0,7 = 0,21 \text{ тощо.}$$

Результати обчислення наведено в табл. 5.4. Видно, що випадкова подія відбувається в першому ж випробуванні з імовірністю 0,7.

Зі зростанням значення випадкової величини його ймовірність швидко спадає, можливість пропустити помилки навіть у 2 документах є подією малоімовірною.

Таблиця 5.4

**Ряд розподілу дискретної випадкової величини,
розподіленої за геометричним законом**

x_i	0	1	2	3	4	5
p_i	0,7000	0,2100	0,0630	0,0189	0,0057	0,0017

Імовірність події "пропустити не більше ніж п'ять помилок" дорівнює сумі ймовірностей перших шести подій (ураховуючи подію "не пропустити жодної помилки"), у цьому разі вона становить 0,9993.

Приклад 5.5. У магазин надійшла партія товару, що становить 20 опалювальних котлів. Під час транспортування 5 із них зазнали незначних пошкоджень корпусу. Із цієї партії для філії було передано 4 котли.

Складімо ряд розподілу випадкової величини, якою вважають кількість пошкоджених котлів серед тих, що відібрані для філії.

Розв'язання. Оскільки з партії товару вибирають її частину, де кожний з об'єктів має однакову ймовірність потрапити до вибірки, то йдеться про гіпергеометричний розподіл. Значення випадкової величини можуть змінюватися від 0 до 4.

За такою формулою:

$$P_N(X = k) = \frac{C_n^k \cdot C_{N-n}^{m-k}}{C_N^m},$$

буде:

$$P_4(X = 0) = \frac{C_5^0 \cdot C_{15}^4}{C_{20}^4} = \frac{5!}{5! \cdot 0!} \cdot \frac{15!}{11! \cdot 4!} \cdot \frac{16! \cdot 4!}{20!} = 0,2817;$$

$$P_4(X = 1) = \frac{C_5^1 \cdot C_{15}^3}{C_{20}^4} = \frac{5!}{4! \cdot 1!} \cdot \frac{15!}{12! \cdot 3!} \cdot \frac{16! \cdot 4!}{20!} = 0,4696;$$

$$P_4(X = 2) = \frac{C_5^2 \cdot C_{15}^2}{C_{20}^4} = \frac{5!}{3! \cdot 2!} \cdot \frac{15!}{13! \cdot 2!} \cdot \frac{16! \cdot 4!}{20!} = 0,2167;$$

$$P_4(X = 3) = \frac{C_5^3 \cdot C_{15}^1}{C_{20}^4} = \frac{5!}{2! \cdot 3!} \cdot \frac{15!}{14! \cdot 1!} \cdot \frac{16! \cdot 4!}{20!} = 0,0310;$$

$$P_4(X = 4) = \frac{C_5^4 \cdot C_{15}^0}{C_{20}^4} = \frac{5!}{1! \cdot 4!} \cdot \frac{15!}{15! \cdot 0!} \cdot \frac{16! \cdot 4!}{20!} = 0,0010.$$

Для перевірки того, що перелічено всі можливі значення випадкової величини, обчислимо суму ймовірностей. Дійсно, $\sum_{i=1}^n p_i = 1$.

Зверніть увагу, що 1 котел із 4 становить ті самі 25 %, що і 5 котлів із 20, тобто випадкова подія $X = 1$ має бути найімовірнішою, але хоча вона і є найімовірнішою, ймовірність того, що випадкова величина набуде значення $X = 1$ становить лише 0,4696.

5.4. Вправи для самостійної роботи

5.1. Складіть закон розподілу кількості влучань в мішень за 4 пострілів, якщо ймовірність улучання за кожного пострілу дорівнює 0,3.

Обчисліть математичне сподівання і дисперсію цієї випадкової величини, користуючись тільки їхніми означеннями, а результати перевірте за формулами цих характеристик випадкової величини, розподіленої за біноміальним законом.

5.2. Виконують 8 пострілів у ціль. Імовірність улучання в ціль за кожного пострілу дорівнює 0,6. Нехай X – кількість улучань у ціль.

Визначте: функцію розподілу випадкової величини X ; імовірність того, що буде не менше ніж 7 улучань у ціль; $M(X)$, $D(X)$.

5.3. Імовірність виграти за лотерейним квитком дорівнює 0,4. Придбано 30 квитків.

Визначте математичне сподівання і дисперсію кількості квитків, за які дістануть виграші.

5.4. Визначте функцію розподілу кількості влучань у ціль, якщо зроблено 6 пострілів, а ймовірність улучання за одного пострілу дорівнює 0,2.

Користаючись цією функцією, визначте ймовірність того, що ціль буде вражено не менше від 1 разу, але менше ніж 5 разів.

5.5. У місті 10 комерційних банків. У кожного ризик банкрутства протягом року становить 10 %.

Складіть ряд розподілу кількості банків, які можуть збанкрутувати протягом наступного року; побудуйте його графік.

Визначте числові характеристики його розподілу.

Запишіть функцію розподілу ймовірностей і побудуйте її графік.

Обчисліть імовірність того, що протягом року збанкрутує не більше від 1 банку.

5.6. Деякий пристрій містить 8 деталей. У пристрої застосовують деталі з імовірністю браку 0,01.

Визначте ймовірність того, що пристрій виявиться придатним, знаючи, що він виходить із ладу, якщо кількість зіпсованих деталей не менша від 2.

5.7. Курс акції протягом дня торгів може піднятися або опуститися на один пункт або залишитися незмінним (усі три варіанти є рівноймовірними).

Визначте розподіл курсу акції за 3 дні.

5.8. Курс акції протягом дня торгів може піднятися на один пункт із ймовірністю 0,5; опуститися на один пункт із ймовірністю 0,2 або залишитися незмінним із ймовірністю 0,3.

Визначте розподіл курсу акції за 3 дні.

Обчисліть $M(X)$ і $D(X)$.

5.9. Підручник видано накладом 100 000 прим. Імовірність того, що підручник зброшуровано неправильно, дорівнює 0,0001.

Визначте математичне сподівання і дисперсію кількості бракованих книг.

5.10. Випадкову величину X розподілено за законом Пуассона з параметром $\lambda = 1$.

Визначте ймовірності таких подій: а) $X = 1$; б) $X \geq 1$; в) $X \leq 2$; г) $X = 0$; ґ) $X > 1$; д) $1 \leq X \leq 3$.

5.11. Верстат-автомат штампує деталі. Імовірність того, що виготовлена деталь буде бракованою, дорівнює 0,01.

Визначте математичне сподівання і дисперсію кількості бракованих деталей.

5.12. Монету кидають 5 разів.

Складіть закон розподілу кількості появ "герба".

Визначте ймовірність того, що "герб" випаде не менше від 3 разів.

5.13. Імовірність того, що аудитор зробить помилку під час перевірки бухгалтерського балансу, дорівнює 0,05. Аудитор має зробити висновки про 2 баланси.

Складіть закон розподілу кількості правильних висновків за балансами, що перевіряють.

Обчисліть основні числові характеристики. Зробіть висновки.

5.14. На телефонну станцію надходять у середньому 2 виклики за 1 хв.

Визначте ймовірність того, що на телефонну станцію: а) протягом 1 хв надійде 3 виклики; б) протягом 0,5 хв точно надійде один виклик; в) протягом 24 с не надійде жодного виклику; г) протягом 2 хв надійде хоча б 1 виклик.

5.15. Середня кількість замовлень на таксі, що надходять на диспетчерський пункт за 1 хв, дорівнює 3.

Визначте ймовірність того, що за 2 хв надійде: а) 4 виклики; б) менше від 4 викликів; в) не менше від 4 викликів.

5.16. Мисливець стріляє по дичині до першого влучання, але встигає зробити не більше ніж 5 пострілів. Імовірність улучання в ціль за 1 пострілу дорівнює 0,8.

Складіть закон розподілу кількості промахів. Обчисліть математичне сподівання і середнє квадратичне відхилення цієї випадкової величини.

5.17. В офісі компанії проводять співбесіду з кандидатами на вакантну посаду. Співбесіду проводять доти, доки не буде визначено відповідного кандидата. Імовірність того, що кандидат є прийнятним, дорівнює 0,3.

Складіть закон розподілу випадкової величини X – кількості кандидатів, із якими провели співбесіду, й обчисліть основні числові характеристики, якщо на співбесіду подали заявки 6 осіб.

5.5. Тестові завдання

5.1. Значення ймовірностей біноміального закону розподілу обчислюють за такою формулою

А	Б	В	Г
$\frac{\lambda^k e^{-\lambda}}{k!}$	$(1-p)^k \cdot p$	$C_n^k p^k q_n^{n-k}$	$\frac{C_n^k \cdot C_{N-n}^{m-k}}{C_N^m}$

5.2. Біноміальний закон розподілу є

А однопараметричним

Б двопараметричним

В трипараметричним

5.3. Значення ймовірностей закону розподілу Пуассона обчислюють за такою формулою

А	Б	В	Г
$\frac{\lambda^k e^{-\lambda}}{k!}$	$(1-p)^k \cdot p$	$C_n^k p^k q_n^{n-k}$	$\frac{C_n^k \cdot C_{N-n}^{m-k}}{C_N^m}$

5.4. Значення ймовірностей геометричного закону розподілу обчислюють за такою формулою

А	Б	В	Г
$\frac{\lambda^k e^{-\lambda}}{k!}$	$(1-p)^k \cdot p$	$C_n^k p^k q_n^{n-k}$	$\frac{C_n^k \cdot C_{N-n}^{m-k}}{C_N^m}$

5.5. Значення ймовірностей гіпергеометричного закону розподілу обчислюють за такою формулою

А	Б	В	Г
$\frac{\lambda^k e^{-\lambda}}{k!}$	$(1-p)^k \cdot p$	$C_n^k p^k q_n^{n-k}$	$\frac{C_n^k \cdot C_{N-n}^{m-k}}{C_N^m}$

5.6. Параметрами гіпергеометричного закону розподілу є

А	Б	В	Г
n, p	N, M, n	N, n	λ

5.7. Параметрами геометричного закону розподілу є

А	Б	В	Г
n, p	N, M, n	N, n	λ

5.8. Математичне сподівання біноміального закону розподілу дорівнює

А	Б	В	Г
np	npq	p	λ

5.9. Дисперсія біноміального закону розподілу дорівнює

А	Б	В	Г
np	npq	p	λ

5.10. Математичне сподівання приблизно дорівнює дисперсії для випадкової величини заданої

- А** біноміальним законом
- Б** законом розподілу Пуассона
- В** геометричним законом
- Г** гіпергеометричним законом

5.6. Запитання для самоперевірки

5.1. Які є способи задавання дискретної випадкової величини?

5.2. У чому полягає різниця між скінченними та нескінченними випадковими величинами?

5.3. Сформулюйте принципи, яким має відповідати ряд розподілу випадкової величини.

5.4. Які основні числові характеристики використовують для дискретних випадкових величин? Який сенс має кожна з основних числових характеристик розподілу і як їх визначають?

5.5. Який закон розподілу дискретної випадкової величини називають біноміальним? Які параметри його характеризують?

5.6. Сформулюйте умови застосування формули Пуассона, поясніть, чому цей розподіл називають законом виключних подій. Скількома параметрами його задають?

5.7. Який закон розподілу дискретної випадкової величини називають геометричним? Скількома параметрами його задають?

5.8. Який закон розподілу дискретної випадкової величини називають гіпергеометричним? Скількома параметрами його задають?

5.7. Висновки за підрозділом 5

Після вивчення підрозділу здобувач вищої освіти має:

чітко визначати закони розподілу випадкових дискретних величин;
складати закон розподілу й обчислювати його основні числові характеристики;

надавати геометричну інтерпретацію побудованого закону розподілу;
робити висновки за побудованим розподілом.

Література: [3; 7; 9; 11; 13; 16].

6. Основні закони розподілу неперервної випадкової величини

6.1. Мета та компетентності

Метою вивчення підрозділу є ознайомлення з основними законами розподілу неперервної випадкової величини та формування системи теоретичних знань, практичних умінь і навичок із визначення функції розподілу та основних числових характеристик неперервної випадкової величини.

Компетентності, що формують під час вивчення підрозділу:

здатність обчислювати основні числові характеристики неперервної (одновимірної) випадкової величини, визначати зв'язок між функціями розподілу та щільності ймовірностей;

знання основних законів розподілу неперервної випадкової величини, функції розподілу щільності ймовірностей, розуміння зв'язку між розподілами, заданими різними законами;

уміння за функцією щільності ймовірностей визначати ймовірність потрапляння значень випадкової величини до заданого інтервалу й інтервал, до якого значення випадкової величини будуть належати із заздалегідь заданою надійністю;

навички із застосовування функцій розподілу та щільності ймовірностей для визначення характеристик розподілу неперервної випадкової величини, використанні табличних даних для здійснення обчислень.

6.2. Основні теоретичні відомості

Для неперервної випадкової величини функція розподілу є неперервною, отже, для задавання неперервної випадкової величини разом із функцією розподілу можна використовувати першу похідну від цієї функції:

$$f(x) = F'(x),$$

де $f(x)$ – щільність імовірностей, або диференціальна функція розподілу (на відміну від функції $F(x)$, яку іноді називають інтегральною функцією розподілу).

Відповідно, функція розподілу є невласним інтегралом зі змінною верхньою межею від щільності ймовірностей:

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

Властивості диференціальної функції розподілу

1. Щільність ймовірностей є невід'ємною:

$$f(x) \geq 0.$$

2. Невласний інтеграл від щільності ймовірностей у межах від $-\infty$ до $+\infty$ дорівнює одиниці:

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

3. Імовірність того, що неперервна випадкова величина набуде значення, яке належить до інтервалу (a, b) , дорівнює визначеному інтегралу від диференціальної функції розподілу, межами якого є межі цього інтервалу:

$$P(a < X < b) = \int_a^b f(x)dx = F(b) - F(a).$$

Основні числові характеристики неперервної випадкової величини

Математичне сподівання неперервної випадкової величини X визначають за таким співвідношенням:

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

а дисперсія за означенням має такий вигляд:

$$DX = \int_{-\infty}^{+\infty} (x - MX)^2 \cdot f(x)dx,$$

де $f(x)$ – щільність ймовірностей.

Формула для обчислення дисперсії має такий вигляд:

$$D(X) = \int_{-\infty}^{+\infty} x^2 \cdot f(x) dx - M^2(X).$$

Поняття початкового та центрального теоретичних моментів, а також моди, медіани, асиметрії й ексцесу для неперервної випадкової величини формально визначають тими самими співвідношеннями, що й для дискретної випадкової величини.

Закони розподілу неперервної випадкової величини

Слід розглянути найбільш поширені закони розподілу неперервної випадкової величини.

Рівномірним називають закон розподілу неперервної випадкової величини, диференціальна функція розподілу якого має такий вигляд:

$$f(x) = \begin{cases} 0, & \text{якщо } x \leq a; \\ \frac{1}{b-a}, & \text{якщо } a < x \leq b; \\ 0, & \text{якщо } x > b. \end{cases}$$

Отже, на проміжку $(a, b]$, до якого належать усі можливі значення випадкової величини, щільність імовірностей є сталою величиною.

Рівномірний закон розподілу визначають за двома параметрами: початком (a) і кінцем (b) інтервалу, до якого належать усі значення неперервної випадкової величини.

Отже, він є двопараметричним.

Функція розподілу випадкової величини, розподіленої за рівномірним законом, має такий вигляд:

$$F(x) = \begin{cases} 0, & \text{якщо } x \leq a; \\ \frac{x-a}{b-a}, & \text{якщо } a < x \leq b; \\ 1, & \text{якщо } x > b. \end{cases}$$

На рис. 6.1 показано графіки щільності ймовірностей та функції розподілу випадкової величини, розподіленої за рівномірним законом.

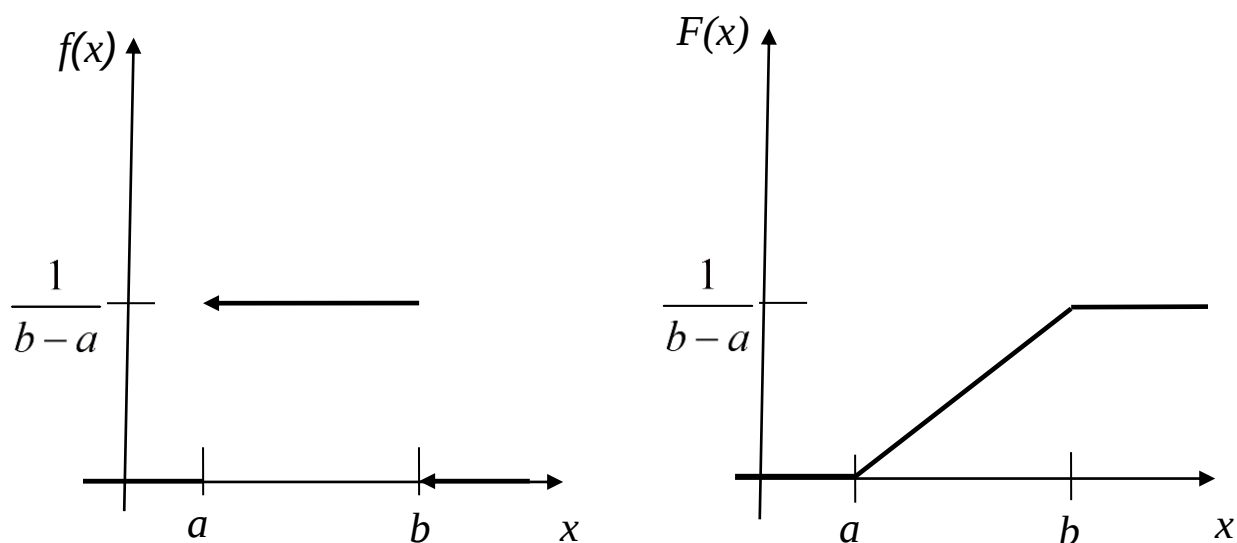


Рис. 6.1. Щільність імовірностей $f(x)$ та функція розподілу $F(x)$ випадкової величини, розподіленої за рівномірним законом

Основні числові характеристики випадкової величини, розподіленої за рівномірним законом такі:

математичне сподівання: $M(X) = \frac{a+b}{2}$;

дисперсія: $D(X) = \frac{(b-a)^2}{12}$;

середнє квадратичне відхилення: $\sigma = \frac{b-a}{2\sqrt{3}}$.

Неперервну випадкову величину X називають розподіленою за **показниковим**, або **експоненціальним законом**, якщо її щільність імовірності має такий вигляд:

$$f(x) = \begin{cases} 0, & \text{якщо } x \leq 0; \\ \lambda e^{-\lambda x}, & \text{якщо } x > 0, \end{cases}$$

де λ – параметр функції показникового розподілу.

Для цього закону функція розподілу має такий вигляд:

$$F(x) = \begin{cases} 0, & \text{якщо } x \leq 0; \\ 1 - e^{-\lambda x}, & \text{якщо } x > 0. \end{cases}$$

Показниковий (експоненціальний) розподіл є однопараметричним із параметром λ .

На рис. 6.2 показано графіки щільності ймовірностей та функції розподілу неперервної випадкової величини, розподіленої за експоненціальним законом.

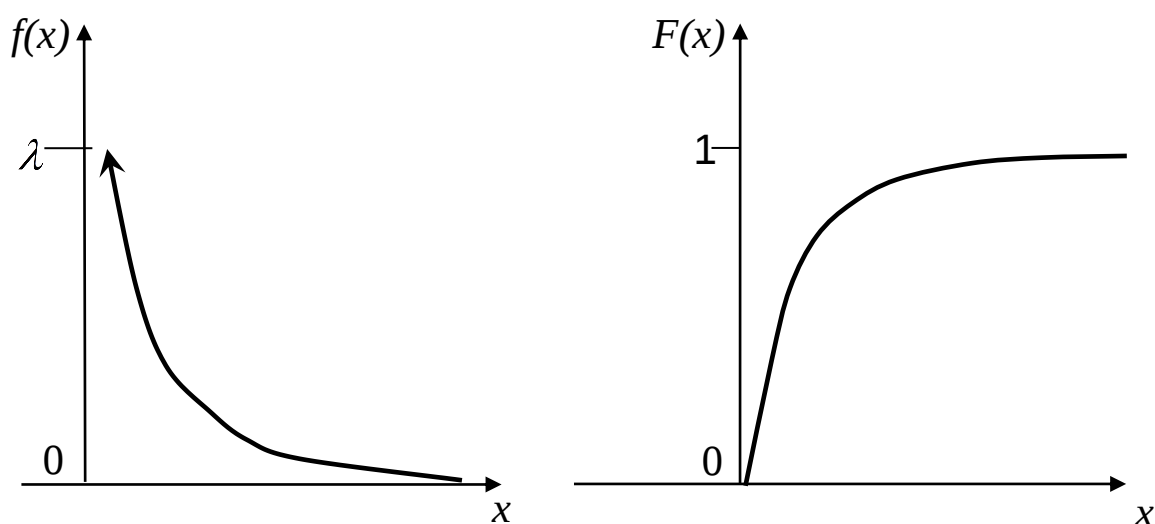


Рис. 6.2. Щільність імовірностей $f(x)$ та функція розподілу $F(x)$ випадкової величини, розподіленої за експоненціальним законом

Основні числові характеристики випадкової величини, розподіленої за експоненціальним законом такі:

математичне сподівання: $M(X) = \frac{1}{\lambda}$;

дисперсія: $D(X) = \frac{1}{\lambda^2}$;

середнє квадратичне відхилення: $\sigma = \frac{1}{\lambda}$.

Отже, для експоненціального закону має місце таке співвідношення:

$$M(X) = \sigma.$$

Відповідно, коефіцієнт варіації дорівнює 100 %:

$$V = \frac{\sigma}{M(X)} \cdot 100 \% = 100 \%.$$

Найбільш поширеним законом розподілу неперервної випадкової величини є нормальний розподіл.

Випадкову величину X називають розподіленою за **нормальним законом**, якщо її диференціальна функція розподілу (щільність імовірності) має такий вигляд:

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

де a – математичне сподівання випадкової величини X ;

σ – середнє квадратичне відхилення випадкової величини.

Основні числові характеристики випадкової величини, розподіленої за нормальним законом такі:

математичне сподівання: $M(X) = a$;

дисперсія: $D(X) = \sigma^2$.

Функція розподілу для нормального закону має такий вигляд:

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

Отже, нормальний закон розподілу є двопараметричним, його параметрами є математичне сподівання a та середнє квадратичне відхилення σ . Випадкову величину, розподілену за нормальним законом, позначають так:

$$X \sim \mathcal{N}(a, \sigma^2),$$

де a і σ – параметри розподілу.

На рис. 6.3 показано графіки інтегральної та диференціальної функцій розподілу для нормального закону.

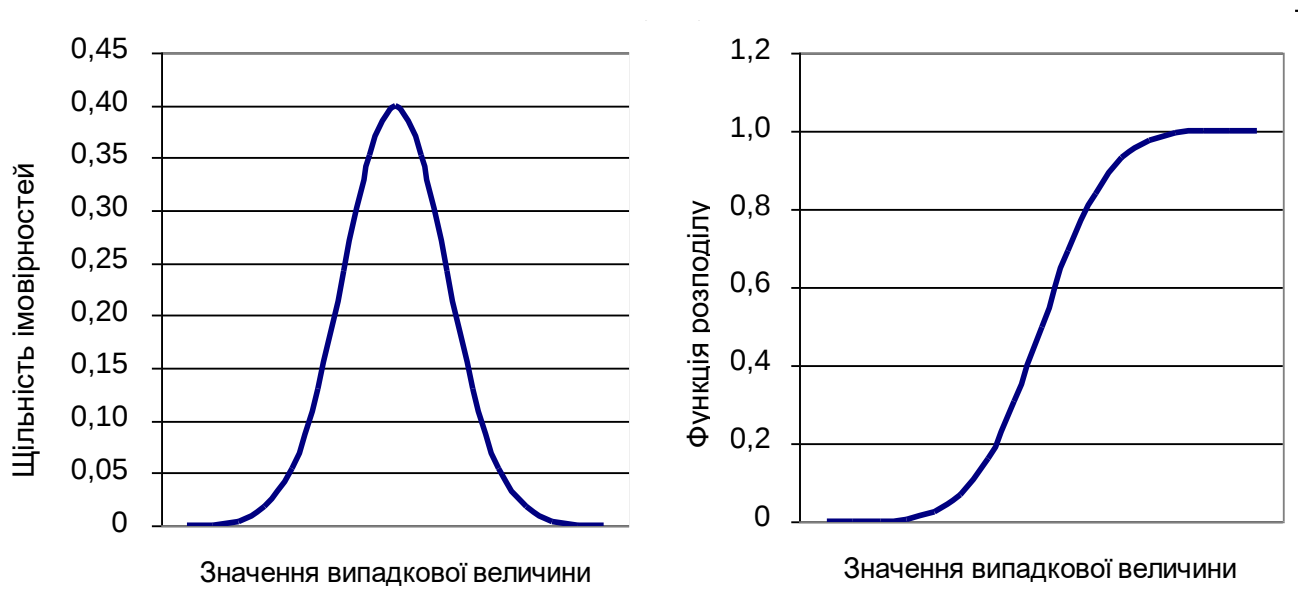


Рис. 6.3. Графіки функції щільності ймовірностей $f(x)$ та функції розподілу $F(x)$ для нормального закону

Головна особливість нормального розподілу полягає в тому, що цей закон є граничним, тобто до нього за певних умов наближаються інші закони розподілу.

Оскільки крива є симетричною, то для нормального закону мода випадкової величини дорівнює медіані й обидві вони збігаються з математичним сподіванням: $Mo(X) = Me(X) = M(X) = a$.

Зміна параметра a призводить до зсуву кривої вздовж осі OX .

Параметр σ характеризує розпорошення випадкової величини навколо математичного сподівання. Середнє квадратичне відхилення дорівнює відстані від точок перегину на кривій $f(x)$ до точки максимуму, яка відповідає значенню математичного сподівання.

Отже, чим більше значення має σ , тим ширшою є крива і, відповідно, тим нижчим максимум, оскільки площа під кривою $f(x)$ є сталою величиною й дорівнює одиниці.

За $a = 0$ та $\sigma = 1$ функція щільності ймовірностей має такий вигляд:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Отже, функція $\varphi(x)$ є диференціальною функцією розподілу випадкової величини X , розподіленої за нормальним законом, що має числові характеристики $M(X) = 0$ та $D(X) = 1$. Вона ще має назву **нормованої** кривої, або **функції Гаусса**.

Функція розподілу випадкової величини, розподіленої за нормальним законом, має такий вигляд:

$$F(x) = 0,5 + \Phi(x).$$

Для випадкової величини, розподіленої за нормальним законом, коефіцієнт асиметрії $A_s = 0$ і ексцес $E_k = 0$.

У тому разі, якщо закон розподілу є наближеним до нормального, за коефіцієнтом асиметрії порівнюють внески значень випадкової величини, які є меншими та більшими за математичне сподівання.

Якщо $A_s > 0$, то внесок значень випадкової величини, більших за математичне сподівання є більш вагомим, отже, крива розподілу є розмитою в бік більших значень випадкової величини. Якщо $A_s < 0$, то крива розподілу є розмитою в бік менших значень.

На рис. 6.4 показано вплив асиметрії на вигляд диференціальної функції розподілу.

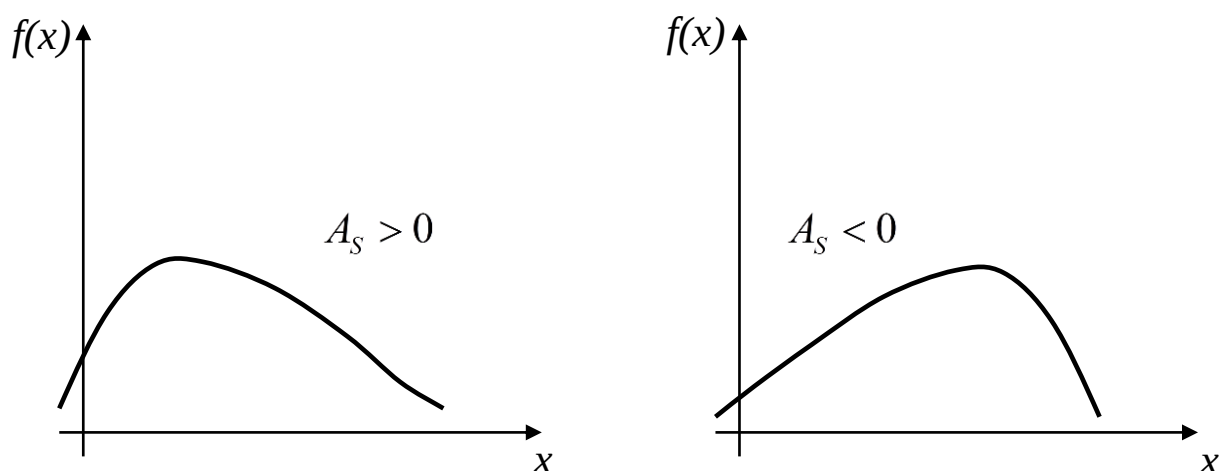


Рис. 6.4. Вигляд диференціальної функції розподілу, залежно від коефіцієнта асиметрії

Ексцес характеризує гостровершинність кривої розподілу, порівняно із кривою Гаусса.

Якщо $E_k > 0$, то вершина кривої розподілу є більш гострою, ніж для кривої Гаусса, якщо $E_k < 0$, то є більш пологою.

Імовірність того, що значення випадкової величини, розподіленої за нормальним законом, буде належати певному проміжку (α, β) , визначають за такою формулою:

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

Імовірність відхилення випадкової величини від свого математичного сподівання на величину, що не перевищує δ , обчислюють за такою формулою:

$$P(|X - a| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right).$$

Відповідно, відхилення випадкової величини від свого математичного сподівання на величину, що є більшою за δ , становить:

$$P(|X - a| \geq \delta) = 1 - 2\Phi\left(\frac{\delta}{\sigma}\right).$$

Нехай $\delta = 3\sigma$. Тоді, за правилом трьох сигм, буде:

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

Отже, 99,73 % значень випадкової величини належать проміжку $(a - 3\sigma, a + 3\sigma)$.

Згідно із цим правилом, відхилення випадкової величини, розподіленої за нормальним законом, від свого математичного сподівання на величину, більшу за 3σ , є подією малої ймовірності.

Звідси впливає важливий практичний висновок. Якщо невідомо, який саме закон розподілу має випадкова величина, то, якщо правило трьох сигм порушено, є підстави вважати, що цей закон не відповідає нормальному.

На рис. 6.5 показано графік диференціальної функції розподілу випадкової величини, розподіленої за нормальним законом.

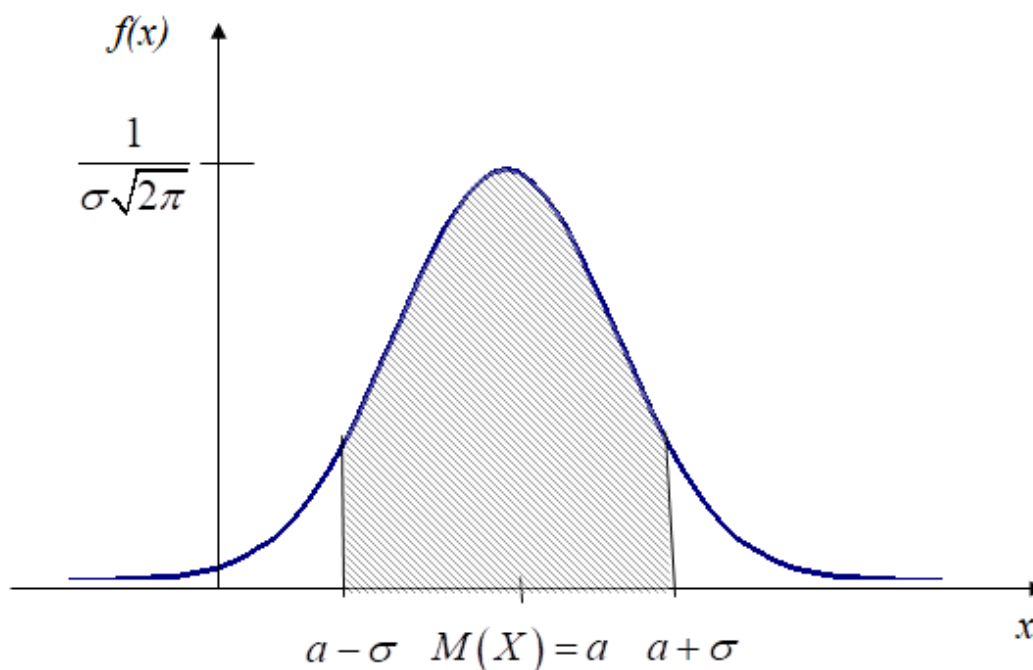


Рис. 6.5. Загальний вигляд функції щільності ймовірностей для нормального закону розподілу

Площа заштрихованої фігури відповідає ймовірності випадкової події, яка полягає в тому, що відхилення значення випадкової величини від її математичного сподівання не перевищує її середнього квадратичного відхилення.

6.3. Приклади розв'язування задач

Приклад 6.1. Задано інтегральну функцію випадкової величини X :

$$F(x) = \begin{cases} 0, & \text{якщо } x \leq 0; \\ \sin 2x, & \text{якщо } 0 < x \leq \frac{\pi}{4}; \\ 1, & \text{якщо } x > \frac{\pi}{4}. \end{cases}$$

Визначмо диференціальну функцію $f(x)$.

Розв'язання.

$$f(x) = F'(x) = \begin{cases} 0, & \text{якщо } x \leq 0; \\ 2 \cos 2x, & \text{якщо } 0 < x \leq \frac{\pi}{4}; \\ 0, & \text{якщо } x > \frac{\pi}{4}. \end{cases}$$

Приклад 6.2. Випадкову величину X задано функцією щільності

ймовірностей $f(x) = \frac{a}{1+x^2}$, $-\infty < x < +\infty$.

Визначмо коефіцієнт a функції $f(x)$ і $F(x)$, імовірність того, що випадкова величина набуде будь-якого значення, не меншого за 0, але не більшого за 5.

Розв'язання. Із рівності $\int_{-\infty}^{+\infty} \frac{a}{1+x^2} dx = 1$ слід обчислити a .

$$a \cdot \arctg x \Big|_{-\infty}^{+\infty} = 1 \Rightarrow a \cdot \left(\frac{\pi}{2} + \frac{\pi}{2} \right) = 1, \text{ звідси } a = \frac{1}{\pi}, \text{ отже, } f(x) = \frac{1}{\pi(1+x^2)}.$$

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(x) dx = \int_{-\infty}^x \frac{1}{\pi(1+x^2)} dx = \frac{\arctg x}{\pi} \Big|_{-\infty}^x = \frac{\arctg x}{\pi} + \frac{1}{\pi} \cdot \frac{\pi}{2} = \\ &= \frac{1}{\pi} \arctg x + \frac{1}{2}. \end{aligned}$$

$$P(0 \leq X \leq 5) = \int_0^5 f(x) dx = \int_0^5 \frac{1}{\pi(1+x^2)} dx = \frac{\arctg x}{\pi} \Big|_0^5 \approx 0,435.$$

Приклад 6.3. Задано диференціальну функцію неперервної випадкової величини X :

$$f(x) = \begin{cases} 0, & \text{якщо } x \leq \frac{\pi}{6}; \\ 3\sin 3x, & \text{якщо } \frac{\pi}{6} < x \leq \frac{\pi}{3}; \\ 0, & \text{якщо } x > \frac{\pi}{3}. \end{cases}$$

Визначмо інтегральну функцію $F(x)$.

Розв'язання. Для визначення функції $F(x)$ слід використати таку формулу:

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

Якщо $x \leq \frac{\pi}{6}$, то $f(x) = 0$, а, отже,

$$F(x) = \int_{-\infty}^x 0 \cdot dx = 0.$$

Якщо $\frac{\pi}{6} < x \leq \frac{\pi}{3}$, то

$$F(x) = \int_{-\infty}^{\frac{\pi}{6}} 0 \cdot dx + \int_{\frac{\pi}{6}}^x 3\sin 3x dx = \left. \frac{-3\cos 3x}{3} \right|_{\frac{\pi}{6}}^x = -\cos 3x.$$

Якщо $x > \frac{\pi}{3}$, то

$$F(x) = \int_{-\infty}^{\frac{\pi}{6}} 0 \cdot dx + \int_{\frac{\pi}{6}}^{\frac{\pi}{3}} 3\sin 3x dx + \int_{\frac{\pi}{3}}^x 0 \cdot dx = \left. \frac{-3\cos 3x}{3} \right|_{\frac{\pi}{6}}^{\frac{\pi}{3}} = -\cos 3x \Big|_{\frac{\pi}{6}}^{\frac{\pi}{3}} = 1.$$

$$\text{Отже, } F(x) = \begin{cases} 0, & \text{якщо } x \leq \frac{\pi}{6}; \\ -\cos 3x, & \text{якщо } \frac{\pi}{6} < x \leq \frac{\pi}{3}; \\ 1, & \text{якщо } x > \frac{\pi}{3}. \end{cases}$$

Приклад 6.4. Обчислимо математичне сподівання, дисперсію випадкової величини X та побудуємо графіки інтегральної та диференціальної функцій розподілу, якщо

$$F(x) = \begin{cases} 0, & \text{якщо } x \leq -1; \\ \frac{x+1}{3}, & \text{якщо } -1 < x \leq 2; \\ 1, & \text{якщо } x > 2. \end{cases}$$

Розв'язання. Диференціальна функція має такий вигляд:

$$f(x) = F'(x) = \begin{cases} 0, & \text{якщо } x \leq -1; \\ \frac{1}{3}, & \text{якщо } -1 < x \leq 2; \\ 0, & \text{якщо } x > 2. \end{cases}$$

Для обчислення математичного сподівання та дисперсії слід застосувати такі формули:

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

$$M(X) = \int_{-\infty}^{-1} x \cdot 0 dx + \int_{-1}^2 x \cdot \frac{1}{3} dx + \int_2^{+\infty} x \cdot 0 dx = \frac{x^2}{6} \Big|_{-1}^2 = \frac{4}{6} - \frac{1}{6} = \frac{1}{2};$$

$$D(X) = \int_{-1}^2 x^2 \frac{1}{3} dx - \frac{1}{4} = \frac{x^3}{9} \Big|_{-1}^2 - \frac{1}{4} = \left(\frac{8}{9} - \frac{1}{9} \right) - \frac{1}{4} = \frac{19}{36}.$$

На рис. 6.6 та 6.7 показано графіки інтегральної функції розподілу $F(x)$ і щільності розподілу $f(x)$.

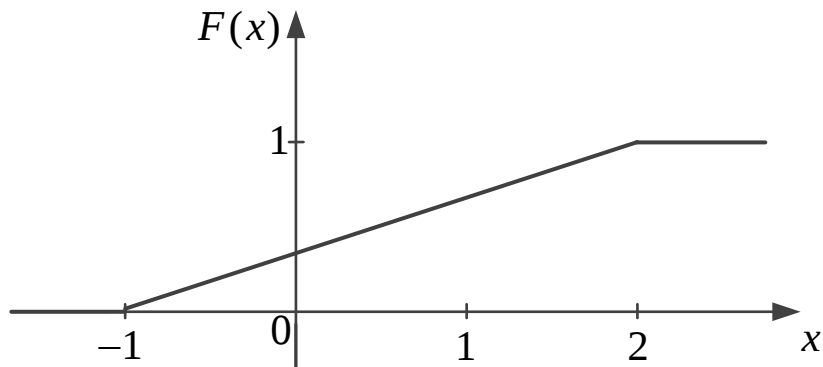


Рис. 6.6. Графік інтегральної функції розподілу

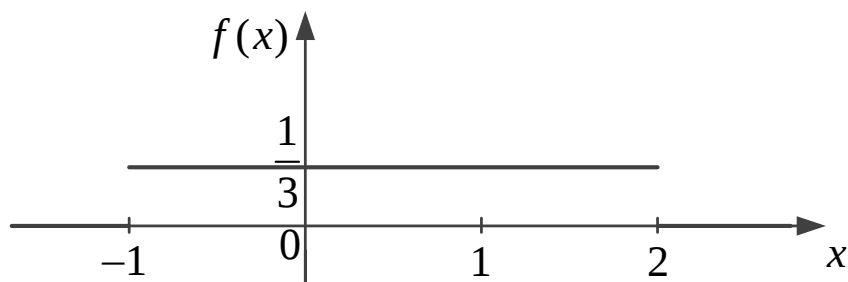


Рис. 6.7. Графік диференціальної функції розподілу

Приклад 6.5. Випадкова величина є розподіленою на проміжку $(0, 24]$ за рівномірним законом.

Визначмо її основні числові характеристики.

Розв'язання. За умовою $a = 0$ та $b = 24$. Обчислимо основні числові характеристики розподілу:

$$M(X) = \frac{a+b}{2} = \frac{0+24}{2} = 12;$$

$$D(X) = \frac{(b-a)^2}{12} = \frac{(24-0)^2}{12} = 48;$$

$$\sigma = \sqrt{D(X)} = \sqrt{48} \approx 6,93.$$

Приклад 6.6. Усі значення рівномірно розподіленої випадкової величини належать відрізку $[2, 8]$.

Визначмо ймовірність того, що випадкова величина набуде значення з інтервалу $(3, 5)$.

Розв'язання. Для визначення ймовірностей застосуємо таку умову:

$$P(\alpha < X < \beta) = F(\beta) - F(\alpha).$$

Для рівномірно розподіленої випадкової величини:

$$F(x) = \begin{cases} 0, & \text{якщо } x \leq a; \\ \frac{x-a}{b-a}, & \text{якщо } a < x \leq b; \\ 1, & \text{якщо } x > b. \end{cases}$$

Тому в цьому прикладі:

$$F(x) = \begin{cases} 0, & \text{якщо } x \leq 2; \\ \frac{x-2}{6}, & \text{якщо } 2 < x \leq 8; \\ 1, & \text{якщо } x > 8. \end{cases}$$

$$\text{Звідси } P(3 < X < 5) = F(5) - F(3) = \frac{5-2}{6} - \frac{3-2}{6} = \frac{3}{6} - \frac{1}{6} = \frac{1}{2}.$$

Приклад 6.7. За розкладом потяги метро рухаються з інтервалом 5 хв. Пасажир приходить на станцію.

Яка ймовірність появи пасажира не раніше ніж через 1 хв після відходу попереднього, але не пізніше ніж за 2 хв до підходу наступного потяга?

Розв'язання. Час прибуття пасажира на станцію можна розглядати як випадкову величину X , розподілену рівномірно в інтервалі руху, тривалістю 5 хв:

$$f(x) = \frac{1}{5}.$$

Треба обчислити $P(1 < X < 3)$.

За формулою $P(\alpha < X < \beta) = \int_{\alpha}^{\beta} f(x)dx$, буде:

$$P(1 < X < 3) = \int_1^3 f(x)dx = \int_1^3 \frac{1}{5} dx = \frac{1}{5} x \Big|_1^3 = \frac{1}{5} (3-1) = \frac{2}{5}.$$

Приклад 6.8. Неперервну випадкову величину X задано диференціальною функцією розподілу:

$$f(x) = \begin{cases} 0, & \text{якщо } x < 0; \\ 5e^{-5x}, & \text{якщо } x \geq 0. \end{cases}$$

Визначмо параметр λ показникового розподілу.

Розв'язання. Для визначення параметра λ слід використати таку формулу:

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

Тоді $\lambda = 5$.

Приклад 6.9. Неперервна випадкова величина X має показниковий розподіл із функцією:

$$f(x) = \begin{cases} 0, & \text{якщо } x < 0; \\ 0,04 \cdot e^{-0,04x}, & \text{якщо } x \geq 0. \end{cases}$$

Визначмо: $F(x)$, $M(X)$, $\sigma(X)$, $P(1 < X < 2)$.

Розв'язання.

$$F(x) = \begin{cases} 0, & \text{якщо } x < 0; \\ 1 - 0,04 \cdot e^{-0,04x}, & \text{якщо } x \geq 0. \end{cases}$$

За умовою $\lambda = 0,04$, тому $M(X) = \sigma(X) = \frac{1}{0,04} = 25$.

$$P(1 < X < 2) = e^{-0,04 \cdot 2} - e^{-0,04 \cdot 1} \approx 0,038.$$

Приклад 6.10. Випадкова величина є розподіленою за показниковим законом із параметром $\lambda = 2$.

Визначмо ймовірність того, що випадкова величина набуде значення, меншого за її математичне сподівання.

Розв'язання. Оскільки закон є показниковим, то $M(X) = \frac{1}{\lambda} = 0,5$.

Потрібно визначити ймовірність події, яка полягає в тому, що значення випадкової величини будуть належати відріzkу $[0; 0,5]$.

$$P(0 < X < 0,5) = F(0,5) - F(0) = 1 - e^{-0,5} - (1 - e^{-0}) = 1 - e^{-0,5} \approx 0,3935.$$

Отже, майже 40 % значень випадкової величини є меншими за її математичне сподівання.

Приклад 6.11. Випадкова величина X є розподіленою за показниковим законом, до того ж $\lambda = 1$.

Обчислімо ймовірність потрапляння випадкової величини X в інтервали: 1) $(0, 2)$; 2) $(3, 4)$.

Розв'язання. Для визначення ймовірності потрапляння випадкової величини X в інтервал слід застосовувати таку формулу:

$$P(\alpha < X < \beta) = e^{-\lambda\alpha} - e^{-\lambda\beta}.$$

Тоді:

$$1) P(0 < X < 2) = e^{-1 \cdot 0} - e^{-1 \cdot 2} = 1 - e^{-2} = 1 - 0,1353 = 0,8647;$$

$$2) P(3 < X < 4) = e^{-1 \cdot 3} - e^{-1 \cdot 4} = e^{-3} - e^{-4} = 0,0498 - 0,0183 = 0,0315.$$

Приклад 6.12. Час безвідмовної роботи елемента розподілено за показниковим законом $f(x) = 0,02e^{-0,02t}$ за $t \geq 0$ (t – час у год).

Визначмо ймовірність того, що елемент пропрацює безвідмовно 100 год.

Розв'язання. Імовірність того, що елемент пропрацює безвідмовно певний час можна визначити за допомогою інтегральної функції розподілу, якщо змінну x замінити на t .

За умовою $\lambda = 0,02$, тому маємо таке:

$$F(t) = 1 - e^{-0,02t}.$$

Для того щоб визначити ймовірність того, що елемент пропрацює безвідмовно 100 год, потрібно замість t підставити 100.

Тоді

$$F(100) = 1 - e^{-0,02 \cdot 100} = 1 - e^{-2} = 0,1353.$$

Приклад 6.13. Математичне сподівання нормально розподіленої випадкової величини X дорівнює $a = 15$ і середнє квадратичне відхилення $\sigma = 2$.

Напишімо диференціальну функцію випадкової величини X .

Розв'язання. Для того щоб записати диференціальну функцію нормально розподіленої випадкової величини X , застосуємо таку формулу:

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

Отже,
$$f(x) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-15)^2}{8}}.$$

Приклад 6.14. Деталі, що випускають, за розміром діаметра розподіляють за нормальним законом за такими параметрами: математичне сподівання дорівнює 5 см, дисперсія – $0,81 \text{ см}^2$.

Визначмо:

- а) імовірність того, що діаметр навмання взятої деталі від 4 до 7 см;
- б) імовірність того, що діаметр деталі відрізняється від математичного сподівання не більше ніж на 2 см;
- в) межі, у яких варто очікувати розмір діаметра деталі, щоб імовірність невиходу за ці межі дорівнювала 0,95.

Розв'язання. Використовуючи такі формули, визначмо:

$$1. P(4 < X < 7) = \Phi\left(\frac{7-5}{\sqrt{0,81}}\right) - \Phi\left(\frac{4-5}{\sqrt{0,81}}\right) \approx \Phi(2,22) + \Phi(1,11) = 0,8533.$$

$$2. P(|X - 5| < 2) = 2\Phi\left(\frac{2}{0,9}\right) \approx 2\Phi(2,22) = 0,9736.$$

$$3. 2\Phi\left(\frac{\varepsilon}{\sigma}\right) = 0,95; \quad \Phi\left(\frac{\varepsilon}{\sigma}\right) = 0,475.$$

За таблицею значень функції $\Phi(x)$ (див. додаток В) визначмо, що $\Phi(x) = 0,475$ за $x = 1,96$.

Тоді, $\frac{\varepsilon}{\sigma} = 1,96$, звідси $\varepsilon = 0,9 \cdot 1,96 \approx 1,8$ см.

Отже, з імовірністю 0,95 можна стверджувати, що діаметр деталі міститься в межах $5 \pm 1,8$ см або від 3,2 до 6,8 см.

Приклад 6.15. Розглядаймо можливість придбання акцій 2 видів – I та II. Їхні прибутковості розподілено за нормальним законом з однаковими середніми значення 20 %. Середнє квадратичне відхилення для першої акції – $\sigma = 1,5$ %, для другої акції – $\sigma = 2$ %.

Визначмо інтервал прибутковості кожної акції з надійністю 95 %. Яка акція є кращою і чому?

Розв'язання. Нехай випадкова величина X_1 – прибутковість акцій I виду, а випадкова величина X_2 – прибутковість акцій II виду. Для обчислення інтервалу прибутковості використаємо таку формулу:

$$P(|X - a| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right).$$

Звідси $|X_1 - a| < \delta$, або $-\delta + a < X_1 < \delta + a$;

$|X_2 - a| < \delta$, або $-\delta + a < X_2 < \delta + a$.

За умовою $a = 20 \%$, отже, треба обчислити δ .

Відомо, що $P = 2\Phi\left(\frac{\delta}{\sigma}\right) = 0,95$; звідси $\Phi\left(\frac{\delta}{\sigma}\right) = 0,475$ та $\frac{\delta}{\sigma} = 1,96$

(див. додаток В);

за умовою для випадкової величини X_1 $\sigma = 1,5 \%$, тоді $\sigma = 1,96 \cdot 1,5 = 2,94$ та $-2,94 \% + 20 \% < X < 20 \% + 2,94 \%$,
отже, $17,06 \% < X < 22,94 \%$.

Тому з імовірністю 95 % можна стверджувати, що прибутковість акцій I виду міститься в межах (17 %; 23 %).

За умовою для випадкової величини X_2 $\sigma = 2 \%$, тоді $\sigma = 1,96 \cdot 2 = 3,92$ та $-3,92 \% + 20 \% < X < 20 \% + 3,92 \%$.

Отже, $16,08 \% < X < 23,92 \%$.

Тому прибутковість акцій II виду від 16 до 24 % можна гарантувати з імовірністю 95 %.

Відповідь на запитання: кращою є перша акція, бо вона є менш ризиковою (середнє квадратичне відхилення в неї є меншим).

Приклад 6.16. Величину X відхилення довжини виготовлюваних деталей від стандарту розподілено за нормальним законом із математичним сподіванням 40 см, і середнім квадратичним відхиленням 0,4 см.

Яку точність довжини виробу можна гарантувати з імовірністю 0,8?

Розв'язання. Визначмо додатне δ , для якого

$$P(|X - 40| \leq \delta) = 0,8.$$

З умови $P(|X - 40| \leq \delta) = 2\Phi\left(\frac{\delta}{0,4}\right) = 0,8$, буде $\Phi\left(\frac{\delta}{0,4}\right) = 0,4$.

Із додатка В $\frac{\delta}{0,4} = 1,28$, звідси $\delta = 0,512$.

Отже, з імовірністю 80 % можна гарантувати, що довжина виробу відхиляється від 40 см менше ніж на 0,512 см.

Приклад 6.17. Випадкову величину розподілено за нормальним законом із математичним сподіванням $a = 20$. Імовірність потрапляння її в інтервал $(20, 30)$ дорівнює $0,477\ 2$.

Визначмо ймовірність потрапляння випадкової величини в інтервал $(10, 25)$.

Розв'язання. Для визначення ймовірності потрапляння випадкової величини в інтервал (α, β) використаємо таку формулу:

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

$$\text{тобто } P(10 < X < 25) = \Phi\left(\frac{25 - 20}{\sigma}\right) - \Phi\left(\frac{10 - 20}{\sigma}\right) = \Phi\left(\frac{5}{\sigma}\right) + \Phi\left(\frac{10}{\sigma}\right).$$

Отже, треба визначити σ .

За умовою $P(20 < X < 30) = 0,477\ 2$.

Тоді

$$P(20 < X < 30) = \Phi\left(\frac{30 - 20}{\sigma}\right) - \Phi\left(\frac{20 - 20}{\sigma}\right) = \Phi\left(\frac{10}{\sigma}\right) - \Phi(0) = \Phi\left(\frac{10}{\sigma}\right) - 0.$$

Тепер буде $\Phi\left(\frac{10}{\sigma}\right) = 0,477\ 2$.

За допомогою таблиці значень функції Лапласа (див. додаток В) визначмо, що $\frac{10}{\sigma} = 2$, звідси $\sigma = 5$.

$$\begin{aligned} \text{Обчислімо } P(10 < X < 25) &= \Phi\left(\frac{25 - 20}{5}\right) - \Phi\left(\frac{10 - 20}{5}\right) = \\ &= \Phi(1) - \Phi(-2) = \Phi(1) + \Phi(2) = 0,3413 + 0,477\ 2 = 0,818\ 5. \end{aligned}$$

Отже, випадкова величина потрапить в інтервал $(10, 25)$ з імовірністю $81,85\ \%$.

6.4. Вправи для самостійної роботи

6.1. Функція розподілу $F(x)$ випадкової величини X має такий вигляд:

$$F(x) = \begin{cases} 0, & \text{якщо } x \leq 1; \\ \frac{(x-1)^2}{4}, & \text{якщо } 1 < x \leq 3; \\ 1, & \text{якщо } x > 3. \end{cases}$$

Визначте:

а) щільність імовірності $f(x)$;

б) $P(1 < X \leq 2,5)$;

в) $P(2,5 < X < 3,5)$.

Побудуйте графіки $f(x)$ і $F(x)$.

6.2. Відомо, що щільність імовірності $f(x)$ величини X визначено за допомогою таких виразів:

$$f(x) = \begin{cases} A \cos x, & \text{якщо } |x| \leq \frac{\pi}{2}; \\ 0, & \text{якщо } |x| > \frac{\pi}{2}. \end{cases}$$

Визначте коефіцієнт A і ймовірність потрапляння X в інтервал $\left(-\frac{\pi}{4}, \frac{\pi}{4}\right)$.

6.3. Щільність імовірності $f(x)$ випадкової величини X задано такою рівністю: $f(x) = \frac{a}{e^x + e^{-x}}$, де a – стала.

Визначте цю сталу та функцію розподілу $F(x)$ величини X .

6.4. Щільність імовірності $f(x)$ випадкової величини X задано та-

кою рівністю: $f(x) = \frac{ae^x}{(e^x + 1)^2}$, де a – стала.

Визначте цю сталу і функцію розподілу $F(x)$ величини X та ймовірність виконання нерівності $|X| < 1$.

6.5. Диференціальна функція випадкової величини X $f(x) = ax$ в інтервалі $(0, 2)$, поза цим інтервалом $f(x) = 0$.

Визначте a , математичне сподівання і медіану X .

6.6. Щільність імовірності випадкової величини X дорівнює 0 за $|x| > 4$ і дорівнює $\frac{1}{\pi\sqrt{16-x^2}}$ за $|x| < 4$.

Обчисліть математичне сподівання і медіану X .

6.7. Щільність імовірностей випадкової величини X задано виразами:

$$f(x) = \begin{cases} a(x^2 + 2x), & \text{якщо } 0 \leq x \leq 1; \\ 0, & \text{якщо } x < 0, x > 1. \end{cases}$$

Обчисливши параметр a , визначте $M(X)$.

6.8. Обчисліть математичне сподівання та дисперсію випадкової величини X за заданою функцією розподілу:

$$\text{а) } F(x) = \begin{cases} 0, & \text{якщо } x \leq 0; \\ \frac{1}{4}, & \text{якщо } 0 < x \leq 4; \\ 1, & \text{якщо } x > 4. \end{cases} \quad \text{б) } F(x) = \begin{cases} 0, & \text{якщо } x \leq -2; \\ \frac{x}{4} + \frac{1}{2}, & \text{якщо } -2 < x \leq 2; \\ 1, & \text{якщо } x > 2. \end{cases}$$

6.9. Щільність імовірностей випадкової величини X $f(x) = a \sin x$ в інтервалі $(0, \pi)$, поза цим інтервалом $f(x) = 0$.

Визначивши параметр a , обчисліть математичне сподівання, моду, медіану і дисперсію величини X .

6.10. Випадкову величину X задано диференціальною функцією $f(x) = a \cos 2x$ в інтервалі $\left(0, \frac{\pi}{4}\right)$, поза цим інтервалом $f(x) = 0$.

Визначивши параметр a , обчисліть математичне сподівання, моду, медіану і дисперсію величини X .

6.11. Випадкову величину X в інтервалі $(2, 4)$ задано диференціальною функцією $f(x) = -\frac{3}{4}x^2 + \frac{9}{2}x - 6$, поза цим інтервалом $f(x) = 0$.

Обчисліть математичне сподівання, моду і медіану величини X .

6.12. Усі значення рівномірно розподіленої випадкової величини належать відрізку $(12, 18)$.

Визначте щільність розподілу цієї випадкової величини, а також імовірність того, що випадкова величина набуде значення з інтервалу $(14, 17)$.

6.13. Визначте математичне сподівання випадкової величини X , яка є рівномірно розподіленою в інтервалі $(4, 10)$.

6.14. Визначте дисперсію і середнє квадратичне відхилення випадкової величини X , рівномірно розподіленої в інтервалі $(2, 10)$.

6.15. Неперервна випадкова величина X є рівномірно розподіленою на відрізку $(1, 3)$.

Потрібно:

1) визначити диференціальну й інтегральну функції розподілу, а також побудувати їхні графіки;

2) обчислити математичне сподівання і середнє квадратичне відхилення;

3) визначити ймовірність того, що X набуде якого-небудь значення з інтервалу $(15, 2)$.

6.16. Автобуси деякого маршруту йдуть за розкладом з інтервалом руху 6 хв.

Визначте ймовірність того, що пасажир буде очікувати автобус менше ніж 2 хв.

6.17. Випадкова величина X має рівномірний розподіл із математичним сподіванням $M(X) = 7$ і дисперсією $D(X) = \frac{4}{3}$.

Визначте інтегральну функцію розподілу випадкової величини X .

6.18. Випадкова величина X має рівномірний розподіл із математичним сподіванням $M(X) = 9$ і середнім квадратичним відхиленням $\sigma(X) = \frac{2\sqrt{3}}{3}$.

Визначте щільність рівномірного розподілу.

6.19. На перехресті доріг рух регулюють автоматичним світлофором, що вмикає зелене світло через кожні 2 хв. Час простою в цього світлофора автомобіля, що зупинився на червоне світло, є випадковою величиною, що розподілена рівномірно на інтервалі $(0, 2)$ хв.

Визначте середній час простою та середнє квадратичне відхилення.

6.20. Визначте математичне сподівання, дисперсію, середнє квадратичне відхилення випадкової величини X , якщо інтегральна функція показникового розподілу має такий вигляд:

$$f(x) = \begin{cases} 0, & \text{якщо } x < 0; \\ 1 - e^{-2x}, & \text{якщо } x \geq 0. \end{cases}$$

6.21. Випадкова величина X , дисперсія якої $D(X) = \frac{1}{25}$, є розподіленою за показниковим законом.

Обчисліть імовірність потрапляння випадкової величини X в інтервал: а) $(0, 1)$; б) $(1, 3)$.

6.22. Неперервна випадкова величина X є розподіленою за показниковим законом із функцією розподілу $F(x) = 1 - e^{-0,2x}$ за $x \geq 0$; $F(x) = 0$ за $x < 0$.

Визначте ймовірність виконання нерівності $2 < X < 4$.

6.23. Визначте функцію розподілу $F(x)$ та обчисліть імовірність потрапляння випадкової величини X , що є розподіленою за показниковим законом в інтервал $(1, 3)$, якщо її щільність імовірності має такий вигляд:

$$f(x) = \begin{cases} 0, & \text{якщо } x < 0; \\ 0,03e^{-0,03x}, & \text{якщо } x \geq 0. \end{cases}$$

6.24. Обчисліть математичне сподівання, дисперсію, середнє квадратичне відхилення випадкової величини зі щільністю ймовірності:

$$f(x) = \begin{cases} 0, & \text{якщо } x < 0; \\ 3e^{-3x}, & \text{якщо } x \geq 0. \end{cases}$$

6.25. Випадкова величина X є розподіленою за показниковим законом із функцією розподілу $F(x) = 1 - e^{-0.2x}$ за $x \geq 0$; $F(x) = 0$ за $x < 0$.

Обчисліть:

- а) математичне сподівання величини X ;
- б) дисперсію величини X ;
- в) середнє квадратичне відхилення величини X ;
- г) ймовірність виконання нерівності $1 \leq X \leq 2$.

6.26. Випробовують 2 елементи, які працюють незалежно один від одного. Тривалість безвідмовної роботи першого має показниковий розподіл із функцією розподілу $F_1(t) = 1 - e^{-0.1t}$, другого – $F_2(t) = 1 - e^{-0.05t}$.

Визначте ймовірність того, що протягом 20 год:

- а) обидва елементи будуть працювати;
- б) відмовить тільки 1 елемент;
- в) відмовить хоча б 1 елемент;
- г) відмовлять обидва елементи.

6.27. Фахівець рекламної агенції стверджує, що час, протягом якого телеглядачі пам'ятають зміст комерційного рекламного ролика, є підпорядкованим експоненційному закону з $\lambda = 0,25$ днів.

Визначте частку глядачів, здатних згадати рекламу через 10 днів.

6.28. Строк служби жорсткого диска комп'ютера – випадкова величина, що підпорядкована експоненційному розподілу з параметром 12 000 год.

Визначте частку жорстких дисків, строк служби яких буде перевищувати 20 000 год.

6.29. Математичне сподівання нормально розподіленої випадкової величини X $a = 2$, а дисперсія $D(X) = 9$.

Напишіть диференціальну функцію випадкової величини.

6.30. Диференціальна функція нормально розподіленої випадкової величини X має такий вигляд:

$$f(x) = \frac{1}{6\sqrt{2\pi}} e^{-\frac{(x-10)^2}{72}}.$$

Визначте математичне сподівання та дисперсію.

6.31. Випадкова величина X є розподіленою за нормальним законом із математичним сподіванням $a = 40$ і дисперсією $D(X) = 200$.

Обчисліть імовірність потрапляння випадкової величини в інтервал $(30, 80)$.

6.32. Вагу риб, що виловлюють у ставку, підпорядковано нормальному закону розподілу з математичним сподіванням $a = 1\,000$ г і середнім квадратичним відхиленням $\sigma = 150$ г.

Визначте імовірність того, що вага спійманої риби буде:

- 1) від 900 до 1 300 г;
- 2) не менше ніж 800 г;
- 3) не більше ніж 1 500 г;
- 4) відрізнятися від середньої ваги за модулем не більше ніж на 200 г.

6.33. Керівник банку встановив, що тривалість обслуговування клієнта в черзі підпорядковано нормальному закону. Математичне сподівання дорівнює 3 хв, середнє квадратичне відхилення – 1 хв.

Визначте ймовірність того, що клієнт перебуває в черзі:

- 1) від 2 до 3,5 хв;
- 2) менше ніж 1 хв;
- 3) понад 5 хв.

6.34. Установлено, що діаметром деталі є нормально розподілена випадкова величина з дисперсією 4 мм^2 .

Яка ймовірність браку, якщо бракують деталі, діаметр яких відхиляється від норми (математичного сподівання) більш ніж на 3,5 мм?

6.35. Автомат виготовляє деталь, яку вважають стандартною, якщо відхилення діаметра деталі від проектного розміру не перевищує 2 мм. Випадкове відхилення діаметра підпорядковано нормальному закону

із середнім квадратичним відхиленням $\sigma = 1,6$ мм і математичним сподіванням $a = 0$.

Скільки відсотків стандартних деталей виготовляє автомат?

6.36. Деталі за розміром розподілено за нормальним законом із середнім значенням 20 см і дисперсією $0,04 \text{ см}^2$.

Яку точність виробу можна гарантувати з імовірністю 0,95?

6.37. Математичне сподівання та середнє квадратичне відхилення нормально розподіленої випадкової величини X , відповідно, дорівнюють 20 і 5.

Визначте ймовірність того, що величина X набуде значення у проміжку (15, 25).

6.38. Зріст юнаків, яких забирають до війська за віком, передбачають нормально розподіленим із середнім значенням $a = 170$ см і середнім квадратичним відхиленням $\sigma = 5$ см.

Визначте відсоток юнаків, що мають такий зріст:

а) нижчий за 160 см;

б) вищий за 180 см;

в) від 160 до 175 см.

6.39. Діаметр деталі є випадковою величиною, розподіленою за нормальним законом з $a = 40$ см і $\sigma = 0,4$ см.

Яку точність ε діаметра можна гарантувати з імовірністю $P = 0,8$?

6.40. Цукор пакують у пакети автоматично, їхня середня маса дорівнює 1,06 кг.

Визначте стандартне відхилення, якщо 5 % пакетів мають масу, меншу за 1 кг. Передбачено, що вагу пакета розподілено за нормальним законом.

6.41. Діаметр деталей є випадковою величиною з нормальним законом розподілу. Математичне сподівання діаметру дорівнює 2,5 см, а дисперсія – $0,0001 \text{ см}^2$.

У яких межах можна практично гарантувати діаметр деталі з імовірністю $P = 0,9973$?

6.42. Випадкова величина X є розподіленою нормально з $a = 25$. Імовірність $P(35 < X < 40) = 0,2$.

Обчисліть імовірність $P(10 < X < 15)$.

6.43. Математичне сподівання і дисперсія нормально розподіленої випадкової величини X , відповідно, дорівнюють 1 і 4.

Визначте ймовірності того, що:

а) $|X - 1| < 2$;

б) $1 < X < 4$;

в) $X > 2$;

г) $X < 0$.

6.44. Випадкова величина є розподіленою за нормальним законом із параметрами $a = 5$, $\sigma = 0,3$.

Визначте ймовірність того, що випадкова величина набуде значення, що буде перевищувати $x = 5,5$.

6.45. За даними бухгалтерії, заробітна плата має нормальний розподіл із середнім значенням 9 500 грн та середнім квадратичним відхиленням 1 500 грн.

Визначте ймовірність того, що випадково вибраний працівник підприємства отримує таку заробітну плату:

- 1) меншу за 8 000 грн;
- 2) від 4 500 до 6 500 грн.

6.46. Кількість покупців, яких обслуговує продавець протягом години, є випадковою величиною, розподіленою за нормальним законом із параметрами $a = 7$, $\sigma = 2$.

Визначте проміжок, до якого з імовірністю 0,9 будуть належати значення цієї випадкової величини.

6.47. Випадкова величина є розподіленою за нормальним законом із параметрами $a = 5$, $\sigma = 1$.

Порівняйте ймовірності потрапляння значення випадкової величини до інтервалів $x \in [2, 4]$ та $x \in [5, 6]$.

6.48. На ринок надійшла велика партія яловичини. Вага туш є випадковою величиною, що розподілена за нормальним законом із математичним сподіванням 500 кг і середнім квадратичним відхиленням 150 кг.

Визначте ймовірність того, що вага випадково відібраної туші:

- а) буде більшою за 550 кг;
- б) буде меншою за 450 кг;

в) буде міститися в межах 450 і 700 кг;

г) відхилиться від математичного сподівання менше ніж на 50 кг.

6.49. Компанія, що продає товари за каталогом, отримує замовлення поштою. Щомісяця кількість цих замовлень є нормально розподіленою випадковою величиною із середнім квадратичним відхиленням 560. У 90 % випадків кількість щомісячних заказів перевищує 12 500.

Визначте щомісячну середню кількість замовлень, що надходить до компанії.

6.5. Тестові завдання

6.1. Щільністю розподілу неперервної випадкової величини X називають функцію

А	Б	В	Г
$f(x) = F(x)$	$f(x) = F'(x)$	$F(x) = f'(x)$	$f(x) = F''(x)$

6.2. Математичне сподівання $M(X)$ неперервної випадкової величини X , можливі значення якої належать інтервалу (x_1, x_2) , дорівнює

А	Б	В	Г
$\int_{x_1}^{x_2} x^2 f(x) dx$	$\int_{x_1}^{x_2} x dx$	$\int_{x_1}^{x_2} f(x) dx$	$\int_{x_1}^{x_2} x f(x) dx$

6.3. Дисперсія $D(X)$ неперервної випадкової величини X , можливі значення якої належать інтервалу (x_1, x_2) , дорівнює

А	Б	В	Г
$\int_{x_1}^{x_2} x f(x) dx - M(X)$	$\int_{x_1}^{x_2} x^2 f(x) dx - (M(X))^2$	$\int_{x_1}^{x_2} x^2 f(x) dx$	$\int_{x_1}^{x_2} f(x) dx - (M(X))^2$

6.4. Якщо $f(x)$ – щільність розподілу випадкової величини X , то ймовірність того, що випадкова величина X набуває значення в інтервалі (x_1, x_2) , дорівнює

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

Б $P(x_1 < X < x_2) = \int_{x_1}^{x_2} x^2 f(x) dx$

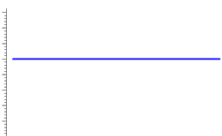
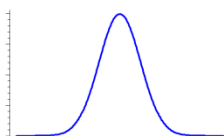
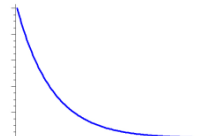
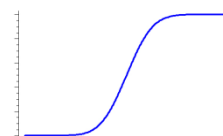
В $P(x_1 < X < x_2) = \int_{x_1}^{x_2} x f(x) dx$

Г $P(x_1 < X < x_2) = \int_{x_1}^{x_2} x^2 f(x) dx - (M(X))^2$

6.5. Щільність розподілу нормально розподіленої випадкової величини X дорівнює

А	Б	В	Г
$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$	$f(x) = e^{-\frac{(x-a)^2}{2\sigma^2}}$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-a)^2}$	$f(x) = e^{-\frac{(x-a)^2}{\sigma}}$

6.6. На якому рисунку зображено графік щільності розподілу нормально розподіленої випадкової величини X

А	Б	В	Г
			

6.7. Числові характеристики нормально розподіленої випадкової величини X дорівнюють

А	Б	В	Г
$M(X) = a,$ $D(X) = \sigma^2$	$M(X) = a,$ $D(X) = \sigma^2$	$M(X) = a,$ $D(X) = \sigma$	$M(X) = a,$ $D(X) = \sqrt{\sigma}$

6.8. Імовірність $P(x_1 < X < x_2)$ того, що випадкова величина, розподілена за нормальним законом, набуває значення в інтервалі (x_1, x_2) , визначають за такою формулою:

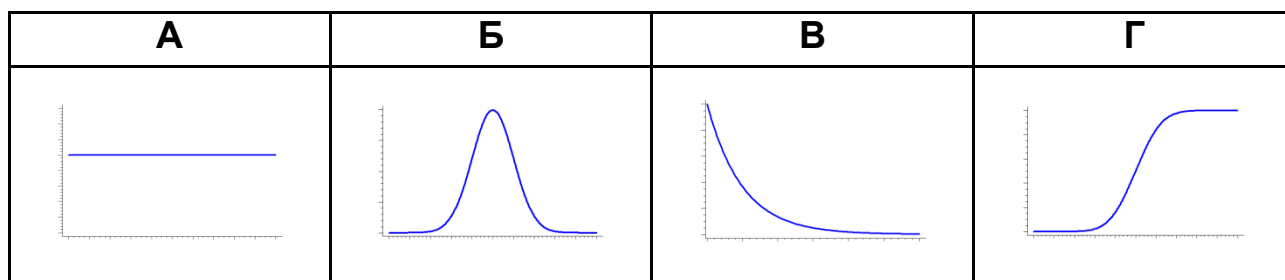
А $P(x_1 < X < x_2) = \Phi(x_2) - \Phi(x_1)$

Б $P(x_1 < X < x_2) = \Phi\left(\frac{x_2 - a}{\sigma}\right) - \Phi\left(\frac{x_1 - a}{\sigma}\right)$

В $P(x_1 < X < x_2) = \Phi\left(\frac{x_1 - a}{\sigma}\right) - \Phi\left(\frac{x_2 - a}{\sigma}\right)$

Г $P(x_1 < X < x_2) = \Phi\left(\frac{x_2 - a}{\sigma}\right) + \Phi\left(\frac{x_1 - a}{\sigma}\right)$

6.9. На якому рисунку зображено графік щільності розподілу випадкової величини X , розподіленої за рівномірним законом



6.10. Якщо $a < x \leq b$, то щільність розподілу випадкової величини, розподіленої за рівномірним законом дорівнює

А	Б	В	Г
$f(x) = b - a$	$f(x) = \frac{1}{a - b}$	$f(x) = \frac{1}{b - a}$	$f(x) = \frac{a}{b - a}$

6.11. Числові характеристики рівномірно розподіленої випадкової величини X дорівнюють

А	Б	В	Г
$M(X) = \frac{a+b}{2},$ $D(X) = \frac{(b-a)^2}{12}$	$M(X) = a + b,$ $D(X) = \frac{b-a}{12}$	$M(X) = \frac{a+b}{2},$ $D(X) = \frac{(b-a)^2}{12}$	$M(X) = \frac{a+b}{2},$ $D(X) = \frac{b-a}{12}$

6.12. Для випадкової величини, розподіленої за показниковим законом, виконано умову

А	Б	В	Г
$M(X) = \sigma(X) = \frac{1}{\lambda}$	$M(X) = \sigma^2(X)$	$D(X) = \lambda$	$M(X) = \sigma(X) = \lambda$

6.6. Запитання для самоперевірки

6.1. Які є способи подання неперервної випадкової величини?

6.2. Чому неможливо побудувати диференціальну функцію розподілу для дискретної випадкової величини?

6.3. Яка відмінність між визначенням числових характеристик розподілу дискретної та неперервної випадкових величин?

6.4. Які є властивості диференціальної функції розподілу?

6.5. Укажіть зв'язок між функцією розподілу і щільністю ймовірностей.

6.6. Який закон розподілу неперервної випадкової величини називають рівномірним? Поясніть сенс цієї назви.

6.7. Назвіть параметри рівномірного закону та наведіть формули для обчислення його основних числових характеристик.

6.8. Який закон розподілу неперервної випадкової величини називають експоненціальним?

6.9. Укажіть кількість параметрів експоненціального закону і наведіть формули для обчислення його основних числових характеристик.

6.10. Який закон розподілу неперервної випадкової величини називають нормальним? Поясніть сенс цієї назви.

6.11. Які є параметри нормального закону і який їхній вплив на вигляд диференціальної функції розподілу?

6.12. Які значення мають додаткові числові характеристики випадкової величини, розподіленої за нормальним законом?

6.13. Як визначають імовірність потрапляння значення випадкової величини до певного інтервалу за нормального закону розподілу?

6.14. Як формулюють правило трьох сигм та як його можна застосовувати?

6.7. Висновки за підрозділом 6

Після вивчення підрозділу здобувач вищої освіти має вміти:

чітко визначати закони розподілу неперервних випадкових величин;
застосовувати функції розподілу та функції щільності ймовірностей для визначення характеристик розподілу неперервної випадкової величини;

надавати геометричну інтерпретацію побудованого закону розподілу;
робити висновки за побудованим розподілом.

Література: [3; 7; 9; 11; 13; 16].

7. Багатовимірні випадкові величини

7.1. Мета та компетентності

Метою вивчення цього підрозділу є ознайомлення з основними законами розподілу двовимірної дискретної випадкової величини, поняттям функції випадкового аргументу та формуванням системи теоретичних знань, практичних умінь і навичок із визначення умовних законів розподілу та обчислення їхніх основних числових характеристик.

Компетентності, що формують під час вивчення підрозділу:

уявлення про закон розподілу та умовні закони розподілу її компонентів;

знання основних числових характеристик системи двох випадкових величин, означення кореляційного моменту, його властивостей, числових характеристик умовних розподілів;

уміння будувати умовні закони розподілу двовимірної дискретної випадкової величини, обчислювати її основні числові характеристики, визначати щільність кореляційного зв'язку;

здатність самостійно визначати етапи дослідження двовимірної випадкової величини для обчислення щільності кореляційного зв'язку її компонентів;

розуміння сутності поняття функції, аргументом якої є випадкова величина;

уміння будувати закон розподілу функції одного випадкового аргументу за відомим розподілом її аргументу та визначати основні числові характеристики цієї функції;

формування схильності до самостійного пошуку засобів до розв'язання задачі;

відповідальність за точність і коректність розрахунків імовірності значень функції випадкового аргументу та її числових характеристик.

7.2. Основні теоретичні відомості

Випадкові величини, можливі значення яких визначають двома числами, називають **двовимірними**. Двовимірну випадкову величину (X, Y) визначають за двома складовими X і Y , які утворюють систему двох випадкових величин.

Геометрична інтерпретація двовимірної величини – це випадкова точка $M(X, Y)$ на площині. Якщо X і Y є дискретними, то й двовимір-на величина є дискретною.

Законом розподілу двовимірної випадкової величини називають відповідність між можливими значеннями цієї величини, тобто парами чисел (x_i, y_i) , та ймовірністю $p(x_i, y_i)$, із якою випадкова величина на-буває цих значень.

Закон розподілу задають:

а) аналітично, наприклад, у вигляді інтегральної функції;

б) таблицею з подвійним входом (табл. 7.1), де $x_1 < x_2 < \dots < x_n$, $y_1 < y_2 < \dots < y_m$, $p(x_i, y_j)$ – імовірність події, яка полягає в одночасному виконанні рівностей $X = x_i, Y = y_j$.

$$\text{Водночас } \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) = 1.$$

Таблиця 7.1

Закон розподілу двовимірної випадкової величини

$Y \backslash X$	x_1	x_2	...	x_i	...	x_n
y_1	$p(x_1, y_1)$	$p(x_2, y_1)$	...	$p(x_i, y_1)$	...	$p(x_n, y_1)$
y_2	$p(x_1, y_2)$	$p(x_2, y_2)$	...	$p(x_i, y_2)$	...	$p(x_n, y_2)$
...	...	...	...	...	...	...
y_j	$p(x_1, y_j)$	$p(x_2, y_j)$	...	$p(x_i, y_j)$	...	$p(x_n, y_j)$
...	...	...	...	...	...	...
y_m	$p(x_1, y_m)$	$p(x_2, y_m)$	...	$p(x_i, y_m)$	...	$p(x_n, y_m)$

Аналогічно тому, як це визначали для одновимірної випадкової величини, закон розподілу двовимірної випадкової величини будь-якої при-роди (як дискретну, так і неперервну) можна задати за допомогою функ-ції розподілу $F(x, y)$, яка визначає ймовірність такої випадкової події, що компонента X двовимірної випадкової величини набуде значення, меншого за певне значення x , яке не обов'язково належить множині значень цієї випадкової величини, одночасно з тим, що компонента Y

набуває значення, меншого за певне значення y . Цю функцію визначають таким співвідношенням:

$$F(x, y) = P(X < x; Y < y),$$

де $F(x, y)$ – функція розподілу двовимірної випадкової величини, визначеної на всій множині дійсних чисел, що називають **інтегральною функцією розподілу ймовірностей двовимірної випадкової величини** (X, Y) (як для дискретної, так і неперервної).

Функції розподілу $F(x, y)$ можна надати геометричне тлумачення (рис. 7.1).

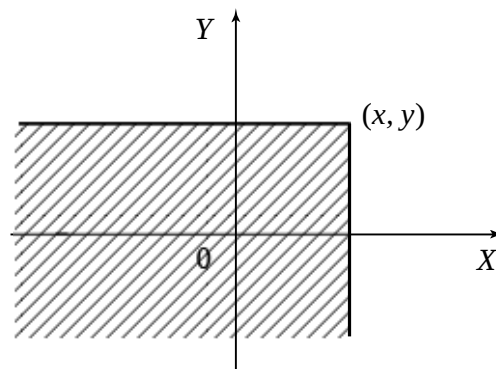


Рис. 7.1. Геометричне тлумачення функції розподілу $F(x, y)$

Властивості інтегральної функції розподілу:

1. Значення інтегральної функції задовольняють подвійну нерівність:

$$0 \leq F(x, y) \leq 1.$$

2. Інтегральна функція є неспадною функцією за кожним аргументом:

$$F(x_2, y) \geq F(x_1, y), \text{ якщо } x_2 > x_1;$$

$$F(x, y_2) \geq F(x, y_1), \text{ якщо } y_2 > y_1.$$

3. Виконано такі граничні співвідношення:

$$1) F(-\infty, y) = 0; \quad 3) F(-\infty, -\infty) = 0;$$

$$2) F(x, -\infty) = 0; \quad 4) F(-\infty, +\infty) = 1.$$

4. а) якщо $y = \infty$, то інтегральна функція системи стає інтегральною функцією складової X : $F(x, \infty) \leq F_1(x)$;

б) якщо $x = \infty$, то інтегральна функція системи стає інтегральною функцією складової Y : $F(\infty, y) \leq F_2(y)$.

Використовуючи інтегральну функцію, можна визначити ймовірність потрапляння випадкової точки (X, Y) у напівсмугу $x_1 < X < x_2$ і $Y < y$ (рис. 7.2):

$$P(x_1 < X < x_2, Y < y) = F(x_2, y) - F(x_1, y),$$

або в напівсмугу $X < x$ і $y_1 < Y < y_2$ (рис. 7.3):

$$P(X < x, y_1 < Y < y_2) = F(x, y_2) - F(x, y_1).$$

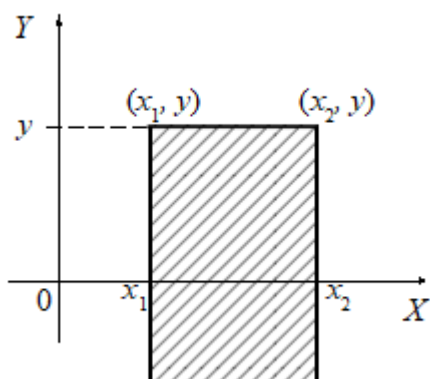


Рис. 7.2. Імовірність потрапляння випадкової точки (X, Y) у напівсмугу $x_1 < X < x_2$ і $Y < y$

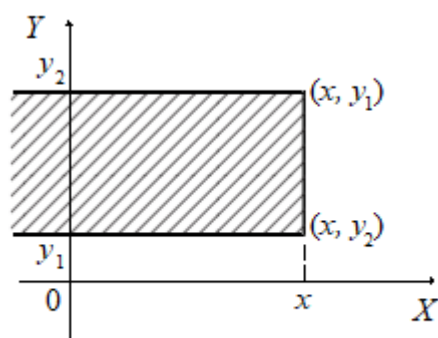


Рис. 7.3. Імовірність потрапляння випадкової точки (X, Y) у напівсмугу $X < x$ і $y_1 < Y < y_2$

Імовірність потрапляння випадкової точки в прямокутник $x_1 < X < x_2$, $y_1 < Y < y_2$ (рис. 7.4) дорівнює:

$$P(x_1 < X < x_2, y_1 < Y < y_2) = [F(x_2, y_2) - F(x_1, y_2)] - [F(x_2, y_1) - F(x_1, y_1)].$$

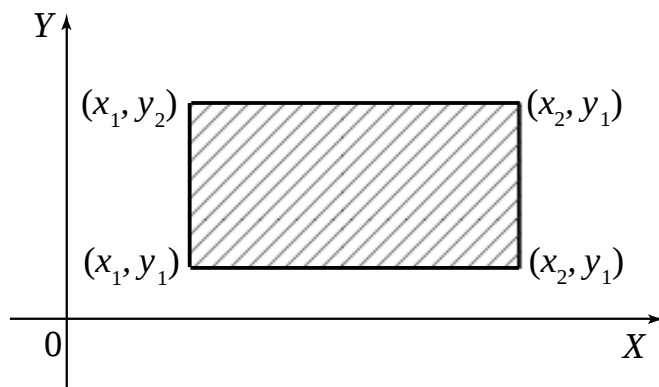


Рис. 7.4. Імовірність потрапляння випадкової точки в прямокутник $x_1 < X < x_2$, $y_1 < Y < y_2$

Нехай складові X і Y мають такі значення:

$$x_1, x_2, \dots, x_n; \quad y_1, y_2, \dots, y_m.$$

Умовним розподілом складової X за $Y = y_j$ називають сукупність значень X і відповідні їм умовні ймовірності $P(x_1|y_j), P(x_2|y_j), \dots, P(x_n|y_j)$, які обчислюють за такою формулою:

$$P(x_i|y_j) = \frac{P(x_i, y_j)}{P(y_j)} \quad (i = 1, 2, \dots, n),$$

де $P(y_j)$ – сума ймовірностей у j -му рядку.

Умовним розподілом складової Y за $X = x_i$ називають сукупність значень Y і відповідні їм умовні ймовірності $P(y_1|x_i), P(y_2|x_i), \dots, P(y_m|x_i)$, які обчислюють за такою формулою:

$$P(y_j|x_i) = \frac{P(x_i, y_j)}{P(x_i)} \quad (j = 1, 2, \dots, m),$$

де $P(x_i)$ – сума ймовірностей в i -му рядку.

Способи подання неперервної випадкової двовимірної величини

Двовимірну випадкову величину можна подати за допомогою інтегральної функції. Неперервну двовимірну випадкову величину також можна задати, використовуючи диференціальну функцію розподілу (щільність ймовірностей).

Диференціальною функцією розподілу $f(x, y)$ двовимірної неперервної випадкової величини (X, Y) називають другу мішану частинну похідну від інтегральної функції:

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \cdot \partial y}.$$

Графічне зображення функції $f(x, y)$ має назву **поверхні розподілу**.

Функцію розподілу $F(x, y)$ визначають за щільністю ймовірностей таким співвідношенням:

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

Властивості диференціальної функції двовимірної величини:

1. Диференціальна функція є невід'ємною: $f(x, y) \geq 0$.

2. Подвійний невластний інтеграл на нескінчених проміжках від диференціальної функції дорівнює одиниці: $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$.

Якщо $(X, Y) \subset D$, то властивість (2) набуває такого вигляду: $\iint_D f(x, y) dx dy = 1$.

Імовірність потрапляння випадкової величини (X, Y) в область D визначають за такою рівністю: $P((X, Y) \subset D) = \iint_D f(x, y) dx dy$.

Геометрично цю формулу можна розглядати як об'єм тіла, обмеженого зверху поверхнею $z = f(x, y)$, в основі якого лежить проєкція цієї поверхні на площину xOy .

Диференціальна функція однієї зі складових дорівнює невластному інтегралу на нескінченному проміжку від диференціальної функції системи зі змінною інтегрування, що відповідає другій складовій:

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy; \quad f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

Умовною диференціальною функцією $\varphi(x|y)$ складової X за заданого значення $Y = y$ називають відношення диференціальної функції $f(x, y)$ до диференціальної функції $f_2(y)$ складової Y :

$$\varphi(x|y) = \frac{f(x, y)}{f_2(y)} = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y) dx}.$$

Умовною диференціальною функцією $\psi(y|x)$ складової Y за заданого значення $X = x$ називають відношення диференціальної функції $f(x, y)$ до диференціальної функції $f_1(x)$ складової X :

$$\psi(y|x) = \frac{f(x, y)}{f_1(x)} = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y) dy}.$$

Властивості умовних диференціальних функцій:

$$1. \varphi(x|y) \geq 0, \quad \psi(y|x) \geq 0.$$

$$2. \int_{-\infty}^{\infty} \varphi(x|y) dx = 1, \quad \int_{-\infty}^{\infty} \psi(y|x) dy = 1.$$

Теорема. Для того щоб випадкові величини X і Y були незалежними, необхідно і достатньо, щоб інтегральна функція системи (X, Y) дорівнювала добутку інтегральних функцій таких складових:

$$F(x, y) = F_1(x) \cdot F_2(y),$$

$$\text{де } F_1(x) = F(x, \infty), \quad F_2(y) = F(\infty, y).$$

Висновок:

$$f(x, y) = f_1(x) \cdot f_2(y).$$

Основні числові характеристики розподілу двовимірної випадкової величини

Математичні сподівання дискретних випадкових величин X і Y , що входять до системи, визначають за такими формулами:

$$M(X) = \sum_{i=1}^n \sum_{j=1}^m x_i p_{ij} \quad \text{та} \quad M(Y) = \sum_{i=1}^n \sum_{j=1}^m y_j p_{ij},$$

а математичні сподівання неперервних випадкових величин X і Y визначають за такими формулами:

$$M(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dx dy \quad \text{та} \quad M(Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x, y) dx dy,$$

а також:

$$M(X) = \int_{-\infty}^{\infty} x f_1(x) dx \text{ та } M(Y) = \int_{-\infty}^{\infty} y f_2(y) dy.$$

Точку $(M(X), M(Y))$ називають **центром розсіювання системи випадкових величин (X, Y)** .

Для обчислення дисперсії дискретних випадкових величин X і Y застосовують такі формули:

$$D(X) = \sum_{i=1}^n \sum_{j=1}^m p_{ij} (x_i - M(X))^2 \text{ та } D(Y) = \sum_{i=1}^n \sum_{j=1}^m p_{ij} (y_j - M(Y))^2,$$

а дисперсії неперервних випадкових величин X і Y , що входять до системи, визначають за такими формулами:

$$D(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - M(X))^2 f(x, y) dx dy$$

$$\text{та } D(Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - M(Y))^2 f(x, y) dx dy.$$

Середні квадратичні відхилення випадкових величин X і Y :

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

Умовним математичним сподіванням дискретної випадкової величини Y за $X = x$ (x – певне можливе значення X) називають суму добутків можливих значень Y на їхні умовні ймовірності:

$$M(Y | X = x) = \sum_{j=1}^m y_j p(y_j | x).$$

Умовним математичним сподіванням дискретної випадкової величини X за $Y = y$ називають суму добутків можливих значень X на їхні умовні ймовірності:

$$M(X | Y = y) = \sum_{i=1}^n x_i p(x_i | y).$$

Умовним математичним сподіванням неперервної випадкової величини Y за $X = x$ називають невластний інтеграл із нескінченними межами інтегрування від добутку змінної y на умовну щільність випадкової величини Y за $X = x$:

$$M(Y|X=x) = \int_{-\infty}^{\infty} y\psi(y|x)dy.$$

Умовним математичним сподіванням неперервної випадкової величини X за $Y = y$ називають невластний інтеграл із нескінченними межами інтегрування від добутку змінної x на умовну щільність випадкової величини X за $Y = y$:

$$M(X|Y=y) = \int_{-\infty}^{\infty} x\varphi(x|y)dx.$$

Початковим моментом $\nu_{k,s}$ порядку k,s системи (X, Y) називають математичне сподівання добутку $X^k Y^s$:

$$\nu_{k,s} = M(X^k Y^s).$$

Окремим випадком є $\nu_{1,0} = M(X)$, $\nu_{0,1} = M(Y)$.

Центральним моментом $\mu_{k,s}$ порядку k,s системи (X, Y) називають математичне сподівання добутку відхилень, відповідно, k -го і μ -го ступеня:

$$\mu_{k,s} = M((X - M(X))^k (Y - M(Y))^s).$$

Окремим випадком є $\mu_{2,0} = M(X - M(X))^2 = D(X)$,

$$\mu_{0,2} = M(Y - M(Y))^2 = D(Y).$$

Кореляційним моментом μ_{xy} системи (X, Y) називають центральний момент $\mu_{1,1}$ порядку 1,1, що дорівнює математичному сподіванню добутку відхилень цих величин:

$$\mu_{xy} = M((X - M(X))(Y - M(Y))).$$

Для дискретних величин формула набуває такого вигляду:

$$\mu_{xy} = \sum_{i=1}^n \sum_{j=1}^m (x_i - M(X))(y_j - M(Y))p(x_i, y_j),$$

а для неперервних величин:

$$\begin{aligned} \mu_{xy} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - M(X))(y - M(Y))f(x, y)dx dy = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y)dx dy - M(X) \cdot M(Y). \end{aligned}$$

Кореляційний момент також має назву **коефіцієнт коваріації**, тобто одночасної варіації випадкових величин X і Y . Відповідно, його позначають K_{xy} , або $\text{cov}(X, Y)$.

Властивості коефіцієнта коваріації:

1. Коефіцієнт коваріації не змінюється, якщо випадкові величини поміняти місцями:

$$\text{cov}(X, Y) = \text{cov}(Y, X).$$

2. Кореляційний момент між однаковими компонентами двовимірної випадкової величини дорівнює дисперсії цього компонента:

$$\mu_{xx} = M((X - M(X))(X - M(X))) = D(X);$$

$$\mu_{yy} = M((Y - M(Y))(Y - M(Y))) = D(Y).$$

3. **Коваріаційною матрицею двовимірної випадкової величини** називають матрицю $C = (c_{ij})_{n \times n}$, де $c_{ij} = \text{cov}(X_i, X_j)$, тобто для двовимірної випадкової величини коваріаційна матриця має такий вигляд:

$$C = \begin{pmatrix} D(X) & \text{cov}(X, Y) \\ \text{cov}(Y, X) & D(Y) \end{pmatrix}.$$

Коваріаційна матриця є симетричною відносно головної діагоналі. Вона є невід'ємно визначеною, звідси випливає, що

$$D(X) \cdot D(Y) - \text{cov}(X, Y) \cdot \text{cov}(Y, X) \geq 0.$$

Отже, кореляційний момент є обмеженим за модулем:

$$|\mu_{xy}| \leq \sqrt{D(X) \cdot D(Y)}.$$

4. Для дисперсії суми залежних випадкових величин X і Y буде таке співвідношення:

$$D(X + Y) = D(X) + D(Y) + 2\text{cov}(X, Y).$$

Якщо випадкові величини є незалежними, то $\text{cov}(X, Y) = 0$, і означають таке співвідношення:

$$D(X + Y) = D(X) + D(Y).$$

У ході обчислення кореляційного моменту, замість означення, зручно використовувати таку перетворену формулу:

$$\mu_{xy} = M(XY) - M(X) \cdot M(Y).$$

Теорема. Кореляційний момент двох незалежних випадкових величин X і Y дорівнює нулю.

Отже, якщо кореляційний момент не дорівнює нулю, то це є ознакою наявності зв'язку між випадковими величинами X і Y .

Коефіцієнтом кореляції називають числову характеристику, яку обчислюють за такою формулою:

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

Властивості коефіцієнта кореляції:

1. Для незалежних випадкових величин $r_{xy} = 0$, оскільки в цьому випадку $\mu_{xy} = 0$.

2. За абсолютною величиною коефіцієнт кореляції є не більшим за одиницю: $-1 \leq r_{xy} \leq 1$.

3. Оскільки для кореляційного моменту $\text{cov}(X, Y) = \text{cov}(Y, X)$, то для коефіцієнта кореляції теж $\rho_{xy} = \rho_{yx}$.

Випадкові величини, для яких $r_{xy} = 0$, називають **некорельованими**.

Функції випадкового аргументу

Якщо кожному можливому значенню випадкової величини X відповідає одне можливе значення випадкової величини Y , то Y називають **функцією випадкового аргументу X** і записують так: $Y = f(X)$.

Якщо аргумент X – дискретна випадкова величина і різним можливим значенням аргументу X відповідають різні можливі значення функції Y , то ймовірності відповідних значень X і Y є однаковими між собою, тобто $P(Y = y_i) = P(X = x_i)$.

Якщо $x_1, x_2, \dots, x_n$ – можливі значення дискретної випадкової величини X , імовірності яких дорівнюють, відповідно, $p_1, p_2, \dots, p_n$, то для випадкової величини X закон розподілу записують у такому вигляді (табл. 7.2):

Таблиця 7.2

Закон розподілу дискретної випадкової величини X

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

Якщо різним значенням X відповідають однакові значення Y , то ймовірності значення Y , що повторюють, додають.

Нехай $Y = f(X)$, де аргумент X – дискретна випадкова величина, тоді випадкова величина Y також є дискретною випадковою величиною з можливими значеннями $y_1 = f(x_1)$, $y_2 = f(x_2)$, $\dots$, $y_n = f(x_n)$.

Для того щоб випадкова величина Y набула значення $y_1 = f(x_1)$ потрібно, щоб випадкова величина X набула значення x_1 . Випадкова величина X набуває можливого значення x_1 з імовірністю p_1 . Аналогічно, для того щоб випадкова величина Y набула значення $y_2 = f(x_2)$ слід, щоб випадкова величина X набула значення x_2 . Випадкова величина X набуває можливого значення x_2 з імовірністю p_2 .

Отже, можна зробити висновок, що закон розподілу випадкової величини Y має такий вигляд (табл. 7.3):

Таблиця 7.3

Закон розподілу випадкової величини Y

$Y = f(X)$	$f(x_1)$	$f(x_2)$	$\dots$	$f(x_n)$
$P(Y)$	p_1	p_2	$\dots$	p_n

Якщо аргумент X – це неперервна випадкова величина, задана диференціальною функцією (щільністю розподілу) $f(x)$, і якщо $y = \varphi(x)$ – диференційована строго зростаюча або строго спадна функція, обернена функція якої $x = \psi(y)$, то диференціальну функцію $g(y)$ випадкової величини Y визначають із такої рівності:

$$g(y) = f(\psi(y)) \cdot |\psi'(y)|.$$

Примітка. Якщо функція $y = \varphi(x)$ не є монотонною в інтервалі можливих значень X , то потрібно цей інтервал розподілити на такі інтервали, у яких функція $y = \varphi(x)$ буде монотонною, і визначити диференціальні функції $g_i(y)$ у кожному із цих інтервалів, а потім обчислити їхню суму:

$$g(y) = \sum_{i=1}^n g_i(y) = \sum_{i=1}^n f(\psi_i(y)) \cdot |\psi_i'(y)|.$$

Числові характеристики функції дискретного випадкового аргументу

До числових характеристик дискретної випадкової величини Y належать математичне сподівання, дисперсія та середнє квадратичне відхилення. Їх обчислюють за такими формулами:

$$\text{математичне сподівання: } M(Y) = \sum_{k=1}^n f(x_k) p_k;$$

$$\text{дисперсію: } D(Y) = M(Y^2) - M^2(Y) = \sum_{k=1}^n (f(x_k))^2 p_k - M^2(Y);$$

$$\text{середнє квадратичне відхилення: } \sigma(Y) = \sqrt{D(Y)}.$$

*Числові характеристики функції
неперервного випадкового аргументу*

Нехай аргумент X – це неперервна випадкова величина, тоді математичне сподівання функції Y обчислюють за такою формулою:

$$M(Y) = M(\varphi(X)) = \int_{-\infty}^{\infty} \varphi(X) f(x) dx,$$

якщо $x \in (a, b)$, то:

$$M(Y) = \int_a^b \varphi(X) f(x) dx.$$

Дисперсію функції Y обчислюють за такою формулою:

$$\begin{aligned} D(Y) &= D(\varphi(X)) = \int_{-\infty}^{\infty} (\varphi(X) - M(\varphi(X)))^2 f(x) dx = \\ &= \int_{-\infty}^{\infty} \varphi^2(X) f(x) dx - (M(\varphi(X)))^2. \end{aligned}$$

Якщо $x \in (a, b)$, то:

$$D(Y) = \int_a^b (\varphi(X) - M(\varphi(X)))^2 f(x) dx = \int_a^b \varphi^2(X) f(x) dx - (M(\varphi(X)))^2.$$

Примітка. Якщо використовувати диференціальну функцію $g(y)$ випадкової величини Y , то математичне сподівання обчислюють за такою формулою:

$$M(Y) = \int_c^d y \cdot g(y) dy,$$

де c і d – межі інтегрування для Y ,

а дисперсію – за такою формулою:

$$D(Y) = \int_c^d y^2 \cdot g(y) dy - (M(Y))^2.$$

Функції двох випадкових аргументів

Означення. Якщо кожній парі можливих значень випадкових величин X та Y відповідає одне можливе значення випадкової величини Z , то Z є **функцією двох випадкових аргументів X та Y** : $Z = \varphi(X, Y)$.

Якщо X та Y є дискретними випадковими величинами, то й Z буде дискретною, а якщо X та Y є неперервними випадковими величинами, то й Z буде неперервною.

Диференціальну функцію суми незалежних випадкових величин називають **композицією**.

Якщо X та Y – це неперервні випадкові величини, то щільність розподілу $g(z)$ суми $Z = X + Y$ може бути визначено за допомогою такої рівності:

$$g(z) = \int_{-\infty}^{\infty} f_1(x) f_2(z-x) dx, \text{ або } g(z) = \int_{-\infty}^{\infty} f_1(z-y) f_2(y) dy,$$

де f_1, f_2 – щільності розподілу аргументів.

Якщо можливі значення аргументів є невід'ємними, то $g(z)$ визначають за такою формулою:

$$g(z) = \int_0^z f_1(x) f_2(z-x) dx, \text{ або } g(z) = \int_0^z f_1(z-y) f_2(y) dy.$$

Якщо обидві диференціальні функції $f_1(x)$ і $f_2(y)$ задано на обмежених інтервалах, то для обчислення диференціальної функції $Z = X + Y$ доцільно спочатку визначити інтегральну функцію $\zeta(z)$, а потім диференціювати її за Z :

$$g(z) = \zeta'(z),$$

$$\text{де } \zeta(z) = \iint f(x, y) dx dy = \int_{-\infty}^{\infty} dy \int_{-\infty}^{z-y} f(x, z-x) dx = \int_{-\infty}^{\infty} dx \int_{-\infty}^{z-x} f(z-y, y) dy,$$

а $f(x, y)$ – диференціальна функція системи випадкових величин (X, Y) .

Якщо X і Y є незалежними випадковими величинами, що мають диференціальні функції $f_1(x)$ і $f_2(y)$, то ймовірність потрапляння випадкової точки в область D дорівнює:

$$P((X, Y) \in D) = \iint_D f_1(x) f_2(y) dx dy.$$

7.3. Приклади розв'язування задач

Приклад 7.1. Задано закон розподілу ймовірностей дискретної двовимірної випадкової величини (табл. 7.4).

Визначмо закони розподілу X і Y .

Таблиця 7.4

**Закон розподілу ймовірностей
дискретної двовимірної випадкової величини**

$Y \backslash X$	10	15	20
2	0,03	0,09	0,11
4	0,10	0,15	0,24
5	0,12	0,13	0,03

Розв'язання. Слід визначити закон розподілу для складової X . Вона набуває таких значень: 10, 15, 20. Тепер обчислимо відповідні ймовірності, для цього треба підсумувати ймовірності за стовпцями:

$$P(10) = 0,03 + 0,10 + 0,12 = 0,25;$$

$$P(15) = 0,09 + 0,15 + 0,13 = 0,37;$$

$$P(20) = 0,11 + 0,24 + 0,03 = 0,38.$$

Перевірка: $0,25 + 0,37 + 0,38 = 1.$

Отже, закон розподілу складової X має такий вигляд (табл. 7.5):

Таблиця 7.5

Закон розподілу для складової X

X	10	15	20
P	0,25	0,37	0,38

Визначмо закон розподілу для складової Y , значення якої 2, 4, 6.

Для обчислення відповідних імовірностей складемо ймовірності за рядками:

$$P(2) = 0,03 + 0,09 + 0,11 = 0,23;$$

$$P(4) = 0,10 + 0,15 + 0,24 = 0,49;$$

$$P(6) = 0,12 + 0,13 + 0,03 = 0,28.$$

Перевірка: $0,23 + 0,49 + 0,28 = 1.$

Отже, закон розподілу складової Y має такий вигляд (табл. 7.6):

Таблиця 7.6

Закон розподілу для складової Y

Y	2	3	6
P	0,23	0,49	0,28

Приклад 7.2. Задано закон розподілу ймовірностей дискретної двовимірної випадкової величини (табл. 7.7).

Визначмо:

а) безумовні закони розподілу складових X і Y ;

б) умовний закон розподілу складової X за умови, що складова Y набула значення $y_1 = 10$;

в) умовний закон розподілу складової Y за умови, що складова X набула значення $x_2 = 7$.

Таблиця 7.7

**Закон розподілу ймовірностей
дискретної двовимірної випадкової величини**

$Y \backslash X$	5	7	10
10	0,31	0,12	0,19
20	0,11	0,15	0,12

Розв'язання:

а) складімо закон розподілу для складової X . Для цього підрахуймо такі ймовірності:

$$P(5) = 0,31 + 0,11 = 0,42;$$

$$P(7) = 0,12 + 0,15 = 0,27;$$

$$P(10) = 0,19 + 0,12 = 0,31.$$

Перевірка: $0,42 + 0,27 + 0,31 = 1$.

Отже, закон розподілу складової X має такий вигляд (табл. 7.8):

Таблиця 7.8

Закон розподілу для складової X

X	5	7	10
P	0,42	0,27	0,31

Для складової Y такі ймовірності:

$$P(10) = 0,31 + 0,12 + 0,19 = 0,62;$$

$$P(20) = 0,11 + 0,15 + 0,12 = 0,38.$$

Перевірка: $0,62 + 0,38 = 1$.

Отже, закон розподілу складової Y має такий вигляд (табл. 7.9):

Таблиця 7.9

Закон розподілу для складової Y

Y	10	20
P	0,62	0,38

б) тепер складемо умовний закон розподілу складової X за умови, що складова Y набула значення $y_1 = 10$. Для цього обчислімо умовні ймовірності можливих значень X за такою формулою:

$$P(x_i | y_i) = \frac{P(x_i, y_i)}{P(y_i)};$$

$$P(y_1) = P(10) = 0,62.$$

Тоді

$$P(x_1 | y_1) = \frac{P(x_1, y_1)}{P(y_1)} = \frac{0,31}{0,62} = \frac{1}{2};$$

$$P(x_2 | y_1) = \frac{P(x_2, y_1)}{P(y_1)} = \frac{0,12}{0,62} = \frac{6}{31};$$

$$P(x_3 | y_1) = \frac{P(x_3, y_1)}{P(y_1)} = \frac{0,19}{0,62} = \frac{19}{62}.$$

Отже, умовний закон розподілу складової X за умови, що складова Y набула значення $y_1 = 10$, має такий вигляд (табл. 7.10):

Таблиця 7.10

Закон розподілу для складової X

X	5	7	10
$P(X y_1)$	$\frac{1}{2}$	$\frac{6}{31}$	$\frac{19}{62}$

Перевірка: $\frac{1}{2} + \frac{6}{31} + \frac{19}{62} = 1;$

в) для визначення умовного закону розподілу складової Y за умови, що складова X набула значення $x_2 = 7$, обчислимо умовні ймовірності можливих значень Y за такою формулою:

$$P(y_i | x_i) = \frac{P(x_i, y_i)}{P(x_i)};$$

$$P(x_2) = P(7) = 0,12 + 0,15 = 0,27.$$

Далі

$$P(y_1 | x_2) = \frac{P(x_2, y_1)}{P(x_2)} = \frac{0,12}{0,27} = \frac{12}{27} = \frac{4}{9};$$

$$P(y_2 | x_2) = \frac{P(x_2, y_2)}{P(x_2)} = \frac{0,15}{0,27} = \frac{15}{27} = \frac{5}{9}.$$

Отже, умовний закон розподілу складової Y за умови, що складова X набула значення $x_2 = 7$, має такий вигляд (табл. 7.11):

Таблица 7.11

Закон розподілу для складової Y

Y	10	20
$P(Y x_2)$	$\frac{4}{9}$	$\frac{5}{9}$

Перевірка: $\frac{4}{9} + \frac{5}{9} = 1.$

Приклад 7.3. Задано інтегральну функцію розподілу ймовірностей двовимірної випадкової величини (X, Y) :

$$F(x, y) = \left(\frac{1}{4\pi} \arctg \frac{x}{4} + \frac{1}{8} \right) \left(\frac{1}{5\pi} \arctg \frac{y}{5} + \frac{1}{10} \right).$$

Обчислимо імовірність того, що в результаті випробування складова X набуде значення $X \leq 4$, а складова Y – значення $Y < 5$.

Розв'язання. Цю ймовірність визначаймо за такою формулою:

$$F(x, y) = P(X < x, Y < y).$$

Тому

$$\begin{aligned} P(X < x, Y < y) &= F(4, 5) = \left(\frac{1}{4\pi} \arctg \frac{4}{4} + \frac{1}{8} \right) \left(\frac{1}{5\pi} \arctg \frac{5}{5} + \frac{1}{10} \right) = \\ &= \left(\frac{1}{4\pi} \cdot \frac{\pi}{4} + \frac{1}{8} \right) \left(\frac{1}{5\pi} \cdot \frac{\pi}{4} + \frac{1}{10} \right) = \left(\frac{1}{16} + \frac{1}{8} \right) \left(\frac{1}{20} + \frac{1}{10} \right) = \frac{3}{16} \cdot \frac{3}{20} = \frac{9}{320}. \end{aligned}$$

Приклад 7.4. Задано інтегральну функцію двовимірної випадкової величини:

$$F(x, y) = \begin{cases} \frac{1}{54}(x^2y + xy^2), & \text{за } 0 \leq x \leq 3, 0 \leq y \leq 3; \\ 0, & \text{за межами квадрата.} \end{cases}$$

Визначмо ймовірність того, що в результаті випробування випадкова точка (X, Y) потрапить у квадрат, який обмежено лініями $x_1 = 1, x_2 = 2, y_1 = 2, y_2 = 3$.

Розв'язання. Для визначення ймовірності використаємо таку формулу:

$$P(x_1 < X < x_2, y_1 < Y < y_2) = [F(x_2, y_2) - F(x_1, y_2)] - [F(x_2, y_1) - F(x_1, y_1)].$$

Оскільки, за умовою $x_1 = 1, x_2 = 2, y_1 = 2, y_2 = 3$, то:

$$\begin{aligned} P(1 < X < 2, 2 < Y < 3) &= F(2, 3) - F(1, 3) - F(2, 2) + F(1, 2) = \\ &= \frac{1}{54}(2^2 \cdot 3 + 2 \cdot 3^2 - 1^2 \cdot 3 - 1 \cdot 3^2 - 2^2 \cdot 2 - 2 \cdot 2^2 + 1^2 \cdot 2 + 1 \cdot 2^2) = \\ &= \frac{1}{54}(12 + 18 - 3 - 9 - 8 - 8 + 2 + 4) = \frac{8}{54} = \frac{4}{27}. \end{aligned}$$

Приклад 7.5. Задано інтегральну функцію двовимірної випадкової величини (X, Y) :

$$F(x, y) = \begin{cases} (1 - e^{-5x})(1 - e^{-2y}), & \text{за } x > 0, y > 0; \\ 0, & \text{за } x \leq 0 \text{ або } y \leq 0. \end{cases}$$

Визначмо диференціальну функцію розподілу величини (X, Y) .

Розв'язання. За означенням, буде:

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}.$$

Спочатку, уважаючи, що $y = \text{const}$, визначмо частинну похідну за X від функції $F(x, y) = (1 - e^{-5x})(1 - e^{-2y})$:

$$\frac{\partial F(x, y)}{\partial x} = (1 - e^{-2y})(0 - e^{-5x}(-5)) = 5e^{-5x}(1 - e^{-2y}).$$

Тепер визначмо частинну похідну за y від функції $\frac{\partial F(x, y)}{\partial x}$, уважаючи, що $x = \text{const}$:

$$\frac{\partial^2 F(x, y)}{\partial x \partial y} = 5e^{-5x}(0 - e^{-2y}(-2)) = 10e^{-5x} \cdot e^{-2y} = 10e^{-(5x+2y)}.$$

Отже,

$$f(x, y) = \begin{cases} 10e^{-(5x+2y)}, & \text{за } x > 0, y > 0; \\ 0, & \text{за } x \leq 0 \text{ або } y \leq 0. \end{cases}$$

Приклад 7.6. Визначмо a , якщо диференціальна функція системи випадкових величин (X, Y) має такий вигляд:

$$f(x, y) = \begin{cases} a \cos x \cos y, & \text{за } 0 \leq x \leq \frac{\pi}{4}, 0 \leq y \leq \frac{\pi}{4}; \\ 0, & \text{за межами квадрата.} \end{cases}$$

Розв'язання. Для того щоб функція $f(x, y)$ була диференціальною функцією системи випадкових величин, потрібно виконання двох умов:

$$\text{а) } f(x, y) \geq 0; \quad \text{б) } \iint_D f(x, y) dx dy = 1.$$

Тоді:

$$\text{а) } f(x, y) = a \cos x \cos y, \quad \cos x \geq 0, \quad \cos y \geq 0, \quad \text{отже } a \geq 0;$$

$$\text{б) } \iint_D f(x, y) dx dy = 1;$$

$$\begin{aligned}
 a \iint_D \cos x \cos y dx dy &= a \int_0^{\frac{\pi}{4}} \int_0^{\frac{\pi}{4}} \cos x \cos y dx dy = a \int_0^{\frac{\pi}{4}} \cos x dx \int_0^{\frac{\pi}{4}} \cos y dy = \\
 &= a \int_0^{\frac{\pi}{4}} \cos x dx \sin y \Big|_0^{\frac{\pi}{4}} = a \int_0^{\frac{\pi}{4}} \cos x \left(\sin \frac{\pi}{4} - \sin 0 \right) dx = \frac{a\sqrt{2}}{2} \int_0^{\frac{\pi}{4}} \cos x dx = \\
 &= \frac{a\sqrt{2}}{2} \sin x \Big|_0^{\frac{\pi}{4}} = a \frac{\sqrt{2}}{2} \left(\sin \frac{\pi}{4} - 0 \right) = a \frac{\sqrt{2}}{2} \cdot \frac{\sqrt{2}}{2} = \frac{a}{2}.
 \end{aligned}$$

Отже, $\frac{a}{2} = 1$, звідси $a = 2$.

Приклад 7.7. Закон розподілу дискретної двовимірної випадкової величини задано таблицею (табл. 7.12).

Таблиця 7.12

Закон розподілу двовимірної дискретної випадкової величини

$Y \backslash X$	1	2	$\sum_{i=1}^n p_{ij}$
2	0,4	0,3	0,7
4	0,2	0,1	0,3
$\sum_{j=1}^m p_{ij}$	0,6	0,4	1 = 1

Визначмо кореляційний момент компонентів X і Y двовимірної випадкової величини.

Розв'язання. Обчислимо математичні сподівання випадкових величин X і Y за їхніми розподілами.

Для випадкової величини X за першим і четвертим рядками табл. 7.12 визначмо:

$$M(X) = \sum_{i=1}^n x_i p_i = 1 \cdot 0,6 + 2 \cdot 0,4 = 1,4.$$

Для випадкової величини Y за першим і останнім стовпчиками табл. 7.12 обчислимо математичне сподівання:

$$M(Y) = \sum_{j=1}^m y_j p_j = 2 \cdot 0,7 + 4 \cdot 0,3 = 2,6.$$

Визначмо математичне сподівання двовимірної випадкової величини, яка є добутком випадкових величин X і Y :

$$M(XY) = \sum_{i=1}^n \sum_{j=1}^m x_i y_j p_{ij} = 1 \cdot 2 \cdot 0,4 + 2 \cdot 2 \cdot 0,3 + 4 \cdot 1 \cdot 0,2 + 4 \cdot 2 \cdot 0,1 = 3,6.$$

За формулою $\mu_{xy} = M(XY) - M(X) \cdot M(Y)$, обчислимо кореляційний момент:

$$\mu_{xy} = 3,6 - 1,4 \cdot 2,6 = -0,04.$$

Отже, кореляційний момент не дорівнює нулю, відповідно, компоненти X і Y двовимірної випадкової величини не можна вважати незалежними.

Приклад 7.8. Задано закон розподілу дискретної випадкової величини X (табл. 7.13):

Таблиця 7.13

Закон розподілу дискретної випадкової величини X

X	1	3	4	6
$P(X)$	0,1	0,3	0,2	0,4

Побудуємо закон розподілу ймовірностей для випадкової величини $Y = 4X^2 + 3$.

Розв'язання. Оскільки $Y = 4X^2 + 3$, то закон розподілу дискретної випадкової величини Y має такий вигляд (табл. 7.14):

Закон розподілу дискретної випадкової величини Y

$Y = 4X^2 + 3$	7	39	67	147
$P(Y)$	0,1	0,3	0,2	0,4

Приклад 7.9. Напишімо закон розподілу функції $Z = X + Y$, якщо

X	1	2	3
p	0,2	0,5	0,3

та

Y	4	5	6
p	0,1	0,6	0,3

Розв'язання. Можливі значення випадкової величини Z дорівнюють сумі кожного можливого значення випадкової величини X з усіма значеннями Y , тобто:

$$z_1 = 1 + 4 = 5, \quad z_2 = 1 + 5 = 6, \quad z_3 = 1 + 6 = 7,$$

$$z_4 = 2 + 4 = 6, \quad z_5 = 2 + 5 = 7, \quad z_6 = 2 + 6 = 8,$$

$$z_7 = 3 + 4 = 7, \quad z_8 = 3 + 5 = 8, \quad z_9 = 3 + 6 = 9.$$

Визначмо ймовірності цих можливих значень. Для того щоб $Z = 5$ достатньо, щоб випадкова величина X набула значення $x_1 = 1$, а випадкова величина Y – значення $y_1 = 4$. Імовірності цих можливих значень, відповідно, дорівнюють 0,2 та 0,1.

Величини X та Y є незалежними, тому, за теоремою множення, імовірність їхньої сумісної появи дорівнює $0,2 \cdot 0,1 = 0,02$.

Аналогічно визначаймо:

$$P(Z = 1 + 5) = 0,2 \cdot 0,6 = 0,12;$$

$$P(Z = 1 + 6) = 0,2 \cdot 0,3 = 0,06;$$

$$P(Z = 2 + 4) = 0,5 \cdot 0,1 = 0,05;$$

$$P(Z = 2 + 5) = 0,5 \cdot 0,6 = 0,30;$$

$$P(Z = 2 + 6) = 0,5 \cdot 0,3 = 0,15;$$

$$P(Z = 3 + 4) = 0,3 \cdot 0,1 = 0,03;$$

$$P(Z = 3 + 5) = 0,3 \cdot 0,6 = 0,18;$$

$$P(Z = 3 + 6) = 0,3 \cdot 0,3 = 0,09.$$

Отже, розподіл має такий вигляд (табл. 7.15):

Таблиця 7.15

Закон розподілу функції Z

Z	5	6	7	6	7	8	7	8	9
p	0,02	0,12	0,06	0,05	0,30	0,15	0,03	0,18	0,09

Додавання ймовірності несумісних подій:

$$Z = z_2, Z = z_4 \quad 0,12 + 0,05 = 0,17;$$

$$Z = z_3, Z = z_5, Z = z_7 \quad 0,06 + 0,3 + 0,03 = 0,39;$$

$$Z = z_6, Z = z_8 \quad 0,15 + 0,18 = 0,33.$$

Остаточно, закон розподілу функції $Z = X + Y$ набуває такого вигляду (табл. 7.16):

Таблиця 7.16

Закон розподілу функції Z

Z	5	6	7	8	9
p	0,02	0,17	0,39	0,33	0,09

$$\text{Перевірка: } 0,02 + 0,17 + 0,39 + 0,33 + 0,09 = 1.$$

Приклад 7.10. Випадкова величина X є розподіленою за нормальним законом із математичним сподіванням, що дорівнює нулю. Напишімо закон розподілу функції $Y = X^5$.

Розв'язання. Диференціальна функція нормального закону розподілу неперервної величини X має такий вигляд:

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

Функція $Y = X^5$ є диференційованою, $Y' = 5X^4 > 0$, тому вона зростає для всіх $x \in (-\infty, +\infty)$.

Оскільки $Y = X^5$, то $X = \sqrt[5]{Y}$, тоді $\psi(y) = \sqrt[5]{y} = y^{\frac{1}{5}}$.

Отже, буде:

$$g(y) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{y^{\frac{2}{5}}}{2\sigma^2}} \cdot \left| \left(y^{\frac{1}{5}} \right)' \right| = \frac{e^{\frac{y^{\frac{2}{5}}}{2\sigma^2}}}{5\sigma\sqrt{2\pi} \cdot y^{\frac{4}{5}}}.$$

Приклад 7.11. Дискретну випадкову величину задано законом розподілу (табл. 7.17):

Таблиця 7.17

Закон розподілу дискретної випадкової величини X

X	0	1	4	9
$P(X)$	0,729	0,243	0,027	0,001

Обчислімо математичне сподівання випадкової величини $Y = X^2 + 1$.

Розв'язання. Закон розподілу випадкової величини $Y = X^2 + 1$ має такий вигляд (табл. 7.18):

Таблиця 7.18

Закон розподілу дискретної випадкової величини Y

$Y = X^2 + 1$	$0^2 + 1 = 1$	$1^2 + 1 = 2$	$4^2 + 1 = 17$	$9^2 + 1 = 82$
$P(Y)$	0,729	0,243	0,027	0,001

Математичне сподівання випадкової величини $Y = X^2 + 1$ дорівнює:

$$M(Y) = 1 \cdot 0,729 + 2 \cdot 0,243 + 17 \cdot 0,027 + 82 \cdot 0,001 = 1,756.$$

Для обчислення дисперсії побудуємо закон розподілу випадкової величини Y^2 (табл. 7.19):

Таблиця 7.19

Закон розподілу дискретної випадкової величини Y^2

Y^2	1	4	289	6 724
$P(Y)$	0,729	0,243	0,027	0,001

Тоді

$$\begin{aligned} M(Y^2) &= 1 \cdot 0,729 + 4 \cdot 0,243 + 289 \cdot 0,027 + 6\,724 \cdot 0,001 = \\ &= 0,729 + 0,972 + 7,803 + 6,724 = 16,228. \end{aligned}$$

Отже,

$$D(Y) = M(Y^2) - M(Y)^2 = 16,228 - (1,756)^2 = 13,144.$$

Середнє квадратичне відхилення дорівнює:

$$\sigma(Y) = \sqrt{D(Y)} = \sqrt{13,144} \approx 3,626.$$

Приклад 7.12. Неперервну випадкову величину X задано щільністю розподілу ймовірностей:

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

Визначмо числові характеристики випадкової величини $Y = 3X^2$.

Розв'язання. За умовою $f(x) = \sin x$, тоді

$$M(Y) = M(3X^2) = \int_0^{\frac{\pi}{2}} 3x^2 \sin x \, dx;$$

$$\int_0^{\frac{\pi}{2}} 3x^2 \sin x dx = 3 \left(-x^2 \cos x \Big|_0^{\frac{\pi}{2}} + 2 \int_0^{\frac{\pi}{2}} x \cos x dx \right) = 3 \left(x \sin x \Big|_0^{\frac{\pi}{2}} - \int_0^{\frac{\pi}{2}} \sin x dx \right) =$$

$$= 3 \left(\pi + 2 \cos x \Big|_0^{\frac{\pi}{2}} \right) = 3(\pi - 2) = 3\pi - 6.$$

Отже, $M(Y) = 3\pi - 6$.

Обчислимо дисперсію функції $Y = 3X^2$:

$$D(Y) = M(Y^2) - M^2(Y) = \int_0^{\frac{\pi}{2}} 9x^4 \sin x dx - M^2(Y);$$

$$M(Y^2) = \int_0^{\frac{\pi}{2}} 9x^4 \sin x dx = 9 \left(-x^4 \cos x \Big|_0^{\frac{\pi}{2}} + 4 \int_0^{\frac{\pi}{2}} x^3 \cos x dx \right) =$$

$$= 36 \left(x^3 \sin x \Big|_0^{\frac{\pi}{2}} - 3 \int_0^{\frac{\pi}{2}} x^2 \sin x dx \right) = 36 \left(\frac{\pi^3}{8} - 3(\pi - 2) \right) = \frac{36}{8} \pi^3 - 108\pi + 216;$$

$$D(Y) = \frac{36}{8} \pi^3 - 108\pi + 216 - (3\pi - 6)^2 = \frac{36}{8} \pi^3 - 108\pi + 216 - 9\pi^2 + 36\pi -$$

$$- 36 = \frac{36}{8} \pi^3 - 9\pi^2 - 72\pi + 180 \approx 4,4997;$$

$$\sigma(Y) = \sqrt{D(Y)} = \sqrt{4,4997} \approx 2,1213.$$

Приклад 7.13. Випадкові величини X та Y мають такі закони розподілу:

X	2	4	6	6
P_I	0,3	0,4	0,1	0,2

та

Y	0	1	2
P_{II}	0,5	0,3	0,2

Складімо закон розподілу для: 1) $X + Y$; 2) $X \cdot Y$.

Розв'язання.

Випадкова величина $X + Y$ набуває таких значень:

2 – з імовірністю 0,15;

3 – з імовірністю 0,19;

4 – якщо $X + Y = 2 + 2$ або $X + Y = 4 + 0$ з імовірністю $0,06 + 0,20 = 0,26$ (для несумісних подій імовірності додають);

5 – з імовірністю 0,12;

6 – якщо $X + Y = 4 + 2$ або $X + Y = 6 + 0$ з імовірністю $0,08 + 0,05 = 0,13$;

7 – з імовірністю 0,03;

8 – якщо $X + Y = 6 + 2$ або $X + Y = 8 + 0$ з імовірністю $0,10 + 0,02 = 0,12$;

9 – з імовірністю 0,06;

10 – з імовірністю 0,04.

Запишімо ці розрахунки у вигляді таблиці (табл. 7.20):

Таблица 7.20

Розрахункова таблиця

X	Y	$X + Y$	$P_I \cdot P_{II}$
2	0	$2 + 0 = 2$	$0,3 \cdot 0,5 = 0,15$
2	1	$2 + 1 = 3$	$0,3 \cdot 0,3 = 0,09$
2	2	$2 + 2 = 4$	$0,3 \cdot 0,2 = 0,06$
4	0	$4 + 0 = 4$	$0,4 \cdot 0,5 = 0,20$
4	1	$4 + 1 = 5$	$0,4 \cdot 0,3 = 0,12$
4	2	$4 + 2 = 6$	$0,4 \cdot 0,2 = 0,08$
6	0	$6 + 0 = 6$	$0,1 \cdot 0,5 = 0,05$
6	1	$6 + 1 = 7$	$0,1 \cdot 0,3 = 0,03$
6	2	$6 + 2 = 8$	$0,1 \cdot 0,2 = 0,02$
8	0	$8 + 0 = 8$	$0,2 \cdot 0,5 = 0,10$
8	1	$8 + 1 = 9$	$0,2 \cdot 0,3 = 0,06$
8	2	$8 + 2 = 10$	$0,2 \cdot 0,2 = 0,04$

Перевірка:

$$0,15 + 0,09 + 0,26 + 0,12 + 0,13 + 0,03 + 0,12 + 0,06 + 0,04 = 1.$$

Отже, остаточно буде такий ряд розподілу для випадкової величини $X + Y$ (табл. 7.21):

Таблиця 7.21

Закон розподілу випадкової величини $X + Y$

$X + Y$	2	3	4	5	6	7	8	9	10
p	0,15	0,09	0,26	0,12	0,13	0,03	0,12	0,06	0,04

Складімо закон розподілу випадкової величини $X \cdot Y$.

Випадкова величина $X \cdot Y$ набуває таких значень:

0, якщо $X \cdot Y = 2 \cdot 0$, або $X \cdot Y = 4 \cdot 0$, або $X \cdot Y = 6 \cdot 0$, або $X \cdot Y = 8 \cdot 0$ з імовірністю $0,15 + 0,20 + 0,05 + 0,10 = 0,50$ (для несумісних подій імовірності додають);

2 – з імовірністю 0,09;

4 – якщо $X \cdot Y = 2 \cdot 2$ або $X \cdot Y = 4 \cdot 1$ з імовірністю $0,06 + 0,12 = 0,18$;

6 – з імовірністю 0,03;

8 – якщо $X \cdot Y = 4 \cdot 2$ або $X \cdot Y = 8 \cdot 1$ з імовірністю $0,08 + 0,06 = 0,14$;

12 – з імовірністю 0,02;

16 – з імовірністю 0,04.

Запишімо ці розрахунки у вигляді таблиці (табл. 7.22):

Таблиця 7.22

Розрахункова таблиця

X	Y	$X \cdot Y$	$P_I \cdot P_{II}$
2	0	$2 \cdot 0 = 0$	$0,3 \cdot 0,5 = 0,15$
2	1	$2 \cdot 1 = 2$	$0,3 \cdot 0,3 = 0,09$
2	2	$2 \cdot 2 = 4$	$0,3 \cdot 0,2 = 0,06$
4	0	$4 \cdot 0 = 0$	$0,4 \cdot 0,5 = 0,20$
4	1	$4 \cdot 1 = 4$	$0,4 \cdot 0,3 = 0,12$
4	2	$4 \cdot 2 = 8$	$0,4 \cdot 0,2 = 0,08$
6	0	$6 \cdot 0 = 0$	$0,1 \cdot 0,5 = 0,05$
6	1	$6 \cdot 1 = 6$	$0,1 \cdot 0,3 = 0,03$
6	2	$6 \cdot 2 = 12$	$0,1 \cdot 0,2 = 0,02$
8	0	$8 \cdot 0 = 0$	$0,2 \cdot 0,5 = 0,10$
8	1	$8 \cdot 1 = 8$	$0,2 \cdot 0,3 = 0,06$
8	2	$8 \cdot 2 = 16$	$0,2 \cdot 0,2 = 0,04$

Перевірка:

$$0,50 + 0,09 + 0,18 + 0,03 + 0,14 + 0,02 + 0,04 = 1.$$

Отже, буде такий ряд розподілу для випадкової величини $X \cdot Y$ (табл. 7.23):

Таблиця 7.23

Закон розподілу випадкової величини $X \cdot Y$

$X \cdot Y$	0	2	4	6	8	12	16
p	0,50	0,09	0,18	0,03	0,14	0,02	0,04

Приклад 7.14. Незалежні випадкові величини X та Y задано за допомогою щільності розподілу:

$$f(x) = \frac{1}{4} e^{-\frac{x}{4}} \quad (0 \leq x < \infty); \quad f(y) = \frac{1}{3} e^{-\frac{y}{3}} \quad (0 \leq y < \infty).$$

Визначмо композицію цих законів.

Розв'язання. Визначмо щільність розподілу випадкової величини $Z = X + Y$.

Можливі значення аргументів є невід'ємними, тому можна застосувати таку формулу:

$$g(z) = \int_0^z f_1(x) f_2(z-x) dx.$$

Отже,

$$\begin{aligned} g(z) &= \int_0^z f_1(x) f_2(z-x) dx = \int_0^z \left(\frac{1}{4} e^{-\frac{x}{4}} \right) \left(\frac{1}{3} e^{-\frac{z-x}{3}} \right) dx = \frac{1}{12} e^{-\frac{z}{3}} \int_0^z e^{-\frac{x}{12}} dx = \\ &= e^{-\frac{z}{3}} \left(1 - e^{-\frac{z}{12}} \right). \end{aligned}$$

7.4. Вправи для самостійної роботи

7.1. Задано закон розподілу ймовірностей дискретної двовимірної випадкової величини (табл. 7.24).

Складіть закон розподілу складових X і Y .

Таблиця 7.24

Закон розподілу двовимірної дискретної випадкової величини

$Y \backslash X$	5	15	0
2	0,15	0,21	0,05
3	0,16	0,30	0,13

7.2. Складіть закон розподілу складових X і Y , якщо закон розподілу ймовірностей дискретної двовимірної випадкової величини має такий вигляд (табл. 7.25):

Таблиця 7.25

Закон розподілу двовимірної дискретної випадкової величини

$Y \backslash X$	10	15	20	25
-1	0,05	0,08	0,11	0,12
0	0,04	0,12	0,10	0,08
1	0,06	0,09	0,07	0,08

7.3. Задано закон розподілу ймовірностей дискретної двовимірної випадкової величини (табл. 7.26).

Визначте:

- а) безумовні закони розподілу складових X і Y ;
- б) умовний закон розподілу складової X за умови, що складова Y набуде значення $y_2 = 2,8$;
- в) умовний закон розподілу складової Y за умови, що складова X набуде значення $x_2 = 45$.

Таблиця 7.26

Закон розподілу двовимірної дискретної випадкової величини

$Y \backslash X$	40	45	50	55
1,5	0,07	0,08	0,09	0,12
2,8	0,06	0,12	0,10	0,08
3,3	0,08	0,02	0,06	0,02

7.4. Задано закон розподілу ймовірностей дискретної двовимірної випадкової величини (табл. 7.27).

Таблиця 7.27

Закон розподілу двовимірної дискретної випадкової величини

$Y \backslash X$	11	15	17
2	0,06	0,05	0,04
9	0,09	0,09	0,11
14	0,10	0,12	0,06
17	0,07	0,09	0,12

Визначте:

- а) безумовні закони розподілу складових X і Y ;
- б) умовний закон розподілу складової X за умови, що складова Y набуде значення $y_3 = 14$;
- в) умовний закон розподілу складової Y за умови, що складова X набуде значення $x_1 = 11$.

7.5. На біржі продають акції п'яти компаній, для яких складено прогноз можливого підвищення їхньої вартості. Перші три компанії є будівельними. Для акцій першої компанії ймовірність зростання вартості становить 0,2; для другої – 0,1; для третьої – 0,4. Четверта та п'ята компанії є виробниками продовольчих товарів. Для четвертої компанії ймовірність зростання вартості становить 0,3; для п'ятої – 0,1. Клієнт придбав 2 акції

четвертої компанії, 5 акцій третьої компанії та по 1 акції другої та п'ятої компанії.

Складіть закон розподілу, розглядаючи двовимірну випадкову величину, одним компонентом якої є кількість акцій будівельних компаній, ціна на які може зрости, а другим компонентом – кількість акцій компаній із виготовлення харчових продуктів, вартість яких буде зростати, якщо підвищення вартості акції різних компаній є незалежними подіями.

7.6. Інтегральна функція розподілу ймовірностей двовимірної випадкової величини (X, Y) має такий вигляд:

$$F(x, y) = \left(\frac{1}{\pi} \arctg 2x + \frac{1}{2} \right) \left(\frac{1}{\pi} \arctg 3y + \frac{1}{2} \right).$$

Визначте імовірність того, що складова X набуде значення $X < \frac{1}{2}$

та водночас складова Y – значення $Y < \frac{1}{3}$.

7.7. Перевірте, чи становлять дані, наведені в таблиці (табл. 7.28), закон розподілу двовимірної випадкової величини.

Таблиця 7.28

Вихідні дані

$Y \backslash X$	0	2	4	6
5	0,15	0,10	0,10	0,05
10	0,20	0,25	0,10	0,05

Якщо в таблиці дійсно наведено закон розподілу двовимірної випадкової величини (X, Y) , побудуйте умовний розподіл випадкової величини X , якщо $Y = 10$.

7.8. За умовами задачі 7.7, обчисліть основні числові характеристики двовимірної випадкової величини (X, Y) . За значенням коефіцієнта кореляції зробіть висновок про наявність зв'язку між її компонентами.

7.9. За умовами задачі 7.8, складіть коваріаційну матрицю випадкової величини (X, Y) і перевірте, чи є ця матриця невід'ємно визначеною.

7.10. Задано розподіл імовірностей дискретної двовимірної випадкової величини (табл. 7.29):

Таблиця 7.29

Закон розподілу двовимірної дискретної випадкової величини

$Y \backslash X$	1,4	1,6	2,0
0	0,11	0,12	0,07
4,2	0,18	0,11	0,11
6,2	0,10	0,08	0,12

Обчисліть:

- а) математичне сподівання складових;
- б) дисперсії випадкових;
- в) середні квадратичні відхилення;
- г) коефіцієнт кореляції.

7.11. Задано розподіл імовірностей дискретної двовимірної випадкової величини (табл. 7.30):

Таблиця 7.30

Закон розподілу двовимірної дискретної випадкової величини

$Y \backslash X$	10	15	20	25
0	0,10	0,15	0,14	0,11
4	0,09	0,08	0,07	0,06
6	0,04	0,06	0,08	0,02

Обчисліть:

- 1) умовні математичні сподівання складової X за $Y = y_2 = 4$, а також складової Y за $X = x_3 = 20$;

2) умовні математичні сподівання складової X за $Y = y_1 = 0$, а також складової Y за $X = x_2 = 15$.

7.12. Систему двох випадкових величин (X, Y) рівномірно розподілено в трикутнику, обмеженому такими прямими: $y = x$, $y = 0$, $x = 2$.

Визначте:

- 1) $f(x, y)$;
- 2) математичні сподівання складових X і Y ;
- 3) дисперсії складових X і Y ;
- 4) коефіцієнт кореляції.

7.13. Диференціальна функція системи двох випадкових величин (X, Y) :

$$f(x, y) = \begin{cases} a(x + y), & \text{в області } D; \\ 0, & \text{за межами області } D, \end{cases}$$

де область D обмежено такими лініями: $x = 0$, $x = 3$, $y = 0$, $y = 3$.

Визначте:

- 1) a ;
- 2) математичні сподівання складових X і Y ;
- 3) дисперсії складових X і Y ;
- 4) коефіцієнт кореляції.

7.14. Задано щільність сумісного розподілу неперервної випадкової величини (X, Y) :

$$f(x, y) = \begin{cases} 2 \cos x \cos y, & \text{в області } D: 0 \leq x \leq \frac{\pi}{2}, 0 \leq y \leq \frac{\pi}{2}; \\ 0, & \text{за межами області } D. \end{cases}$$

Обчисліть математичні сподівання і дисперсії складових X і Y .

7.15. Задано щільність сумісного розподілу неперервної випадкової величини (X, Y) :

$$f(x, y) = \begin{cases} 2 \sin(x + y), & \text{в області } D: 0 \leq x \leq \frac{\pi}{2}, 0 \leq y \leq \frac{\pi}{2}; \\ 0, & \text{за межами області } D. \end{cases}$$

Визначте:

- 1) коефіцієнт a ;
- 2) $M(X), M(Y)$;
- 3) $D(X), D(Y)$;
- 4) коефіцієнт кореляції.

7.16. Задано диференціальну функцію неперервної двовимірної випадкової величини (X, Y) :

$$f(x, y) = \begin{cases} \frac{1}{2} \sin x \sin y, & \text{в області } D: 0 \leq x \leq \pi, 0 \leq y \leq \pi; \\ 0, & \text{за межами області } D. \end{cases}$$

Обчисліть:

- 1) математичні сподівання складових X і Y ;
- 2) дисперсії складових X і Y ;
- 3) коефіцієнт кореляції.

7.17. Дискретну випадкову величину X задано законом розподілу (табл. 7.31):

Таблиця 7.31

Закон розподілу дискретної випадкової величини X

X	-1	1	2	3
$P(X)$	0,4	0,3	0,2	0,1

Визначте закон розподілу випадкової величини Y :

- 1) $Y = 3X$; 2) $Y = 3X - 1$; 3) $Y = X^2$; 4) $Y = |X|$; 5) $Y = 2^X$.

7.18. Задано розподіли випадкових величин X та Y :

X	3	4	6
p	0,2	0,3	0,5

 та

Y	1	8
p	0,7	0,3

Складіть розподіл величини $Z = X + Y$ і визначте математичне сподівання та дисперсію.

7.19. Випадкову величину X розподілено рівномірно на інтервалі $(0, 3)$.

Обчисліть математичне сподівання та дисперсію випадкової величини $Y = 3 + 2X^2$.

7.20. Випадкову величину X розподілено нормально з параметрами $a = 2$, $\sigma = 3$ на інтервалі $(0, 3)$.

Визначте щільність розподілу випадкової величини $Y = 2X^2$.

7.21. Випадкову величину X розподілено рівномірно на інтервалі $(1, 4)$.

Визначте $M(3X^2 - 2X + 4)$.

7.22. Випадкова величина X має рівномірний розподіл в інтервалі $(1, 2)$.

Визначте математичне сподівання і дисперсію випадкової величини $Y = \frac{1}{X}$.

7.23. Диференціальна функція випадкової величини X :

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

Обчисліть математичне сподівання і дисперсію випадкової величини $Y = X^2$:

- 1) без використання диференціальної функції Y ;
- 2) використовуючи диференціальну функцію $g(y)$.

7.24. Неперервну випадкову величину X розподілено за показниковим законом:

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

Визначте математичне сподівання і дисперсію випадкової величини $Y = e^{-X}$.

7.25. Випадкові збільшення цін акцій двох компаній X та Y протягом дня мають сумісний розподіл, що наведено в табл. 7.32:

Таблиця 7.32

Закон розподілу двовимірної дискретної випадкової величини

$Y \backslash X$	-1	+1
-1	0,1	0,4
1	0,3	0,2

Перевірте чи є ці величини взаємно незалежними.

7.26. Сумісний розподіл двовимірної дискретної випадкової величини наведено в табл. 7.33:

Таблиця 7.33

Закон розподілу двовимірної дискретної випадкової величини

$Y \backslash X$	-1	+1
-1	$\frac{1}{4}$	$\frac{1}{2} - a$
1	$\frac{1}{4}$	a

Визначте значення a , за якого випадкові величини є незалежними.

7.27. Розподіл двовимірної дискретної випадкової величини наведено в табл. 7.34:

Таблиця 7.34

Закон розподілу двовимірної дискретної випадкової величини

$Y \backslash X$	1	4	7	9
2	0,14	0,15	0,14	0,11
5	0,09	0,14	0,15	0,08

Визначте:

- 1) умовний розподіл та умовне математичне сподівання Y за умови, що $X = 4$;
- 2) умовне математичне сподівання X за умови, що $Y = 5$.

7.28. Неперервні випадкові величини X та Y мають нормальний розподіл із такими математичними сподіваннями та дисперсіями: $a_1 = 4$; $a_2 = 8$; $D_1 = 0,5$; $D_2 = 0,8$ та $a_1 = 5$; $a_2 = 6$; $D_1 = 0,6$; $D_2 = 1,2$ відповідно.

Визначте математичне сподівання і дисперсію випадкової величини $Z = X + Y$.

7.5. Тестові завдання

7.1. Інтегральною функцією розподілу ймовірностей двовимірної випадкової величини (X, Y) називають функцію $F(x, y)$ таку, що

А $F(x, y) = P(X \leq x, Y < y)$

Б $F(x, y) = P(X < x, Y \leq y)$

В $F(x, y) = P(X < x, Y < y)$

Г $F(x, y) = P(X \leq x, Y \leq y)$

7.2. Значення інтегральної функції задовольняють подвійну нерівність

А	Б	В	Г
$0 \leq F(x, y) < 1$	$0 \leq F(x, y) \leq 1$	$0 < F(x, y) \leq 1$	$0 < F(x, y) < 1$

7.3. Умовним розподілом складової X за $Y = y_j$ називають сукупність значень X і відповідні їм умовні ймовірності, які обчислюють за такою формулою

$$\mathbf{А} \quad P(x_i | y_j) = \frac{P(x_i, y_j)}{P(y_j)}$$

$$\mathbf{Б} \quad P(y_j | x_i) = \frac{P(x_i, y_j)}{P(x_i)}$$

$$\mathbf{В} \quad P(x_i | y_j) = \frac{P(x_i, y_j)}{P(x_i)}$$

$$\mathbf{Г} \quad P(y_j | x_i) = \frac{P(x_i, y_j)}{P(y_j)}$$

7.4. Умовним розподілом складової Y за $X = x_j$ називають сукупність значень Y і відповідні їм умовні ймовірності, які обчислюють за такою формулою

$$\mathbf{А} \quad P(x_i | y_j) = \frac{P(x_i, y_j)}{P(y_j)}$$

$$\mathbf{Б} \quad P(y_j | x_i) = \frac{P(x_i, y_j)}{P(x_i)}$$

$$\text{В } P(x_i | y_j) = \frac{P(x_i, y_j)}{P(y_j)}$$

$$\text{Г } P(y_j | x_i) = \frac{P(x_i, y_j)}{P(x_i)}$$

7.5. Математичні сподівання дискретних випадкових величин X і Y , що входять до системи, визначають за такими формулами

$$\text{А } M(X) = \sum_{i=1}^n \sum_{j=1}^m x_i p_{ij}, \quad M(Y) = \sum_{i=1}^n \sum_{j=1}^m y_j p_{ij}$$

$$\text{Б } M(X) = \sum_{i=1}^n \sum_{j=1}^m x_j p_{ij}, \quad M(Y) = \sum_{i=1}^n \sum_{j=1}^m y_j p_{ij}$$

$$\text{В } M(X) = \sum_{i=1}^n \sum_{j=1}^m x_i p_{ij}, \quad M(Y) = \sum_{i=1}^n \sum_{j=1}^m y_i p_{ij}$$

$$\text{Г } M(X) = \sum_{j=1}^m \sum_{i=1}^n x_i p_{ij}, \quad M(Y) = \sum_{i=1}^n \sum_{j=1}^m y_j p_{ij}$$

7.6. Для обчислення дисперсії дискретних випадкових величин X і Y застосовують такі формули

$$\text{А } D(X) = \sum_{i=1}^n \sum_{j=1}^m p_{ij} (x_j - M(X))^2, \quad D(Y) = \sum_{i=1}^n \sum_{j=1}^m p_{ij} (y_j - M(Y))^2$$

$$\text{Б } D(X) = \sum_{i=1}^n \sum_{j=1}^m p_{ij} (x_i - M(X))^2, \quad D(Y) = \sum_{i=1}^n \sum_{j=1}^m p_{ij} (y_j - M(Y))^2$$

$$\text{В } D(X) = \sum_{i=1}^n \sum_{j=1}^m p_{ij} (x_i - M(X))^2, \quad D(Y) = \sum_{i=1}^n \sum_{j=1}^m p_{ij} (y_i - M(Y))^2$$

$$\text{Г } D(X) = \sum_{j=1}^m \sum_{i=1}^n p_{ij} (x_i - M(X))^2, \quad D(Y) = \sum_{i=1}^n \sum_{j=1}^m p_{ij} (y_j - M(Y))^2$$

7.7. Кореляційним моментом μ_{xy} системи (X, Y) називають

А $\mu_{xy} = M((Y - M(X))(X - M(Y)))$

Б $\mu_{xy} = M((X - M(Y))(Y - M(X)))$

В $\mu_{xy} = M((X - M(X))(X - M(Y)))$

Г $\mu_{xy} = M((X - M(X))(Y - M(Y)))$

7.8. Коефіцієнт кореляції обчислюють за такою формулою

А	Б	В	Г
$r_{xy} = \frac{\mu_{xy}}{D_x D_y}$	$r_{xy} = \frac{\mu_{xy}}{\sigma_x \sigma_y}$	$r_{xy} = \frac{\rho_{xy}}{\sigma_x \sigma_y}$	$r_{xy} = \frac{v_{xy}}{\sigma_x \sigma_y}$

7.9. Коефіцієнт кореляції набуває таких значень

А	Б	В	Г
$ r_{xy} \geq 1$	$ r_{xy} \leq 1$	$r_{xy} \leq 0$	$r_{xy} \geq 0$

7.10. Для двовимірної випадкової величини коваріаційна матриця має такий вигляд

А $C = \begin{pmatrix} D(Y) & \text{cov}(X, Y) \\ \text{cov}(Y, X) & D(X) \end{pmatrix}$

Б $C = \begin{pmatrix} D(X) & \text{cov}(Y, X) \\ \text{cov}(X, Y) & D(Y) \end{pmatrix}$

$$\text{В } C = \begin{pmatrix} D(X) & \text{cov}(X, Y) \\ \text{cov}(Y, X) & D(Y) \end{pmatrix}$$

$$\text{Г } C = \begin{pmatrix} D(X) & \text{cov}(Y, X) \\ \text{cov}(Y, X) & D(Y) \end{pmatrix}$$

7.11. Якщо аргумент X – це дискретна випадкова величина і різним можливим значенням аргументу X відповідають однакові можливі значення функції Y , то ймовірності значення Y , що повторюються

А	Б	В	Г
додають	віднімають	множать	ділять

7.12. Якщо аргумент X – це неперервна випадкова величина, задана щільністю розподілу $f(x)$, то диференціальну функцію $g(y)$ випадкової величини Y визначають із такої рівності

А	Б	В
$g(y) = f(\psi(y)) \cdot \psi'(y) $	$g(y) = f(\psi(y)) \cdot \psi(y) $	$g(y) = \psi'(y) $

7.13. Нехай аргумент X – це дискретна випадкова величина, а задана функція $Y = f(X)$ випадкового аргументу X , тоді математичне сподівання функції Y дорівнює

А	Б	В	Г
$M(f(X)) = \sum_{i=1}^n f(x_i) p_i$	$M(Y) = \sum_{i=1}^n f(x_i)$	$M(f(X)) = \sum_{i=1}^n p_i$	інше

7.14. Нехай аргумент X – це неперервна випадкова величина, причому $x \in (a, b)$, тоді математичне сподівання функції Y дорівнює

А	Б	В
$M(Y) = \int_a^b f(x) dx$	$M(Y) = \int_a^b \varphi(X) f(x) dx$	$M(Y) = \int_a^b \varphi(X) f(y) dx$

7.15. Якщо використовувати диференціальну функцію $g(y)$ випадкової величини Y , то математичне сподівання визначають за такою формулою

А	Б	В
$M(Y) = \int_d^c yg(y)dy$	$M(Y) = \int_c^d yg(y)dy$	$M(Y) = \int_c^d g(y)dy$

7.16. Диференціальну функцію суми незалежних випадкових величин називають

А	Б	В	Г
композицією	комбінацією	сполученням	інакше

7.6. Запитання для самоперевірки

7.1. Дайте означення багатовимірної випадкової величини.

7.2. Якими бувають багатовимірні величини?

7.3. Наведіть геометричну інтерпретацію двовимірної випадкової величини.

7.4. Як можна задати закон розподілу двовимірної дискретної величини?

7.5. Що таке "умовний розподіл"? Як побудувати умовний розподіл, користуючись табличними даними про розподіл випадкової величини?

7.6. Як за табличного задавання розподілу дискретної двовимірної величини дістати інформацію про розподіл її компонентів?

7.7. Наведіть способи задавання двовимірної неперервної випадкової величини.

7.8. Як можна надати геометричну інтерпретацію розподілу неперервної випадкової величини?

7.9. Як визначають щільність імовірностей і які властивості вона має?

7.10. Які є основні числові характеристики розподілу двовимірної випадкової величини?

7.11. Якому теоретичному моменту відповідає коефіцієнт?

7.12. Що таке "коваріаційна матриця"?

7.13. Які властивості має коваріаційна матриця?

7.14. У чому полягають переваги використання коефіцієнта кореляції як характеристики щільності кореляційного зв'язку перед кореляційним моментом?

7.15. Чи може коефіцієнт кореляції бути від'ємним? Якщо так, то що це означає?

7.16. Чи можна вважати некорельовані випадкові величини незалежними?

7.17. Дайте означення функції розподілу випадкової величини.

7.18. Сформулюйте закон розподілу функції дискретного випадкового аргументу.

7.19. Назвіть основні числові характеристики функції дискретного випадкового аргументу.

7.20. Який вигляд має закон розподілу функції неперервного випадкового аргументу?

7.21. Назвіть основні числові характеристики функції неперервного випадкового аргументу.

7.22. Дайте означення функції двох випадкових аргументів.

7.23. Що називають композицією двох функцій?

7.7. Висновки за підрозділом 7

Після вивчення підрозділу здобувач вищої освіти має вміти:

складати закон розподілу двовимірної дискретної випадкової величини;

визначати умовні закони розподілу дискретних і неперервних двовимірних випадкових величин;

уміти обчислювати основні числові характеристики дискретних і неперервних двовимірних випадкових величин;

розуміти особливості побудови розподілів функції одного випадкового аргументу (дискретного чи неперервного);

розуміти особливості побудови розподілів функції двох випадкових аргументів (дискретних чи неперервних);

визначати основні числові характеристики функцій випадкового аргументу та робити висновки.

Література: [3; 7; 9; 11; 13; 16].

Розділ 2

Математична статистика

8. Граничні теореми теорії ймовірностей

8.1. Мета та компетентності

Метою вивчення підрозділу є ознайомлення зі змістом граничних теорем теорії ймовірностей та формування компетентностей щодо використання їх на практиці.

Компетентності, що формують під час вивчення підрозділу:

знання особливостей застосування теорем закону великих чисел і центральної граничної теореми;

уміння застосовувати граничні теореми теорії ймовірностей для обчислення ймовірностей відхилення між числовими характеристиками випадкової величини та результатами їхнього статистичного оцінювання.

8.2. Основні теоретичні відомості

Граничні теореми теорії ймовірностей можна умовно розподілити на два класи: закон великих чисел і граничні теореми.

Закон великих чисел математично описує стійкість середніх значень масових випадкових явищ. Закон великих чисел становлять декілька теорем, у кожній із яких за певних умов затверджений факт збіжності середніх характеристик під час проведення великої кількості випробувань до певних не випадкових, постійних величин.

Нерівність Маркова. Якщо випадкова величина набуває лише невід'ємних значень та має математичне сподівання, то для будь-якого $\varepsilon > 0$ має місце така нерівність:

$$P(X > \varepsilon) \leq \frac{M(X)}{\varepsilon}.$$

Нерівність Чебишова. Якщо випадкова величина має математичне сподівання та обмежену дисперсію, то для будь-якого $\varepsilon > 0$ має місце така нерівність:

$$P(|X - M(X)| \leq \varepsilon) \geq 1 - \frac{D(X)}{\varepsilon^2}.$$

Нерівність Чебишова дає можливість оцінити помилку, яку допускають, якщо математичне сподівання замінюють середнім значенням обмеженої вибірки.

Оцінювання відхилення відносної частоти появи випадкової величини від імовірності її появи в окремому випробуванні виконують за такою формулою:

$$P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) \geq 1 - \frac{pq}{n\varepsilon^2}.$$

Теорема Чебишова. Нехай $X_1, X_2, \dots, X_n$ – попарно незалежні випадкові величини з обмеженими дисперсіями, тобто існує таке D , що $D(X_1) \leq D, D(X_2) \leq D, \dots, D(X_n) \leq D$. Тоді для будь-якого $\varepsilon > 0$ ймовірність відхилення середнього арифметичного значень цих величин від середнього арифметичного їхніх математичних сподівань на величину, не більшу ніж ε , є скільки завгодно близькою до одиниці, якщо кількість випадкових величин є достатньо великою:

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - \frac{M(X_1) + M(X_2) + \dots + M(X_n)}{n}\right| \leq \varepsilon\right) = 1.$$

Теорему Чебишова можна розглядати як *необхідну і достатню умову* граничної збіжності середніх значень незалежних випадкових величин до середніх значень їхніх математичних сподівань. Сенс теореми полягає в тому, що середнє арифметичне достатньо великої кількості незалежних випадкових величин утрачає характер випадкової величини.

Теорема Бернуллі. Якщо ймовірність події в n випробуваннях є сталою і дорівнює p , то ймовірність відхилення відносної частоти від цієї ймовірності на величину, що не перевершує $\varepsilon > 0$, прагне до одиниці в разі достатньо великої кількості випробувань n :

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{m}{n} - p\right| \leq \varepsilon\right) = 1.$$

Теорема Бернуллі стверджує стійкість відносних частот за умови, що випробування є однорідними.

Якщо умови випробувань змінюють, тобто змінюють ймовірність появи події в одиничному випробуванні, то теж має місце аналогічна стійкість результатів.

Теорема Хінчина. Якщо $X_1, X_2, \dots, X_n$ – незалежні однаково розподілені випадкові величини зі скінченним математичним сподіванням a , то

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - a\right| < \varepsilon\right) = 1.$$

Установлені раніше зв'язки між послідовностями випадкових величин і їхніми математичними сподіваннями насамперед належать до *дискретних законів розподілу*.

Якщо розглянути випадкові величини з *неперервними законами розподілу*, то послідовності незалежних випадкових величин за певних умов також виявляють тенденцію до утворення граничних залежностей.

Граничні теореми, що встановлюють граничні закони розподілу випадкових величин, об'єднують загальною назвою **центральна гранична теорема**. Різні форми центральної граничної теореми відрізняються умовами, за яких виникає нормальний закон розподілу. Серед цих теорем важливе місце посідає теорема Ляпунова.

Теорема. Якщо випадкові величини $X_1, X_2, \dots, X_n$ у послідовності є незалежними, однаково розподіленими та для них існують моменти 2-го порядку і

$$S_n = \sum_{i=1}^n X_i; \quad D(X_i) = b_i^2; \quad B_n^2 = \sum_{i=1}^n b_i^2; \quad S_n^* = \frac{S_n - M(S_n)}{B_n},$$

то

$$\lim_{n \rightarrow \infty} P(S_n^* < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt,$$

тобто граничним розподілом для S_n^* є нормальний закон розподілу з нульовим математичним сподіванням та одиничною дисперсією.

Теорема Ляпунова. Якщо $X_1, X_2, \dots, X_n$ – незалежні випадкові величини, у кожній з яких математичне сподівання $M(X_i) = a_i$, дисперсія $D(X_i) = \sigma_i^2$, абсолютний центральний момент 3-го порядку $M(|X_i - a_i|^3) = m_i$

та $\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n m_i}{\left(\sum_{i=1}^n \sigma_i^2\right)^{3/2}} = 0$, то закон розподілу суми $Y = X_1 + X_2 + \dots + X_n$

є необмежено наближенням до нормального з математичним сподіванням $\sum_{i=1}^n a_i$ та дисперсією $\sum_{i=1}^n \sigma_i^2$.

Сенс теореми Ляпунова для її використання на практиці полягає в такому: якщо випадкова величина є сумою дуже великої кількості взаємно незалежних випадкових величин, вплив кожної з яких на суму є малим, то вона має розподіл, близький до нормального.

8.3. Приклади розв'язування задач

Приклад 8.1. Кількість палива, що витрачають на фірмі протягом доби, є випадковою величиною, математичне сподівання якої дорівнює 6 т.

Оцінімо ймовірність того, що протягом найближчої доби витрати палива перевищать 10 т.

Розв'язання. За умовою задачі, маємо: $M(X) = 6$, $\varepsilon = 10$.

За нерівністю Маркова, буде: $P(X > 10) \leq \frac{6}{10}$.

Отже, з імовірністю, яка не перевищує 0,6, можна стверджувати, що витрати палива будуть більшими за 10 т.

Приклад 8.2. Електрична мережа має 18 000 ламп, імовірність увімкнення кожної з яких – 0,9.

Оцінімо ймовірність того, що кількість увімкнених ламп відхиляється від свого математичного сподівання на величину, не меншу ніж 200.

Розв'язання. За умовою задачі, маємо: $n = 18\,000$, $p = 0,9$.

Тоді

$$M(X) = np = 18\,000 \cdot 0,9 = 16\,200;$$

$$D(X) = npq = 18\,000 \cdot 0,9 \cdot 0,1 = 1\,620.$$

Згідно з нерівністю Чебишова, буде:

$$P(|X - 16\,200| \leq 200) \geq 1 - \frac{1\,620}{200^2} = 1 - 0,045 = 0,955;$$

$$P(|X - 16\,200| \geq 200) \geq 1 - 0,955 = 0,045.$$

Отже, з ймовірністю понад 0,045 можна стверджувати, що кількість увімкнених ламп відрізняється від математичного сподівання величиною, не меншою ніж 200.

Приклад 8.3. Оцінімо ймовірність того, що з 800 малих підприємств збанкрутує не менше ніж 50, якщо ймовірність банкрутства будь-якого навання взятого малого підприємства є сталою і дорівнює 0,05.

Розв'язання. Нехай випадкова величина X – це кількість збанкрутілих підприємств. Цю випадкову величину підпорядковано біноміальному розподілу з параметрами $n = 800$, $p = 0,05$. Тоді, згідно з формулою обчислення математичного сподівання випадкової величини, що має біноміальний розподіл, буде: $M(X) = np = 800 \cdot 0,05 = 40$.

Використаймо нерівність Маркова для $\varepsilon = 50$:

$$P(X > \varepsilon) \leq \frac{40}{50} = \frac{4}{5}.$$

Приклад 8.4. Середня температура в приміщенні в період опалювального сезону дорівнює 20°C , а середнє квадратичне відхилення дорівнює 2°C .

За допомогою нерівності Чебишова оцінімо ймовірність того, що температура у квартирі відхилиться від середньої менше ніж на 4 °С.

Розв'язання. Нехай випадкова величина X – це температура в приміщенні.

За умови задачі, маємо: $M(X) = 20$, $\sigma(X) = 2$, $\varepsilon = 4$.

Із нерівності Чебишова, буде:

$$P(|X - 20| < 4) \geq 1 - \frac{4}{16} = \frac{3}{4}.$$

Приклад 8.5. Імовірність того, що телевізор витримає гарантійний строк роботи дорівнює 0,95 для всіх 200 телевізорів, які обслуговує гарантійна майстерня.

Оцінімо ймовірність того, що кількість телевізорів, які витримають гарантійний строк роботи, буде в межах [185; 195].

Розв'язання. Дискретна випадкова величина X – це кількість телевізорів, які витримають гарантійний строк роботи, має біноміальний розподіл імовірності, тому

$$M(X) = np = 200 \cdot 0,95 = 190; \quad D(X) = npq = 200 \cdot 0,05 \cdot 0,95 = 9,5.$$

Обчислімо $\varepsilon = 195 - 190 = |185 - 190| = 5$ і використаємо нерівність Чебишова:

$$P(|X - 190| < 5) \geq 1 - \frac{9,5}{25} = 0,62.$$

Приклад 8.6. Середнє споживання електроенергії протягом травня в місті дорівнює 360 000 кВт год.

Оцінімо ймовірність того, що споживання електроенергії в травні поточного року перевищить 1 000 000 кВт год.

Оцінімо ту саму ймовірність за умови, що середнє квадратичне відхилення споживання електроенергії за травень дорівнює 40 000 кВт · год.

Розв'язання. Випадкова величина X – це споживання електроенергії – набуває невід'ємних значень. Математичне сподівання її дорівнює 360 000.

Оцінімо ймовірність за допомогою нерівності Маркова:

$$P(X \geq 1\,000\,000) \leq \frac{360\,000}{1\,000\,000} = 0,36.$$

Оцінімо цю саму нерівність, якщо відоме середнє квадратичне відхилення X .

Скористаймося нерівністю Чебишова:

$$\begin{aligned} P(X \geq 1\,000\,000) &= 1 - P(X < 1\,000\,000) = 1 - P(0 < X < 1\,000\,000) = \\ &= 1 - P(-360\,000 < X - MX < 640\,000) = 1 - P(|X - MX| < 640\,000) \leq \\ &\leq 1 - 1 + \frac{(40\,000)^2}{(640\,000)^2} = \frac{1}{256}. \end{aligned}$$

Приклад 8.7. Оцінімо кількість деталей, потрібну для того щоб з імовірністю, не нижчою ніж 0,98, можна було очікувати, що абсолютна величина відхилення відносної частоти стандартних деталей від імовірності, яка дорівнює $p = 0,95$, була нижчою за 0,05.

Розв'язання. Застосуємо нерівність Чебишова.

У цьому прикладі маємо: $p = 0,95$, $P \geq 0,98$, $\varepsilon = 0,05$.

Визначмо n за такою нерівністю:

$$0,98 \leq 1 - \frac{0,95 \cdot 0,05}{n \cdot (0,05)^2} \Rightarrow n \geq \frac{0,95 \cdot 0,05}{0,02(0,05)^2}.$$

Отже, $n \geq 950$, тобто потрібно не менше ніж 950 деталей щоб з імовірністю, не нижчою ніж 0,98, можна було очікувати, що абсолютна величина відхилення відносної частоти стандартних деталей від $p = 0,95$ є нижчою за 0,05.

Приклад 8.8. Контролер перевіряє деякі вироби. На першому етапі перевірки, який триває 10 с, він або відразу оцінює виріб, або вирішує, що перевірку треба повторити. Повторна перевірка триває 10 с, у результаті чого обов'язково ухвалюють рішення про якість продукції.

Визначмо ймовірність того, що за семигодинний робочий день контролер перевірить: понад 1 800 виробів; понад 1 640 виробів; не менш як 1 500 виробів.

Передбачаймо, що кожний виріб, незалежно від інших, з ймовірністю 0,5 проходить повторну перевірку.

Розв'язання. Нехай X_i – час (хв), потрібний для перевірки i -го виробу. Визначмо послідовність незалежних однаково розподілених випадкових величин $X_1, X_2, \dots, X_n, \dots$.

Кожна з них має такий закон розподілу:

$$x_1 = \frac{1}{6}; p(x_1) = \frac{1}{2};$$

$$x_2 = \frac{1}{3}; p(x_2) = \frac{1}{2}.$$

Числові характеристики такі:

$$M(X_i) = \frac{1}{12} + \frac{1}{6} = \frac{1}{4}; D(X_i) = \frac{1}{72} + \frac{1}{18} - \frac{1}{16} = \frac{1}{144}.$$

Загальний час S_n , що його витрачає робітник на перевірку n виробів, є сумою n -х незалежних однаково розподілених величин:

$$S_n = \sum_{i=1}^n X_i.$$

Згідно з теоремою, величина S_n має закон розподілу, близький до нормального.

Якщо $n=1800$, то $M(S_n)=450$, $D(S_n)=12,5$. Тривалість зміни становить 420 хв.

Тоді, за інтегральною теоремою Лапласа, буде:

$$P(S_n < 420) = \Phi\left(\frac{420 - 450}{\sqrt{12,5}}\right) - \Phi\left(\frac{0 - 420}{\sqrt{12,5}}\right) = -\Phi(6\sqrt{2}) + \Phi(90\sqrt{2}) = 0.$$

Якщо $n = 1\,640$, то $M(S_n) = 410$, $D(S_n) = \frac{205}{18} \approx 11,39$.

У такому разі:

$$P(S_n < 420) = \Phi\left(\frac{420 - 410}{\sqrt{11,39}}\right) - \Phi\left(\frac{0 - 410}{\sqrt{11,39}}\right) = \Phi(2,96) + \Phi(121,5) \approx 0,9985.$$

Без обчислень можна стверджувати, що за $n = 1\,500$

$$P(S_n < 420) = 1.$$

8.4. Вправи для самостійної роботи

8.1. Середнє значення ваги виробу – 50 г. За допомогою нерівності Маркова оцініть імовірність того, що навмання взятий виріб має вагу, меншу за 90 г.

8.2. Використовуючи нерівність Чебишова, оцініть імовірність того, що $|X - M(X)| < 0,3$, якщо $D(X) = 0,002\,5$.

8.3. Відомо, що $P\{|X - M(X)| < \varepsilon\} \geq 0,9$ і $D(X) = 0,001$.

Скориставшись нерівністю Чебишова, обчисліть ε .

8.4. Випадкову величину X задано законом розподілу (табл. 8.1). Використовуючи нерівність Чебишова, оцініть імовірність того, що: абсолютна різниця між значенням випадкової величини X і її математичним сподіванням є меншою за 5;

абсолютна різниця між значенням випадкової величини X і її математичним сподіванням є не меншою ніж 2.

Таблиця 8.1

Закон розподілу випадкової величини X

X	1	3	5	6
$P(X = x_i)$	0,1	0,3	0,4	0,2

8.5. Випадкову величину X задано законом розподілу (табл. 8.2).

Використовуючи нерівність Чебишова, оцініть імовірність того, що абсолютна різниця між значенням випадкової величини X і її математичним сподіванням є меншою за 2.

Таблиця 8.2

Закон розподілу випадкової величини X

X	3	5
$P(X = x_i)$	0,4	0,6

8.6. Із 5 000 проведених випробувань у 2 000 імовірність появи події A дорівнює 0,2; у 1 400 – 0,5 і в 1 600 – 0,6.

Визначте межі, у яких має міститися частота появи події A , якщо є потрібно гарантувати з імовірністю 0,95.

8.7. Імовірність того, що під час одиничного звернення банкомат спрацює правильно, дорівнює 0,95.

Оцініть імовірність того, що:

у разі 2 500 звернень відносна частота випадків правильної роботи банкомата відхилиться (за абсолютною величиною) від її ймовірності менше ніж на 0,02;

у разі 2 000 звернень кількість випадків правильної роботи банкомата є в межах від 1 860 до 1 940.

8.8. Довжина виробів, що виробляють у цеху, – це випадкова величина, середнє значення якої дорівнює 90 см. Дисперсія цієї величини – 0,02.

Оцініть імовірність того, що:

а) відхилення довжини виготовленого виробу від його середнього значення за абсолютною величиною буде не більшим ніж 0,4;

б) довжину виробу виражено числом, що міститься в межах $[89,4; 90,6]$.

8.9. Імовірність народження дівчинки приблизно дорівнює 0,485.

Оцініть імовірність того, що кількість дівчат серед 3 000 новонароджених буде відрізнятися від математичного сподівання цієї кількості за абсолютною величиною менше ніж на 55 дівчат.

8.10. Дисперсія кожної із 2 500 незалежних випадкових величин не перевищує 5.

Оцініть імовірність того, що відхилення середнього арифметичного цих випадкових величин від середнього арифметичного їхніх математичних сподівань не перевищить 0,4.

8.11. Відомо, що цех виготовляє 80 % продукції вищого ґатунку.

Оцініть імовірність того, що відносна частота виробів вищого ґатунку серед 20 000 виготовлених буде відрізнятися від імовірності виготовлення виробу вищого ґатунку не більше ніж на 0,02.

8.12. Імовірність настання події A в кожному випробуванні $p = 1/3$.

Яку найменшу кількість випробувань потрібно виконати, щоб з імовірністю, не меншою за 0,99 можна було стверджувати, що частість настання події A відхилялася за абсолютною величиною від її ймовірності не більш ніж на 0,01? Для розв'язування скористайтесь: а) нерівністю Чебишова; б) інтегральною теоремою Лапласа.

8.13. За результатами дослідження в середньому 80 % клієнтів певної страхової компанії оминають страховий випадок і не потребують виплат страхової суми.

Оцініть імовірність того, що із 600 клієнтів цієї страхової компанії частка тих, хто не буде потребувати виплат, буде відрізнятися від імовірності не більше, ніж на 0,05 (за абсолютною величиною).

8.14. Середньодобове споживання електроенергії в певному населеному пункті дорівнює 20 000 кВт/год, а середнє квадратичне відхилення – 200 кВт/год.

Який обсяг споживання електроенергії в цьому населеному пункті можна очікувати в найближчу добу з імовірністю, не нижчою за 0,96?

8.15. Визначте ймовірність того, що грошей у банкоматі вистачить для обслуговування 250 клієнтів, якщо в банкомат завантажено 80 000 грн, суми, узяті з банкомату клієнтами є випадковими величинами з математичним сподіванням 400 грн та середнім квадратичним відхиленням 150 грн.

8.16. Імовірність того, що платіжний термінал спрацьовує під час подавання купюри з імовірністю 0,99.

За допомогою теореми Бернуллі визначте відхилення відносної частоти кількості випадків, якщо термінал спрацьовує, від їхньої ймовірності під час подавання 2 000 купюр, якщо результат потрібно гарантувати з імовірністю, не нижчою ніж 0,95.

Визначте межі, у яких перебуває кількість випадків правильної роботи терміналу.

8.17. На пакеті гречаної крупи надруковано, що її маса $1\,000\text{ г} \pm 1,5\%$. Для перевірки ваги було взято 30 пакетів крупи та виконано контрольне зважування. У результаті було виявлено, що дві упаковки мають вагу по 985 г, три упаковки – по 990 г, вісім упаковок – по 999 г, дванадцять упаковок – по 1 000 г, дві упаковки – по 1 005 г та три упаковки – по 1 010 г.

За допомогою нерівності Чебишова оцініть, чи варто довіряти зазначеному на пакеті напису?

8.18. Було здійснено фінансовий аналіз діяльності певної компанії. У результаті чого встановлено, що середньомісячні витрати на рекламу становлять 25 000 грн.

Потрібно оцінити ймовірність того, що наступного місяця витрати на рекламу не вийдуть за межі 23 000 – 27 000 грн, якщо середнє квадратичне відхилення дорівнює 4 500 грн.

8.19. На підприємстві, що виготовляє батони, відділ контролю за якістю продукції перевірів партію випечених батонів. У результаті перевірки було виявлено, що з імовірністю 0,02 батон має брак, пов'язаний із неякісними інгредієнтами, які входять до складу тіста, та з імовірністю 0,03 батон має брак, пов'язаний із процесом випікання.

Визначте межі, у яких майже напевно (з імовірністю 0,99) буде кількість бракованих батонів, якщо перевіряють партію з 800 батонів?

8.20. Банк провів опитування своїх клієнтів щодо задоволення від їхньої роботи з банком. У результаті опитування встановлено, що клієнти не є задоволеними компетентністю співробітників банку з імовірністю 0,03 та з імовірністю 0,01 клієнти не є задоволеними відсотком, який пропонує банк за депозитним рахунком. Випадковим способом було відібрано 1 000 клієнтів.

Визначте межі, у яких майже напевно (з імовірністю 0,99) буде кількість невдоволених клієнтів банку.

8.5. Тестові завдання

8.1. Сенс закону великих чисел полягає у

А визначенні числових характеристик випадкових величин у разі великої кількості спостережуваних даних

Б поведінці числових характеристик і законів розподілу спостережуваних значень випадкових величин

В визначенні області застосовності нормального закону розподілу випадкових величин під час складання великої кількості випадкових величин

Г поведінці числових характеристик і законів розподілу випадкових величин у разі збільшення кількості спостережень і дослідів.

8.2. У запису теореми Чебишова вставити пропущений символ:

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - \frac{M(X_1) + M(X_2) + \dots + M(X_n)}{n}\right| \leq \varepsilon\right) = \dots$$

А	Б	В	Г
0	1	$\frac{1}{n}$	інше

8.3. Теорему Чебишова можна розглядати як

А достатню умову граничної збіжності середніх значень незалежних випадкових величин до середніх значень їхніх математичних сподівань

Б необхідну умову граничної збіжності середніх значень незалежних випадкових величин до середніх значень їхніх математичних сподівань

В необхідну і достатню умову граничної збіжності середніх значень незалежних випадкових величин до середніх значень їхніх математичних сподівань

8.4. У запису теореми Хінчина вставити пропущений символ:

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - a\right| < \varepsilon\right) = \dots$$

А	Б	В	Г
0	1	$\frac{1}{n}$	інше

8.5. Уставити пропущений символ у нерівності Маркова: $P(x > \varepsilon) \leq \frac{\dots}{\varepsilon}$

А	Б	В	Г
$D(X)$	1	$M(X)$	інше

8.6. Уставити пропущений символ у нерівності Чебишова:

$$P(|X - M(X)| \leq \varepsilon) \geq 1 - \frac{\dots}{\varepsilon^2}$$

А	Б	В	Г
$D(X)$	1	$M(X)$	інше

8.7. Для того щоб оцінити ймовірність відхилення випадкової величини від її математичного сподівання, потрібно використати

А	Б	В	Г
теорему Бернуллі	теорему Чебишова	центральну граничну теорему	нерівність Чебишова

8.8. Для того щоб оцінити за нерівністю Чебишова відхилення на ε випадкової величини від математичного сподівання, потрібно знати значення

А	Б	В	Г
випадкової величини	математичного сподівання	дисперсії	інше

8.6. Запитання для самоперевірки

8.1. Які теореми теорії ймовірностей називають граничними?

8.2. За яким принципом граничні теореми розподіляють на два класи?

8.3. Сформулюйте загальний сенс закону великих чисел.

8.4. Сформулюйте центральну граничну теорему Ляпунова.

Які її основні умови?

8.5. Наведіть приклад застосування центральної граничної теореми в реальному дослідженні (наприклад, у соціології або економіці).

8.6. Дайте означення закону великих чисел у формі теореми Чебишова.

8.7. Наведіть теорему Бернуллі та розкрийте її зв'язок із статистичним означенням імовірності.

8.8. Яка різниця між теоремою Бернуллі та теоремою Хінчина щодо закону великих чисел?

8.9. Поясніть сенс центральної граничної теореми та розкрийте її значення для математичної статистики.

8.7. Висновки за підрозділом 8

Граничні теореми не лише підтверджують стійкість середніх результатів великої кількості випробувань, а й пояснюють механізми такої стійкості, що надає їм особливого значення для математичної статистики та прикладних досліджень. Їх умовно розподіляють на два класи: перший описує збіжність характеристик випадкових величин (зокрема, середнього значення) до сталих детермінованих величин – це так звані закони великих чисел; другий клас – це граничні теореми у вузькому розумінні, які встановлюють граничні закони розподілу випадкових величин за певних умов, часто вказуючи на збіжність до нормального розподілу.

Закон великих чисел у різних формулюваннях (Бернуллі, Чебишова, Хінчина) доводить, що зі зростанням кількості спостережень середнє арифметичне випадкових величин наближається до їхнього математичного сподівання.

Центральна гранична теорема в різних її варіаціях (зокрема, теорема Ляпунова), стверджує, що, незалежно від закону розподілу окремих доданків, за певних умов розподіл їхньої суми або середнього наближається до нормального.

Література: [1 – 3; 5; 8; 10; 11; 13; 15; 16].

9. Первинне опрацювання статистичних даних

9.1. Мета та компетентності

Метою вивчення підрозділу є ознайомлення із завданнями математичної статистики, особливостями вибіркового методу та формування компетентностей щодо можливостей його застосування до дослідження характеристик об'єктів у генеральній сукупності.

Компетентності, що формують під час вивчення підрозділу:

знання основних методів відбору об'єктів генеральної сукупності у вибіркoву сукупність;

розуміння впливу методів формування вибіркової сукупності на результати статистичних досліджень;

уміння робити вибір методу формування вибіркової сукупності, відповідно до завдань дослідження та гіпотетичних властивостей досліджуваної випадкової величини.

9.2. Основні теоретичні відомості

Математична статистика розглядає такі групи завдань:

розроблення й обґрунтування способів відбору об'єктів для формування вибіркової сукупності та визначення обсягу вибіркової сукупності;

статистичне оцінювання числових характеристик випадкових величин та параметрів їхнього розподілу в генеральній сукупності за результатами дослідження вибірки;

перевірку статистичних гіпотез щодо значущості числових характеристик випадкової величини довільної вимірності, а також параметрів її розподілу та закону розподілу за результатами дослідження вибірки.

Первинне опрацювання статистичних даних виконують за таким планом: 1) формулюють завдання дослідження і визначають обсяг, місце та час потрібної вибірки; 2) збирають потрібні дані та наочно подають (аналітично, таблично, графічно); 3) виконують опрацювання зібраних статистичних даних та формулюють висновок.

Словом "сукупність" у статистиці називають множину об'єктів, із яких беруть вибірку. Усю сукупність, що вивчають, називають **генеральною сукупністю**.

Генеральну сукупність можна вивчати за допомогою суцільного вивчення всіх об'єктів або спостереження за частиною об'єктів. Частину об'єктів, яку дістають із генеральної сукупності, називають **вибіркою** або **вибірковою сукупністю**. Повну кількість об'єктів генеральної сукупності чи вибіркової сукупності називають їхнім **обсягом**. Обсяг генеральної сукупності позначають N , а обсяг вибіркової сукупності – n . Результати досліджень будь-якої ознаки генеральної сукупності будуть більш достовірними, якщо вибірку створити випадково.

Вибірку називають **випадковою**, якщо з генеральної сукупності елемент брати навмання і кожен із них може потрапити до неї з однаковою ймовірністю. На практиці випадкову вибірку можна скласти так: усі елементи нумерують від 1 до N , після чого вибирають послідовність n випадкових чисел ($1 \leq n \leq N$). Вибірку також можна вибрати і використавши датчик випадкових чисел.

Відбір об'єктів може бути повторним або безповторним.

Повторним називають відбір, якщо відібраний об'єкт повертають у генеральну сукупність (до відбору наступного об'єкта).

Безповторним називають відбір, якщо відібраний об'єкт не повертають до генеральної сукупності. Відбір об'єктів відбувається випадковим чином.

Якщо випадкова вибірка достатньо повно характеризує генеральну сукупність, то її називають **репрезентативною**.

Для складання репрезентативної вибірки слід: забезпечити кожному елементу генеральної сукупності однакову ймовірність потрапити у вибірку; відібрати таку кількість елементів генеральної сукупності, за якої можна забезпечити потрібну точність характеристик генеральної сукупності.

Нехай є вибірка з генеральної сукупності $x_1, x_2, \dots, x_n$. Якщо записати цю вибірку у вигляді зростаючої або спадної послідовностей, то буде дискретний *варіаційний ряд* або *емпіричний розподіл*.

Деякі значення дискретної випадкової величини можуть повторюватися, тоді варіаційний ряд подають у вигляді таблиці (табл. 9.1), де визначають можливі значення (x_i) випадкової величини X та їхні частоти (m_i), $\sum_{i=1}^k m_i = n$ – обсяг вибірки.

Дискретний варіаційний ряд

x_i	x_1	x_2	...	x_k
m_i	m_1	m_2	...	m_k

Числові значення $x_1, x_2, \dots, x_n$, яких набуває досліджувана ознака, називають **варіантами**.

Частотою (m_i) називають кількість появи окремих значень (x_i) досліджуваної ознаки (варіант).

У дискретному варіаційному ряді, замість частот, можна вказувати відносні частоти.

Відносною частотою w_i називають відношення частоти появи ознаки до загального обсягу вибірки:

$$w_i = \frac{m_i}{n}, \quad \sum_{i=1}^k w_i = 1.$$

Відносна частота характеризує частку сукупності членів з однаковими значеннями ознаки.

Якщо обсяг вибірки є великим ($n \geq 30$), то користуватися дискретним варіаційним рядом незручно. У цьому разі, а також у разі, якщо дані визначено в результаті вимірювання неперервної випадкової величини, за вибіркою складають *інтервальний варіаційний ряд*. Для його побудови весь інтервал розподілу ознаки розподіляють на k однакових частин довжиною h . Кількість k таких інтервалів, на які розподіляють вибіркочну сукупність обсягом n за первинного групування, оцінюють за формулою Стерджеса:

$$k = 1 + 3,322 \cdot \lg n,$$

де кількість інтервалів округлюють до найближчого цілого числа. Зазвичай це число є в межах від 8 до 12.

Для того щоб розподілити за інтервалами результати вимірювань неперервної випадкової величини, що становлять вибіркочну сукупність

насамперед потрібно визначити найменше x_{min} та найбільше x_{max} значення варіант. Різницю між ними $R = x_{max} - x_{min}$ називають **розмахом** (що можна розглядати як додаткову числову характеристику емпіричного розподілу випадкової величини). Розмах відображає тільки граничні відхилення й не відображає відхилень усіх варіант у ряду, тому він є головним показником у тих випадках, якщо кожному варіанту спостерігають один раз. В інших випадках розмах варіації застосовують для того, щоб сформулювати загальне уявлення про варіацію ознаки у всій сукупності.

Поділивши розмах на кількість інтервалів, визначають довжину інтервалу, тобто **крок** $h = \frac{R}{k}$. Крок визначає відстань між верхньою і нижньою межами кожного інтервалу:

$$h = x_{i+1} - x_i.$$

Щоб оптимізувати обчислення, значення кроку доцільно округлити таким способом, щоб межі інтервалів і їхні середини були зручними для обчислення. Оскільки це призводить до збільшення розмаху, зміна значень x_1 та x_{k+1} не має перевищувати $0,5 \cdot h$.

Отже, межі граничних інтервалів слід зсунути: $x_1 < x_{min}$ та $x_{k+1} > x_{max}$. Для кожного інтервалу частоту m_i визначають як кількість значень випадкової величини, що потрапили до цього інтервалу. Кожний з інтервалів є напіввідкритим, тобто відкритим із боку верхньої межі.

Поставивши кожному з інтервалів у відповідність частоту m_i потраплянь у заданий інтервал або відносну частоту $w_i = \frac{m_i}{n}$, визначають статистичний розподіл неперервної випадкової величини, який задано у формі таблиці (табл. 9.2).

Таблиця 9.2

Інтервальний варіаційний ряд

$[x_i; x_{i+1})$	$[x_1; x_2)$	$[x_2; x_3)$	...	$[x_k; x_{k+1})$
$m_i (w_i)$	$m_1 (w_1)$	$m_2 (w_2)$	...	$m_k (w_k)$

Якщо варіаційний ряд задано у вигляді табл. 9.1 і 9.2, то його називають **статистичним рядом розподілу**.

Статистичний ряд розподілу випадкової величини X , визначений за емпіричними даними, називають також **емпіричним законом розподілу**. Такий ряд можна зобразити графічно. Для цього на осі абсцис наносять варіанти x_i , а на осі ординат – частоти m_i (або відносні частоти w_i), тобто є певні точки $M_i(x_i, m_i)$. З'єднавши точки M_i відповідних варіант і частот, визначають ламану лінію, яку називають **полігоном розподілу**.

Емпіричною функцією розподілу вибірки називають функцію $F^*(x)$, яка для будь-якого значення x визначає відносну частоту події, що задовольняє умову $X < x$, тобто випадкова величина набуде значення, меншого за x :

$$F^*(x) = \frac{m_x}{n}, \quad (9.1)$$

де m_x – сума частот варіант для значень аргументу, менших за x ;

n – обсяг вибірки.

Емпіричну функцію розподілу визначають за допомогою послідовного додавання відносних частот варіант, менших за x . Графіком емпіричної функції розподілу є **кумулята**, або графік накопичених відносних частот. Цей графік – східчаста фігура, яка має точки розриву, якщо значення абсцис дорівнюють числовим значенням, яких набуває випадкова величина.

Для генеральної сукупності функцію розподілу позначають $F(x)$. Різниця між емпіричною та теоретичною функціями полягає в тому, що *теоретична функція* $F(x)$ визначає ймовірність події $X < x$, а *емпірична функція* $F^*(x)$ визначає відносну частоту цієї події.

Згідно з теоремою Бернуллі, можна сказати, що в разі великої кількості n значення $F^*(x)$ і $F(x)$ майже не відрізняються одне від одного в тому сенсі, що $\lim_{n \rightarrow \infty} P(|F(x) - F^*(x)| < \varepsilon) = 1 \quad (\varepsilon > 0)$.

Отже, емпіричну функцію розподілу вибірки може бути використано для оцінювання теоретичної функції розподілу генеральної сукупності.

Функція $F^*(x)$ має такі самі властивості, що й функція $F(x)$, тобто: значення емпіричної функції належить відрізку $[0; 1]$;

$F^*(x)$ – неспадна функція: $F^*(x_2) > F^*(x_1)$, якщо $x_2 > x_1$;

якщо x_1 – найменша варіанта, то $F^*(x) = 0$ за $x \leq x_1$;

якщо x_k – найбільша варіанта, то $F^*(x) = 1$ за $x > x_k$.

У разі інтервальних варіаційних рядів, визначених для неперервної випадкової величини, графічним зображенням є *гістограма*.

Розрізняють гістограму частот і гістограму відносних частот.

Гістограмою частот називають східчасту фігуру, яку становлять прямокутники, площа кожного з яких дорівнює частоті потрапляння випадкової величини у відповідний інтервал. Основами прямокутників є інтервали, довжиною h , а висоти дорівнюють щільності частоти $\frac{m_i}{h}$.

Для побудови гістограми на осі абсцис відкладають часткові інтервали, а над ними проводять відрізки, які є паралельними осі абсцис на відстані $\frac{m_i}{h}$. Площа i -го окремого прямокутника дорівнює $\frac{m_i}{h} \cdot h = m_i$ – сумі частот варіант i -го інтервалу. Отже, площа гістограми частот дорівнює сумі всіх частот, тобто обсягу вибірки.

Гістограмою відносних частот називають східчасту фігуру, яку становлять прямокутники, основою яких є інтервали довжиною h , а висоти дорівнюють відношенню $\frac{w_i}{h}$. Для побудови гістограм відносних частот на осі абсцис відкладають часткові інтервали, а над ними проводять відрізки, паралельні осі абсцис на відстані $\frac{w_i}{h}$ від неї. Площа гістограми відносних частот дорівнює сумі всіх відносних частот, тобто одиниці.

9.3. Приклади розв'язування задач

Приклад 9.1. Протягом місяця магазин із продажу канцелярських товарів вів облік щоденної кількості покупців, які використовують платіжні картки. Результати спостережень наведено в табл. 9.3.

Таблиця 9.3

Кількість користувачів платіжними картками

12	15	10	8	12	14	11	12	13	12
11	14	14	9	7	15	16	10	11	13
8	11	15	17	10	13	9	12	12	9

За вибірковою сукупністю побудуємо варіаційний ряд, а також полігон частот.

Розв'язання. Оскільки серед результатів 30 вимірювань є тільки 11 різних варіант, розташуймо їх у порядку зростання та підрахуймо, скільки разів кожен з них спостерігали у вибірковій сукупності.

Результати надаймо у вигляді таблиці (табл. 9.4).

Таблиця 9.4

Дискретний ряд

$X = x_i$	7	8	9	10	11	12	13	14	15	16	17
m_i	1	2	3	3	4	6	3	3	3	1	1

Перевірмо, що $\sum_{i=1}^{11} m_i = 30$, та побудуємо полігон частот (рис. 9.1).

Частота

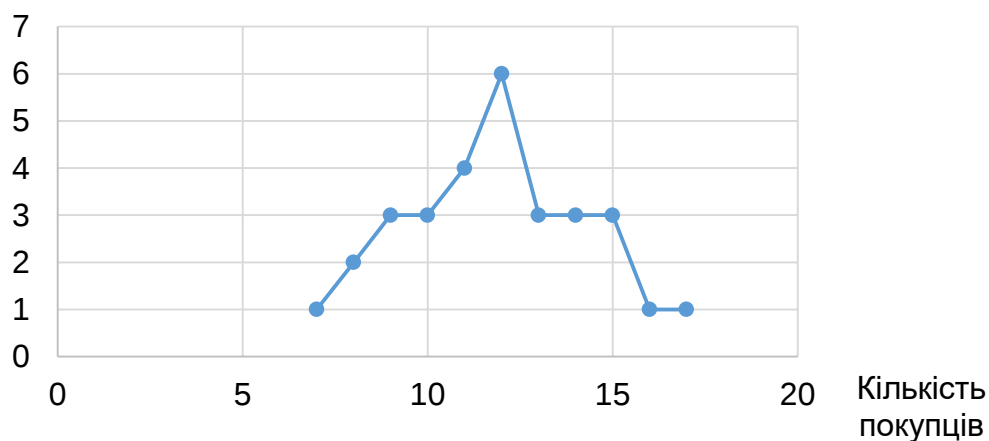


Рис. 9.1. Полігон частот за даними прикл. 9.1

Приклад 9.2. Подано розподіл 70 філій ІТ-компанії за кількістю працівників на них (табл. 9.5).

Таблиця 9.5

Ряд розподілу філій ІТ-компанії

x_i	15	19	21	25	35	37
m_i	2	8	15	22	13	10

Побудуємо полігон розподілу частот.

Розв'язання. Ознака X – кількість працівників в окремій філії компанії, ця ознака є дискретною.

Перевірмо, що $\sum_{i=1}^6 m_i = 70$, та побудуємо полігон частот (рис. 9.2).

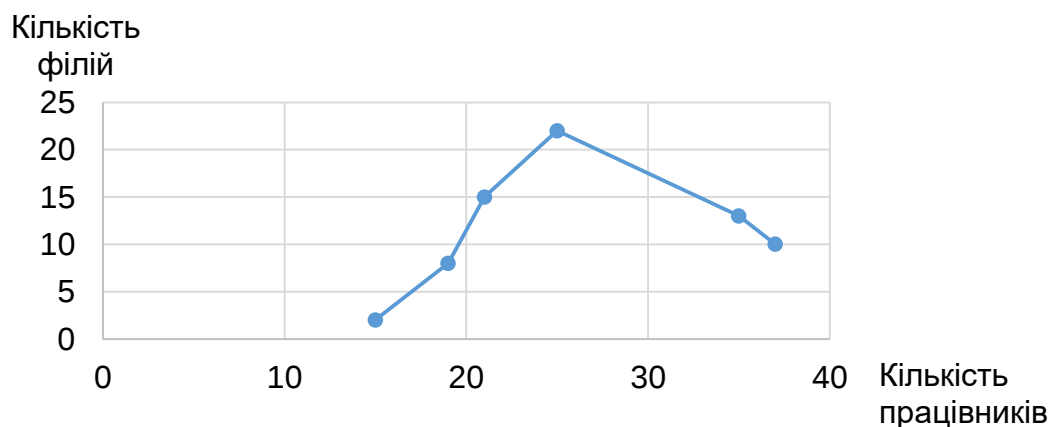


Рис. 9.2. Полігон частот за даними прикл. 9.2

Полігон показує, що розподіл є несиметричним, зміщеним у бік філій із більшою кількістю працівників.

Приклад 9.3. Задано вибірку рядом розподілу (табл. 9.6).

Таблиця 9.6

Ряд розподілу

x_i	1	6	11	16
m_i	20	10	40	30

Визначмо та побудуємо емпіричну функцію розподілу $F^*(x)$.

Розв'язання. Визначмо обсяг вибірки:

$$n = \sum_{i=1}^4 m_i = 20 + 10 + 40 + 30 = 100.$$

Найменше значення величини X : $x_1 = 1$, тому

$$F^*(1) = 0, \text{ якщо } X \leq 1.$$

Значення $X < 6$ спостерігали 20 разів, отже,

$$F^*(x) = \frac{20}{100} = 0,2, \text{ якщо } 1 < x \leq 6.$$

Значення $X < 11$, тобто 1 і 6, спостерігали $20 + 10 = 30$ разів, отже,

$$F^*(x) = \frac{30}{100} = 0,3, \text{ якщо } 6 < x \leq 11.$$

Значення $X < 16$, тобто 1, 6 і 11, спостерігали $20 + 10 + 40 = 70$ разів, отже,

$$F^*(x) = \frac{70}{100} = 0,7, \text{ якщо } 11 < x \leq 16.$$

Найбільшим значенням величини X є $x_4 = 16$, тому $F^*(x) = 1$, якщо $X > 16$.

Отже, можна записати емпіричну функцію розподілу вибірки в такому вигляді:

$$F^*(x) = \begin{cases} 0, & x \leq 1; \\ 0,2, & 1 < x \leq 6; \\ 0,3, & 6 < x \leq 11; \\ 0,7, & 11 < x \leq 16; \\ 1, & x > 16. \end{cases}$$

Графік цієї функції зображено на рис. 9.3.

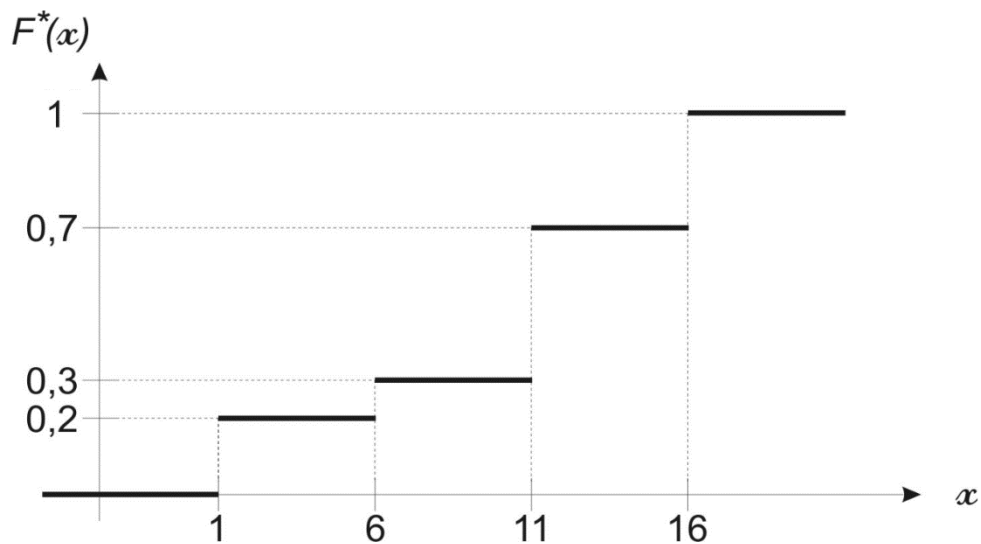


Рис. 9.3. Емпірична функція розподілу

Приклад 9.4. Проведено 40 спостережень швидкості автобусів на певній ділянці дороги (км/год), результати яких наведено в табл. 9.7. Складімо інтервальний ряд статистичного розподілу.

Таблиця 9.7

Дані спостережень швидкості автобусів

41	41	29	25	41	43	42	34	41	30
33	48	50	36	35	46	28	46	50	41
50	27	43	53	48	47	34	35	29	42
30	35	38	41	36	38	45	59	44	43

Розв'язання. Оскільки обсяг вибіркової сукупності є великим ($n = 40$), виконаймо групування варіант заданої вибірки за інтервалами.

Отже,

$$x_{\min} = 25, \quad x_{\max} = 59, \quad R = x_{\max} - x_{\min} = 34.$$

Для визначення кількості інтервалів застосуємо ф-лу (9.1):

$$k = 1 + 3,322 \cdot \lg 40 \approx 6.$$

Довжину інтервалу обчислимо за такою формулою:

$$h = \frac{x_{\max} - x_{\min}}{k}.$$

Тоді буде:

$$h = \frac{59 - 25}{1 + 3,322 \cdot \lg 40} = 5,52.$$

Оскільки початок першого інтервалу можна зсунути на півкроку в бік менших значень, то виберімо крок $h = 5$, а початком першого інтервалу вважаємо $x_1 = x_{\min} - 0,5h$, тобто $x_1 = 25 - 2,5$. Межі останнього інтервалу мають задовольняти умову: $x_k < x_{\max} < x_{k+1}$.

Підрахуймо частоти m_i улучання значень випадкової величини в кожний інтервал і результати запишімо в таблицю (табл. 9.8).

Таблиця 9.8

Інтервальний варіаційний ряд

$[x_i; x_{i+1})$	$[22,5; 27,5)$	$[27,5; 32,5)$	$[32,5; 37,5)$	$[37,5; 42,5)$	$[42,5; 47,5)$	$[47,5; 52,5)$	$[52,5; 57,5)$	$[57,5; 62,5)$	Сума
m_i	2	5	8	11	7	5	1	1	$n = 40$

Приклад 9.5. Для інтервального варіаційного ряду (табл. 9.9) з обсягом вибірки $n = 20$ побудуємо гістограму частот.

Таблиця 9.9

Вихідний інтервальний варіаційний ряд

$(x_{i-1} \div x_i]$	9 – 10	10 – 11	11 – 12	12 – 13	13 – 14	14 – 15	15 – 16
m_i	2	6	4	3	3	1	1

Розв'язання. Для побудови гістограми частот (рис. 9.4) на осі абсцис відкладімо проміжки та на них побудуємо прямокутники, висоти яких дорівнюють $\frac{m_i}{h}$ ($h = 1$ – довжини інтервалів).

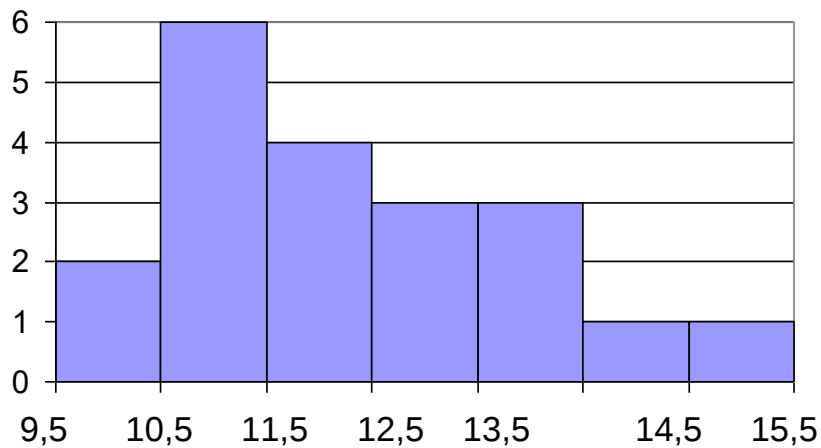


Рис. 9.4. Гістограма частот

Площа побудованих прямокутників дорівнює обсягу вибірки $n = 20$.

9.4. Вправи для самостійної роботи

9.1. За результатами дослідження середньомісячної потреби в картоплі (у кг) серед основних споживачів визначено таке: 200, 150, 250, 100, 150, 200, 200, 300, 500, 300.

Складіть дискретний статистичний розподіл частот та відносних частот.

9.2. Обстеження величини рівня прибутковості (рентабельності, у %) підприємства протягом року показало такі результати: 98, 100, 102, 96, 99, 100, 83, 81, 80, 104, 102, 100.

Потрібно:

- 1) скласти дискретний варіаційний ряд частот та відносних частот;
- 2) побудувати полігон частот і відносних частот;
- 3) визначити емпіричну функцію розподілу та побудувати її графік.

9.3. Обстеження часу (с), протягом якого виконують обслуговування одного клієнта на терміналі, показало такі значення: 34, 34, 34, 33, 35, 32, 34, 32, 37, 36, 34, 35, 35, 32, 33, 35, 34, 36, 38, 33, 35, 34, 36, 34, 33,

35, 32, 37, 32, 34, 36, 32, 34, 33, 35, 35, 33, 34, 35, 38, 36, 34, 35, 33, 37, 35, 34, 37, 33, 35.

Потрібно:

1) скласти дискретний варіаційний ряд, який характеризує розподіл строку обслуговування клієнтів;

2) скласти інтервальний ряд, узявши за довжину інтервалу величину 2 с.

9.4. У цеху встановлено 5 верстатів. Протягом 25 днів реєстрували кількість верстатів, які не працювали. Здобуто такі значення: 0, 1, 2, 1, 1, 2, 3, 2, 1, 4, 2, 0, 0, 2, 2, 3, 3, 1, 0, 1, 2, 1, 3, 5, 0.

Побудуйте статистичну функцію розподілу.

9.5. Для дискретного варіаційного ряду (табл. 9.10) визначте відносні частоти (частоті) та накопичені частоті; запишіть функцію розподілу; побудуйте полігон відносних частот і кумуляту.

Таблиця 9.10

Дискретний варіаційний ряд до вправи 9.3

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

9.6. Кількість балів (x_i), що набрали абітурієнти однієї зі шкіл на тестуванні з математики, наведено в табл. 9.11.

Таблиця 9.11

Дискретний варіаційний ряд кількості балів на тестуванні

x_i	115	125	135	145	155	165	175	185	195
m_i	2	10	15	20	25	15	8	2	3

Побудуйте полігон частот і кумуляту.

9.7. У результаті обстеження фермерських господарств однієї із країн Європи станом на початок року сформовано дані щодо розміру їхньої земельної площі (га): 4,0; 6,4; 8,5; 7,5; 4,5; 6,4; 6,5; 4,5; 6,0; 8,0; 12,0; 4,0; 10,0; 9,5; 12,0; 9,5; 6,5; 9,0; 6,0; 7,2; 8,3; 7,0; 6,8; 11,4; 9,3.

Побудуйте інтервальний ряд розподілу фермерських господарств за розміром земельної площі, утворивши чотири групи з однаковими інтервалами. Результат групування подайте в табличній формі.

9.8. Проведено дослідження річного обсягу споживання хліба (у кг на одну особу) за регіонами. Визначено такі результати: 101; 102,5; 101; 103,5; 104; 100; 103; 107,5; 110; 102,5; 108; 105; 103; 104; 102; 107; 101,5; 105; 102; 104; 105,5; 106; 105,5; 107; 108.

Потрібно:

- 1) скласти інтервальний статистичний розподіл частот та відносних частот (кількість інтервалів визначити за формулою Стерджеса);
- 2) побудувати гістограму частот і відносних частот;
- 3) визначити емпіричну функцію розподілу та побудувати її графік.

9.9. За інтервальним розподілом відсотків преміювання робітників підприємства (табл. 9.12) побудуйте гістограму частот та графік емпіричної функції розподілу.

Сформулюйте гіпотезу про закон розподілу сукупності.

Таблица 9.12

Інтервальний варіаційний ряд відсотків преміювання робітників підприємства

$[x_i, x_{i+1})$	[2, 7)	[7, 12)	[12, 17)	[17, 22)	[22, 27)
m_i	5	10	25	6	4

9.10. У готель заїхали туристи віком 24, 21, 22, 26, 28, 32, 38, 40, 27, 41, 44, 48, 52 роки.

Потрібно:

- 1) скласти інтервальний статистичний розподіл частот та відносних частот (кількість інтервалів визначити за формулою Стерджеса);
- 2) побудувати гістограму частот і відносних частот;
- 3) визначити емпіричну функцію розподілу та побудувати її графік.

9.11. Розподіл робітників підприємства за часом (хв), який вони витрачають на оброблення однієї деталі, наведено у вигляді інтервального статистичного ряду (табл. 9.13).

Інтервальний варіаційний ряд часу оброблення деталі

$[x_i, x_{i+1})$	[2, 4)	[4, 6)	[6, 8)	[8, 10)	[10, 12)
Кількість робітників	42	73	154	205	26

Побудуйте гістограму відносних частот та кумуляту.

9.5. Тестові завдання

9.1. Задано таку вибірку: 2, 4, 5, 3, 5, 7, 5, 6, 8, 9, 3, 7, 6, 7. Її обсяг дорівнює

А 7 **Б** 9 **В** 14 **Г** 15

9.2. Для того щоб задану в завд. 9.1 вибірку перетворити на варіаційний ряд, потрібно

А виписати один раз ті варіанти, які спостерігають у вибірці

Б записати всі варіанти в порядку зростання

В нічого не змінювати

9.3. У результаті дослідження 25 здобувачів вищої освіти щодо часу (хв) на відповідь теоретичного питання з вищої математики отримано такі значення: 15, 4, 17, 11, 16, 17, 7, 5, 6, 23, 19, 22, 20, 7, 9, 13, 19, 18, 21, 25, 1, 15, 11, 5, 21. Інтервальний ряд для п'яти однакових інтервалів $(x_i; x_{i+1}]$ має такий вигляд

А

Інтервали	(1, 4]	(4, 8]	(8, 13]	(13, 19]	(19, 25]
Частоти	2	7	3	6	7

Б

Інтервали	(1, 4]	(4, 8]	(8, 13]	(13, 19]	(19, 25]
Частоти	4	5	4	7	5

В

Інтервали	(0, 5]	(5, 10]	(10, 15]	(15, 20]	(20, 25]
Частоти	4	4	5	7	5

Г

Інтервали	(1, 6]	(6, 10]	(10, 15]	(15, 19]	(19, 25]
Частоти	3	8	4	5	5

Д

Інтервали	(0, 5]	(5, 10]	(10, 15]	(15, 20]	(20, 25]
Частоти	2	5	6	7	3

9.4. Гістограму будують для

- А** дискретного розподілу
- Б** інтервального розподілу

9.5. Ламаною може бути графік

- А** кумуляти
- Б** полігону
- В** гістограми

9.6. Частотою називають

- А** кількість появи окремих значень випадкової величини
- Б** повну кількість об'єктів вибіркової сукупності
- В** різницю між найменшим x_{min} та найбільшим x_{max} значен-

нями варіант

9.7. Сума частот усіх варіант вибірки дорівнює

- А** одиниці
- Б** обсягу вибірки
- В** кількості різних варіант вибірки

9.8. Відносна частота – це

- А** відношення частоти до максимальної частоти
- Б** відношення частоти до обсягу вибірки
- В** відношення максимальної частоти до обсягу вибірки

9.9. Сума відносних частот усіх варіант вибірки дорівнює

- А** одиниці
- Б** обсягу вибірки
- В** кількості різних варіант вибірки

9.10. Сума площ усіх прямокутників, із яких складається гістограма частот статистичного розподілу, дорівнює

- А** одиниці
- Б** обсягу вибірки
- В** кількості різних варіант вибірки

9.6. Запитання для самоперевірки

- 9.1. Сформулюйте, у чому полягають завдання математичної статистики.
- 9.2. Поясніть, у чому полягає сенс вибіркового методу та чому виникає потреба у його застосуванні.
- 9.3. Яким вимогам має відповідати вибірка сукупність?
- 9.4. Які способи відбору застосовують для створення вибірки?
- 9.5. Дайте означення генеральної та вибіркової сукупностей.
- 9.6. Що таке "варіанта"? Як утворюють варіаційний ряд?
- 9.7. Які є способи подання емпіричного закону розподілу дискретної випадкової величини?
- 9.8. Які властивості має емпірична функція розподілу?
- 9.9. Що таке "інтервальний варіаційний ряд"? Які принципи його формування?
- 9.10. Які є способи подання емпіричного закону розподілу неперервної випадкової величини?
- 9.11. Яким вимогам мають відповідати гістограма частот та гістограма відносних частот?
- 9.12. Як побудувати полігон для неперервної випадкової величини?
- 9.13. Які особливості розподілу випадкової величини можна спостерігати за гістограмою?

9.7. Висновки за підрозділом 9

Математична статистика розробляє методи збирання та аналізу даних для формування наукових і практичних висновків. Її предмет – вивчення випадкових величин за результатами спостережень і експериментів. У статистичних дослідженнях застосовують вибірковий метод – із генеральної сукупності вибирають меншу вибірку сукупність, для якої вивчають кількісні або якісні ознаки. Дискретний розподіл зображають полігоном частот або відносних частот, а неперервний – гістограмою.

Література: [1 – 3; 5; 8; 10; 11; 13; 15; 16].

10. Статистичне оцінювання параметрів розподілу

10.1. Мета та компетентності

Метою вивчення підрозділу є ознайомлення з методами визначення точкових та інтервальних оцінок основних числових характеристик одновимірної випадкової величини за вибірковими даними.

Компетентності, що формують під час вивчення підрозділу:

знання методів визначення числових характеристик емпіричного закону розподілу;

розуміння точності вибіркових оцінок та особливостей обчислення довірчого інтервалу для основних числових характеристик;

уміння здійснювати ідентифікацію кількісних характеристик економічних процесів за допомогою вибіркового методу.

10.2. Основні теоретичні відомості

Одним із завдань математичної статистики є статистичне оцінювання числових характеристик випадкових величин та параметрів їхнього розподілу в генеральній сукупності за результатами дослідження вибірки. За результатами вибірки можна визначити не значення певного параметра θ , що характеризує випадкову величину в генеральній сукупності, а лише його *статистичну оцінку* θ^* , яка є випадковою величиною.

Для того щоб статистична оцінка достатньо точно відображала значення параметра, для якого здійснюють оцінювання, вона має відповідати таким вимогам, як: незсунутість, ефективність та ґрунтовність (спроможність).

Статистична оцінка θ^* є **незсунутою оцінкою** параметра θ , якщо математичне сподівання цієї оцінки за будь-якого обсягу вибіркової сукупності дорівнює параметру, що оцінюють: $M(\theta^*) = \theta$. Якщо математичне сподівання статистичної оцінки не дорівнює параметру, який оцінюють, то таку статистичну оцінку називають **зсунутою**.

Ефективною називають статистичну оцінку, якій для заданого обсягу вибірки відповідає найменша з можливих дисперсій. Цю вимогу

можна зрозуміти так: недостатньо дати точкову оцінку деякого параметра, потрібно також, щоб інтервал, до якого буде належати цей параметр генеральної сукупності, був якомога меншим, а надійність, із якою гарантовано потрапляння параметра до цього інтервалу, якомога вищою.

Статистична оцінка є **ґрунтовною (спроможною)**, якщо в разі необмеженого зростання обсягу вибірки вона за ймовірністю наближається до параметра, для якого це оцінювання здійснюють:

$$\lim_{n \rightarrow \infty} P(|\theta^* - \theta| < \varepsilon) = 1,$$

де ε – заздалегідь задане як завгодно мале додатне число $\varepsilon > 0$.

Для будь-якої одновимірної випадкової величини, незалежно від її природи, основними числовими характеристиками статистичного оцінювання є її математичне сподівання і середнє квадратичне відхилення. У зв'язку із цим розглядають точкові та інтервальні статистичні оцінки. Якщо як статистичну оцінку розглядають тільки одне число, то таку оцінку називають **точковою**.

У деяких випадках доцільно можливі значення параметра, що підлягає оцінюванню, характеризувати двома числами – початком і кінцем інтервалу, до якого цей параметр θ буде належати з певною надійністю (довірчою ймовірністю). Таку статистичну оцінку, що визначають двома числами, називають **інтервальною**.

За даними статистичного розподілу визначають основні точкові числові характеристики випадкової величини.

Вибіркова середня $\bar{x}$ визначає центр вибіркової сукупності. За незгрупованими даними вибірку середню визначають як середнє арифметичне вибіркових даних:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i. \quad (10.1)$$

За варіаційним рядом вибірку середню обчислюють як середнє виважене варіант, кожному з яких беруть із вагою, що відповідає її відносній частоті:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^k x_i m_i. \quad (10.2)$$

Вибіркова середня є незсунутою оцінкою математичного сподівання випадкової величини, оскільки правильним є таке співвідношення:

$$M(\bar{x}) = M(X).$$

Вибіркову дисперсію D^* за варіаційним рядом визначають як середнє виважене квадратів відхилення значень випадкової величини від вибіркової середньої:

$$D^* = \frac{1}{n} \cdot \sum_{i=1}^k (x_i - \bar{x})^2 m_i, \text{ або } D^* = \frac{1}{n} \cdot \sum_{i=1}^k x_i^2 m_i - (\bar{x})^2. \quad (10.3)$$

Вибіркова дисперсія є зсунутою оцінкою дисперсії випадкової величини генеральній сукупності.

Зважаючи на це, для вибіркової дисперсії вводять поправку на зсув, яка дорівнює $\frac{n}{n-1}$.

Для визначення **виправленої дисперсії**, що позначають S_x^2 , використовують таку формулу:

$$S_x^2 = \frac{n}{n-1} D^*. \quad (10.4)$$

Середнє квадратичне відхилення вибірки σ^* визначають як корінь квадратний із вибіркової дисперсії: $\sigma^* = \sqrt{D^*}$. Воно характеризує розсіювання випадкової величини навколо центра вибіркової сукупності. На відміну від дисперсії, його вимірність є вимірністю випадкової величини.

Величина $S_x = \sqrt{S_x^2}$ є **виправленим середнім квадратичним відхиленням**. Саме ця величина і є оцінкою середнього квадратичного відхилення теоретичного розподілу, яке характеризує розсіювання значень випадкової величини навколо центра сукупності.

Разом з основними є такі додаткові числові характеристики статистичного розподілу:

мода – це варіанта, якій відповідає найбільша частота;

медіана – це варіанта, що за кількістю варіант розподіляє навпіл статистичний ряд розподілу;

коефіцієнт варіації: $V^* = \frac{\sigma_{\bar{x}}}{\bar{x}} \cdot 100 \%$;

початкові моменти k -го порядку: $\nu_k = \overline{X^k} = \frac{\sum x_i^k m_i}{n}$;

центральні моменти k -го порядку: $\mu_k = \overline{(X - \bar{X})^k} = \frac{\sum (x_i - \bar{x})^k m_i}{n}$.

Інтервальні оцінки використовують у тих випадках, якщо точкові оцінки недостатньо точно відображають характеристики ознаки. Тоді ознаку, яку вивчають, покривають довірчим інтервалом.

Довірчим інтервалом називають інтервал, що покриває всі значення випадкової величини із заданою ймовірністю або із заданим рівнем значущості. Загальноприйнятими є два рівні довірчої ймовірності: 0,95 та 0,99. Разом із надійністю оцінки γ розглядають *рівень значущості* $\alpha = 1 - \gamma$. Він визначає ймовірність того, що за заданим рівнем надійності значення параметра θ , який оцінюють за вибірковою сукупністю, може вийти за межі довірчого інтервалу.

Стандартна форма надання довірчого інтервалу для оцінювання математичного сподівання:

$$\bar{x} - t_{\gamma} \frac{S_x}{\sqrt{n}} < a < \bar{x} + t_{\gamma} \frac{S_x}{\sqrt{n}}. \quad (10.5)$$

Із довірчою ймовірністю $P = \gamma$ похибка ε , із якою математичне сподівання a оцінюють за вибірковою середньою $\bar{x}$, не перевищує значення:

$$\varepsilon = t_{\gamma} \cdot \frac{S_x}{\sqrt{n}}. \quad (10.6)$$

Зі збільшенням $n \rightarrow \infty$ розподіл стає наближеним до нормального розподілу. Тому для великих вибірових сукупностей величину t_{γ} визначають як аргумент інтегральної функції Лапласа (див. додаток В) або за спеціальною таблицею (табл. Г.1, додаток Г).

Для невеликих за обсягом вибірових сукупностей значення t_γ знаходять за спеціальною таблицею критичних точок розподілу Стюдента (додаток Д), відповідно до рівня довірчої ймовірності γ і кількості ступенів свободи $n - 1$.

За ф-лою (10.6) визначають граничний обсяг вибірки:

$$n \geq \left(\frac{t_\gamma \cdot S_x}{\Delta} \right)^2, \quad n \in N, \quad (10.7)$$

де Δ – напівширина довірчого інтервалу, до якого з довірчою ймовірністю, не меншою за γ , буде належати математичне сподівання випадкової величини, розподіленої за нормальним законом.

Довірчий інтервал для оцінювання середнього квадратичного відхилення:

$$S_x(1 - q) \leq \sigma \leq S_x(1 + q). \quad (10.8)$$

Значення величини q , яка залежить від заданої надійності та обсягу вибірки, визначають у спеціальній таблиці (див. табл. Г.2, додаток Г).

10.3. Приклади розв'язування задач

Приклад 10.1. Подано розподіл 70 філій ІТ-компанії за кількістю працівників на них (табл. 10.1).

Таблиця 10.1

Ряд розподілу філій ІТ-компанії

x_i	12	14	21	28	33	35
m_i	6	8	16	12	8	20

Визначмо числові характеристики розподілу філій за кількістю працівників.

Розв'язання. За даним варіаційним рядом вибіркову середню обчислімо за ф-лою (10.2):

$$\begin{aligned}\bar{x} &= \frac{1}{n} \cdot \sum_{i=1}^k x_i m_i = \frac{1}{70} (12 \cdot 6 + 14 \cdot 8 + 21 \cdot 16 + 28 \cdot 12 + 33 \cdot 8 + 35 \cdot 20) = \\ &= \frac{1820}{70} = 26.\end{aligned}$$

Отже, середня кількість працівників у філіях становить 26 осіб.
Дисперсію обчислімо за ф-лами (10.3):

$$\begin{aligned}D^* &= \frac{1}{n} \cdot \sum_{i=1}^k (x_i - \bar{x})^2 m_i = \frac{1}{70} ((12 - 26)^2 \cdot 6 + (21 - 26)^2 \cdot 16 + \\ &+ (28 - 26)^2 \cdot 12 + (33 - 26)^2 \cdot 8 + (35 - 26)^2 \cdot 20) = \frac{2\,298}{35},\end{aligned}$$

або

$$\begin{aligned}D^* &= \frac{1}{n} \cdot \sum_{i=1}^k x_i^2 m_i - (\bar{x})^2 = \frac{1}{70} (12^2 \cdot 6 + 21^2 \cdot 16 + 28^2 \cdot 12 + \\ &+ 33^2 \cdot 8 + 35^2 \cdot 20) - 26^2 = \frac{2\,298}{35}.\end{aligned}$$

Середнє квадратичне відхилення вибірки σ^* визначмо за такою формулою:

$$\sigma^* = \sqrt{D^*} = \sqrt{\frac{2\,298}{35}} \approx 8$$

(кількість працівників у кожній філії компанії відхиляється від середньої кількості в середньому на 8 осіб).

Коефіцієнт варіації визначмо за такою формулою:

$$V^* = \frac{\sigma^*}{\bar{x}} \cdot 100 \% = \frac{8}{26} \cdot 100 \% \approx 30,8 \%.$$

Оскільки $V^* \approx 30,8 \% < 33 \%$, то досліджувана сукупність філій компанії є однорідною.

Приклад 10.2. Вибіркову сукупність задано таблицею розподілу (табл. 10.2).

Таблиця 10.2

Вихідна вибірка сукупність

x_i	3	3,5	4	4,5	5
m_i	5	4	9	4	8

Визначмо виправлену вибірку дисперсію.

Розв'язання. За ф-лою (10.3) обчислимо виправлену вибірку дисперсію.

Обсяг вибірки: $n = 5 + 4 + 9 + 4 + 8 = 30$.

Визначмо вибірку середню за ф-лою (10.2):

$$\bar{x} = \frac{3 \cdot 5 + 3,5 \cdot 4 + 4 \cdot 9 + 4,5 \cdot 4 + 5 \cdot 8}{30} = 4,1$$

та вибірку дисперсію, за ф-лою (10.3):

$$D = \frac{(3 - 4,1)^2 \cdot 5 + (3,5 - 4,1)^2 \cdot 4 + (4 - 4,1)^2 \cdot 9 + (4,5 - 4,1)^2 \cdot 4 + (5 - 4,1)^2 \cdot 8}{30},$$

$$D = 0,49.$$

Отже, виправлена вибірка дисперсія така:

$$S_x^2 = \frac{30}{29} \cdot 0,49 \approx 0,505.$$

Приклад 10.3. Із метою перевірки стану вкладів у банк, аудитор відібрав 100 рахунків. Середнє значення вкладу $\bar{x} = 12,57$ (тис. грн), середнє

квадратичне відхилення для всієї генеральної сукупності (усіх можливих вкладів) $\sigma = 5$ (тис. грн).

Визначмо 95-відсотковий довірчий інтервал середнього вкладу в банку.

Розв'язання. $P = 2\Phi(t) = 0,95$, тоді $\Phi(t) = 0,475$.

За таблицею значень функції $\Phi(t)$ (див. додаток В) визначають $t = 1,96$.

Тоді, за ф-лою (10.5), буде:

$$12,57 - \frac{5 \cdot 1,96}{\sqrt{100}} \leq a \leq 12,57 + \frac{5 \cdot 1,96}{\sqrt{100}},$$

$$11,59 \leq a \leq 13,56.$$

Отже, із надійністю 95 % довірчий інтервал середнього вкладу для цього банку становить (11,59; 13,56) тис. грн.

Приклад 10.4. Нехай $\bar{x} = 2$, $\varepsilon = 0,2$, $\sigma = 1$.

Визначмо, який обсяг має бути у виборці, щоб побудувати довірчий інтервал для математичного сподівання довжини не більш ніж 2ε із надійністю 99 %.

Розв'язання. $P = 2\Phi(t) = 0,99$, тоді $\Phi(t) = 0,495$.

За таблицею значень функції $\Phi(t)$ (див. додаток В), визначмо $t = 2,58$.

За формулою $\varepsilon = \frac{\sigma t}{\sqrt{n}}$, визначмо $n = \frac{\sigma^2 t^2}{\varepsilon^2}$,

звідси

$$n = \frac{1 \cdot (2,58)^2}{(0,2)^2} = 166,41.$$

Отже, $n = 167$.

Приклад 10.5. Нехай $n = 50$, $s = 2,82$.

Із надійністю 95 % побудуймо довірчий інтервал для $\sigma_{\text{ген.}}$.

Розв'язання. За табл. Г.2 додатка Г, буде:

$$q(0,95; 50) = 0,21.$$

Тоді, за ф-лою (10.8), маємо:

$$2,82(1 - 0,21) \leq \sigma_{\text{ген.}} \leq 2,82(1 + 0,21),$$

$$2,228 \leq \sigma_{\text{ген.}} \leq 3,412.$$

Отже, із надійністю 95 % будь-яке число із цього інтервалу можна взяти за генеральне середнє квадратичне відхилення.

10.4. Вправи для самостійної роботи

10.1. Задано інтервальний ряд розподілу заробітної плати (тис. грн) робітників цеху (табл. 10.3).

Таблиця 10.3

Розподіл заробітної плати

X_i	4,00 – 4,40	4,40 – 4,80	4,80 – 5,20	5,20 – 5,60	5,60 – 6,00
m_i	10	15	20	15	10

Визначте середній заробіток, дисперсію та середнє квадратичне відхилення.

10.2. За результатами дослідження щоденного прибутку (тис. грн) від однієї зали ресторану "Харків" протягом трьох тижнів визначено такі дані: 5,2; 4,9; 6,0; 4,8; 5,9; 8,2; 7,0; 6,4; 5,9; 6,5; 7,4; 6,1; 9,1; 8,9; 5,3; 6,2; 6,1; 7,4; 9,0; 9,8; 7,0.

Побудуйте інтервальний варіаційний ряд.

За варіаційним рядом побудуйте гістограму відносних частот, визначте основні числові характеристики розподілу й обчисліть межі інтервалу,

до якого середня в генеральній сукупності буде належати з надійністю 95 %.

10.3. Інспектор відібрав 30 зразків пакетів вівсянки швидкого приготування для контролю за вагою. Результати зважування такі (г): 28,7; 29,8; 30,6; 31,6; 29,5; 30,1; 30; 31,9; 31,7; 29,4; 29,3; 30,1; 30,4; 29,4; 29,8; 30,9; 30; 30,2; 29,8; 30,7; 30,5; 29,9; 30,6; 30,4; 29,8; 30,8; 30,1; 30,4; 30,7; 30.

Побудуйте інтервальний варіаційний ряд, визначте основні числові характеристики й обчисліть межі інтервалу, до якого середня в генеральній сукупності буде належати з надійністю 95 %.

10.4. Проведено дослідження прибутку (тис. грн.) 50 випадково відібраних малих підприємств.

Із надійністю $\gamma = 0,95$ визначте межі надійного інтервал для оцінювання середнього квадратичного відхилення, якщо $\bar{x} = 240$, $s = 25$.

10.5. Виконано 15 вимірювань деякої величини, у результаті яких було визначено вибірку середню $\bar{x} = 30$ і виправлену дисперсію $S^2 = 36$.

Визначте межі довірчого інтервалу, до якого з надійністю 95 % буде потрапляти істинне значення цієї величини.

10.6. За результатами дослідження обчислено кількість туроператорів (од.) у Харківській області за 2011 – 2020 рр. (табл. 10.4).

Таблиця 10.4

Кількість туроператорів (од.) у Харківській області за 2011 – 2020 рр.

Рік	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Кількість туроператорів (од.)	17	13	13	9	9	11	7	4	8	7

Обчисліть основні числові характеристики розподілу та визначте інтервал, до якого середня в генеральній сукупності буде належати з надійністю 95 %.

10.7. Випадковим способом було відібрано 20 абітурієнтів і визначено їхні бали на тестуванні з математики. Визначені результати дослідження

вибіркової сукупності наведено в табл. 10.5 у вигляді інтервального статистичного розподілу.

Побудуйте довірчий інтервал, до якого з надійністю $\gamma = 99\%$ буде належати середній бал усіх абітурієнтів цього віку.

Таблиця 10.5

Розподіл балів оцінювання з математики

$[x_i; x_{i+1})$	[165,5; 170,5)	[170,5; 175,5)	[175,5; 180,5)	[180,5; 185,5)
m_i	4	6	8	2

10.8. Виробник запчастин бажає оцінити середню вагу виробленої деталі. За результатами обстеження вибіркової сукупності із 40 деталей визначено середню вагу $\bar{x} = 250$ мг та середнє квадратичне відхилення $\sigma^* = 2$ мг.

Визначте довірчий інтервал, до якого середня вага деталі для всієї партії буде належати з надійністю 95 %, і потрібний обсяг вибірки, що забезпечить із надійністю 95 % точність оцінювання $\varepsilon = 1$ мг.

10.9. Визначте мінімальний обсяг вибірки, за якого точність оцінювання математичного сподівання за вибірковою середньою дорівнює значенню $\varepsilon = 0,2$ з надійністю 94 %, якщо відомо середнє квадратичне відхилення $\sigma = 1,5$ нормально розподіленої генеральної сукупності.

10.10. За заданим статистичним розподілом вибірки (табл. 10.6) визначте з надійністю $\gamma = 0,99$ інтервальну оцінку для середнього квадратичного відхилення нормально розподіленої випадкової величини X .

Таблиця 10.6

Статистичний розподіл вибірки

x_i	10	15	20	25	30
m_i	5	10	20	5	10

10.11. Визначте з надійністю $\gamma = 95\%$ довірчий інтервал для генерального середнього квадратичного відхилення нормально розподіленої випадкової величини X , якщо для вибіркової сукупності обсягом $n = 20$ відомо виправлене середнє квадратичне відхилення $S = 2,37$.

10.12. Для виробничої фірми є дані про вік та стаж роботи на-
вмання відібраних 15 працівників (табл. 10.7).

Таблиця 10.7

Дані про працівників фірми

Вік, років	22	25	27	29	30	34	35	39	40	41	45	48	50	52	55
Стаж, років	1	6	8	9	10	12	12	20	22	20	24	27	27	28	33

Визначте, за якою ознакою вибірка працівників фірми є більш одно-
рідною.

10.13. Адміністрація супермаркету здійснює аналіз оптимального
рівня запасів продуктів у торговельному залі, а також середньомісячний
обсяг купівель червоної ікри. Для цього менеджер супермаркету протя-
гом серпня реєстрував частоту купівель 120-грамових банок з ікром і зі-
брав такі дані: 7, 4, 5, 9, 3, 2, 1, 2, 0, 3, 2, 3, 5, 8, 10, 6, 5, 7, 3, 2, 9, 8, 2, 4,
6, 5, 4, 3, 1, 2, 8.

Побудуйте варіаційний ряд, визначте його числові характеристики.

Зробіть висновки щодо заходів для поліпшення роботи супермаркету.

10.14. Лікар-дієтолог здійснює оцінювання втрати середньої ваги
людей, які випробовують нову дієту. Із попереднього досвіду він знає,
що середнє квадратичне відхилення втрат ваги сукупності становить
близько 4 кг.

Яку кількість пацієнтів лікарю потрібно вибрати для дослідження,
щоб оцінити втрату середньої ваги з похибкою 0,8 кг із ймовірністю 0,95?

10.15. Проведено опитування навмання вибраних 100 клієнтів перу-
карні щодо якості обслуговування. Виявилося, що 25 із них залишилися
незадоволеними обслуговуванням.

Визначте 99 % довірчий інтервал для виявлення відсотка незадо-
волених серед усіх клієнтів перукарні.

10.16. Туристична фірма провела опитування 150 випадково відібраних потенційних клієнтів щодо їхньої зацікавленості в новому виді послуг. Результати показали, що 27 % опитаних віддали перевагу новому виду екскурсійного обслуговування.

Побудуйте 95 % довірчий інтервал для частки клієнтів туристичної компанії, які можуть стати споживачами нового виду послуг.

10.17. Авіакомпанія, що відкрила новий авіамаршрут, здійснює оцінювання частки пасажирів, що подорожують у цьому напрямку. Було відібрано 350 пасажирів, які користувалися цим маршрутом, причому 200 з них подорожували в ділові поїздки.

Побудуйте 99 % довірчий інтервал частки пасажирів, що подорожують саме в бізнесових справах.

10.18. Товариство захисту прав споживачів оцінює середні витрати бензину марки А-95 у новій моделі автомобіля. Було перевірено 25 автомобілів. Середнє квадратичне відхилення витрати бензину – 2 л на 100 км траси.

Яка довжина 90 % довірчого інтервалу для середніх витрат бензину?

10.19. Із метою просування сухих сніданків на ринку, було запропоновано новий дизайн упаковки. Товар у новій упаковці було розміщено в 15 магазинах міста. У результаті місячного експерименту обсяг продажу становив 8 000 грн із середнім квадратичним відхиленням 720 грн.

Побудуйте 99 % довірчий інтервал середнього обсягу продажу сухих сніданків у новій упаковці.

10.20. Для нарахування абонплати оператор відстежує середню тривалість телефонної розмови.

Скільки телефонних розмов має бути зафіксовано, щоб з імовірністю 0,997 можна було б стверджувати, що відхилення вибіркової середньої від генеральної середньої не перевищило 10 с, якщо генеральне середнє квадратичне відхилення дорівнює 2,5 хв?

10.5. Тестові завдання

10.1. До оцінок генеральної сукупності ставлять такі вимоги

А оцінка має бути стаціонарною, ергодичною й ефективною

Б оцінка має бути спроможною й ефективною

- В** оцінка має бути спроможною, ефективною і незміщеною
Г оцінка має бути незміщеною, стаціонарною й ефективною

10.2. Статистична оцінка θ^* є незсунутою оцінкою параметра θ , якщо

А для заданого обсягу вибірки їй відповідає найменша з можливих дисперсій

Б $M(\theta^*) = \theta$

В $\lim_{n \rightarrow \infty} P(|\theta^* - \theta| < \varepsilon) = 1, \varepsilon > 0$

Г її визначають двома числами

10.3. Статистична оцінка θ^* є спроможною оцінкою параметра θ , якщо

А для заданого обсягу вибірки їй відповідає найменша з можливих дисперсій

Б $M(\theta^*) = \theta$

В $\lim_{n \rightarrow \infty} P(|\theta^* - \theta| < \varepsilon) = 1, \varepsilon > 0$

Г її визначають двома числами

10.4. Статистична оцінка θ^* є ефективною оцінкою параметра θ , якщо

А для заданого обсягу вибірки їй відповідає найменша з можливих дисперсій

Б $M(\theta^*) = \theta$

В $\lim_{n \rightarrow \infty} P(|\theta^* - \theta| < \varepsilon) = 1, \varepsilon > 0$

Г її визначають двома числами

10.5. Вибіркова середня є

А незміщеною оцінкою генерального середнього

Б зміщеною оцінкою генерального середнього

В оцінкою генерального середнього квадратичного відхилення

10.6. Вибіркову середню за незгрупованими даними визначають за такою формулою

А	Б	В	Г
$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^k x_i m_i$	$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$	$\bar{x} = \sum_{i=1}^k x_i m_i$	$\bar{x} = \sum_{i=1}^k (x_i - x_j)^2 m_i$

10.7. Вибіркову дисперсію визначають за такою формулою

А $D^* = \frac{1}{n} \cdot \sum_{i=1}^k x_i^2 m_i$

Б $D^* = \frac{1}{n} \cdot \sum_{i=1}^k x_i^2 m_i - (\bar{x})^2$

В $D^* = \sum_{i=1}^k x_i^2 m_i - (\bar{x})^2$

10.8. Вибіркова дисперсія є

А незміщеною оцінкою дисперсії генеральної сукупності

Б зміщеною оцінкою дисперсії генеральної сукупності

В оцінкою генерального середнього квадратичного відхилення

10.9. Середину варіаційного ряду визначає

А мода

Б медіана

В вибіркова середня

10.10. Між імовірністю P та рівнем значущості α є таке співвідношення

А $P = 1 - \alpha^2$

Б $P = 1 - \alpha$

В $P = 1 / \alpha$

10.11. Стандартна форма подання довірчого інтервалу для оцінювання математичного сподівання з довірчою ймовірністю $P = \gamma$ має такий вигляд

А $\bar{x} - t_\gamma \frac{S_x}{\sqrt{n}} < a < \bar{x} + t_\gamma \frac{S_x}{\sqrt{n}}$

$$\text{Б } \bar{x} - t_{\gamma} \frac{1}{\sqrt{n}} < a < \bar{x} + t_{\gamma} \frac{1}{\sqrt{n}}$$

$$\text{В } \bar{x} - \frac{S_x}{\sqrt{n}} < a < \bar{x} + \frac{S_x}{\sqrt{n}}$$

10.6. Запитання для самоперевірки

10.1. Які є види статистичних оцінок?

10.2. Сформулюйте вимоги, яким мають відповідати статистичні оцінки, і дайте їхнє аналітичне означення.

10.3. Наведіть приклади числових характеристик випадкової величини, які обчислені за вибірковою сукупністю та є статистичними оцінками відповідних характеристик випадкової величини в генеральній сукупності.

10.4. Що таке "інтервальна оцінка" і якими характеристиками її визначають? У чому полягає потреба в побудові інтервальних статистичних оцінок параметрів розподілу?

10.5. Наведіть формулу для визначення довірчого інтервалу для математичного сподівання випадкової величини, розподіленої за нормальним законом, за результатами вибірки.

10.6. Наведіть формулу для визначення довірчого інтервалу для середнього квадратичного відхилення випадкової величини, розподіленої за нормальним законом, за результатами вибірки.

10.7. Висновки за підрозділом 10

Точкові статистичні оцінки θ^* є випадковими величинами, а тому наближена заміна θ на θ^* часто призводить до істотних похибок, особливо якщо обсяг вибірки є малим.

У цьому разі застосовують інтервальні статистичні оцінки. Інтервал $[\theta^* - \delta; \theta^* + \delta]$, що покриває оцінюваний параметр θ генеральної сукупності із заданою надійністю γ , називають довірчим.

Література: [1 – 3; 5; 8; 10 – 13; 15 – 17].

11. Перевірка статистичних гіпотез

11.1. Мета та компетентності

Метою вивчення підрозділу є ознайомлення з методами перевірки статистичних гіпотез та формування компетентностей щодо застосування статистичних критеріїв для перевірки певних статистичних гіпотез.

Компетентності, що формують під час вивчення підрозділу:

знання основних типів статистичних гіпотез;

розуміння принципів побудови статистичних критеріїв та визначення їхніх критичних точок;

знання методик перевірки статистичних гіпотез;

уміння перевіряти відповідність між числовими характеристиками й законом розподілу випадкової величини в генеральній сукупності та їхніми оцінками за результатами дослідження вибіркової сукупності.

11.2. Основні теоретичні відомості

Під час використання методів математичної статистики можна допустити певний відсоток помилкових рішень, оскільки вони ґрунтуються на випадкових величинах і їх беруть в умовах невизначеності. Частку помилкових рішень, якою можна знехтувати, називають **рівнем значущості**. Найчастіше вона становить 5 або 1 %, тобто $\alpha = 0,05$ або $\alpha = 0,01$.

У процесі порівняння декількох статистичних характеристик, обчислених за результатами статистичного оцінювання вибірок, виникає потреба встановити, чи істотна між ними різниця. **Істотною** називають відмінність, що за величиною перевищує ту, яку можна було б пояснити випадковими коливаннями.

Статистичною гіпотезою називають будь-яке припущення на певному рівні статистичної значущості щодо виду або параметрів невідомого закону розподілу.

Твердження про відсутність істотної відмінності між емпіричною та теоретичною характеристиками називають **нульовою (основною) гіпотезою** (H_0). Уводять *альтернативну (конкурентну) гіпотезу* (H_1) – це логічне заперечення нульової гіпотези.

Статистичні гіпотези можуть бути *спрямованими* та *неспрямованими* (табл. 11.1).

Формулювання спрямованих і неспрямованих гіпотез

Спрямовані	Неспрямовані
$H_0 : \theta_1 \geq \theta_2,$ $H_1 : \theta_1 < \theta_2$	$H_0 : \theta_1 = \theta_2,$ $H_1 : \theta_1 \neq \theta_2$

Статистичні гіпотези розподіляють на *параметричні* (гіпотези щодо невідомого значення параметра розподілу, що входить до параметричної сукупності розподілів) та *непараметричні* (припущення, за яким вид розподілу невідомо).

Правило, за яким вирішують підтвердити або відхилити нульову гіпотезу H_0 , називають **статистичним критерієм**.

Під час перевірки статистичних гіпотез можливі помилки двох родів.

Помилка першого роду полягає в тому, що нульову гіпотезу H_0 відхиляють, тоді як вона насправді є правильною. Беруть рівень значущості 0,01, або надійність 99 %.

Помилка другого роду полягає в тому, що нульову гіпотезу H_0 підтверджують, а в дійсності вона є неправильною (рівень значущості 0,05, або надійність 95 %).

Для визначення кращого критерію перевірки гіпотези H_0 потрібно серед усіх критеріїв, які мають ту саму ймовірність похибки першого роду, вибрати той, для якого ймовірність похибки другого роду є найменшою. Похибку першого роду, яку можна припустити, можна задати заздалегідь.

Множину значень, для яких нульову гіпотезу відхиляють, називають **критичною областю**.

Завдання перевірки гіпотези зведено до визначення критичної області заданого рівня значущості. Ймовірність недопущення помилки другого роду називають **потужністю критерію**.

У разі перевірки спрямованих гіпотез використовують одnobічний критерій (лівобічну або правобічну критичну область), неспрямованих – двобічний, критична область якого є удвічі меншою, ніж для одnobічного.

Схему перевірки статистичних гіпотез можна подати в послідовності таких *процедур*:

формулювання нульової та альтернативної гіпотез H_0 , H_1 ;

вибір рівня значущості α ;

вибір типу статистичного критерію;

розрахунки емпіричного значення статистичного критерію;

визначення критичної області критерію (визначення критичних точок);

порівняння емпіричного та критичного значень критерію;

формулювання статистичних висновків.

Порівняння дисперсій двох нормально розподілених генеральних сукупностей

Нехай є вибірки із двох нормально розподілених генеральних сукупностей. Обсяги вибірок, відповідно, n_1 та n_2 . Визначено виправлені вибіркові дисперсії $S_1^2 > S_2^2$. Треба перевірити, чи суттєвою є їхня різниця.

Гіпотеза H_0 полягає в тому, що різницю дисперсій вважають незначущою, тобто $H_0 : S_1^2 = S_2^2$.

Тоді, якщо гіпотези є спрямованими, то $H_1 : S_1^2 > S_2^2$; якщо гіпотези є неспрямованими, то $H_1 : S_1^2 \neq S_2^2$.

Як емпіричне значення критерію обчислюють статистику *критерію Фішера – Снедекора*:

$$F = \frac{S_1^2}{S_2^2}.$$

За таблицею критичних точок розподілу Фішера – Снедекора (додаток Е), визначають такі теоретичні значення статистики:

$$F^* = F(\alpha, k_1, k_2) \text{ у разі спрямованих гіпотез;}$$

$$F^* = F(\alpha / 2, k_1, k_2) \text{ у разі неспрямованих гіпотез,}$$

де $k_1 = n_1 - 1$, $k_2 = n_2 - 1$, α – рівень значущості.

Якщо $F > F^*$, то за рівнем значущості α гіпотезу H_0 відхиляють; якщо $F < F^*$, то з надійністю $1 - \alpha$ гіпотеза H_0 підтверджується.

Порівняння середніх двох нормально розподілених генеральних сукупностей із відомими дисперсіями (незалежні вибірки)

Нехай є дві генеральні сукупності X_1 та X_2 , розподілені нормально, їхні виправлені дисперсії S_1^2 та S_2^2 відомо. За незалежними вибірками, обсяги яких, відповідно, n_1 та n_2 , вилучених із цих сукупностей, визначено вибіркові середні $\bar{X}_1$ та $\bar{X}_2$. Треба перевірити, чи однакові генеральні середні досліджуваних сукупностей.

Нульова й альтернативні гіпотези такі:

$$H_0 : \bar{X}_1 = \bar{X}_2, H_1 : \bar{X}_1 \neq \bar{X}_2 \text{ (неспрямовані гіпотези),}$$

або

$$H_0 : \bar{X}_1 = \bar{X}_2, H_1 : \bar{X}_1 > \bar{X}_2,$$

або

$$H_0 : \bar{X}_1 = \bar{X}_2, H_1 : \bar{X}_1 < \bar{X}_2 \text{ (спрямовані гіпотези).}$$

Як статистику критерію обчислюють таку величину:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}. \quad (11.1)$$

Якщо $H_1 : \bar{X}_1 \neq \bar{X}_2$, то будують двобічну критичну область. За таблицею інтегральної функції Лапласа (див. додаток В), визначають критичну точку t_α ($2\Phi(t) = 1 - \alpha$). Якщо $|t| < t_\alpha$, то за рівнем значущості α гіпотезу H_0 підтверджують, якщо $|t| \geq t_\alpha$, то гіпотезу H_0 відхиляють із надійністю $P = 1 - \alpha$.

Якщо $H_1 : \bar{X}_1 > \bar{X}_2$, то будують правобічну критичну область. За таблицею інтегральної функції Лапласа (див. додаток В), визначають критичну точку t_α ($2\Phi(t) = 1 - 2\alpha$). Якщо $t < t_\alpha$, то за рівнем значущості α гіпотезу H_0 підтверджують, якщо $t \geq t_\alpha$, то гіпотезу H_0 відхиляють із надійністю $P = 1 - \alpha$.

Якщо $H_1 : \bar{X}_1 < \bar{X}_2$, то будують лівобічну критичну область. За таблицею інтегральної функції Лапласа (див. додаток В), визначають критичну точку t_α ($2\Phi(t) = 1 - 2\alpha$). Якщо $t < t_\alpha$, то за рівнем значущості α гіпотезу H_0 відхиляють, якщо $t \geq t_\alpha$, то гіпотезу H_0 підтверджують із надійністю $P = 1 - \alpha$.

Порівняння середніх двох довільно розподілених генеральних сукупностей (великі незалежні вибірки)

За незалежними великими вибірками, обсяги яких, відповідно, n_1 та n_2 , вилучених із довільно розподілених генеральних сукупностей, знайдено вибіркові середні $\bar{X}_1$ та $\bar{X}_2$.

Треба перевірити, чи суттєвою є їхня різниця.

Нульова й альтернативні гіпотези такі:

$$H_0 : \bar{X}_1 = \bar{X}_2, H_1 : \bar{X}_1 \neq \bar{X}_2,$$

або

$$H_0 : \bar{X}_1 = \bar{X}_2, H_1 : \bar{X}_1 > \bar{X}_2,$$

або

$$H_0 : \bar{X}_1 = \bar{X}_2, H_1 : \bar{X}_1 < \bar{X}_2.$$

Завдяки великим обсягам вибірок, вибіркові дисперсії є гарними оцінками генеральних дисперсій і тому їх можна вважати відомими приблизно. Тоді як приблизну статистику критерію обчислюють таку величину:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}. \quad (11.2)$$

Далі приблизні висновки роблять такі, як і під час порівняння середніх двох нормально розподілених генеральних сукупностей із відомими дисперсіями.

*Порівняння середніх двох нормально розподілених
генеральних сукупностей із невідомими дисперсіями
(малі незалежні вибірки)*

Нехай є дві малі вибірки з генеральних сукупностей, нормально розподілені. Обсяг першої вибірки – n_1 , другої – n_2 . Треба перевірити, чи суттєвою є різниця їхніх вибірових середніх $\bar{X}_1$ та $\bar{X}_2$.

Нульова й альтернативні гіпотези такі:

$$H_0 : \bar{X}_1 = \bar{X}_2, H_1 : \bar{X}_1 \neq \bar{X}_2,$$

або

$$H_0 : \bar{X}_1 = \bar{X}_2, H_1 : \bar{X}_1 > \bar{X}_2,$$

або

$$H_0 : \bar{X}_1 = \bar{X}_2, H_1 : \bar{X}_1 < \bar{X}_2.$$

Додатково слід припустити, що невідомі генеральні дисперсії є однаковими. Інакше слід перевірити гіпотезу про рівність генеральних дисперсій за критерієм Фішера – Снедекора.

Перевірку нульової гіпотези здійснюють за *t-критерієм Стьюдента*, статистика якого має такий вигляд:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}. \quad (11.3)$$

Визначене емпіричне значення порівнюють із критичною точкою розподілу Стьюдента для заданого рівня значущості α і кількості ступенів свободи $k = n_1 + n_2 - 2$ (за таблицею додатка Д визначають t_α).

Якщо $H_1 : \bar{X}_1 \neq \bar{X}_2$, то будують двобічну критичну область. Якщо $|t| < t_\alpha$, то за рівнем значущості α гіпотезу H_0 підтверджують; якщо $|t| \geq t_\alpha$, то гіпотезу H_0 відхиляють із надійністю $P = 1 - \alpha$.

Якщо $H_1: \bar{X}_1 > \bar{X}_2$, то будують правобічну критичну область. Якщо $t < t_\alpha$, то за рівнем значущості α гіпотезу H_0 підтверджують; якщо $t \geq t_\alpha$, то гіпотезу H_0 відхиляють із надійністю $P = 1 - \alpha$.

Якщо $H_1: \bar{X}_1 < \bar{X}_2$, то будують лівобічну критичну область. Якщо $t < t_\alpha$, то за рівнем значущості α гіпотезу H_0 відхиляють; якщо $t \geq t_\alpha$, то гіпотезу H_0 підтверджують із надійністю $P = 1 - \alpha$.

Перевірка гіпотези про значення генеральної середньої

Для перевірки нульової гіпотези $H_0: \bar{X}_T = a$, ($M(x) = a$), де a є певним числом, за заданого рівня значущості α припускають, що дисперсію генеральної сукупності відомо або визначено за вибіркою великого обсягу.

Як статистику критерію обчислюють таку величину:

$$t = \frac{\sqrt{n}(\bar{x}_g - a)}{\sigma_T}. \quad (11.4)$$

Якщо $H_1: \bar{X}_T > a$, то будують правобічну критичну область. За таблицею інтегральної функції Лапласа (див. додаток В), визначають критичну точку t_α ($2\Phi(t) = 1 - 2\alpha$). Якщо $t < t_\alpha$, то за рівнем значущості α гіпотезу H_0 підтверджують; якщо $t \geq t_\alpha$ гіпотезу H_0 відхиляють із надійністю $P = 1 - \alpha$.

Якщо $H_1: \bar{X}_T < a$, то будують лівобічну критичну область. За таблицею інтегральної функції Лапласа (див. додаток В), визначають критичну точку t_α ($2\Phi(t) = 1 - 2\alpha$). Якщо $t < t_\alpha$, то за рівнем значущості α гіпотезу H_0 відхиляють; якщо $t \geq t_\alpha$, то гіпотезу H_0 підтверджують із надійністю $P = 1 - \alpha$.

Якщо $H_1: \bar{X}_T \neq a$, то будують двобічну критичну область. За таблицею інтегральної функції Лапласа (див. додаток В), визначають критичну точку t_α ($2\Phi(t) = 1 - \alpha$). Якщо $|t| < t_\alpha$, то за рівнем значущості α гіпотезу H_0 підтверджують; якщо $|t| \geq t_\alpha$, то гіпотезу H_0 відхиляють із надійністю $P = 1 - \alpha$.

Розглянутий підхід до побудови критичних областей є прийнятним лише за умови, якщо відомо значення середнього квадратичного відхилення σ_T ознаки генеральної сукупності.

У разі, якщо значення σ_T є невідомим, його замінюють таким ви-
правленим середнім квадратичним відхиленням:

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2 n_i}{n-1}}.$$

Тоді статистикою критерію є така випадкова величина:

$$t = \frac{\bar{x}_B - a}{\frac{S}{\sqrt{n}}}. \quad (11.5)$$

Критичні точки в цьому разі визначають за таблицею критичних точок розподілу Стюдента (див. додаток Д) за заданим рівнем значущості α та кількості ступенів свободи $k = n - 1$.

Критерії згоди щодо закону розподілу

Часто закон розподілу випадкової величини в генеральній сукупності є невідомим, але певні припущення щодо його характеру можна зробити, зважаючи на гістограму у вибірковій сукупності. У цьому разі перевіряють нульову гіпотезу H_0 – це генеральна сукупність, розподілена за теоретичним законом.

Тоді H_1 – це розподіл генеральної сукупності, що відрізняється від теоретичного закону.

Критерієм згоди називають критерій перевірки гіпотези щодо закону розподілу.

Згідно із *критерієм згоди Пірсона*, спостережуваний емпіричний розподіл вибіркової сукупності, визначений емпіричними частотами m_i згрупованого варіаційного ряду, порівнюють із допустимим теоретичним розподілом генеральної сукупності, відображеним теоретичними частотами $\tilde{m}_i$.

Емпіричними називають частоти, які спостерігають у процесі аналізу вибірки, а **теоретичними** – ті, які обчислюють за формулами.

Якщо кількість спостережень є дуже великою ($n \rightarrow \infty$), то закон розподілу випадкової величини, незалежно від того, якому закону розподілу підпорядковано генеральну сукупність, наближається до розподілу χ^2 із k ступенями свободи, а сам критерій називають **критерієм згоди "хі-квадрат"**, або **критерієм Пірсона**.

Для перевірки нульової гіпотези H_0 слід обчислити таку статистику:

$$\chi^2 = \sum_{i=1}^s \frac{(m_i - \tilde{m}_i)^2}{\tilde{m}_i}, \quad (11.6)$$

де s – кількість інтервалів згрупованого ряду розподілу;

m_i – емпіричні частоти;

$\tilde{m}_i$ – теоретичні частоти ($\tilde{m}_i = n\tilde{p}_i$);

$\tilde{p}_i = hf_i$ – теоретичні ймовірності;

$$f_i = \frac{\varphi(t_i)}{\sigma};$$

$$t_i = \frac{x_i - \bar{x}}{\sigma} \text{ – стандартизована величина;}$$

$$x_i = \frac{x_{i-1}^* + x_i^*}{2} \text{ – середнє інтервальне;}$$

$\varphi(t_i)$ – диференціальна функція Лапласа.

Слід зауважити, що кількість спостережень m_i у кожному інтервалі має бути не меншою від 5 % від загальної кількості спостережень: $m_i \geq 0,05n$. Якщо їх буде менше, то слід збільшити інтервали.

Обчислену за ф-лою (11.6) величину порівнюють із критичними значеннями $\chi_\alpha^2(k)$, які визначають у спеціальній довідковій таблиці (додаток Ж).

Величину k визначають за такою формулою:

$$k = s - r - 1, \quad (11.7)$$

де s – кількість збільшених інтервалів;

r – кількість параметрів теоретичного закону розподілу (для нормального закону $r = 2$, для показникового – $r = 1$, для рівномірного – $r = 2$).

Величина α визначає рівень значущості. Для критерію Пірсона розглядають два рівні значущості: $\alpha = 0,05$ і $\alpha = 0,01$.

Якщо $\chi^2 < \chi_{0,05}^2(k)$, то нульову гіпотезу H_0 підтверджують, тобто теоретичний закон розподілу відповідає емпіричним даним і помилка у п'яти випадках зі ста в разі є підтвердженням можливо хибної гіпотези (помилка другого роду).

Якщо $\chi^2 > \chi_{0,01}^2(k)$, то нульову гіпотезу слід відхилити, тобто теоретичний закон розподілу не відповідає емпіричним даним, помилка в одному випадку зі ста в разі відхилення можливо правильної гіпотези (помилка першого роду).

Якщо $\chi_{0,05}^2(k) < \chi^2 < \chi_{0,01}^2(k)$, то є невизначеність і слід використати інші критерії.

За критерієм згоди Пірсона за Романовським (критерій Романовського), обчислюють таку величину:

$$t = \left| \frac{\chi^2 - k}{\sqrt{2k}} \right|,$$

де $k = s - 3$.

Якщо $t \leq 1,96$, то з надійністю 95 % гіпотезу H_0 підтверджують, тобто закон розподілу відповідає нормальному;

якщо $t \geq 2,58$, то з рівнем значущості $\alpha = 0,01$ гіпотезу H_0 відхиляють;

якщо $t \in (1,96; 2,58)$, то це область невизначеності.

За критерієм згоди Колмогорова, обчислюють таку величину:

$$\lambda = D\sqrt{n},$$

де $D = \max|F(x) - \tilde{F}(x)|$, що є найбільшою за абсолютною величиною різницею між значеннями емпіричної функції розподілу та відповідної теоретичної функції розподілу.

За спеціальною таблицею (табл. 11.2), визначають відповідне критичне значення λ_α .

Таблиця 11.2

Критичні значення критерію Колмогорова

α	0,400	0,300	0,200	0,100	0,050	0,025	0,010	0,005	0,001	0,0005
λ_α	0,89	0,97	1,07	1,22	1,36	1,48	1,63	1,73	1,95	2,03

Якщо $\lambda > \lambda_\alpha$, то гіпотезу H_0 відхиляють, тобто розбіжність між розподілами вважають значущою і її не може бути викликано випадковими причинами. Якщо $\lambda \leq \lambda_\alpha$, то гіпотезу H_0 погоджують із даними розподілу, тобто закон розподілу відповідає нормальному.

Перевірка гіпотези про однорідність вибірок

Якщо є дві незалежні вибірки з невідомими теоретичними функціями розподілу $F_1(x)$ та $F_2(x)$, то нульова гіпотеза про однорідність заданих вибірок має такий вигляд: $H_0 : F_1(x) = F_2(x)$ за альтернативної $H_1 : F_1(x) \neq F_2(x)$.

Для перевірки нульової гіпотези за критерієм Колмогорова – Смирнова обчислюють таку величину:

$$\lambda = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \cdot \max|F_{n_1}(x) - F_{n_2}(x)|,$$

де $F_{n_1}(x)$ та $F_{n_2}(x)$ – емпіричні функції розподілу, побудовані за двома вибірками обсягів n_1 та n_2 .

Для вибірок великих обсягів за спеціальною таблицею (див. табл. 11.2) визначають відповідне критичне значення λ_α .

Якщо $\lambda > \lambda_\alpha$, то гіпотезу H_0 відхиляють; якщо $\lambda \leq \lambda_\alpha$, то гіпотезу H_0 не відхиляють.

11.3. Приклади розв'язування задач

Приклад 11.1. Є дві вибірки X та Y з нормально розподілених генеральних сукупностей. Для вибірки X : $n_1 = 10$, $S_x = 1,23$. Для вибірки Y : $n_2 = 18$, $S_y = 0,41$.

Із надійністю 95 % перевірмо гіпотезу про значущість різниці двох вибірових дисперсій.

Розв'язання. Гіпотеза H_0 полягає в тому, що різницю дисперсій вважають незначущою. Нехай $H_1: S_x^2 > S_y^2$, тобто гіпотези є спрямованими.

За формулою обчислимо статистику критерію Фішера – Снедекора:

$$F = \frac{S_{max}^2}{S_{min}^2} = \frac{1,23}{0,41} = 3.$$

Обчислимо кількість ступенів свободи $k_1 = 10 - 1 = 9$, $k_2 = 18 - 1 = 17$, якщо за умовою завдання рівень значущості $\alpha = 0,05$.

Тоді з таблиці критичних точок розподілу Фішера – Снедекора (див. додаток Е), визначмо критичну точку $F(0,05; 9,17) = 2,5$.

Оскільки $F > F(0,05; 9,17)$, то гіпотезу H_0 відхиляють, тобто різниця дисперсій є значущою й дисперсія вибірки X є суттєво більшою від дисперсії вибірки Y .

Приклад 11.2. Зроблено два вимірювання температури повітря, по 15 спроб у кожному (табл. 11.3 і 11.4), X – температура повітря.

Перевірмо з надійністю 95 %, чи є суттєво різними вибірові середні.

Таблиця 11.3

Розподіл температури повітря в першому вимірюванні

x_{1i}	-3	-2	-1	0	1	2	3
m_{1i}	1	3	2	2	2	3	2

Таблиця 11.4

Розподіл температури повітря в другому вимірюванні

x_{2i}	-2	-1	0	1	2
m_{2i}	2	5	3	4	1

Розв'язання. Обчислимо:

вибіркові середні $\bar{X}_1 = 0,2$; $\bar{X}_2 = -0,2$;

вибіркові дисперсії $D(X_1) = 3,63$; $D(X_2) = 1,36$;

виправлені вибіркові дисперсії $S_1^2 = 3,89$; $S_2^2 = 1,46$.

Гіпотеза H_0 – розбіжність між вибірковими середніми є неістотною.

$H_0 : \bar{X}_1 = \bar{X}_2$, $H_1 : \bar{X}_1 \neq \bar{X}_2$.

Нехай невідомі генеральні дисперсії є однаковими. Оскільки вибіркові сукупності є малими за обсягом, застосуємо критерій Стюдента.

За ф-лою (11.3), обчислимо статистику критерію Стюдента:

$$t = \frac{|0,2 + 0,2|}{\sqrt{\frac{(15-1) \cdot 3,89 + (15-1) \cdot 1,46}{15+15-2} \left(\frac{1}{15} + \frac{1}{15} \right)}} = 0,67.$$

Емпіричне значення порівняймо із критичною точкою розподілу Стюдента для заданого рівня значущості α і кількості ступенів свободи $k = 15 + 15 - 2 = 28$ (за таблицею додатка Д, визначмо $t_{0,05}(28) = 2,05$).

Оскільки $t = 0,67 < t_{0,05}(28) = 2,05$, то з рівнем значущості 0,05 гіпотезу H_0 про неістотність різниці вибіркових середніх не відхиляють.

Приклад 11.3. Розбіжність вимірів діаметрів металевої деталі є випадковою величиною, що має нормальний закон розподілу з $\sigma_T = 4$ мм.

За рівня значущості $\alpha = 0,01$ перевірмо правильність нульової гіпотези $H_0 : a = 240$ мм за альтернативної гіпотези $H_1 : a > 240$ мм, якщо відомо, що вибіркове середнє значення діаметрів, виміряних у 100 однотипних деталей, $\bar{x}_B = 225$ мм.

Розв'язання. Оскільки $H_1 : a > 240$ мм, будуймо правобічну критичну область. Для цього потрібно визначити критичну точку та побудувати правобічну критичну область.

Для визначення критичної точки застосуємо такий вираз:

$$\Phi(t_\alpha) = \frac{1-2\alpha}{2} = \frac{1-2 \cdot 0,01}{2} = \frac{1-0,02}{2} = \frac{0,98}{2} = 0,49.$$

За значенням $\Phi(t_\alpha) = 0,49$, скориставшись таблицею інтегральної функції Лапласа (див. додаток В), визначмо $t_\alpha \approx 2,34$.

Обчислимо значення статистики критерію за ф-лою (11.4):

$$t = \frac{225 - 240}{\frac{4}{\sqrt{100}}} = \frac{-15}{\frac{4}{10}} = -\frac{15}{0,4} = -\frac{150}{4} = -37,5.$$

Оскільки $t < t_\alpha$, то немає підстав для відхилення нульової гіпотези $H_0 : a = 240$ мм.

Отже, нульову гіпотезу не відхиляють.

Приклад 11.4. За прогнозом середня дебіторська заборгованість малих підприємств однієї області має становити $a = 100$ тис. грн. Вибіркова перевірка 20 підприємств показала, що середня заборгованість становить $\bar{x}_B = 115$ тис грн та середнє квадратичне відхилення – $S = 15$ тис грн.

За рівня значущості $\alpha = 0,05$ перевірмо, чи можна взяти такий прогноз.

Розв'язання. Нульова гіпотеза H_0 – середня дебіторська заборгованість становить 100 тис. грн. $H_0: \bar{x}_B = 100$, альтернативна гіпотеза H_1 – середня дебіторська заборгованість є більшою ніж 100 тис. грн $H_1: \bar{x}_B > 100$.

Генеральна дисперсія є невідомою, тому слід використати ф-лу (11.5):

$$t = \frac{\bar{x}_B - a}{\frac{S}{\sqrt{n}}} = \frac{115 - 100}{15} \cdot \sqrt{20} = 4,47.$$

Оскільки конкурентна гіпотеза $H_1: \bar{x}_B > 100$ є правобічною, то критичні точки визначмо за таблицею критичних точок розподілу Стьюдента (див. додаток Д) за заданим рівнем значущості $\alpha = 0,05$, який розміщений у нижньому рядку таблиці, та кількості ступенів свободи $k = n - 1 = 20 - 1 = 19$:

$$t_{кр}(0,05; 19) = 1,73.$$

Оскільки $t < t_{кр}$, то нульову гіпотезу $H_0: \bar{x}_B = 100$ слід відхилити, тобто за рівня значущості $\alpha = 0,05$ такий прогноз взяти не можна.

Приклад 11.5. На підприємстві здійснено аналіз стажу роботи 100 працівників (табл. 11.5).

Таблиця 11.5

Розподіл стажу роботи працівників

Стажі роботи, роки	[1, 3)	[3, 5)	[5, 7)	[7, 9)	[9, 11)	[11, 13)	[13, 15)	[15, 17)	[17, 19)	[19, 21)	[21, 23)
m_i	2	4	6	10	18	20	16	11	7	5	1

Перевірмо гіпотезу про нормальний розподіл досліджуваної випадкової величини – стажу роботи працівників.

Розв'язання. Нульова гіпотеза H_0 : досліджувану випадкову величину розподілено нормально.

Альтернативна гіпотеза H_1 : розподіл досліджуваної випадкової величини не підпорядковано нормальному закону.

Обчислення виконаймо в табл. 11.6.

Таблиця 11.6

Визначення вирівнювальних частот

№ з/п	$\bar{x}_i$	m_i	$t_i = \frac{\bar{x}_i - \bar{x}}{S_x}$	$\varphi(t_i)$	$\hat{m}_i$	$\frac{(m_i - \hat{m}_i)^2}{\hat{m}_i}$
1	2	12	-2,37	0,0241	11	0,09
2	4		-1,89	0,0669		
3	6		-1,42	0,1456		
4	8	10	-0,95	0,2541	12	0,33
5	10	18	-0,48	0,3555	17	0,06
6	12	20	-0,01	0,3989	19	0,05
7	14	16	0,46	0,3589	17	0,06
8	16	11	0,93	0,2589	12	0,33
9	18	7	1,40	0,1497	7	0,00
10	20	6	1,87	0,0694	5	0,20
11	22		2,34	0,0252		
Сума		100	—	—	100	1,12

Визначмо основні числові характеристики емпіричного розподілу. Оскільки вихідні дані наведено у вигляді інтервального ряду, то для визначення вибіркової середньої застосуємо таку формулу:

$$\bar{x} = \frac{1}{\sum_{i=1}^s m_i} \sum_{i=1}^s \bar{x}_i m_i.$$

Значеннями випадкової величини $\bar{x}_i$ для цього ряду є середини відповідних інтервалів інтервального ряду емпіричного розподілу.

Визначмо $\bar{x} = 12,04$ (р.).

Далі обчислимо виправлену дисперсію за таким співвідношенням:

$$S_x^2 = \frac{n}{n-1} \sum_{i=1}^s (\bar{x}_i - \bar{x})^2 \cdot m_i = 4,26.$$

Перевірку нульової гіпотези виконаймо за критерієм Пірсона.

Після об'єднання малонасичених інтервалів їхня кількість становить $s = 8$, а тоді кількість ступенів свободи $k = s - r - 1 = 8 - 2 - 1 = 5$ (для нормального розподілу).

Для обчислення вирівнювальних частот застосовують таке співвідношення:

$$\hat{m}_i = \frac{n \cdot h}{S_x} \varphi(t_i).$$

Отже, визначмо $\chi_{емп.}^2 = 1,12$.

За таблицею критичних точок розподілу Пірсона (див. додаток Ж) визначмо $\chi_{0,05}^2(5) = 15,1$. Оскільки $\chi_{емп.}^2 < \chi_{0,05}^2(k)$, то немає підстав для відхилення нульової гіпотези.

Отже, закон розподілу стажу роботи працівників досліджуваного підприємства можна вважати нормальним.

Приклад 11.6. За умовами прикл. 11.4, перевірмо гіпотезу про нормальний закон розподілу, використовуючи критерій Романовського.

Розв'язання. Обчислимо емпіричне значення критерію Романовського за такою формулою:

$$\frac{|\chi^2 - k|}{\sqrt{2k}},$$

де k – кількість ступенів свободи.

Для цієї задачі $k = 8 - 3 = 5$.

$$\text{Тоді } \frac{|\chi^2 - k|}{\sqrt{2k}} = \frac{|1,12 - 5|}{\sqrt{10}} = 1,23 < 1,96.$$

Отже, розбіжність між емпіричним і теоретичним розподілами можна вважати статистично незначущою, тобто гіпотезу про нормальний закон розподілу не відхиляють.

Приклад 11.7. За рівнем значущості $\alpha = 0,05$, перевірмо нульову гіпотезу, що кількість верстатів, що не працюють, серед 5 верстатів, які є в цеху, розподілено за біноміальним законом із $p = \frac{1}{3}$, якщо

$$D_n = \max_x |F(x) - F_n^*(x)| = 0,0683 \text{ i } n = 25.$$

Розв'язання. Використаймо для перевірки нульової гіпотези критерій згоди Колмогорова.

За таблицею розподілу Колмогорова (див. табл. 11.2) $\lambda_\alpha = 1,36$.

Обчислімо статистику критерію та порівняймо її із цим критичним значенням:

$$\sqrt{n}D_n = \sqrt{25} \cdot 0,0683 = 0,3415 < \lambda_\alpha = 1,36.$$

Значення статистичної характеристики не належить критичній області, тому гіпотезу про біноміальний закон розподілу в сукупності слід підтвердити.

Приклад 11.8. Протягом місяця вибірково здійснювали перевірку торговельних майданчиків міста із продажу овочів. Результати двох перевірок щодо обважування покупців за одним видом овочів наведено в табл. 11.7.

За рівнем значущості $\alpha = 0,05$, перевірмо нульову гіпотезу про однорідність досліджуваних вибірок.

Результати двох перевірок щодо обважування покупців

Інтервали обважування, г	Кількість під час першої перевірки, шт.	Кількість під час другої перевірки, шт.
0 – 10	3	5
10 – 20	10	12
20 – 30	15	8
30 – 40	20	25
40 – 50	12	10
50 – 60	5	8
60 – 70	25	20
70 – 80	15	7
80 – 90	5	5
Разом	110	100

Розв'язання. Позначмо $n_1 = 110, n_2 = 100$.

Обчислимо накопичені частоти n_{i1}^+, n_{i2}^+ вибірок та значення їхніх емпіричних функцій:

$$F_{n1}(x_i) = n_{i1}^+ / n_1, F_{n2}(x_i) = n_{i2}^+ / n_2.$$

Нульова гіпотеза про однорідність заданих вибірок:

$$H_0 : F_1(x) = F_2(x) \text{ за альтернативної } H_1 : F_1(x) \neq F_2(x).$$

Для перевірки нульової гіпотези за критерієм Колмогорова – Смирнова обчислимо цю величину за такою формулою:

$$\lambda = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \cdot \max |F_{n_1}(x) - F_{n_2}(x)|,$$

де $F_{n_1}(x)$ та $F_{n_2}(x)$ – емпіричні функції розподілу, побудовані за двома вибірками обсягів n_1 та n_2 .

Допоміжні обчислення наведено в табл. 11.8.

Таблица 11.8

Обчислення емпіричних функцій розподілу

x_i	n_{i1}^+	n_{i2}^+	$F_{n1}(x_i)$	$F_{n2}(x_i)$	$ F_{n1}(x_i) - F_{n2}(x_i) $
10	3	5	0,027	0,05	0,023
20	13	17	0,118	0,17	0,052
30	28	25	0,254	0,25	0,004
40	48	50	0,436	0,50	0,064
50	60	60	0,545	0,60	0,055
60	65	68	0,591	0,68	0,089
70	90	88	0,818	0,88	0,072
80	105	95	0,955	0,95	0,050
90	110	100	1,000	1,00	0,000

Отже, буде:

$$\lambda = \sqrt{\frac{110 \cdot 100}{110 + 100}} \cdot 0,089 = 0,644.$$

За таблицею розподілу Колмогорова (див. табл. 11.2) $\lambda_\alpha = 1,36$.

Оскільки $\lambda < \lambda_\alpha$, то гіпотезу H_0 не відхиляють, тобто обважування покупців описують однією функцією розподілу.

Тим самим підтверджують, що обважування покупців є стійким та закономірним процесом із продажу овочів у цьому місті.

11.4. Вправи для самостійної роботи

11.1. Із генеральних сукупностей, розподілених за нормальним законом, було сформовано дві вибіркові сукупності обсягами $n = 12$ і $m = 16$. Для цих вибірок визначено виправлені дисперсії $S_x^2 = 1,52$ та $S_y^2 = 0,76$.

Перевірте, за рівнем значущості $\alpha = 0,05$ для генеральної сукупності, нульову гіпотезу $H_0 : D(X) = D(Y)$ про рівність генеральних дисперсій.

11.2. За рівнем значущості $\alpha = 0,05$, перевірте нульову гіпотезу $H_0 : M(X) = M(Y)$ про рівність генеральних середніх нормально розподілених сукупностей. Вибіркові сукупності надано у вигляді дискретних варіаційних рядів (табл. 11.9 і 11.10).

Таблиця 11.9

Вибіркова сукупність X

x_i	12,3	12,5	12,8	13,0	13,5
n_i	1	2	4	2	1

Таблиця 11.10

Вибіркова сукупність Y

y_i	12,2	12,3	13,0
m_i	6	8	2

11.3. Досліджували продуктивність робітників (грн) двох виробничих філій одного заводу. Результати досліджень подано двома статистичними розподілами (табл. 11.11 і 11.12).

Таблиця 11.11

Продуктивність робітників першої філії

x_i	140,8	160,8	180,8	200,8	220,8
n_i	2	6	32	8	2

Таблиця 11.12

Продуктивність робітників другої філії

y_i	150,6	160,6	170,6	180,6	190,6
m_i	12	28	40	18	2

Ознаки X і Y є незалежними та розподіленими за нормальним законом.

За рівнем значущості $\alpha = 0,01$, перевірте правильність нульової гіпотези $H_0 : M(X) = M(Y)$, якщо $H_1 : M(X) < M(Y)$.

11.4. У двох партіях містяться однотипні підшипники, які виготовлено двома заводами. Вимірювання їхніх діаметрів (мм) дали результати, наведені в табл. 11.13 і 11.14.

Таблиця 11.13

Розподіл діаметрів підшипників першого заводу

x_i	6,58	6,60	6,80	7,00	7,20
n_i	6	8	10	4	2

Таблиця 11.14

Розподіл діаметрів підшипників другого заводу

y_i	6,60	6,70	6,74	6,78	6,82
m_i	2	4	8	6	4

За рівнем значущості $\alpha = 0,01$, перевірте правильність нульової гіпотези $H_0 : M(X) = M(Y)$ за альтернативної $H_1 : M(X) \neq M(Y)$, якщо відомо значення $D_x = 50$; $D_y = 60$.

11.5. За даними завд. 11.3 та 11.4, перевірте гіпотезу про однорідність вибірок.

11.6. Із генеральної сукупності, ознака якої X має нормальний закон розподілу з $\sigma_T = 5$, реалізовано вибірку і побудовано статистичний розподіл (табл. 11.15).

Таблиця 11.15

Вибірковий розподіл

x_i	10,9	11,0	11,2	11,3	11,5	11,6	11,8	11,9
n_i	2	4	1	3	4	1	2	3

За рівнем значущості $\alpha = 0,01$, перевірте правильність нульової гіпотези $H_0 : a = 11,44$, якщо альтернативна гіпотеза $H_1 : a \neq 11,44$.

11.7. Розподіл випадкової величини X за даними вибірки визначено інтервальним варіаційним рядом (табл. 11.16).

Таблиця 11.16

Вихідний інтервальний варіаційний ряд

Інтервали	[3,0; 3,6)	[3,6; 4,2)	[4,2; 4,8)	[4,8; 5,4)	[5,4; 6,0)	[6,0; 6,6)	[6,6; 7,2)
Частоти (m_i)	2	8	35	43	22	15	5

За рівнем значущості $\alpha = 0,01$, перевірте гіпотезу про нормальний розподіл випадкової величини X у генеральній сукупності.

11.8. Як результат опрацювання статистичних даних обчислено значення емпіричних (m_i) і вирівнювальних ($\hat{m}_i$) частот у припущенні, що ознаку розподілено за нормальним законом (табл. 11.17).

Таблиця 11.17

Емпіричні та вирівнювальні частоти за нормального закону

m_i	3	7	11	20	28	19	10	2
$\hat{m}_i$	2	6	14	23	25	18	9	3

За рівнем значущості $\alpha = 0,05$, перевірте гіпотезу про нормальний закон розподілу ознаки генеральної сукупності.

11.9. Як результат опрацювання статистичних даних обчислено значення емпіричних (m_i) і вирівнювальних ($\hat{m}_i$) частот у припущенні, що ознаку розподілено за показниковим законом (табл. 11.18).

Емпіричні та вирівнювальні частоти за показникового закону

m_i	40	30	20	6	4
$\hat{m}_i$	48	25	13	7	7

За рівнем значущості $\alpha = 0,01$, перевірте нульову гіпотезу про показниковий розподіл випадкової ознаки X у генеральній сукупності.

11.10. Компанія-автоперевізник здійснює міжміські пасажирські перевезення. Менеджер компанії перш ніж придбати партію нових мікроавтобусів "Мерседес", вирішив проаналізувати вигідність цього проєкту. У результаті він виявив, що придбання нових мікроавтобусів дасть компанії суттєву вигоду лише тоді, якщо витрата палива на 100 км траси на нових автобусах не перевищить 19 л. Продавець надав компанії-автоперевізнку для перевірки реальної витрати палива мікроавтобус на 24 робочі дні. Результати щоденних прогонів показали, що середня витрата палива на 100 км шляху становила 19,6 л за виправленого середнього квадратичного відхилення 1,6 л.

На рівні значущості $\alpha = 0,05$ перевірте, чи варто компанії купувати нові мікроавтобуси.

11.11. За допомогою критеріїв χ^2 і Колмогорова перевірте гіпотезу про нормальний закон розподілу в сукупності розмірів деталей після шліфування, на підставі даних, які наведено в табл. 11.19.

Розподіл розміру деталей після шліфування

Межі інтервалу	Частоти
3,6 – 3,7	1
3,7 – 3,8	22
3,8 – 3,9	40
3,9 – 4,0	79
4,0 – 4,1	27
4,1 – 4,2	26
4,2 – 4,3	4
4,3 – 4,4	1

Рівень значущості $\alpha = 0,02$. Оцінки для параметрів візьміть на підставі вибірових даних.

11.12. За результатами екзамену з навчальної дисципліни "Теорія ймовірностей та математична статистика" здійснено аналіз балів, що набрали здобувачі вищої освіти (табл. 11.20).

Таблиця 11.20

Розподіл балів, набраних здобувачами вищої освіти на екзамені

Бали	0 – 5	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30	30 – 35	35 – 40
Кількість здобувачів вищої освіти	3	5	8	14	20	25	15	10

Використовуючи критерії згоди Пірсона і Колмогорова, перевірте гіпотезу про нормальний закон розподілу для рівня значущості $\alpha = 0,05$.

11.13. Точність роботи верстата-автомата перевіряють за дисперсією контрольованого розміру виробів, яка не має перевищувати 0,15. Вибірковому контролю було піддано 25 виробів і за результатами визначено оцінку дисперсії 25. Припускають, що розмір виробів – нормально розподілена випадкова величина.

Перевірте гіпотезу, що верстат забезпечує потрібну точність.

11.14. Проведено дослідження ваги пакетів із борошном у супермаркеті. Виявилося, що середня вага 100 навмання відібраних пакетів із борошном становила 996 г із середнім квадратичним відхиленням 5 г.

Чи можна з рівнем значущості $\alpha = 0,05$ стверджувати, що середня вага пакетів із борошном є меншою за 1 000 г чи виявлена різниця у вазі є випадковою?

11.15. Дослідження машинобудівної галузі продемонструвало, що середній річний дохід за акціями підприємств машинобудівної галузі становив 15,5 %. Потенційний інвестор перевіряє, чи відповідає ця заява дійсності. Для цього він зробив випадкову вибірку 50 акцій цієї галузі.

У результаті перевірки обчислено середній річний дохід $\bar{x} = 13\%$ та середнє квадратичне відхилення $\sigma = 3,2\%$.

Перевірте з рівнем значущості $\alpha = 0,05$, чи має інвестор достатньо підстав для того, щоб спростувати результати дослідження.

11.5. Тестові завдання

11.1. Статистична гіпотеза – це

А припущення щодо параметрів і виду закону розподілу генеральної сукупності

Б припущення щодо обсягу генеральної сукупності

В припущення щодо параметрів і виду закону розподілу вибірки

Г припущення щодо статистичного критерію

11.2. Рівень значущості α – це

А імовірність того, що відмінність визнано випадковою, а вона насправді є істотною

Б імовірність того, що відмінність визнано істотною, а вона насправді є випадковою

В імовірність того, що відмінність визнано випадковою, а вона насправді є випадковою

11.3. Помилку відхилення нульової гіпотези, тоді як вона є правильною, називають помилкою

А першого роду

Б другого роду

В відповіді немає

11.4. Помилку підтвердження нульової гіпотези, тоді як вона є неправильною, називають помилкою

А першого роду

Б другого роду

В відповіді немає

11.5. Потужністю критерію називають здатність не припуститися помилки

- А** першого роду
- Б** другого роду
- В** відповіді немає

11.6. Наведені гіпотези $H_0 : \theta_1 \geq \theta_2$, $H_1 : \theta_1 < \theta_2$ є

- А** спрямованими
- Б** неспрямованими
- В** основними

11.7. Моделі статистичних висновків щодо різниці середніх значень сукупностей для інтервальних даних, які мають нормальний розподіл і невідомі дисперсії, використовують

- А** t -критерій Стюдента
- Б** F -критерій Фішера – Снедекора
- В** χ^2 -критерій

11.8. Для оцінювання рівня дисперсій сукупності використовують

- А** t -критерій Стюдента
- Б** F -критерій Фішера – Снедекора
- В** χ^2 -критерій

11.9. χ^2 -критерій застосовують для зіставлення

- А** тільки емпіричного розподілу з теоретичним
- Б** декількох емпіричних розподілів між собою
- В** й емпіричного розподілу з теоретичним, і декількох емпіричних розподілів між собою

11.10. У критерії Колмогорова за міру якості погодження емпіричного та теоретичного розподілу беруть

- А** відносну різницю між теоретичною й емпіричною частотами потрапляння випадкової величини в інтервал

Б максимум різниці за модулем між теоретичною й емпіричною частотами потрапляння випадкової величини в інтервал

В середнє квадратичне відхилення між теоретичною й емпіричною частотами потрапляння випадкової величини в інтервал

Г максимум модуля різниці між емпіричною та теоретичною функціями розподілу

Д максимум модуля різниці між емпіричною й теоретичною функціями щільності ймовірності

11.6. Запитання для самоперевірки

11.1. Поясніть, чому висновки про генеральну сукупність, зроблені за результатами дослідження вибірки, вважають статистичними гіпотезами.

11.2. Дайте означення нульової й альтернативної гіпотез. Яка із цих гіпотез підлягає перевірці?

11.3. Наведіть приклади відомих вам статистичних критеріїв і назовіть типи задач, у яких їх застосовують.

11.4. Чи можна, завдяки застосуванню критеріїв згоди, підтвердити нульову гіпотезу?

11.5. Дайте означення критичної області.

11.6. Які бувають критичні області?

11.7. Що таке "критичні точки" та як їх визначають?

11.8. Наведіть загальний алгоритм перевірки нульової гіпотези.

11.9. Що таке "рівень значущості α " і як його пов'язано з надійністю γ ?

11.10. Що таке "помилки першого та другого роду"?

11.11. Що таке "потужність критерію"?

11.12. Які статистичні критерії застосовують для перевірки нульової гіпотези $H_0 : M(X) = M(Y)$?

11.13. У чому полягає нульова гіпотеза в разі перевірки відповідності закону розподілу у вибірковій сукупності припущенням щодо теоретичного закону розподілу випадкової величини?

11.14. Поясніть, у чому полягають відмінності між теоретичними й емпіричними частотами.

11.15. Як обчислюють емпіричне значення критерію згоди Пірсона і за якими параметрами визначають його теоретичне значення?

11.16. У чому полягає нульова гіпотеза в разі перевірки однорідності двох вибірок?

11.17. Який статистичний критерій вибирають для перевірки правильності гіпотези $H_0 : M(X) = M(Y)$?

11.18. Наведіть алгоритм перевірки правильності нульової гіпотези про рівність двох дисперсій.

11.7. Висновки за підрозділом 11

Головна цінність ухвалення статистичних рішень полягає в тому, що в межах імовірнісних категорій можна об'єктивно виміряти ступінь ризику, що відповідає тому чи тому рішення.

Будь-які статистичні висновки, зроблені на підставі опрацювання вибірки, називають статистичними гіпотезами. Гіпотезу, що підлягає перевірці, називають основною (нульовою). Кожній нульовій гіпотезі можна протиставити кілька альтернативних (конкурентних) гіпотез, які заперечують твердження нульової.

Для перевірки правильності висунутої статистичної гіпотези вибирають статистичний критерій, послуговуючись яким відхиляють або не відхиляють нульову гіпотезу.

Схему перевірки статистичних гіпотез можна подати в послідовності таких процедур:

- формулювання нульової та альтернативної гіпотез H_0 , H_1 ;
- вибору рівня значущості α ;
- вибору типу статистичного критерію;
- розрахунків емпіричного значення статистичного критерію;
- визначення критичної області критерію (критичних точок);
- порівняння емпіричного та критичного значень критерію;
- формулювання статистичних висновків.

Література: [1 – 3; 5; 8; 10; 11; 13; 15; 16].

12. Елементи теорії кореляції

12.1. Мета та компетентності

Метою вивчення підрозділу є ознайомлення з елементами теорії кореляції та формування компетентностей щодо застосування кореляційного аналізу для визначення впливу фактора-аргументу на функціональний фактор.

Компетентності, що формують під час вивчення підрозділу:

знання завдань кореляційного аналізу, означення вибіркового коефіцієнта кореляції та кореляційного відношення;

уміння оцінювати вибіркового коефіцієнт кореляції, кореляційне відношення та коефіцієнт рангової кореляції Спірмена;

навички з визначення точкових оцінок основних числових характеристик двовимірної випадкової величини за вибірковими даними.

12.2. Основні теоретичні відомості

Кореляційний аналіз – це метод багатовимірної статистичного аналізу, який полягає в дослідженні коефіцієнтів кореляції між випадковими величинами. Завданнями кореляційного аналізу є визначення наявності, форми та тісноти кореляційного зв'язку, а також перевірка значущості впливу зміни значень однієї випадкової величини на середні значення другої. Для визначення наявності кореляційного зв'язку емпіричні дані заносять до кореляційної таблиці.

Кореляційну таблицю створюють для згрупованих спостережень. Вона становить інтервали зміни величин X і Y та частоти сумісної появи заданої пари значень x та y .

Слід позначити n – загальну кількість спостережень; m_{ki} – частоту сумісної появи двох випадкових величин x_i і y_k , причому $\sum_k \sum_i m_{ki} = n$.

По горизонталі кореляційної таблиці (табл. 12.1) слід розмістити інтервали фактора X , по вертикалі – інтервали функціональної ознаки Y . На перетині стовпця x_i рядка y_k слід занести частоту m_{ki} . У крайній

стовпець та рядок слід записати суму по рядках і стовпцях $m_{x_i} = \sum_{k=1}^s m_{ki}$

і $m_{y_k} = \sum_{i=1}^l m_{ki}$, де $\sum_{i=1}^l m_{x_i} = \sum_{k=1}^s m_{y_k} = n$. Через x_i і y_k позначити середини інтервалів.

Таблица 12.1

Кореляційна таблиця

	Інтервали X	$x_0^* - x_1^*$	$x_1^* - x_2^*$	...	$x_{i-1}^* - x_i^*$	...	$x_{l-1}^* - x_l^*$	
Інтервали Y	x_i y_k	x_1	x_2	...	x_i	...	x_l	m_{y_k}
$y_0^* - y_1^*$	y_1	m_{11}	m_{12}	...	m_{1i}	...	m_{1l}	m_{y_1}
$y_1^* - y_2^*$	y_2	m_{21}	m_{22}	...	m_{2i}	...	m_{2l}	m_{y_2}
...	...	...	...	...	...	...	...	...
$y_{k-1}^* - y_k^*$	y_k	m_{k1}	m_{k2}	...	m_{ki}	...	m_{kl}	m_{y_k}
...	...	...	...	...	...	...	...	...
$y_{s-1}^* - y_s^*$	y_s	m_{s1}	m_{s2}	...	m_{si}	...	m_{sl}	m_{y_s}
m_{x_i}		m_{x_1}	m_{x_2}	...	m_{x_i}	...	m_{x_l}	n

Кореляційною залежністю називають таку залежність між двома випадковими величинами, за якої зі зміною однієї з них змінюється середнє значення другої.

Умовним середнім $\bar{y}_{x_i}$ називають середнє арифметичне значень Y , які відповідають $x = x_i$:

$$\bar{y}_{x=x_i} = \frac{\sum_{k=1}^s y_k m_{ki}}{m_{x_i}}. \quad (12.1)$$

Аналогічно для $Y = y_k$ умовне середнє $\bar{x}_{y=y_k}$ обчислюють за такою формулою:

$$\bar{x}_{y=y_k} = \frac{\sum_{i=1}^l x_i m_{ki}}{m_{y_k}}. \quad (12.2)$$

Кореляційні таблиці дають можливість визначати наявність кореляційної залежності: якщо $\bar{y}_x$ ($\bar{x}_y$) змінюють від стовпця до стовпця (від рядка до рядка), то між величинами X і Y є кореляційний зв'язок.

Кореляційне поле – це графічне зображення кореляційної таблиці. Кореляційним полем називають точкову діаграму, у якій кожна точка становить результат окремого спостереження над двома змінними величинами. Кореляційне поле можна побудувати за кореляційною таблицею. Якщо є лише спостереження, то кореляційне поле створюють разом зі складанням кореляційної таблиці. Записують інтервали величин X та Y і до кожної клітинки на перетині стовпця і рядка заносять точки, кількість яких відповідає величинам m_{ki} (рис. 12.1).

$y_{s-1} - y_s$					• •
...				•	• *
$y_2 - y_3$			• • •		
$y_1 - y_2$		• •	• •		
$y_0 - y_1$	• •				
$y_{k-1} - y_k$	• •				
$x_{l-1} - x_l$	$x_0 - x_1$	$x_1 - x_2$	$x_2 - x_3$	...	$x_{l-1} - x_l$

Рис. 12.1. Кореляційне поле

Якщо на кореляційному полі центри групування зміщено, то можна судити про наявність кореляційного зв'язку між X та Y .

Статистичні висновки про тісноту лінійного взаємозв'язку між величинами X та Y , розподіленими за нормальним законом, формують на основі значення вибіркового коефіцієнта кореляції Пірсона.

Вибірковий коефіцієнт кореляції Пірсона r є статистичною оцінкою коефіцієнта кореляції генеральної сукупності, його обчислюють за такою формулою:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}},$$

де $\bar{x}$, $\bar{y}$ – вибіркові середні випадкових величин X і Y , відповідно.

Формулу застосовують у тому разі, якщо вибіркові дані є незгрупованими. Для даних, поданих у вигляді кореляційної таблиці, вибірковий коефіцієнт кореляції більш зручно оцінювати за такою формулою:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{(\overline{x^2} - (\bar{x})^2) \cdot (\overline{y^2} - (\bar{y})^2)}}, \quad (12.3)$$

де $\overline{xy}$ – середня добутку випадкових величин;

$\overline{x^2}$, $\overline{y^2}$ – вибіркові середні квадратів випадкових величин X і Y .

Величина $\overline{xy} - \bar{x} \cdot \bar{y}$ має назву **коваріація** і є вибірковим *кореляційним моментом* μ_{xy} .

Властивості коефіцієнта кореляції r_{xy} :

- 1) якщо X і Y є незалежними, то $r_{xy} = 0$ (оскільки $\mu_{xy} = 0$);
- 2) якщо випадкові величини є лінійно залежними, то $r_{xy} = 1$;
- 3) $|r_{xy}| \leq 1$, $-1 \leq r \leq 1$;
- 4) $r_{xy} = r_{yx}$ (за визначенням).

За величиною коефіцієнта кореляції дістають інформацію про статистичну значущість кореляційного зв'язку.

Якщо $|r| < 0,35$, то кореляційний зв'язок можна вважати статистично неістотним;

якщо $|r| \geq 0,35$ – зв'язок вважають істотним,

якщо $|r| \geq 0,7$ – тісним.

У разі перевірки статистичної значущості вибіркового коефіцієнта кореляції (за заданим рівнем значущості) нульова гіпотеза полягає в тому, що коефіцієнт кореляції в генеральній сукупності дорівнює нулю, тобто $H_0: \rho = 0$. Альтернативною є гіпотеза $H_1: \rho \neq 0$, отже, їй відповідає двобічна критична область.

Якщо обсяг вибіркової сукупності є невеликим, то як статистику критерію обчислюють таку величину:

$$t = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}}, \quad (12.4)$$

розподілену за статистикою Стюдента з кількістю ступенів свободи $k = n - 2$.

Якщо $|t| \leq t_\alpha(k)$, то нульову гіпотезу немає підстав відхилити, і кореляційний зв'язок між випадковими величинами X і Y є статистично незначущим; якщо $|t| > t_\alpha(k)$, то гіпотезу H_0 відхиляють, звідси $\rho \neq 0$.

За великого обсягу вибірки для перевірки нульової гіпотези як статистику критерію розглядають таку величину:

$$t = \frac{r \cdot \sqrt{n}}{\sqrt{1-r^2}}, \quad (12.5)$$

розподілену за нормальним законом.

Критичне значення t_α обчислюють, за умовою $\Phi(t_\alpha) = (1 - \alpha) / 2$.

Якщо $|t| \leq t_\alpha$, то нульову гіпотезу немає підстав відхилити, і кореляційний зв'язок між випадковими величинами X і Y є статистично незначущим; якщо $|t| > t_\alpha$, то гіпотезу H_0 відхиляють, звідси $\rho \neq 0$.

Також значущість лінійного кореляційного зв'язку для незгрупованих даних можна перевірити за допомогою критерію Фішера, статистика якого має такий вигляд:

$$F_r = \frac{r^2}{1-r^2} \cdot \frac{n-2}{1}.$$

Емпіричне значення критерію F_r порівнюють із критичними точками статистики Фішера – Снедекора (див. додаток Е) за рівнем значущості 0,05 та 0,01 і кількостями ступенів свободи $k_1 = 1$ і $k_2 = n - 2$.

Якщо $F_r < F_{0,05}(1; n - 2)$, то нульову гіпотезу про те, що в генеральній сукупності $\rho = 0$, немає причин відхилити.

Якщо визначають, що $F_r > F_{0,01}(1; n - 2)$, то з надійністю 99 % гіпотезу H_0 відхиляють.

Показником тісноти зв'язку між фактором X і ознакою Y за будь-якою формою залежності є величина $\eta_{y/x}$ – *кореляційне відношення*:

$$\eta_{y/x} = \sqrt{\frac{S_{\bar{y}_x}^2}{S_y^2}}, \quad (12.6)$$

$$\text{де } S_{\bar{y}_x}^2 = \frac{n}{n-1} \cdot \frac{\sum (y_{x_i} - \bar{y})^2 m_{x_i}}{n} = \frac{n}{n-1} \cdot \left(\frac{\sum y_{x_i}^2 m_{x_i}}{n} - (\bar{y})^2 \right).$$

Величина $\eta_{y/x}^2$ характеризує, яку частину повної мінливості функціональної ознаки Y пояснює вплив фактора-аргументу X .

Кореляційні відношення:

$$\eta_{y/x} = \frac{S_{\bar{y}_x}}{S_y}, \quad \eta_{x/y} = \frac{S_{\bar{x}_y}}{S_x} \quad (12.7)$$

є значеннями, визначеними за емпіричними даними.

Величина η характеризує тісноту зв'язку в кореляційній залежності та має такі *властивості*:

- 1) $\eta \geq 0$ (за визначенням);
- 2) $0 \leq \eta \leq 1$;
- 3) якщо $\eta = 1$, то зв'язок між X і Y є функціональним, тобто немає розсіювання всередині групи (для кожного x_i є одне y_i);
- 4) якщо $\eta = 0$, то між X і Y немає кореляційної залежності;
- 5) якщо величина η є ближчою до нуля, то зв'язок між X і Y є слабким; якщо величина η є ближчою до одиниці, то зв'язок між X і Y є тісним;
- 6) якщо $\eta_{y/x} \approx |r|$, то має місце лінійна кореляційна залежність; якщо $\eta \gg r$, то залежність між X і Y є нелінійною.

Значущість кореляційного зв'язку між X і Y перевіряють за критерієм Фішера – Снедекора, статистика якого має такий вигляд:

$$F_{\eta} = \frac{\eta_{y/x}^2}{1 - \eta_{y/x}^2} \cdot \frac{n - k}{k - 1},$$

де k – кількість груп за аргументом X .

Якщо $F_{\eta} < F_{0,05}(k - 1; n - k)$, то гіпотезу $H_0: \eta = 0$ немає підстав спростовувати, тобто кореляційного зв'язку між факторами X і Y теж немає.

Якщо $F_{\eta} > F_{0,01}(k - 1; n - k)$, то нульову гіпотезу спростовують на користь альтернативної $H_1: \eta \neq 0$, тобто кореляційна залежність є значущою.

Для лінійного кореляційного зв'язку коефіцієнт кореляції за модулем дорівнює кореляційному відношенню. Різниця між цими величинами є мірою відхилення кореляційного зв'язку від лінійної форми.

Отже, нульовою є гіпотеза про те, що немає такої розбіжності, тобто $H_0: r^2 = \eta^2$. Відповідно, альтернативною є гіпотеза $H_1: r < \eta$.

Для перевірки адекватності лінійної моделі розглядають F -статистику, тобто:

$$F_A = \frac{(\eta^2 - r^2)/(k - 2)}{(1 - \eta^2)/(n - k)}, \quad (12.8)$$

де k – кількість груп, у які за кореляційною таблицею об'єднано значення зовнішнього фактора X .

Емпіричне значення F -критерію порівнюють зі значеннями, які визначають за таблицею розподілу Фішера – Снедекора (див. додаток Е) для рівня значущості α й кількості ступенів свободи $k_1 = k - 2$ та $k_2 = n - k$.

Якщо $F_A < F_{0,05}(k - 2; n - k)$, то з надійністю 95 % основну гіпотезу немає підстав відхилити. Тобто немає підстав уважати, що кореляційний зв'язок є нелінійним, отже, лінійна модель є адекватною.

Якщо $F_A > F_{0,01}(k - 2; n - k)$, то нульову гіпотезу відхиляють, тобто лінійна модель є неадекватною, і слід шукати іншу форму кореляційного зв'язку.

Отже, під час дослідження кореляційної залежності між ознаками виконують такі завдання: попередній аналіз властивостей досліджуваних сукупностей; установлення факту наявності зв'язку між ознаками, визначення його напрямку та форми; вимірювання степеня тісноти зв'язку між ознаками; побудову регресійної моделі зв'язку, тобто визначення аналітичного виразу зв'язку; оцінювання адекватності моделі, її інтерпретація та рекомендації щодо практичного використання.

Коефіцієнт кореляції Пірсона можна застосовувати тільки в разі, якщо величини X та Y розподілено за нормальним законом. В інших випадках застосовують коефіцієнт рангової кореляції Спірмена.

Вибірковий коефіцієнт рангової кореляції Спірмена r_s , який є статистичною оцінкою коефіцієнта рангової кореляції Спірмена ρ_s у генеральній сукупності, обчислюють за такою формулою:

$$r_s = 1 - \frac{6 \sum (x'_i - y'_i)^2}{n(n^2 - 1)},$$

де x'_i, y'_i – ранги (номери) елементів, відповідно, x_i ($i = \overline{1, n}$) у варіаційному ряду компоненти X та y_i ($i = \overline{1, n}$) у варіаційному ряду компоненти Y .

Властивості вибіркового коефіцієнта рангової кореляції Спірмена:

1) $|r_s| \leq 1$;

2) якщо ранги компонентів X і Y збігаються для всіх $i = \overline{1, n}$, то вибіркового коефіцієнта рангової кореляції Спірмена дорівнює одиниці, тобто має місце повний прямий зв'язок;

3) якщо випадкові величини X і Y є незалежними, то математичне сподівання випадкової величини, якою є статистична оцінка коефіцієнта рангової кореляції Спірмена, дорівнює нулю $M(r_s) = 0$, а її дисперсія

становить $D(r_s) = \frac{1}{n-1}$.

Для перевірки гіпотези про значущість коефіцієнта рангової кореляції ($H_0 : \rho_s = 0$, $H_1 : \rho_s \neq 0$) для вибірок, обсяг яких відповідає умові $4 \leq n \leq 10$, слід обчислити таку величину:

$$t = \frac{r_s \cdot \sqrt{n-2}}{\sqrt{1-r_s^2}}, \quad (12.9)$$

розподілену за статистикою Стюдента з кількістю ступенів свободи $k = n - 2$.

За рівнем значущості α значення цієї випадкової величини порівнюють із критичною точкою $t_\alpha(k)$ двобічної критичної області.

Якщо $|t| < t_\alpha(k)$, то нульову гіпотезу H_0 немає підстав відхилити, отже, на рівні значущості α зв'язок між рангами факторів X і Y є статистично неістотним; якщо $|t| > t_\alpha(k)$, то нульову гіпотезу H_0 відхиляють на користь альтернативної, отже, $\rho_s \neq 0$.

За великих обсягів вибірки нульову гіпотезу щодо незалежності рангів спростовують, якщо

$$|r_s| > \frac{t_\alpha}{\sqrt{n-1}},$$

де t_α є аргументом функції Лапласа, згідно з умовою $\Phi(t_\alpha) = 1 - 0,5\alpha$.

12.3. Приклади розв'язування задач

Приклад 12.1. За даними табл. 12.2, що містить інформацію про продуктивність праці (X) в у. о. та преміальні виплати (Y) у %, складімо кореляційну таблицю.

Таблиця 12.2

Таблиця вихідних даних

x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i
10,2	3,4	21,8	5,1	24,4	5,7	24,5	5,1	27,2	5,3
9,4	3,7	22,9	4,6	21,5	3,9	21,3	4,4	27,1	6,1
13,8	4,1	17,7	5,8	20,9	4,7	29,9	6,1	16,1	5,2
19,1	3,9	23,3	5,6	27,9	6,4	12,8	4,0	20,6	3,6
19,7	4,6	11,3	4,2	25,3	4,9	26,5	5,9	21,7	4,7
7,6	2,6	17,8	5,3	25,5	5,8	17,9	4,2	22,8	6,2
17,1	3,6	26,7	4,7	22,8	5,0	26,6	5,2	22,4	4,9
14,7	4,6	14,7	3,3	7,9	3,9	19,4	6,1	32,4	6,3
20,6	5,0	22,2	4,9	16,8	4,4	22,5	4,4	20,8	5,3
20,6	4,1	13,1	5,1	19,6	4,7	18,1	4,0	25,3	5,4

Розв'язання. За умовою прикладу X є зовнішнім (незалежним) фактором, а Y – внутрішнім (функціональним) фактором. Для побудови кореляційної таблиці визначаймо інтервали для кожного з них.

Наприклад, для випадкової величини X буде: $x_{\max} = 32,4$; $x_{\min} = 7,6$

Тоді розмах становить $R_X = 32,4 - 7,6 = 24,8$.

Візьмімо 5 інтервалів, тоді крок буде таким:

$$h_X = \frac{R_X}{k_X} \approx \frac{25}{5} = 5.$$

Доцільно взяти довжину кроку $h_X = 5$.

Початок першого інтервалу виберімо таким способом, щоб за такої довжини кроку серединам інтервалів відповідали цілі числа (це зручно для подальших обчислень).

Оскільки початок першого інтервалу в загальному випадку можна зсунути щодо x_{min} у бік менших значень на пів кроку, то виберімо нижньою межею першого інтервалу $x_1 = 7,5$.

Тоді $x_2 = 12,5$; $x_3 = 17,5$; ...; $x_6 = 32,5$.

Для випадкової величини Y буде: $y_{max} = 6,4$; $y_{min} = 2,6$, відповідно, розмах становить $R_Y = 6,4 - 2,6 = 3,8$.

Розподілімо цей розмах на 4 інтервали, тоді довжина кожного інтервалу дорівнює:

$$h_Y \approx \frac{4}{4} = 1.$$

Для того щоб мати цілі значення середини інтервалів, виберімо $y_1 = 2,5$.

Тоді $y_2 = 3,5$; $y_3 = 4,5$; $y_4 = 5,5$; $y_5 = 6,5$.

Якщо розмістити інтервали X по горизонталі, а Y – по вертикалі, то буде складено таблицю, яку розподілено на клітинки (табл. 12.3).

Таблиця 12.3

Кореляційна таблиця

$Y \backslash X$	[7,5; 12,5)	[12,5; 17,5)	[17,5; 22,5)	[22,5; 27,5)	[27,5; 32,5)	m_y
[2,5; 3,5)	2					2
[3,5; 4,5)	3	5	8			16
[4,5; 5,5)		3	10	8		21
[5,5; 6,5)			2	6	3	11
m_x	5	8	20	14	3	50

Тепер для кожної клітинки цієї таблиці обчислімо кількість значень двовимірної випадкової величини $(X; Y)$, які одночасно відповідають умовам $x_i \leq X < x_{i+1}$ та $y_j \leq Y < y_{j+1}$, де $i = \overline{1,5}$; $j = \overline{1,4}$.

Отже, для кожної із клітинок таблиці визначмо частоту m_{ij} потрапляння до неї значень двовимірної випадкової величини, що досліджують. Наприклад, першій парі значень $(10,2; 3,4)$ із табл. 12.2 відповідає клітинка табл. 12.3, що стоїть на перетині першого стовпця і першого рядка; парі $(9,4; 3,7)$ – клітинка на перетині першого стовпця і другого рядка тощо.

Остаточню буде складено кореляційну таблицю (див. табл. 12.3), із якої видно, що частоти утворюють *хмару розсіювання* у формі еліпса, більшу вісь якого спрямовано вздовж головної діагоналі таблиці.

Доцільно припустити, що між факторами X і Y існує додатний кореляційний зв'язок.

Приклад 12.2. Задано кореляційну таблицю, у якій наведено результати іспитів із фізики (y) та математики (x) в одній з академічних груп I курсу (табл. 12.4).

Оцінімо тісноту зв'язку між результатами іспитів із фізики та математики.

Таблиця 12.4

Кореляційна таблиця результатів іспитів із фізики та математики

$y \backslash x$	2	3	4	5	l_k
2	1	2	–	–	3
3	1	6	3	–	10
4	–	5	5	–	10
5	–	–	1	1	2
h_i	2	13	9	1	25

Розв'язання. Обчислимо:

$$\bar{x} = \frac{2 \cdot 2 + 3 \cdot 13 + 4 \cdot 9 + 5 \cdot 1}{25} = \frac{84}{25} = 3,36;$$

$$\bar{y} = \frac{2 \cdot 3 + 3 \cdot 10 + 4 \cdot 10 + 5 \cdot 2}{25} = \frac{86}{25} = 3,44;$$

$$\overline{x^2} = \frac{4 \cdot 2 + 9 \cdot 13 + 16 \cdot 9 + 25 \cdot 1}{25} = \frac{294}{25} = 11,76;$$

$$\overline{y^2} = \frac{4 \cdot 3 + 9 \cdot 10 + 16 \cdot 10 + 25 \cdot 2}{25} = \frac{312}{25} = 12,48;$$

$$\begin{aligned} \overline{xy} &= \frac{\sum_i \sum_k m_{ki} x_i y_k}{n} = \frac{1}{25} ((1 \cdot (2 \cdot 2) + 2 \cdot (3 \cdot 2)) + (1 \cdot (2 \cdot 3) + 6 \cdot (3 \cdot 3) + 3 \cdot (4 \cdot 3)) + \\ &+ (5 \cdot (3 \cdot 4) + 5 \cdot (4 \cdot 4)) + (1 \cdot (4 \cdot 5) + 1 \cdot (5 \cdot 5))) = \frac{1}{25} (16 + 96 + 140 + 45) = \frac{297}{25} = 11,88. \end{aligned}$$

За ф-лою (12.3):

$$r = \frac{11,88 - 3,36 \cdot 3,44}{\sqrt{11,76 - 3,36^2} \cdot \sqrt{12,48 - 3,44^2}} \approx 0,58.$$

Оскільки $0,35 < r < 0,7$, то кореляційний зв'язок між результатами іспитів із фізики та математики є статистично значущим, але не дуже тісним.

Приклад 12.3. У припущенні лінійного кореляційного зв'язку обчислимо коефіцієнт кореляції, якщо за результатами дослідження визначено таке:

$$\sum x = 35; \quad \sum y = 136,0; \quad \sum xy = 2\,806,5; \quad \sum x^2 = 2\,038;$$

$$\sum y^2 = 36\,804,4; \quad n = 50.$$

Розв'язання. Коефіцієнт кореляції обчислімо за ф-лою (12.3).
Для цього визначаймо усі величини, що входять до цієї формули.

Маємо таке:

$$\bar{x} = \frac{\sum x}{n} = \frac{35}{50} = 0,7; \quad \bar{y} = \frac{\sum y}{n} = \frac{136,0}{50} = 2,72; \quad \overline{xy} = \frac{\sum xy}{n} = \frac{2\,806,5}{50} = 56,13;$$

$$\overline{x^2} = \frac{\sum x^2}{n} = \frac{2\,038}{50} = 40,76; \quad \overline{y^2} = \frac{\sum y^2}{n} = \frac{36\,804,4}{50} = 736,088.$$

Далі обчислімо такі числові характеристики:
вибірковий кореляційний момент:

$$\mu_{xy} = \overline{xy} - \bar{x} \cdot \bar{y} = 56,13 - 0,7 \cdot 2,72 = 54,226;$$

вибіркові середні квадратичні відхилення для обох факторів:

$$\sigma_x^* = \sqrt{\overline{x^2} - (\bar{x})^2} = \sqrt{40,76 - 0,7^2} = \sqrt{40,27};$$

$$\sigma_y^* = \sqrt{\overline{y^2} - (\bar{y})^2} = \sqrt{736,088 - 2,72^2} = \sqrt{728,69}.$$

Тоді

$$r = \frac{54,226}{\sqrt{40,27} \cdot \sqrt{728,69}} = 0,317.$$

Оскільки $|r| < 0,35$, то кореляційний зв'язок є статистично незначущим.

Приклад 12.4. Дані вибіркової сукупності наведено в табл. 12.5.

Таблиця 12.5

Вихідна вибірка сукупність

$X = x_i$	1	2	4	6	7	9	10	12	16
$Y = y_i$	10,7	12,9	14,9	19,3	22,9	23,2	26,1	31,7	35,3

Обчислімо коефіцієнт кореляції та за критерієм Стюдента, перевірмо його значущість із надійністю 95 %.

Розв'язання. Коефіцієнт кореляції обчислюймо за ф-лою (12.3).

Допоміжні обчислення наведено в табл. 12.6.

Таблица 12.6

Таблица допоміжних обчислень

№ з/п	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	1	10,7	10,7	1	114,49
2	2	12,9	25,8	4	166,41
3	4	14,9	59,6	16	222,01
4	6	19,3	115,8	36	372,49
5	7	22,9	160,3	49	524,41
6	9	23,2	208,8	81	538,24
7	10	26,1	261,0	100	681,21
8	12	31,7	380,4	144	1004,89
9	16	35,3	564,8	256	1246,09
Сума	67	197,0	1787,2	687	4870,24

За даними рядка "Сума" табл. 12.6, обчислімо такі середні значення:

вибіркову середня випадкової величини X :

$$\bar{x} = \frac{1}{9} \cdot 67 = 7,444.$$

вибіркову середня випадкової величини Y :

$$\bar{y} = \frac{197}{9} = 21,889;$$

вибіркову середня добутку випадкових величин:

$$\overline{xy} = \frac{1787,2}{9} = 198,578;$$

середні квадратів випадкових величин:

$$\overline{x^2} = \frac{687}{9} = 76,333;$$

$$\overline{y^2} = \frac{4\,870,24}{9} = 541,138.$$

Далі обчислимо вибіркові середні квадратичні відхилення:

$$\sigma_x^2 = 76,333 - 7,444^2 = 20,920; \sigma_x^* = \sqrt{20,920} = 4,574$$

та

$$\sigma_y^2 = 541,138 - 21,889^2 = 62,010; \sigma_y^* = \sqrt{62,010} = 7,875.$$

За ф-лою (12.3) обчислимо значення коефіцієнта кореляції:

$$r = \frac{198,578 - 7,444 \cdot 21,889}{4,574 \cdot 7,875} = 0,989.$$

$r > 0,7$, що свідчить про тісний зв'язок між величинами X та Y .

Перевірмо статистичну значущість коефіцієнта кореляції. Згідно з нульовою гіпотезою $H_0: r = 0$. Альтернативною є гіпотеза $H_1: r \neq 0$, тобто критична область є двобічною.

Оскільки обсяг вибірки є невеликим, то для перевірки нульової гіпотези обчислимо емпіричне значення t -критерію за ф-лою (12.4):

$$t = \frac{0,989 \cdot \sqrt{9-2}}{\sqrt{1-0,989^2}} = \frac{0,989 \cdot 2,6458}{\sqrt{0,021879}} = 17,69.$$

За таблицю значень критичних точок розподілу Стюдента (див. додаток Д), за рівнем значущості $\alpha = 0,05$ та кількістю ступенів свободи $k = n - 2$, визначмо $t_{0,05}(7) = 2,36$.

Оскільки $|t| > t_{0,05}(7)$, то нульову гіпотезу слід відхилити.

Отже, із надійністю 95 % кореляційний зв'язок між факторами є статистично значущим.

Приклад 12.5. Визначмо кореляційне відношення $\eta_{y/x}$ та пояснімо результат, якщо $\bar{y} = 33,5$, $S_y^2 = 98,75$ та умовні середні фактора Y для сталого значення фактора X наведено в табл. 12.7.

Таблиця 12.7

Умовні середні фактора Y

$\bar{y}(X = x_i)$	11,00	18,14	30,00	33,80	42,05	51,00
m_{x_i}	4	7	10	57	19	3

Розв'язання. Кореляційне відношення обчислімо за ф-лою (12.6).
Спочатку маємо таке:

$$S_{\bar{y}_x}^2 = \frac{1}{99} ((11 - 33,5)^2 \cdot 4 + (18,14 - 33,5)^2 \cdot 7 + (30 - 33,5)^2 \cdot 10 + (33,8 - 33,5)^2 \cdot 57 + (42,05 - 33,5)^2 \cdot 19 + (51 - 33,5)^2 \cdot 3) = 61,74.$$

Тоді

$$\eta_{y/x} = \sqrt{\frac{61,74}{98,75}} = \sqrt{0,6252} = 0,791,$$

звідси буде:

$$\eta_{y/x}^2 = 0,6252.$$

Це означає, що 62,52 % мінливості фактора Y можна пояснити наявністю кореляційного зв'язку між Y та X , а решту 31,48 % пов'язано із впливом інших факторів, не врахованих цією моделлю.

Приклад 12.6. Перевірмо гіпотезу щодо лінійної форми кореляційного зв'язку між факторами Y та X , за рівнем значущості $\alpha = 0,05$, якщо для вибіркової сукупності обсягом $n = 100$ коефіцієнт кореляції – $r_{xy} = 0,526$, кореляційне відношення – $\eta_{y/x} = 0,625$, кількість груп за фактором X – $k = 15$.

Розв'язання. За нульовою гіпотезою $H_0: r^2 = \eta^2$, альтернативною є гіпотеза $H_1: r < \eta$.

Для перевірки припущення про лінійну форму зв'язку визначте емпіричне значення F -критерію за ф-лою (12.8):

$$F_A = \frac{(\eta^2 - r^2)/(k-2)}{(1-\eta^2)/(n-k)}, \text{ тобто } F_A = \frac{0,625^2 - 0,526^2}{1 - 0,625^2} \cdot \frac{100 - 15}{15 - 2} = 1,22$$

і порівняймо його із критичним.

Відповідне критичне значення за розподілом Фішера – Снедекора, становить $F_{0,05}(13,85) = 1,86$ (див. додаток Е).

Оскільки $F_A < F_{0,05}$, то, із надійністю 95 % лінійна модель є адекватною, тобто відхилення зв'язку від лінійного є статистично несуттєвим.

Отже, із надійністю 95 % кореляційний зв'язок між факторами можна вважати лінійним.

Приклад 12.7. Бали тестування десяти здобувачів вищої освіти з вищої математики (A) та інформатики (B) за стобальною шкалою наведено в табл. 12.8.

Таблиця 12.8

Бали здобувачів вищої освіти із вищої математики та інформатики

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

Визначмо вибірковий коефіцієнт рангової кореляції Спірмена між балами за двома тестами. За рівнем значущості 0,01, перевірмо нульову гіпотезу про те, що генеральний коефіцієнт рангової кореляції Спірмена дорівнює нулю.

Розв'язання. Надаймо ранги x_i балам із вищої математики A .

Ці бали розташовано в спадному порядку, тому їхні ранги x_i дорівнюють порядковим номерам:

Ранги x_i	1	2	3	4	5	6	7	8	9	10
Оцінки A	95	90	86	84	75	70	62	60	57	50

(*)

Надаймо ранги y_i балам з інформатики B , для чого спочатку розташуймо ці бали у спадному порядку та пронумеруймо їх:

1	2	3	4	5	6	7	8	9	10
93	92	83	80	72	70	62	60	55	45

(**)

Визначмо ранг y_1 .

Індекс $i = 1$ указує, що розглядають бал, який посідає в ряду (*) перше місце (цей бал дорівнює 95);

з умови видно, що з інформатики бал 92 в ряду (**) розташовано на другому місці.

Отже, ранг $y_1 = 2$.

Визначмо ранг y_2 .

Індекс $i = 2$ вказує, що розглядають бал, який посідає в ряду (*) друге місце;

з умови видно, бал 93 в ряду (**) розташовано на першому місці.

Отже, ранг $y_2 = 1$.

Аналогічно визначмо інші ранги:

$$y_3 = 3, y_4 = 4, y_5 = 9, y_6 = 8, y_7 = 10, y_8 = 5, y_9 = 7, y_{10} = 6.$$

Записуймо послідовності рангів x_i та y_i :

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

(***)

Обчислюймо різниці рангів: $d_1 = x_1 - y_1 = 1 - 2 = -1$.

Аналогічно, $d_2 = 1$, $d_3 = 0$, $d_4 = 0$, $d_5 = -4$, $d_6 = -2$, $d_7 = -3$, $d_8 = 3$, $d_9 = 2$, $d_{10} = 4$.

Тоді сума квадратів різниць рангів є такою:

$$\sum d_i^2 = 1 + 1 + 16 + 4 + 9 + 9 + 4 + 16 = 60.$$

Визначмо коефіцієнт рангової кореляції Спірмена, зважаючи на те, що $n = 10$:

$$r_s = 1 - \frac{6 \sum d_i^2}{n^3 - n} = 1 - \frac{6 \cdot 60}{10^3 - 10} = 0,64.$$

Отже, $r_s = 0,64$.

Перевірмо, чи є значущою рангова кореляційна залежність між балами із двох тестів.

Визначмо критичну точку двобічної критичної області розподілу Стюдента, за рівнем значущості $\alpha = 0,01$ та кількістю ступенів свободи $k = n - 2 = 10 - 2 = 8$ (див. додаток Д): $t_{кр.}(0,01; 8) = 3,36$.

Обчислімо значення статистики критерію:

$$t = \frac{r_s \cdot \sqrt{n-2}}{\sqrt{1-r_s^2}} = \frac{0,64 \cdot \sqrt{8}}{\sqrt{1-0,64^2}} = 2,36.$$

Оскільки $t < t_{кр.}$, то немає підстав відхилити нульову гіпотезу про те, що генеральний коефіцієнт рангової кореляції Спірмена дорівнює нулю.

Отже, ранговий кореляційний зв'язок між балами із двох тестів є незначущим, тобто бали є незалежними між собою.

12.4. Вправи для самостійної роботи

12.1. Залежність швидкості руху людини Y (у. о.) від віку X (роки) наведено статистичним розподілом вибірки (табл. 12.9).

Таблица 12.9

Залежність швидкості руху від віку

Y	4	3	4	5	4	5	6	5	7	8	6	8	9	8	11	8	7	6	4	3
X	10	10	10	10	20	20	20	30	30	30	40	40	50	50	60	60	60	70	70	80

Побудуйте кореляційну таблицю та кореляційне поле залежності Y від X ; обчисліть коефіцієнт кореляції та поясніть його значення.

12.2. Залежність товарообігу Y (у. о.) від суми витрат X (у. о.) в магазині наведено статистичним розподілом вибірки (табл. 12.10).

Таблиця 12.10

Залежність товарообігу від суми витрат у магазині

Y	480	510	530	540	555	564	570	575	580	585	590	596	605	618	625	635	640	650
X	30	25	31	32	38	41	40	46	49	54	58	60	64	75	78	82	83	85

Побудуйте кореляційну таблицю та кореляційне поле залежності Y від X ; обчисліть коефіцієнт кореляції та поясніть його значення.

12.3. За статистичними даними залежності обсягу реалізованої продукції Y (тис. у. о.) від витрат підприємства на інновації X (тис. у. о.), що наведено в табл. 12.11, побудуйте кореляційне поле, оцініть тісноту зв'язку між факторами X та Y .

Таблиця 12.11

Статистичні дані

Y	2,1	3,3	1,4	1,8	1,9	2,4	2,7	2,0	3,1	2,6	3,0
X	1,2	1,5	0,7	0,7	0,8	1,4	1,8	1,1	1,3	1,2	1,6

12.4. Сумісний розподіл неперервних випадкових величин X, Y задано кореляційною таблицею (табл. 12.12).

Таблиця 12.12

Розподіл неперервних випадкових величин

$Y \backslash X$	0 – 2	2 – 4	4 – 6
8 – 10	1	5	–
10 – 12	1	7	–
12 – 14	–	3	3

Обчисліть коефіцієнт кореляції та поясніть його значення.

12.5. Обчисліть коефіцієнт кореляції Пірсона, якщо $\sum x = 324$, $\sum y = 1975$, $\sum xy = 4197$, $\sum x^2 = 2324$, $\sum y^2 = 54834$, $n = 100$.

12.6. За даними обчислень, визначте кореляційне відношення, якщо $\bar{y} = 0,36$; $\sum \bar{y}_x^2 = 136,94$; $S_y^2 = 1,985$; $n = 100$.

12.7. Сумісний розподіл неперервних випадкових величин X – ціни автомобіля (тис. у. о.) та Y – пробігу (тис. км) задано кореляційною таблицею (табл. 12.13).

Таблиця 12.13

Розподіл неперервних випадкових величин

$Y \backslash X$	1 – 3	3 – 5	5 – 7	7 – 9
0 – 25	–	–	2	5
25 – 50	–	3	7	–
50 – 75	–	5	7	–
75 – 100	1	2	–	–

Обчисліть коефіцієнт кореляції, поясніть його значення та оцініть значущість за $\alpha = 0,01$.

12.8. Залежність між собівартістю X та продуктивністю Y наведено в табл. 12.14.

Таблиця 12.14

Залежність між собівартістю та продуктивністю

$Y = y_i$, тис. грн	2,2	3,5	3,7	3,8	4,5	5,7
$X = x_i$, тис. шт.	1,5	1,4	1,2	1,1	0,9	0,8

Обчисліть коефіцієнт кореляції та (за критерієм Стюдента) перевірте його значущість із надійністю 95 %.

12.9. Визначте значення кореляційного відношення $\eta_{y/x}$ та поясніть його, якщо $\bar{y} = 37,8$, $S_y^2 = 102,63$ і задано умовні середні (табл. 12.15).

Таблиця 12.15

Вихідні умовні середні

$\bar{y} (X = x_i)$	17,9	22,9	35,1	41,1	46,9	59,0
m_{x_i}	3	15	19	47	13	3

12.10. Перевірте адекватність лінійної кореляційної моделі, якщо за вибірковими даними $\eta_{y/x} = 0,618$; $r = 0,503$; $n = 60$, а кількість груп за зовнішнім фактором X становить $k = 6$.

12.11. Перевірте адекватність лінійної кореляційної моделі, якщо $r = -0,386$; $S_y^2 = 270,1$ та задано умовні середні (табл. 12.16).

Таблиця 12.16

Вихідні умовні середні

$\bar{y} (X = x_i)$	128,0	133,7	125,1	121,3	115,1	114,1	135,2
m_{x_i}	3	29	12	33	34	15	4

12.12. Вимірювання довжини руки X (см) і кількості влучань м'яча в кільце Y (разів) у 16 учасників спортивної секції з баскетболу дали результати, наведені в табл. 12.17.

Таблиця 12.17

Залежність між кількістю влучань та довжиною руки

X	66	61	67	73	51	59	48	47	58	44	41	54	52	47	51
Y	38	31	36	43	29	33	28	25	36	26	21	30	20	27	28

Визначте вибірковий коефіцієнт рангової кореляції Спірмена між цими показниками та перевірте його значущість за рівнем значущості $\alpha = 0,05$.

12.13. Два викладачі оцінили знання 12 здобувачів вищої освіти за стобальною шкалою та поставили їм бали, наведені в табл. 12.18. У першому рядку вказано кількість балів, поставлених першим викладачем, а в другому – другим.

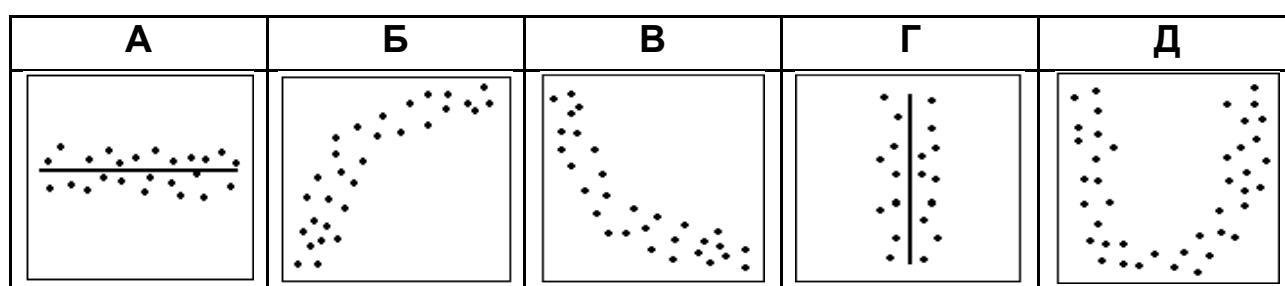
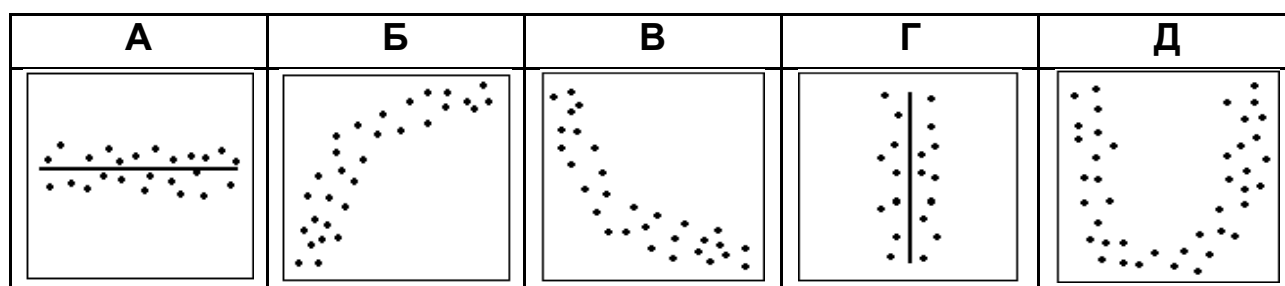
Оцінювання знань першим і другим викладачами

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

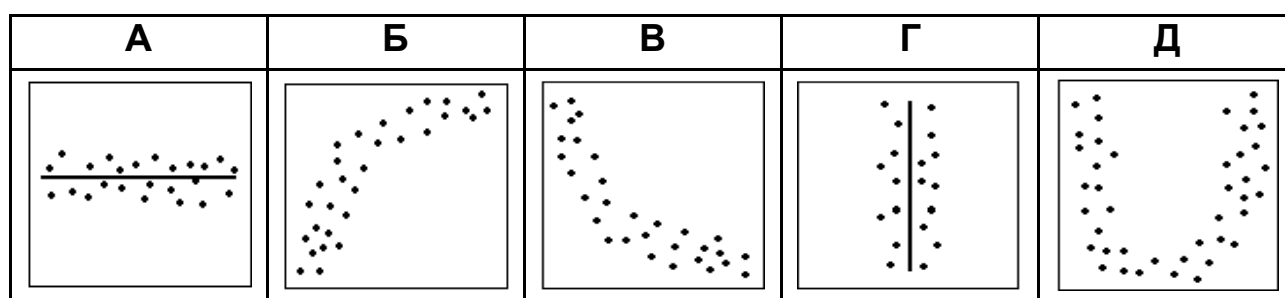
Визначте вибіркового коефіцієнт рангової кореляції Спірмена між балами двох викладачів. За рівнем значущості $\alpha = 0,05$, перевірте нульову гіпотезу про те, що генеральний коефіцієнт рангової кореляції Спірмена дорівнює нулю.

12.5. Тестові завдання

12.1. Форма зв'язку є тільки спадною в разі

12.2. Кореляційного зв'язку $\bar{Y}$ від X немає в разі

12.3. Коефіцієнт кореляції набуває додатного значення в разі



12.4. Вибірковий коефіцієнт кореляції оцінюють за такою формулою

A
$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\left(\overline{x^2} - (\bar{x})^2\right) \cdot \left(\overline{y^2} - (\bar{y})^2\right)}}$$

Б
$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\left(\bar{x} - x^2\right)^2 \cdot \left(\bar{y} - y^2\right)^2}}$$

В
$$r = \frac{\bar{x} \cdot \bar{y} - \overline{xy}}{\sqrt{\left(\overline{x^2} - (\bar{x})^2\right) \cdot \left(\overline{y^2} - (\bar{y})^2\right)}}$$

12.5. Якщо випадкові величини X і Y є незалежними, то

A $r_{xy} = 0$

Б $r_{xy} = 1$

В $-1 \leq r \leq 1$

12.6. Якщо випадкові величини X і Y є лінійно залежними, то

A $r_{xy} = 0$

Б $r_{xy} = 1$

В $-1 \leq r \leq 1$

12.7. Якщо коефіцієнт кореляції дорівнює приблизно 0,68, то

A можна визнати факт наявності прямого кореляційного зв'язку

Б потрібно перевірити значущість коефіцієнта кореляції та після цього формулювати статистичні висновки

В є тісний кореляційний зв'язок

12.8. Якщо $\eta = 1$, то

A зв'язок між X і Y є функціональним

Б між X і Y немає кореляційної залежності

В відповіді немає

12.9. Якщо $\eta = 0$, то

А зв'язок між X і Y є функціональним

Б між X і Y немає кореляційної залежності

В відповіді немає

12.6. Запитання для самоперевірки

12.1. Розкрийте зміст кореляційного аналізу та його завдання.

12.2. Що таке "статистична залежність", у чому полягають її особливості, порівняно з функціональною залежністю?

12.3. Як побудувати кореляційне поле? Що таке "хмара розсіювання"?

12.4. Які висновки про вид кореляційного зв'язку можна зробити за формою цієї хмари?

12.5. Що таке "кореляційна залежність"?

12.6. Який вигляд має кореляційна таблиця? Наведіть способи її побудови.

12.7. Як визначають вибіркового коефіцієнт кореляції Пірсона та які він має властивості?

12.8. Які є критерії для визначення значущості кореляційного зв'язку за значенням вибіркового коефіцієнта кореляції Пірсона?

12.9. Що таке "кореляційне відношення" і як його визначають?

12.10. Як перевірити статистичну значущість кореляційного відношення?

12.11. За яким критерієм здійснюють оцінювання адекватності лінійної моделі?

12.12. Наведіть формулу для обчислення коефіцієнта рангової кореляції Спірмена.

12.13. Як перевірити статистичну значущість коефіцієнта рангової кореляції Спірмена?

12.7. Висновки за підрозділом 12

Кореляційний аналіз – це метод багатовимірного статистичного аналізу, який полягає в дослідженні коефіцієнтів кореляції між випадковими величинами. Завданнями кореляційного аналізу є визначення форми та тісноти кореляційного зв'язку, а також перевірка значущості впливу зміни значень однієї випадкової величини на середні значення іншої.

Література: [1 – 3; 5; 8; 10 – 13; 15 – 17].

13. Елементи дисперсійного аналізу

13.1. Мета та компетентності

Метою вивчення підрозділу є ознайомлення з основними принципами й особливостями дисперсійного аналізу та формування компетентностей щодо застосування дисперсійного аналізу до дослідження економічних процесів.

Компетентності, що формують під час вивчення підрозділу:

розуміння змісту дисперсійного аналізу;

знання можливостей дисперсійного аналізу щодо визначення тісноти статистичного зв'язку;

уміння визначати тісноту статистичного зв'язку між компонентами двовимірної випадкової величини за допомогою дисперсійного аналізу.

13.2. Основні теоретичні відомості

Дисперсійний аналіз призначено для перевірки гіпотез про зв'язок між деякою ознакою та факторами, які не мають кількісного опису, установлення ступеня впливу факторів на цю ознаку та їхньої взаємодії.

Сутність **дисперсійного аналізу** полягає в тому, що загальну дисперсію досліджуваної ознаки розподіляють на окремі компоненти, зумовлені впливом певних факторів. Істотність їхнього впливу на цю ознаку визначають методом дисперсійного аналізу.

У разі здійснення дисперсійного аналізу досліджуваний масив даних, визначених під час експерименту, розподіляють на певні групи, які відрізняє дія на результати експерименту певних рівнів факторів.

Уважають, що досліджувана ознака має нормальний закон розподілу, а дисперсії в кожній окремій групі здобутих значень ознаки є однаковими. Ці припущення потрібно перевірити.

Дисперсійний аналіз, який досліджує наявність зв'язку між компонентами двовимірної випадкової величини, називають **однофакторним**, тобто він розглядає вплив на ознаку лише одного фактора.

У процесі однофакторного дисперсійного аналізу масив емпіричних даних становлять значення однієї з компонент двовимірної випадкової величини, яку називають **ознакою**. Ці значення розподіляють на групи за рівнем дії фактора, що впливає на ознаку, – це друга компонента двовимірної випадкової величини.

Результати експерименту зручно подавати у вигляді таблиці (табл. 13.1).

Таблиця 13.1

Таблиця вихідних даних однофакторного дисперсійного аналізу

Рівні фактора (номери груп)	Емпіричні значення якісної ознаки X за групами	Групові середні значення	Загальні (вибіркові) середні ознаки X
1	$x_{11}, x_{12}, x_{13}, \dots, x_{1n_1}$	$\bar{x}_1 = \frac{1}{n_1} \sum_{j=1}^{n_1} x_{1j}$	$\bar{x} = \frac{\sum_{i=1}^p \sum_{j=1}^{n_i} x_{ij}}{\sum_{i=1}^p n_i}$
2	$x_{21}, x_{22}, x_{23}, \dots, x_{2n_2}$	$\bar{x}_2 = \frac{1}{n_2} \sum_{j=1}^{n_2} x_{2j}$	$\sum_{i=1}^p n_i = N$
...	...	...	
p	$x_{p1}, x_{p2}, x_{p3}, \dots, x_{pn_p}$	$\bar{x}_p = \frac{1}{n_p} \sum_{j=1}^{n_p} x_{pj}$	

Згідно з положеннями дисперсійного аналізу, *загальну суму квадратів відхилень* значень випадкової величини X від її вибіркової середньої, що характеризує загальне розсіювання значень випадкової величини, подають у вигляді суми двох компонентів:

$$SST = SSR + SSE, \quad (13.1)$$

де SST – загальна сума квадратів відхилень значень випадкової величини від її вибіркової середньої (*sum of squares total*);

SSR – сума квадратів відхилень, пов'язаних із регресією (факторна), тобто із впливом досліджуваного фактора (*sum of squares by regression*);

SSE – сума квадратів помилок (залишкова), тобто відхилень, пов'язаних із факторами, які не розглядають у межах заданої моделі та які вважають похибками (*sum of squares by errors*).

Загальну суму квадратів відхилень значень випадкової величини від її вибіркової середньої визначають за такою формулою:

$$SST = \sum_{i=1}^p \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2, \quad (13.2)$$

де $\bar{x}$ – загальна (вибіркова) середня, тобто середня вибіркової сукупності випадкових величин $X_1, X_2, \dots, X_p$ обсягом $N = \sum_{i=1}^p n_i$.

Суму квадратів відхилень, пов'язаних із дією досліджуваного фактора, визначають як виважену суму квадратів відхилень вибіркової середньої кожної з випадкових величин X_i ($i = \overline{1, p}$) від загальної середньої ознаки $\bar{X}$:

$$SSR = \sum_{i=1}^p n_i \cdot (\bar{x}_i - \bar{x})^2, \quad (13.3)$$

де $\bar{x}_i$ – вибіркова середня випадкової величини X_i ($i = \overline{1, p}$).

Суму квадратів помилок визначають за такою формулою:

$$SSE = \sum_{i=1}^p \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2. \quad (13.4)$$

Для перевірки статистичної гіпотези щодо однорідності всієї вибіркової сукупності, тобто для перевірки нульової гіпотези H_0 , розглядають таку випадкову величину:

$$F = \frac{SSR/(p-1)}{SSE/(N-p)}, \quad (13.5)$$

що має розподіл Фішера – Снедекора.

За значеннями α , $k_1 = p-1$, $k_2 = N-p$ визначають критичну точку $F_\alpha(k_1; k_2)$ (див. додаток Е).

Якщо емпіричне значення критерію, що обчислюють за ф-лою (13.5), не перевищує критичне значення $F_{\alpha}(k_1; k_2)$, то нульову гіпотезу про рівність групових середніх немає підстав відхилити, отже, вплив зовнішнього фактора на ознаку X слід уважати статистично незначущим.

Якщо емпіричне значення критерію перевищує критичне, то з надійністю $\gamma = 1 - \alpha$ нульову гіпотезу відхиляють, тобто підтверджують наявність впливу певного зовнішнього фактора на ознаку X .

Результати однофакторного дисперсійного аналізу зручно подавати у вигляді таблиці (табл. 13.2).

Таблиця 13.2

Форма подання результатів однофакторного дисперсійного аналізу

Види варіацій ознаки	Суми квадратів відхилень	Кількість ступенів свободи	Статистичні оцінки дисперсії	Значення критерію Фішера – Снедекора	
				емпіричне	критичне
Міжгрупова	$\sum_{i=1}^p n_i \cdot (\bar{x}_i - \bar{x})^2$	$p - 1$	$\frac{SSR}{p - 1}$	F	$F_{\alpha}(k_1; k_2)$
Внутрішньогрупова	$\sum_{i=1}^p \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$	$N - p$	$\frac{SSE}{N - p}$		
Загальна	$\sum_{i=1}^p \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$	$N - 1$	$\frac{SST}{N - 1}$		

Нехай потрібно дослідити наявність впливу факторів A і B на ознаку X .

Для цього застосовують *двофакторний дисперсійний аналіз*. Розв'язання цієї задачі передбачає виконання певного обсягу обчислень. Результати вимірювань та обчислень двофакторного аналізу зручно подавати в таблицях (табл. 13.3 і 13.4).

Середні значення двофакторного дисперсійного аналізу

Рівні фактора A	Рівні фактора B				Середні величини за рядками	Загальна середня величина
	B_1	B_2	...	B_q		
	блокові середні	блокові середні	...	блокові середні		
A_1	$\bar{x}_{11} = \frac{\sum_{i=1}^n x_{i11}}{n}$	$\bar{x}_{12} = \frac{\sum_{i=1}^n x_{i12}}{n}$	...	$\bar{x}_{1q} = \frac{\sum_{i=1}^n x_{i1q}}{n}$	$\bar{y}_1 = \frac{\sum_{i=1}^n \sum_{k=1}^q x_{i1k}}{nq}$	$\bar{x} = \frac{\sum_{i=1}^n \sum_{j=1}^p \sum_{k=1}^q x_{ikj}}{npq}$
A_2	$\bar{x}_{21} = \frac{\sum_{i=1}^n x_{i21}}{n}$	$\bar{x}_{22} = \frac{\sum_{i=1}^n x_{i22}}{n}$	...	$\bar{x}_{2q} = \frac{\sum_{i=1}^n x_{i2q}}{n}$	$\bar{y}_2 = \frac{\sum_{i=1}^n \sum_{k=1}^q x_{i2k}}{nq}$	
	...	...	...	...	...	
A_p	$\bar{x}_{p1} = \frac{\sum_{i=1}^n x_{ip1}}{n}$	$\bar{x}_{p2} = \frac{\sum_{i=1}^n x_{ip2}}{n}$	...	$\bar{x}_{pq} = \frac{\sum_{i=1}^n x_{ipq}}{n}$	$\bar{y}_p = \frac{\sum_{i=1}^n \sum_{k=1}^q x_{ipk}}{nq}$	
Середні величини за стовпцями	$\bar{z}_1 = \frac{\sum_{i=1}^n \sum_{j=1}^p x_{i1j}}{np}$	$\bar{z}_2 = \frac{\sum_{i=1}^n \sum_{j=1}^p x_{i2j}}{np}$	...	$\bar{z}_q = \frac{\sum_{i=1}^n \sum_{j=1}^p x_{iqj}}{np}$	x_{ikj} – конкретне значення ознаки X , якого вона набуває в разі i -го експерименту, за j -м рівнем фактора A і k -м рівнем фактора B	

Результати обчислень двофакторного дисперсійного аналізу

Джерела, що спонукають до розсіювання	Суми квадратів відхилень	Кількість ступенів свободи	Статистичні оцінки дисперсій (виправлені дисперсії)
Фактор A	$Q_1 = nq \sum_{j=1}^p (\bar{y}_j - \bar{x})^2$	$k_1 = p - 1$	$S_1^2 = \frac{Q_1}{p - 1}$
Фактор B	$Q_2 = np \sum_{j=1}^q (\bar{z}_j - \bar{x})^2$	$k_2 = q - 1$	$S_2^2 = \frac{Q_2}{q - 1}$
Одночасна дія факторів A і B	$Q_3 = \sum_{j=1}^q \sum_{k=1}^p (\bar{x}_{kj} - \bar{z}_j - \bar{y}_i + \bar{x})^2$	$k_3 = (p - 1)(q - 1)$	$S_3^2 = \frac{Q_3}{(p - 1)(q - 1)}$
Дія випадкової компоненти	$Q_4 = \sum_{i=1}^n \sum_{j=1}^q \sum_{k=1}^p (x_{ikj} - \bar{x}_{kj})^2$	$k_4 = N - pq$	$S_4^2 = \frac{Q_4}{N - pq}$
Загальне відхилення	$Q = \sum_{i=1}^n \sum_{j=1}^q \sum_{k=1}^p (x_{ikj} - \bar{x})^2$	$N - 1$	$S^2 = \frac{Q}{N - 1}$

Спочатку для кожної пари значень факторів A і B обчислюють блокові середні, а також визначають середні значення ознаки за стовпцями та рядками. Ці значення порівнюють із загальною вибірковою середньою ознаки. За всіма блоковими середніми обчислюють виправлені дисперсії. Відповідно визначають таке: виправлену дисперсію, зумовлену впливом на ознаку X лише фактора A ; виправлену дисперсію, зумовлену впливом на ознаку X лише фактора B ; виправлену дисперсію, зумовлену одночасним впливом на ознаку X і фактора A , і фактора B ; виправлену дисперсію, зумовлену дією на ознаку X факторів, вплив яких не розглядають у межах заданої моделі.

Останній рядок табл. 13.4 містить формули для обчислення загальної суми квадратів відхилень, яка має дорівнювати сумі всіх квадратів відхилень, пов'язаних із впливом на ознаку кожного з факторів окремо, обох факторів у сукупності, а також розсіювання ознаки під впливом факторів, які не враховано в межах заданої моделі.

Має бути виконано таку тотожність:

$$S^2 = S_1^2 + S_2^2 + S_3^2 + S_4^2.$$

За даними табл. 13.4, визначають емпіричні значення критерію Фішера:

$$F_A = \frac{S_1^2}{S_4^2}, \quad F_B = \frac{S_2^2}{S_4^2}, \quad F_{AB} = \frac{S_3^2}{S_4^2}.$$

Емпіричні значення критерію Фішера для перевірки значущості впливу факторів A і B на ознаку X порівнюють із критичними точками статистики Фішера – Снедекора.

Критичні точки визначають за довідковою таблицею (див. додаток Е) за рівнем значущості α і кількістю ступенів свободи, відповідно: $F_\alpha(k_1; k_4)$; $F_\alpha(k_2; k_4)$ та $F_\alpha(k_3; k_4)$.

Якщо $F_A > F_\alpha(k_1; k_4)$, то нульову гіпотезу про те, що немає впливу на ознаку X фактора A , відхиляють.

Якщо $F_B > F_\alpha(k_2; k_4)$, то нульову гіпотезу про те, що немає впливу фактора B на ознаку X , відхиляють.

Якщо $F_{AB} > F_\alpha(k_3; k_4)$, то нульову гіпотезу про те, що немає спільного впливу факторів A і B , відхиляють.

13.3. Приклади розв'язування задач

Приклад 13.1. Ступінь впливу мінеральних добрив на врожайність наведено в табл. 13.5.

Таблиця 13.5

Аналіз впливу мінеральних добрив на врожайність

Ступені впливу мінеральних добрив	Урожайності, т
1	3,2; 3,1; 3,1; 2,8; 3,3; 3,0
2	2,6; 3,1; 2,7; 2,9; 2,7; 2,8
3	2,9; 2,6; 3,0; 3,1; 3,0; 2,8
4	3,7; 3,4; 3,2; 3,3; 3,5; 3,3
5	3,0; 3,4; 3,2; 3,5; 2,9; 3,1

З'ясуємо, чи істотно впливають мінеральні добрива на врожайність за $\alpha = 0,01$.

Розв'язання. Використовуючи табл. 13.1 і 13.2, виконаймо відповідні обчислення. Результати наведено в табл. 13.6, 13.7.

Таблиця 13.6

Допоміжні обчислення

Ступені впливу мінеральних добрив	Спостережувані значення (урожайності)	Групові середні	Загальна середня
1	3,2; 3,1; 3,1; 2,8; 3,3; 3,0	$\bar{x}_1 = 3,083$	$\bar{x} = 3,073$
2	2,6; 3,1; 2,7; 2,9; 2,7; 2,8	$\bar{x}_2 = 2,800$	
3	2,9; 2,6; 3,0; 3,1; 3,0; 2,8	$\bar{x}_3 = 2,900$	
4	3,7; 3,4; 3,2; 3,3; 3,5; 3,3	$\bar{x}_4 = 3,400$	
5	3,0; 3,4; 3,2; 3,5; 2,9; 3,1	$\bar{x}_5 = 3,180$	

Таблиця 13.7

Дисперсійний аналіз впливу мінеральних добрив на врожайність

Види варіацій ознаки	Суми квадратів відхилень	Кількість ступенів свободи	Статистичні оцінки дисперсій
Міжгрупова	$\sum_{i=1}^5 n_i \cdot (\bar{x}_i - \bar{x})^2 = 1,338$	$p - 1 =$ $= 5 - 1 = 4$	$\frac{SSR}{p - 1} = 0,335$
Внутрішньо-групова	$\sum_{i=1}^5 \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 = 0,927$	$N - p =$ $= 30 - 5 = 25$	$\frac{SSE}{N - p} = 0,037$

За ф-лою (13.5), обчислімо:

$$F = \frac{0,335}{0,037} = 9,02.$$

Критичне значення: $F_{кр.}(\alpha = 0,01; k_1 = 4; k_2 = 25) = 6,6$.

Оскільки $F > F_{кр.}$, то вплив мінеральних добрив на врожайність є істотним.

Приклад 13.2. Для перевірки впливу витрат на рекламу на обсяги продажів, проведено дослідження, результати якого подано в табл. 13.8.

Таблиця 13.8

Аналіз впливу витрат на рекламу на обсяги продажів

Витрати на рекламу, тис. грн	Обсяги продажів, млн грн
150 – 200	6,6; 5,9; 7,4
200 – 250	6,6; 7,9; 7,4; 7,3
250 – 300	7,6; 7,9; 7,4; 8,3; 8,0
300 – 350	8,2; 7,9; 8,4; 7,7

З'ясуємо, чи істотно впливають витрати на рекламу на обсяги продажів за $\alpha = 0,05$.

Розв'язання. Використовуючи табл. 13.1 і 13.2, виконаймо відповідні обчислення. Результат наведено в табл. 13.9.

Таблиця 13.9

Дисперсійний аналіз впливу витрат на рекламу на обсяги продажів

Витрати на рекламу	Спостережувані значення (обсяги продажів)	Групові середні	Загальна середня
150 – 200	6,6; 5,9; 7,4	$\bar{x}_1 = 6,63$	$\bar{x} = 7,53$
200 – 250	6,6; 7,9; 7,4; 7,3	$\bar{x}_2 = 7,30$	
250 – 300	7,6; 7,9; 7,4; 8,3; 8,0	$\bar{x}_3 = 7,84$	
300 – 350	8,2; 7,9; 8,4; 7,7	$\bar{x}_4 = 8,05$	
Види варіацій ознаки	Суми квадратів відхилень	Кількість ступенів свободи	Статистичні оцінки дисперсій
Міжгрупова	$\sum_{i=1}^4 n_i \cdot (\bar{x}_i - \bar{x})^2 = 4,186$	$p - 1 = 4 - 1 = 3$	$\frac{SSR}{p - 1} = 1,395$
Внутрішньо-групова	$\sum_{i=1}^4 \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 = 2,769$	$N - p = 16 - 4 = 12$	$\frac{SSE}{N - p} = 0,231$

За ф-лою (13.5) обчислімо:

$$F = \frac{1,395}{0,231} = 6,04.$$

Критичне значення: $F_{кр.}(\alpha = 0,05; k_1 = 3; k_2 = 12) = 3,49$.

Оскільки $F > F_{кр.}$, то вплив витрат на рекламу на обсяги продажів є істотним.

Приклад 13.3. У табл. 13.10 наведено відсотки браку шліфування 36 деталей, залежно від швидкості (фактор A) та потужності (фактор B) шліфувального верстата.

За рівнем значущості $\alpha = 0,05$ перевірмо, чи є статистично значущим вплив факторів A і B окремо та спільний вплив факторів A і B на результат шліфування деталей.

Таблиця 13.10

Брак шліфування деталей, %

Ступені впливу фактора A	Ступені впливу фактора B		
	B_1	B_2	B_3
A_1	3,6; 3,9; 4,1	2,9; 3,1; 3,0	2,7; 2,5; 2,9
A_2	4,2; 4,0; 4,1	3,3; 2,9; 3,2	3,7; 3,5; 3,6
A_3	3,8; 3,5; 3,6	3,6; 3,7; 3,5	3,2; 3,0; 3,4
A_4	3,4; 3,2; 3,2	3,4; 3,6; 3,5	3,6; 3,8; 3,7

Розв'язання. Використовуючи дані табл. 13.10, обчислимо і наведемо у вигляді таблиці значення блокових середніх $\bar{x}_{ij}$, умовних середніх $\bar{y}_i$ ($i = \overline{1, 4}$), що відповідають сталому рівню фактора A , та умовних середніх $\bar{z}_j$ ($j = \overline{1, 3}$), що відповідають сталому рівню фактора B , а також загальної вибіркової середньої $\bar{x}$ (табл. 13.11).

Таблиця 13.11

Результати обчислення середніх

Ступені впливу фактора A	Ступені впливу фактора B			$\bar{y}_i$
	B_1	B_2	B_3	
A_1	3,86(6)	3,000	2,7	3,189
A_2	4,100	3,13(3)	3,6	3,611
A_3	3,63(3)	3,600	3,2	3,478
A_4	3,26(6)	3,500	3,7	3,442
$\bar{z}_j$	3,717	3,308	3,3	–
$\bar{x}$	3,442			

За табл. 13.11 визначмо виправлені дисперсії, формули яких наведено в табл. 13.4:

$$S_1^2 = \frac{nq}{p-1} \sum_{i=1}^p (\bar{y}_i - \bar{x})^2 = 0,2884;$$

$$S_2^2 = \frac{np}{q-1} \sum_{j=1}^q (\bar{z}_j - \bar{x})^2 = 0,6808;$$

$$S_3^2 = \frac{n}{(p-1)(q-1)} \sum_{i=1}^p \sum_{j=1}^q (\bar{x}_{ij} - \bar{y}_i - \bar{z}_j + \bar{x})^2 = 0,4790;$$

$$S_4^2 = \frac{1}{N - pq} \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (x_{ijk} - \bar{x}_{ij})^2 = 0,0236;$$

$$S^2 = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (x_{ijk} - \bar{x})^2 = 5,6675.$$

Перевірмо виконання тотожності:

$$S^2 = S_1^2 + S_2^2 + S_3^2 + S_4^2.$$

Дійсно, $S^2 = 5,6675 = 0,2884 + 0,6808 + 0,4790 + 0,0236$.

Отже, обчислення не містять помилок.

Тепер визначмо емпіричні значення критерію Фішера – Снедекора за кожним із факторів A , B та за їхньою сумісною дією:

$$F_A = \frac{S_1^2}{S_4^2} = \frac{0,2884}{0,0236} = 12,22;$$

$$F_B = \frac{S_2^2}{S_4^2} = \frac{0,6808}{0,0236} = 28,84;$$

$$F_{AB} = \frac{S_3^2}{S_4^2} = \frac{0,4790}{0,0236} = 20,30.$$

За таблицею критичних точок розподілу Фішера – Снедекора (див. додаток Е), визначмо критичні точки статистичного критерію:

$$F_{0,05}(2; 24) = 3,40; F_{0,05}(3; 24) = 3,01; F_{0,05}(6; 24) = 2,51.$$

Отже, $F_A > F_{0,05}(3; 24)$, $F_B > F_{0,05}(2; 24)$, $F_{AB} > F_{0,05}(6; 24)$, тому з надійністю 95 % вплив фактора A і фактора B окремо, а також вплив факторів A і B у сукупності слід uważати суттєвими.

13.4. Вправи для самостійної роботи

13.1. У результаті проведення досліду, із метою з'ясування впливу чорного пару на врожайність пшениці з ділянки 9 га (3 га були під чорним паром; 3 га – під картоплею; 3 га – під кормовими травами), дістали такі результати (табл. 13.12).

Таблиця 13.12

Вплив чорного пару на врожайність пшениці

Фактори	Урожайності, ц/га
Чорний пар	26,6; 26,6; 30,6
Площа під картоплею	24,3; 25,2; 25,2
Площа під кормовими травами	26,6; 28,0; 31,0

За рівнем значущості $\alpha = 0,01$, перевірте істотність впливу чорного пару на врожайність пшениці.

13.2. Експериментально досліджували вплив на зносостійкість колінчастих валів технології їхнього виготовлення: вплив фактора A , який має чотири рівні, тобто застосовували чотири технології виготовлення валів.

Визначені результати наведено в табл. 13.13.

Таблиця 13.13

Вплив технології виготовлення на зносостійкість колінчастих валів

Ступені впливу фактора A	Кількості відпрацьованих місяців
A_1	9; 8; 10; 12
A_2	10; 12; 11; 8
A_3	8; 16; 10; 18
A_4	9; 18; 10; 8

За рівнем значущості $\alpha = 0,01$, перевірте вплив технологій на зносостійкість валів.

13.3. Для перевірки впливу методики навчання виробничих навичок на якість підготовки з випускників виробничо-технічного училища навмання вибирають чотири групи учнів, яких після закінчення навчання за різними методиками тестували на кількість виготовлених однотипних деталей протягом робочої зміни.

Результати тестування наведено в табл. 13.14.

Таблиця 13.14

Вплив методики навчання на якість підготовки

Ступені впливу фактора A (методики)	Кількості виготовлених деталей за робочу зміну
A_1	60; 80; 75; 70; 85; 70
A_2	75; 66; 85; 80; 75; 80; 90
A_3	60; 80; 65; 70; 86; 75
A_4	95; 85; 100; 80; 85

За рівнем значущості $\alpha = 0,05$, з'ясуйте вплив методики навчання на якість підготовки учнів.

13.4. Із кожної з 8 партій однотипних заготовок навмання брали заготовки, які обробляли на трьох верстатах-автоматах різної модифікації. Кількість деталей, виготовлених верстатами, досліджували на стандартність. Результати досліджень наведено в табл. 13.15.

Таблиця 13.15

Вплив модифікації верстатів-автоматів на якість деталей

Фактор A (типи верстатів-автоматів)	Кількості деталей, виготовлених верстатами-автоматами, що відповідають стандарту
A_1	100; 86; 90; 89; 95; 22; 80; 79
A_2	99; 82; 98; 88; 100; 96; 98; 100
A_3	100; 88; 86; 98; 98; 100; 99; 99

За рівнем значущості $\alpha = 0,01$, перевірте вплив модифікації верстатів-автоматів на якість виготовлених деталей.

13.5. Досліджують вплив на зносостійкість деталей таких факторів: A – матеріал для виготовлення деталей (застосовували три види сталі) та B – технологія виготовлення деталей (дві технології). Результати досліджень наведено в табл. 13.16.

Таблиця 13.16

Аналіз факторів впливу на зносостійкість деталей

Фактор B	Фактор A		
	A_1	A_2	A_3
B_1	10; 7; 8; 6; 12; 8; 11; 10; 14; 13	8; 14; 6; 10; 16; 14; 13; 12; 11; 15	15; 12; 11; 9; 8; 13; 11; 12; 16; 14
B_2	12; 13; 6; 9; 8; 11; 10; 10; 13; 17	11; 12; 12; 16; 13; 8; 10; 9; 8; 15	13; 12; 14; 8; 6; 8; 16; 12; 14; 16

За рівнем значущості $\alpha = 0,05$, визначте вплив факторів A і B та їхній спільний вплив на досліджувану ознаку.

13.6. Досліджують вплив факторів A і B на кількість виготовлених втулок зі 100 взятих, які відповідають нормам стандарту: A – використали дві технології виготовлення; B – заготовки надходили із трьох заводів. Результати досліджень наведено в табл. 13.17.

Таблиця 13.17

Аналіз факторів впливу на кількість виготовлення деталей

Фактор B	Фактор A	
	A_1	A_2
B_1	90; 88; 90; 96; 98; 76; 80; 95; 85; 80	100; 99; 82; 98; 95; 80; 96; 95; 99; 91; 89; 90
B_2	79; 88; 92; 76; 80; 83; 85; 90; 96; 75	81; 82; 100; 98; 89; 85; 96; 98; 75; 97
B_3	82; 78; 75; 79; 80; 81; 86; 89; 75; 90	80; 86; 90; 91; 78; 76; 75; 82; 73; 82

За рівнем значущості $\alpha = 0,05$, визначте вплив факторів A і B та їхній сумісний вплив на досліджувану ознаку.

13.7. Досліджували, скільки часу на добу здобувачі вищої освіти проводять в інтернеті. В опитуванні взяли участь 197 здобувачів вищої освіти. За результатами дослідження опитаних розподілили на чотири групи: юнаків, які займаються спортом, і юнаків, які не займаються спортом; дівчат, які займаються спортом, і дівчат, які не займаються спортом. Для дисперсійного аналізу з кожної групи випадковим чином було відібрано по 10 анкет. Дані з них про час, що здобувачі вищої освіти проводять в інтернеті, наведено в табл. 13.18.

Таблиця 13.18

Аналіз часу проведення в інтернеті

Фактор B	Фактор A	
	дівчата	юнаки
Спортсмени	1; 1,5; 2; 4; 1; 1; 3; 2,5; 2; 3	1; 2; 2; 3; 2; 1; 3; 1; 1; 3
Неспортсмени	1; 4; 2; 5; 2; 1; 3; 5; 2; 3	2; 4; 2; 5; 4; 1; 3; 4; 2; 3

З'ясуйте, чи пов'язаний час, що здобувачі вищої освіти проводять в інтернеті, зі статтю та заняттями спортом.

13.5. Тестові завдання

13.1. У моделі однофакторного дисперсійного аналізу міжгрупову дисперсію зумовлено

- А** впливом досліджуваного фактора на ознаку X
- Б** впливом інших випадкових факторів
- В** впливом ознаки на X на досліджуваний фактор

13.2. У моделі однофакторного дисперсійного аналізу внутрішньогрупову дисперсію зумовлено

- А** впливом досліджуваного фактора на ознаку X
- Б** впливом інших випадкових факторів
- В** впливом ознаки на X на досліджуваний фактор

13.3. Якщо досліджуваний фактор не впливає на значення ознаки X , то в цьому разі

А $\frac{SSR}{p-1}$ і $\frac{SSE}{N-p}$ можна розглядати як незалежні оцінки загальної дисперсії D

Б вибірки слід уважати здійсненими з різних сукупностей

В $\frac{SSR}{p-1}$ і $\frac{SSE}{N-p}$ можна розглядати як залежні оцінки загальної дисперсії D

13.4. Варіацію досліджуваної ознаки, зумовлену дією неконтрольованих у дослідженні факторів, оцінює

- А** факторна дисперсія
- Б** загальна дисперсія
- В** залишкова дисперсія

13.5. Варіацію досліджуваної ознаки, зумовлену дією контрольованих у дослідженні факторів, оцінює

- А** факторна дисперсія
- Б** загальна дисперсія
- В** залишкова дисперсія

13.6. Варіацію досліджуваної ознаки, зумовлену дією як контрольованих, так і неконтрольованих у дослідженні факторів, оцінює

- А** факторна дисперсія
- Б** загальна дисперсія
- В** залишкова дисперсія

13.7. Які можливості двофакторного дисперсійного аналізу

- А** дозволяє оцінити лише вплив кожного з факторів
- Б** дозволяє оцінити лише взаємодію факторів
- В** дозволяє оцінити вплив кожного з факторів і їхню взаємодію

13.8. Якщо $F_A^* > F_{кр.}(\alpha; k_4, k_1)$, то

- А** нульову гіпотезу про відсутність впливу фактора A відхиляють
- Б** нульову гіпотезу про відсутність впливу фактора A не відхиляють
- В** нульову гіпотезу про відсутність спільного впливу факторів A і B відхиляють

13.9. Якщо $F_B^* > F_{кр.}(\alpha; k_3, k_1)$, то

- А** нульову гіпотезу про відсутність впливу фактора B відхиляють
- Б** нульову гіпотезу про відсутність впливу фактора B не відхиляють
- В** нульову гіпотезу про відсутність спільного впливу факторів A і B відхиляють

13.10. Якщо $F_{AB}^* > F_{кр.}(\alpha; k_2, k_1)$, то

- А** нульову гіпотезу про відсутність впливу фактора B відхиляють
- Б** нульову гіпотезу про відсутність спільного впливу факторів A і B не відхиляють
- В** нульову гіпотезу про відсутність спільного впливу факторів A і B відхиляють

13.6. Запитання для самоперевірки

13.1. Розкрийте можливості дисперсійного аналізу, його основні принципи.

13.2. Що таке "вибіркова", "групова" та "загальна" середні?

13.3. Що таке "групова", "внутрішньогрупова", "міжгрупова" та "загальна" дисперсії?

13.4. Сформулюйте загальний принцип обчислення статистичного критерію, за яким перевіряють нульову гіпотезу в дисперсійному аналізі.

13.5. Суми квадратів яких відхилень розглядають в однофакторному дисперсійному аналізі? Поясніть їхній зв'язок із питомими дисперсіями.

13.6. Яку математичну модель має двофакторний дисперсійний аналіз?

13.7. Наведіть формули для обчислення блокових середніх і середніх значень ознаки за рядками та стовпцями у двофакторному аналізі.

13.8. За якими формулами обчислюють спостережувані значення статистичних критеріїв за кожним фактором A і B та їхню спільну дію?

13.7. Висновки за підрозділом 13

Дисперсійний аналіз є методом багатовимірної статистики, який дозволяє проводити дослідження впливу якісних факторів та аналізу значущості цього впливу. Основною метою дисперсійного аналізу є дослідження значущості відмінності між середніми декількох груп даних або змінних. Сенс методу полягає в тому, що загальна дисперсія ознак, що характеризують певний процес, є сумою двох компонентів: дисперсії, зумовленої впливом контрольованих факторів, і дисперсії, пов'язаної з похибками моделі, тобто зумовленої впливом факторів, які є неконтрольованими в межах цієї моделі.

Однофакторний дисперсійний аналіз використовують у дослідженнях зміни результативної ознаки під впливом зміни умов або градацій фактора. Двофакторний дисперсійний аналіз застосовують у дослідженнях одночасної дії двох факторів на різні вибірки об'єктів.

Література: [1 – 3; 5; 8; 10; 13; 15; 16].

14. Елементи теорії регресії

14.1. Мета та компетентності

Метою вивчення підрозділу є ознайомлення із принципами побудови моделі парної регресії та формування компетентностей щодо застосування регресійного аналізу для побудови моделей економічних явищ і прогнозування за цими моделями.

Компетентності, що формують під час вивчення підрозділу:

розуміння змісту економічних величин, що входять до складу моделі парної регресії;

знання методів перевірки значущості параметрів моделі регресії й оцінювання адекватності моделі;

уміння будувати рівняння парної лінійної регресії, здійснювати прогноз за цим рівнянням і визначати його точність.

14.2. Основні теоретичні відомості

У разі кореляційної залежності Y від X будують рівняння умовного середнього $\bar{y}_x$ від x : $\bar{y}_x = f(x)$.

За результатами одного й того самого випробування може бути визначено дві залежності: $\bar{y}_x$ та $\bar{x}_y$, тобто треба вказати, яку зі змінних уважати фактором, а яку функціональною ознакою (де причина, а де наслідок).

Ламану лінію ($\bar{y}_x$), яка з'єднує точки з координатами $(x_i, \bar{y}_{x_i})$, називають **емпіричною лінією регресії Y на X** .

Ламану лінію ($\bar{x}_y$), яка з'єднує точки з координатами $(\bar{x}_{y_k}, y_k)$, називають **емпіричною лінією регресії X на Y** .

Граничне положення емпіричної лінії регресії, до якої вона наближається за необмеженої кількості спостережень, називають **граничною теоретичною лінією регресії**.

Рівняння $\bar{y}_x = f(x)$ називають **рівнянням регресії Y на X** , функцію $f(x)$ – **регресією Y на X** , а графік – **лінією регресії Y на X** . Можна ввести й іншу залежність: $\bar{x}_y = \varphi(y)$ – **рівняння регресії X на Y** .

Поняття регресії та кореляції є безпосередньо пов'язаними між собою. Тоді як у кореляційному аналізі оцінюють силу стохастичного зв'язку, у регресійному – досліджують його форму. Обидва види аналізу призначено для встановлення причинних співвідношень між явищами та означення наявності або відсутності зв'язку.

Розрізняють такі види кореляції та регресії:

1) за **характером**: *додатна* регресія (коли зі зростанням (зменшенням) аргументу x зростає (зменшується) значення y) і *від'ємна* регресія (коли зі зростанням (зменшенням) значення аргументу x значення y зменшуються (зростають));

2) за **кількістю змінних**: *парна* та *множинна* кореляція й регресія;

3) за **формою зв'язку**: *лінійна* та *нелінійна* кореляція й регресія.

Під час установлення форми кореляційного зв'язку між двома змінними перевагу віддають лінійній залежності:

$$\hat{y}_x = b_1x + b_0 \text{ або } \hat{x}_y = a_1y + a_0.$$

Параметри рівняння регресії $\hat{y}_x = b_1x + b_0$ визначають за *методом найменших квадратів* так, щоб сума квадратів відхилень $\hat{y}_i$ від $y_{i \text{ емп.}}$ була найменшою.

Параметр $b_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{x^2 - (\bar{x})^2}$ називають **коефіцієнтом регресії Y**

на X і позначають $\rho_{y/x}$. Коефіцієнт регресії за своїм знаком є збіжним зі знаком коваріації $\mu_{xy} = \overline{xy} - \bar{x} \cdot \bar{y}$: якщо знак $\rho_{y/x}$ "+", то регресія є додатною, якщо знак "-", то регресія є від'ємною.

Отже,

$$\rho_{y/x} = \frac{\mu_{xy}}{\sigma_x^2}, \quad b_0 = \bar{y} - \rho_{y/x}\bar{x}. \quad (14.1)$$

Коефіцієнт регресії $\rho_{y/x}$ – це розмірна величина. Він показує, на скільки одиниць свого вимірювання зменшиться середнє значення функціональної ознаки $\hat{y}_x$, якщо фактор X збільшиться на одиницю свого вимірювання.

Аналогічно можна визначити теоретичне рівняння регресії X на Y :

$$\hat{x} = \rho_{x/y}y + a_0, \quad (14.2)$$

$$\text{де } \rho_{x/y} = \frac{\mu_{xy}}{\sigma_y^2}; \sigma_y^2 = \overline{y^2} - (\bar{y})^2; a_0 = \bar{x} - \rho_{x/y}\bar{y}.$$

Емпіричну та теоретичну лінії регресії зображують на кореляційному полі. Якщо параметри теоретичного рівняння регресії обчислено правильно, то помилки є незначними.

Точка $M(\bar{x}, \bar{y})$ є центром регресії, крізь який проходять обидві теоретичні лінії регресії.

Оскільки

$$\rho_{y/x} = \frac{\mu_{xy}}{\sigma_x^2}, \rho_{x/y} = \frac{\mu_{xy}}{\sigma_y^2}, \text{ то } \rho_{y/x} \cdot \rho_{x/y} = \frac{\mu_{xy}^2}{\sigma_x^2 \cdot \sigma_y^2} = r^2.$$

Звідси:

$$r_{xy} = \pm \sqrt{\rho_{y/x} \cdot \rho_{x/y}}, \quad (14.3)$$

де знак "+", якщо $\rho_{y/x}, \rho_{x/y}$ є додатними,

знак "-", якщо $\rho_{y/x}, \rho_{x/y}$ є від'ємними.

Кут між теоретичними лініями регресії $\hat{y}_x$ та $\hat{x}_y$ визначають за такою формулою:

$$\operatorname{tg} \varphi = \frac{1 - r_{xy}^2}{r_{xy}} \cdot \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}. \quad (14.4)$$

Якщо $|r|=1$, то $\operatorname{tg} \varphi = 0$, $\varphi = 0$, і лінії регресії $\hat{y}_x$ та $\hat{x}_y$ збігаються (зв'язок є лінійним);

якщо $r=0$, то $\operatorname{tg} \varphi = \infty$, $\varphi = \pi/2$, і лінії регресії є взаємно перпендикулярними (немає кореляційного зв'язку);

чим більшим є значення коефіцієнта кореляції, тим ближче одна до одної розміщено спряжені теоретичні лінії регресії;

чим слабкішою є кореляційна залежність, тим більшим є кут між спряженими теоретичними лініями регресії $\hat{y}_x$ та $\hat{x}_y$.

Для кожного з параметрів моделі $\hat{y}_x = b_1 x + b_0$ виконують перевірку, чи є параметр, значення якого обчислювали за вибірковою сукупністю, *статистично значущим*.

Нульову гіпотезу $H_0: b_1 = 0$ за альтернативної гіпотези $H_1: b_1 \neq 0$ перевіряють за допомогою статистики:

$$t_{b_1} = \frac{b_1}{\sigma_{b_1}}, \quad (14.5)$$

де $\sigma_{b_1} = \sqrt{\frac{\sigma_e^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$ – стандартне відхилення для параметра b_1 ;

$\sigma_e = \sqrt{\frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$ – незсунута оцінка середнього квадратичного відхилення випадкової величини e (помилки моделі).

Критичні значення випадкової величини t в разі невеликого обсягу вибірки визначають за таблицею розподілу Стюдента (див. додаток Д), відповідно до рівня значущості α та кількості ступенів свободи $k = n - 2$. Будують двобічну критичну область.

Зі збільшенням обсягу вибірки критичне значення t_α є аргументом функції Лапласа $\Phi(t_\alpha)$ (див. додаток В), відповідно до умови $\Phi(t_\alpha) = \frac{1-\alpha}{2}$.

Якщо $|t_{b_1}| < t_{0,05}$, то, за рівнем значущості $\alpha = 0,05$, статистичну гіпотезу H_0 немає підстав відхилити, тоді як, за умови $|t_{b_1}| > t_{0,01}$, гіпотезу H_0 спростовують на користь альтернативної з надійністю 99 %.

Якщо гіпотезу H_0 немає підстав відхилити, то це означає, що вибіркове рівняння набуває такого вигляду: $y = \bar{y}$, тобто випадкова вели-

чина Y , з огляду на її розсіювання під впливом факторів, на які не зважає регресійна модель, є сталою величиною на рівні свого середнього значення.

Аналогічно перевіряють статистичну значущість параметра b_0 . За нульовою гіпотезою $H_0: b_0 = 0$, а за альтернативною – $H_1: b_0 \neq 0$.

Для перевірки нульової гіпотези обчислюють таке емпіричне значення:

$$t_{b_0} = \frac{b_0}{\sigma_{b_0}}, \quad (14.6)$$

$$\text{де } \sigma_{b_0} = \sqrt{\frac{\sigma_e^2 \cdot \sum_{i=1}^n x_i^2}{n \cdot \sum_{i=1}^n (x_i - \bar{x})^2}} - \text{стандартне відхилення для параметра } b_0.$$

Критичні значення t визначають за таблицею розподілу Стюдента (див. додаток Д), відповідно до заданого рівня значущості α та кількості ступенів свободи $k = n - 2$.

Якщо нульову гіпотезу відхиляють, то рівняння регресії буде містити вільний член: $b_0 \neq 0$.

Довірчі інтервали для параметрів регресійної моделі визначають, відповідно до вибраного рівня значущості за їхніми стандартними відхиленнями.

Довірчі інтервали для параметрів b_i , ($i = 0, 1$), за рівнем значущості α , визначають за такою формулою:

$$[b_i - t_\alpha \cdot \sigma_{b_i}; b_i + t_\alpha \cdot \sigma_{b_i}], \quad (14.7)$$

де для малих вибірових сукупностей величину t_α визначають за таблицею розподілу Стюдента (див. додаток Д) для заданого рівня значущості α , відповідно до кількості ступенів свободи $k = n - 2$;

для великих вибірок t_α визначають як аргумент функції Лапласа (див. додаток В), відповідно до умови $2\Phi(t_\alpha) = 1 - \alpha$.

Для лінії регресії, побудованої за відповідним вибірковим рівнянням, визначають *довірчу смугу*, до якої із заданою надійністю буде належати теоретична лінія регресії.

Для кожного значення зовнішнього фактора $X = x$ верхня та нижня межі довірчої смуги, до якої з надійністю $\gamma = 95\%$ буде належати теоретична лінія регресії, обчислюють, відповідно до такої формули:

$$\Delta y(x) = \pm t_{0,05} \cdot S_y \cdot \sqrt{\frac{1-r^2}{n-2}} \cdot \sqrt{1 + \frac{(x - \bar{x})^2}{S_x^2}}. \quad (14.8)$$

Для побудови *довірчого інтервалу прогнозу* за функцією регресії для кожного індивідуального значення x_{n+1} визначають межі інтервалу за таким співвідношенням:

$$\hat{y}(x_{n+1}) \pm t_\alpha(k) \cdot S_{\text{прогн.}}, \quad (14.9)$$

де $t_\alpha(k)$ – критичне значення статистики Стюдента (див. додаток Д) для рівня значущості α і кількості ступенів свободи $k = n - 2$;

$$S_{\text{прогн.}}^2 = \sigma_e^2 \left(1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) - \text{дисперсія прогнозу.}$$

14.3. Приклади розв'язування задач

Приклад 14.1. Задано кореляційну таблицю (табл. 14.1), у якій наведено результати іспитів з інформатики (y) та вищої математики (x) в одній з академічних груп І курсу.

Результати іспитів з інформатики та математики

$y \backslash x$	2	3	4	5	l_k
2	1	2	–	–	3
3	1	6	3	–	10
4	–	5	5	–	10
5	–	–	1	1	2
h_i	2	13	9	1	25

У припущенні лінійної залежності $\hat{y}_x = \rho_{y/x}x + b$ між X і Y визначмо коефіцієнти $\rho_{y/x}$ і b .

Розв'язання. Обчислимо всі види середніх:

$$\bar{x} = \frac{2 \cdot 2 + 3 \cdot 13 + 4 \cdot 9 + 5 \cdot 1}{25} = \frac{84}{25} = 3,36;$$

$$\bar{y} = \frac{2 \cdot 3 + 3 \cdot 10 + 4 \cdot 10 + 5 \cdot 2}{25} = \frac{86}{25} = 3,44;$$

$$\overline{x^2} = \frac{4 \cdot 2 + 9 \cdot 13 + 16 \cdot 9 + 25 \cdot 1}{25} = \frac{294}{25} = 11,76;$$

$$\overline{y^2} = \frac{4 \cdot 3 + 9 \cdot 10 + 16 \cdot 10 + 25 \cdot 2}{25} = \frac{312}{25} = 12,48;$$

$$\begin{aligned} \overline{xy} &= \frac{\sum_i \sum_k m_{ki} x_i y_k}{n} = \frac{1}{25} ((1 \cdot (2 \cdot 2) + 2 \cdot (3 \cdot 2)) + (1 \cdot (2 \cdot 3) + 6 \cdot (3 \cdot 3) + 3 \cdot (4 \cdot 3)) + \\ &+ (5 \cdot (3 \cdot 4) + 5 \cdot (4 \cdot 4)) + (1 \cdot (4 \cdot 5) + 1 \cdot (5 \cdot 5))) = \frac{1}{25} (16 + 96 + 140 + 45) = \frac{297}{25} = 11,88. \end{aligned}$$

Тепер за ф-лою (14.1) буде:

$$\rho_{y/x} = \frac{11,88 - 3,36 \cdot 3,44}{11,76 - 3,36^2} = \frac{11,88 - 11,56}{11,76 - 11,29} = \frac{0,32}{0,47} \approx 0,68;$$

$$\bar{b} = 3,44 - 0,68 \cdot 3,36 = 3,44 - 2,28 = 1,16.$$

Отже, теоретичне рівняння регресії має такий вигляд:

$$\hat{y}_x = 0,68x + 1,16.$$

Зробимо перевірку:

$$x=3: \quad \hat{y}_x = 0,68 \cdot 3 + 1,16 = 3,2;$$

$$\bar{y}_x = \frac{2 \cdot 2 + 3 \cdot 6 + 4 \cdot 5}{13} = 3,23; \quad \varepsilon_1 = 0,03.$$

$$x=4: \quad \hat{y}_x = 0,68 \cdot 4 + 1,16 = 3,88;$$

$$\bar{y}_x = \frac{3 \cdot 3 + 4 \cdot 5 + 5 \cdot 1}{9} = 3,78; \quad \varepsilon_1 = 0,1.$$

Приклад 14.2. Залежність між змінними величинами X (кількість унесених добрив) і Y (урожайність) наведено в табл. 14.2.

Таблица 14.2

Залежність між кількістю внесених добрив і урожайністю

x_i	1,0	1,5	2,0	2,5	3,0
y_i	2,15	2,30	2,60	2,80	2,50

У припущенні лінійної залежності $\hat{y}_x = \rho_{y/x} x + b$ між X і Y визначмо коефіцієнти $\rho_{y/x}$ і b .

Розв'язання. Вихідні дані й результати обчислімо в табл. 14.3.

Таблиця 14.3

Результати обчислень

n	x_i	y_i	x_i^2	$x_i y_i$
1	1,0	2,15	1,00	2,15
2	1,5	2,30	2,25	3,45
3	2,0	2,60	4,00	5,20
4	2,5	2,80	6,25	7,00
5	3,0	2,50	9,00	7,50
Σ	10,0	12,35	22,50	25,30

Виконуючи обчислення, визначмо:

$$\bar{x} = \frac{10}{5} = 2; \bar{y} = \frac{12,35}{5} = 2,47; \overline{x^2} = \frac{22,5}{5} = 4,5; \overline{xy} = \frac{25,3}{5} = 5,06.$$

Тоді

$$\rho_{y/x} = 0,24; b = 1,99.$$

Теоретичне рівняння лінійної регресії має такий вигляд:

$$\hat{y}_x = 0,24x + 1,99.$$

На рис. 14.1 зображено емпіричні точки та визначено теоретичну лінію регресії $\hat{y}_x$.

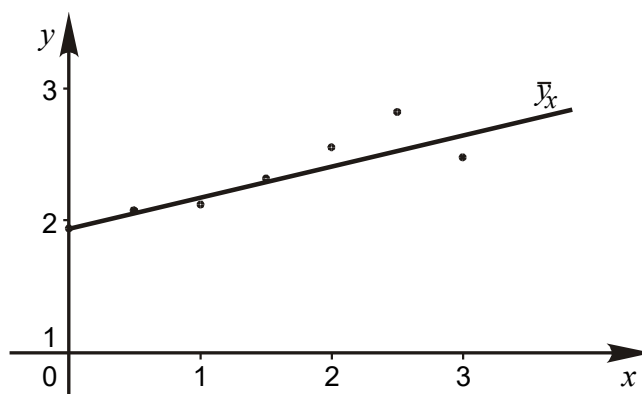


Рис. 14.1. Теоретична лінія регресії $\hat{y}_x$ та емпіричні точки

Перевірмо рівняння:

$$\hat{y}_{x=1,5} = 0,24 \cdot 1,5 + 1,99 = 2,35;$$

$$\hat{y}_{x=2,5} = 0,24 \cdot 2,5 + 1,99 = 2,6.$$

Відхилення:

$$\varepsilon_1 = |2,35 - 2,3| = 0,05; \quad \varepsilon_2 = |2,8 - 2,6| = 0,2.$$

Приклад 14.3. Середню температуру окремого регіону вимірювали протягом 13 днів у січні (A) та лютому (B). Результати вимірювання наведено в табл. 14.4.

Таблиця 14.4

Середня температура в січні (A) та лютому (B)

$A(X = x_i)$	-19,2	-14,8	-19,6	-11,1	-9,4	-16,9	-13,7
$B(Y = y_i)$	-21,8	-15,4	-20,8	-11,3	-11,6	-19,2	-13,0

$A(X = x_i)$	-4,9	-13,9	-9,4	-8,3	-7,9	-5,3
$B(Y = y_i)$	-7,4	-15,1	-14,4	-11,1	-10,5	-7,2

Потрібно:

- 1) скласти рівняння парної лінійної регресії Y на X ;
- 2) побудувати довірчі інтервали для параметрів теоретичного рівняння лінійної регресії з надійністю 95 %;
- 3) за рівнем значущості $\alpha = 0,05$, перевірити значущість параметрів теоретичного рівняння лінійної регресії;
- 4) із надійністю 99 %, побудувати довірчий інтервал для прогнозованого індивідуального значення за умови, що $x = -22,5$.

Розв'язання.

1. Для обчислення значень параметрів теоретичного рівняння лінійної регресії використаємо табл. 14.5.

Допоміжні обчислення

№ з/п	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	-19,2	-21,8	418,56	368,64	475,24
2	-14,8	-15,4	227,92	219,04	237,16
3	-19,6	-20,8	407,68	384,16	432,64
4	-11,1	-11,3	125,43	123,21	127,69
5	-9,4	-11,6	109,04	88,36	134,56
6	-16,9	-19,2	324,48	285,61	368,64
7	-13,7	-13,0	178,10	187,69	169,00
8	-4,9	-7,4	36,26	24,01	54,76
9	-13,9	-15,1	209,89	193,21	228,01
10	-9,4	-14,4	135,36	88,36	207,36
11	-8,3	-11,1	92,13	68,89	123,21
12	-7,9	-10,5	82,95	62,41	110,25
13	-5,3	-7,2	38,16	28,09	51,84
Σ	-154,4	-178,8	2 385,96	2 121,68	2 720,36

Обчислімо середні:

$$\bar{x} = \frac{\sum x_i}{n} = -\frac{154,4}{13} = -11,88;$$

$$\bar{y} = \frac{\sum y_i}{n} = -\frac{178,8}{13} = -13,75;$$

$$\overline{x^2} = \frac{\sum x_i^2}{n} = \frac{2\,121,68}{13} = 163,21;$$

$$\overline{y^2} = \frac{\sum y_i^2}{n} = \frac{2\,720,36}{13} = 209,26;$$

$$\overline{xy} = \frac{\sum x_i y_i}{n} = \frac{2\,385,96}{13} = 183,54.$$

Визначмо параметри теоретичного рівняння лінійної регресії:

$$b_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2} = \frac{183,54 - (-11,88)(-13,75)}{163,21 - (-11,88)^2} = \frac{20,19}{22,08} = 0,92;$$

$$b_0 = \bar{y} - b_1 \bar{x} = -13,75 - 0,92(-11,88) = -13,75 + 10,93 = -2,82.$$

Отже, рівняння регресії має такий вигляд:

$$\hat{y} = -2,82 + 0,92x.$$

2. Для визначення довірчих інтервалів параметрів теоретичного рівняння лінійної регресії потрібно обчислити таке значення:

$$\sigma_e = \sqrt{\frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2},$$

що є точковою незміщеною статистичною оцінкою середнього квадратичного відхилення випадкового фактора e (помилки моделі).

Допоміжні обчислення наведено в табл. 14.6.

Таблиця 14.6

Обчислення довірчих інтервалів

x_i	y_i	$\hat{y}_i = b_0 + b_1 x_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$
-19,2	-21,8	-20,484	-1,316	1,732
-14,8	-15,4	-16,436	1,036	1,073
-19,6	-20,8	-20,852	0,052	0,003
-11,1	-11,3	-13,032	1,732	2,999
-9,4	-11,6	-11,462	-1,138	0,003
-16,9	-19,2	-18,368	-0,832	0,692
-13,7	-13,0	-15,424	2,424	5,876
-4,9	-7,4	-7,328	-0,072	0,005
-13,9	-15,1	-15,608	0,508	0,258
-9,4	-14,4	-11,468	-2,932	8,597
-8,3	-11,1	-10,456	-0,644	0,415
-7,9	-10,5	-10,088	-0,412	0,169
-5,3	-7,2	-7,696	0,496	0,246
-154,4	-178,8	—	—	22,068

Отже, буде:

$$\sigma_e = \sqrt{\frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2} = \sqrt{\frac{22,068}{13-2}} = \sqrt{2,006} = 1,416.$$

Далі обчислимо стандартні відхилення кожного параметра:

$$\sigma_{b_1} = \sqrt{\frac{\sigma_e^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} = \sqrt{\frac{\sigma_e^2}{\sum_{i=1}^n x_i^2 - n(\bar{x})^2}} = \sqrt{\frac{2,006}{2 \cdot 121,68 - 13 \cdot (-11,88)^2}} =$$

$$= \sqrt{\frac{2,006}{286,93}} = \sqrt{0,007} = 0,084;$$

$$\sigma_{b_0} = \sqrt{\frac{\sigma_e^2 \cdot \sum_{i=1}^n x_i^2}{n \cdot \sum_{i=1}^n (x_i - \bar{x})^2}} = \sqrt{\frac{\sigma_e^2 \cdot \sum_{i=1}^n x_i^2}{n \cdot \left(\sum_{i=1}^n x_i^2 - n(\bar{x})^2 \right)}} = \sqrt{\frac{2,006 \cdot 2 \cdot 121,68}{13 \cdot (2 \cdot 121,68 - 13 \cdot (-11,88)^2)}} =$$

$$= \sqrt{\frac{4 \cdot 256,1}{13 \cdot 286,93}} = \sqrt{1,141} = 1,068.$$

За таблицею критичних точок розподілу Стюдента (див. додаток Д), за заданим значенням $\gamma = 0,95$ і $k = n - 2 = 13 - 2 = 11$, визначмо:

$$t_\alpha = t(\gamma = 1 - \alpha = 0,95; k = 11) = 2,201.$$

Довірчі інтервали для параметрів b_i , ($i = 0,1$), за рівнем значущості α , визначмо за такою формулою:

$$[b_i - t_\alpha \cdot \sigma_{b_i}; b_i + t_\alpha \cdot \sigma_{b_i}].$$

Отже, із надійністю 95 %, довірчий інтервал для параметра b_0 становить:

$$[-2,82 - 2,201 \cdot 1,068; -2,82 + 2,201 \cdot 1,068] = [-5,171; 0,469];$$

для параметра b_1 :

$$[0,92 - 2,201 \cdot 0,084; 0,92 + 2,201 \cdot 0,084] = [0,735; 1,105].$$

3. За рівнем значущості $\alpha = 0,05$, перевірмо значущість параметрів теоретичного рівняння лінійної регресії.

Нульову гіпотезу $H_0 : b_i = 0$, за альтернативної гіпотези $H_1 : b_i \neq 0$, перевіримо за допомогою статистики: $t_{b_i} = \frac{b_i}{\sigma_{b_i}}$.

Визначмо:

$$t_{b_0} = \frac{b_0}{\sigma_{b_0}} = \frac{-2,82}{1,068} = -2,64; \quad t_{b_1} = \frac{b_1}{\sigma_{b_1}} = \frac{0,92}{0,084} = 10,95.$$

Отже, $|t_{b_i}| > t_{0,05}(11) = 2,201$, тобто нульову гіпотезу H_0 спростовують на користь альтернативної, із надійністю 95 %. Це означає, що обидва параметри теоретичного рівняння лінійної регресії є значущими.

4. Із надійністю 99 %, побудуймо довірчий інтервал для прогнозованого індивідуального значення за умови, що температура в січні буде $x = -22,5$ °C.

Визначмо межі інтервалу за співвідношенням (14.9):

$$\hat{y}(x_{n+1}) \pm t_{\alpha}(k) \cdot S_{\text{прогн.}},$$

де прогнозоване індивідуальне значення –

$$\hat{y}(x_{n+1}) = \hat{y}(-22,5) = -2,82 + 0,92 \cdot (-22,5) = -23,52;$$

критичне значення статистики Стьюдента (див. додаток Д) –

$$t_{\alpha}(k) = t_{0,01}(11) = 3,11;$$

дисперсія прогнозу –

$$S_{\text{прогн.}}^2 = \sigma_e^2 \left(1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) = 2,006 \left(1 + \frac{1}{13} + \frac{(-22,5 + 11,88)^2}{286,93} \right) =$$

$$= 2,949 \Rightarrow S_{\text{прогн.}} = \sqrt{2,949} = 1,717.$$

Отже, із надійністю 99 %, прогнозоване індивідуальне значення температури в лютому буде належати такому довірчому інтервалу:

$$[-23,52 - 3,11 \cdot 1,717; -23,52 + 3,11 \cdot 1,717] = [-28,86; -18,18].$$

Приклад 14.4. Складімо рівняння парної лінійної регресії Y (середньодобове перероблення сировини, тис. т) на X (основні виробничі фонди, млн грн), якщо за вибірковими даними буде:

$$\bar{x} = 1,381; \bar{y} = 2,905; \overline{xy} = 12,014; n = 100; \sigma_x^2 = 5,33; \sigma_y^2 = 10,24.$$

Побудуємо вибірку лінію регресії та довірчу смугу, до якої, із надійністю 95 %, буде належати лінія регресії генеральної сукупності.

Розв'язання. Рівняння регресії визначають у такому вигляді:

$$\hat{y}_x = b_1 x + b_0.$$

Обчислімо вибірковий коефіцієнт регресії – $b_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x^2}$:

$$b_1 = \frac{12,014 - 2,381 \cdot 2,905}{5,33} = 0,956.$$

Тоді

$$b_0 = 2,905 - 0,956 \cdot 1,381 = 0,629.$$

Отже, вибіркове рівняння регресії таке:

$$\hat{y} = 0,629 + 0,956x.$$

Оскільки графіком лінії регресії є пряма, то для його побудови достатньо обчислити за вибірковим рівнянням координати двох будь-яких точок, що належать цій прямій.

Якщо $x_1 = 0$, тоді $\hat{y}_1 = 0,629$; якщо $x_2 = 3$, тоді $\hat{y}_2 = 3,497$.

Пряму за цими точками побудовано на рис. 14.2.

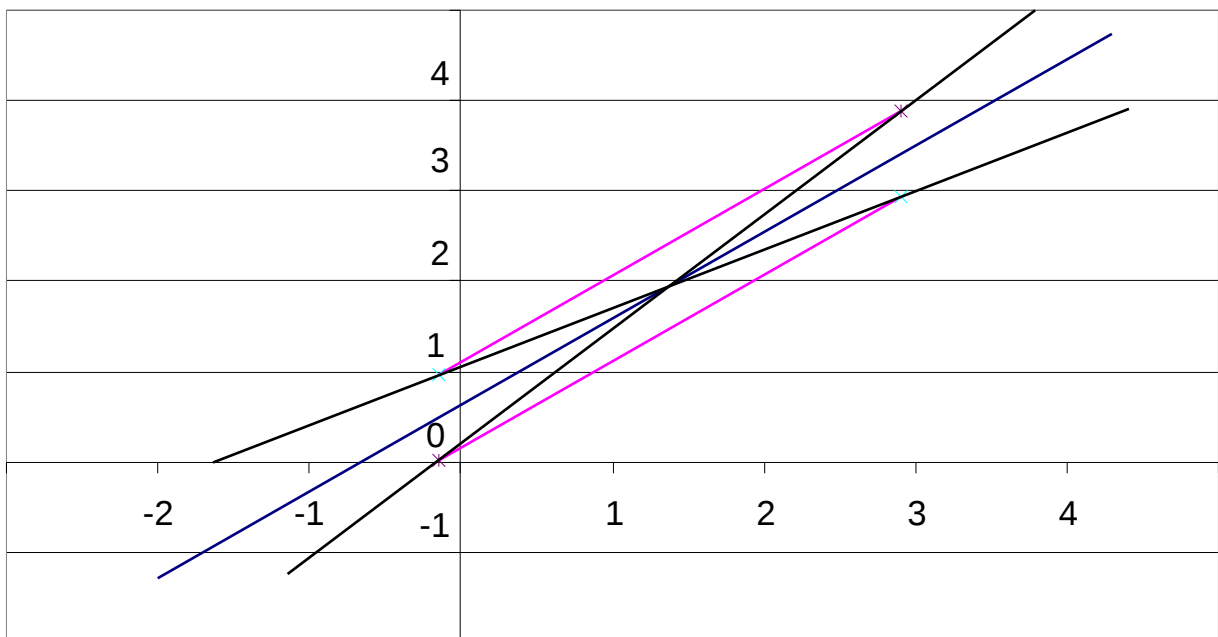


Рис. 14.2. Схематичне подання довірчої смуги лінії регресії

Тепер схематично побудуємо довірчу смугу для лінії регресії.
Для цього спочатку визначмо вибірковий коефіцієнт кореляції:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y}; \quad r = \frac{12,014 - 2,381 \cdot 2,905}{\sqrt{5,33} \cdot \sqrt{10,24}} = 0,69.$$

Далі обчислимо напівширину довірчої смуги для лінії регресії, яка відповідає проміжку, якщо відхилення значення фактора X від його вибіркової середньої не перевищує середнього квадратичного відхилення,

тобто

$$\Delta y(x) \approx t_{0,05} \cdot S_y \cdot \sqrt{\frac{1-r^2}{n-2}},$$

де $S_y = \sqrt{\frac{n}{n-1} \cdot \sigma_y^2}$ – це виправлене середнє квадратичне відхилення фактора Y :

$$S_y = \sqrt{10,24 \cdot \frac{100}{99}} = 3,216.$$

Для $\alpha = 0,05$ та $k = n - 2$, за таблицею критичних точок розподілу Стьюдента (див. додаток Д), визначмо $t_{0,05}(98) = 1,99$.

Отже, буде:

$$\Delta y(x) \approx 1,99 \cdot 3,216 \cdot \sqrt{\frac{1-0,69^2}{98}} = 0,468.$$

Для побудови проміжку $[\bar{x} - S_x; \bar{x} + S_x]$ визначмо виправлене середнє квадратичне відхилення для фактора X :

$$S_x = \sqrt{\frac{n}{n-1} \cdot \sigma_x^2} \Rightarrow S_x = \sqrt{5,33 \cdot \frac{100}{99}} = 1,523.$$

На проміжку $x \in [-0,142; 2,904]$ угору і вниз від лінії регресії відкла-
даймо $\Delta y(x) = 0,468$.

Визначмо паралелограм, у якому проведемо діагоналі та продов-
жмо їх за цей паралелограм.

Продовження діагоналей і є межами 95 % довірчої смуги для $|x - \bar{x}| > S_x$ (див. рис. 14.2).

14.4. Вправи для самостійної роботи

14.1. У припущенні, що кореляційний зв'язок між випадковими вели-
чинами X і Y є лінійним, визначте вибіркове рівняння регресії Y на X
для даних табл. 14.7.

Таблиця 14.7

Вихідні дані залежності

X	0,9	2,5	3,4	5,6	6,9	9,0	9,8	11,2
Y	13,1	14,8	18,7	20,3	26,0	28,1	32,2	33,0

14.2. Складіть рівняння прямої регресії Y на X , якщо за вибірковими даними визначено такі результати: $\bar{x} = 14$; $\bar{y} = 1,8$ та $b_1 = 1,1$.

14.3. Визначте рівняння прямої регресії Y на X за результатами обчислення вибірових даних: $\bar{x} = 3,5$; $\bar{y} = 21,8$; $\mu_{xy} = 2,86$ та $\sigma_x^* = 1,41$.

14.4. У припущенні, що кореляційний зв'язок між випадковими величинами X і Y є лінійним, складіть рівняння спряжених ліній регресії, якщо за даними обчислення 100 вимірювань визначено такі результати:

$$\sum x = 421; \sum y = 34,1; \sum x^2 = 68\,315; \sum y^2 = 576,4; \sum xy = 9\,876,5.$$

14.5. У припущенні лінійного кореляційного зв'язку між факторами X та Y складіть спряжені вибіркові рівняння регресії.

Визначте кут між лініями регресії, а також обчисліть вибірковий коефіцієнт кореляції, якщо $n = 100$; $\sum x = 324$; $\sum y = 1975$; $\sum xy = 4197$; $\sum x^2 = 2\,324$; $\sum y^2 = 54\,834$.

14.6. Складіть рівняння спряжених прямих регресії за результатами вимірювання значень двовимірної випадкової величини (X ; Y), наведених у вигляді кореляційної таблиці (табл. 14.8).

Таблиця 14.8

Вихідна кореляційна таблиця

$Y \backslash X$	1	2	3	4	5	m_y
1	2	—	—	—	—	2
2	1	7	2	—	—	10
3	—	2	5	1	—	8
4	—	—	1	2	2	5
m_x	3	9	8	3	2	25

14.7. Показники товарообігу Y та суми витрат X , які досліджували в 20 магазинах, наведено в табл. 14.9.

Таблиця 14.9

Залежність товарообігу від витрат

$Y = y_i$, грн	480	510	530	540	555	564	570	575	580	585
$X = x_i$, грн	30	25	31	32	38	41	40	46	49	54

$Y = y_i$, грн	590	596	605	618	625	635	640	650	660
$X = x_i$, грн	58	60	64	75	78	82	83	85	90

Потрібно:

- 1) скласти рівняння парної лінійної регресії Y на X ;
- 2) побудувати довірчі інтервали для параметрів теоретичного рівняння лінійної регресії, із надійністю 95 %;
- 3) за рівнем значущості $\alpha = 0,05$, перевірити значущість параметрів теоретичного рівняння лінійної регресії;
- 4) із надійністю 99 %, побудувати довірчий інтервал для прогнозованого індивідуального значення за умови $x = 92$;
- 5) побудувати вибірккову лінію регресії та довірчу смугу, до якої, із надійністю 95 %, буде належати лінія регресії генеральної сукупності.

14.5. Тестові завдання

14.1. Завданнями регресійного аналізу є

А виявлення зв'язку між випадковими величинами та оцінювання їхньої тісноти

Б виявлення зв'язку між випадковими величинами та їхніми числовими характеристиками

В визначення рівняння зв'язку між випадковими величинами

Г визначення рівняння зв'язку між випадковою величиною та не випадковими незалежними змінними й оцінювання невідомих значень залежної змінної

Д визначення рівняння зв'язку між не випадковою залежною величиною та випадковими незалежними змінними й оцінювання невідомих значень залежної змінної

14.2. Ламану лінію $(\bar{y}_x)$, яка з'єднує точки з координатами $(x_i; \bar{y}_{x_i})$, називають

А емпіричною лінією регресії Y на X

Б емпіричною лінією регресії X на Y

В теоретичною лінією регресії

14.3. За формою зв'язку розрізняють регресію

А додатну і від'ємну

Б парну та множинну

В лінійну й нелінійну

14.4. За кількістю змінних розрізняють регресію

А додатну і від'ємну

Б парну та множинну

В лінійну й нелінійну

14.5. Параметри теоретичного рівняння регресії $\hat{y}_x = b_1x + b_0$ визначають за такими формулами

А $b_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2}; \quad b_0 = \bar{y} - b_1\bar{x}$

Б $b_0 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2}; \quad b_1 = \bar{y} - b_0\bar{x}$

В $b_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}}; \quad b_0 = \bar{y} - b_1\bar{x}$

14.6. Чим більшим є значення коефіцієнта кореляції

А тим більшим є кут між спряженими теоретичними лініями регресії $\hat{y}_x$ та $\hat{x}_y$

Б тим ближче одна до одної розміщено спряжені теоретичні лінії регресії

В тим ближче є лінії регресії до взаємної перпендикулярності

14.7. Вільний член у рівнянні регресії $\hat{y}_x = b_1x + b_0$ – це

А точка, у якій лінія регресії перетинає вісь Y

Б показник зв'язку між незалежною та залежною змінними

В величина, яка завжди дорівнює одиниці

14.8. Нульову гіпотезу $H_0 : b_i = 0$, за альтернативної гіпотези $H_1 : b_i \neq 0$, перевіряють за допомогою такої статистики

А $t_{b_i} = \frac{\sigma_{b_i}}{b_i}$, де σ_{b_i} – стандартне відхилення для параметра b_i

Б $t_{b_i} = \frac{b_i}{\sigma_{b_i}}$, де σ_{b_i} – стандартне відхилення помилки

В $t_{b_i} = \frac{b_i}{\sigma_{b_i}}$, де σ_{b_i} – стандартне відхилення для параметра b_i

14.6. Запитання для самоперевірки

14.1. Охарактеризуйте регресійний аналіз та наведіть його завдання.

14.2. Що таке "спряжені емпіричні лінії регресії"?

14.3. Які висновки про форму кореляційного зв'язку можна зробити за виглядом емпіричних ліній регресії?

14.4. Про що свідчить взаємне розташування емпіричних ліній регресії?

14.5. Розкрийте сутність методу найменших квадратів.

14.6. Наведіть формули, за якими здійснюють оцінювання параметрів рівняння парної лінійної регресії.

14.7. Як перевірити значущість кожного з параметрів рівняння парної регресії?

14.8. Як побудувати довірчі інтервали для оцінювання параметрів лінії регресії за вибірковими даними?

14.9. Як визначити довірчу смугу, до якої теоретична лінія регресії буде належати із заданою надійністю?

14.10. Запишіть формулу для визначення меж довірчого інтервалу для прогнозного значення за вибірковим рівнянням регресії.

14.7. Висновки за підрозділом 14

Процес кореляційного і регресійного аналізу становлять такі послідовні етапи: попереднє угруповання статистичних даних і виявлення форми зв'язку; складання рівнянь парної регресії за кожним фактором; оцінювання тісноти зв'язку, надійності й достовірності побудованої залежності; розроблення регресійної моделі явища, що вивчають, оцінювання її точності та визначення сили впливу врахованих факторів; аналіз досліджуваного явища за допомогою складеної моделі.

Регресійну модель здебільшого розглядають як інструмент аналізу, планування та управління. Звідси особливо жорсткі вимоги ставлять до надійності, адекватності й точності кожного коефіцієнта моделі.

Література: [1 – 3; 5; 8; 10 – 13; 15 – 17].

Рекомендована література

1. Барковський В. В. Теорія ймовірностей та математична статистика / В. В. Барковський, Н. В. Барковська, О. К. Лопатін. – 5-те вид. – Київ : Центр учбової літератури, 2010. – 424 с.
2. Горбачук В. М. Теорія ймовірностей та математична статистика [Електронний ресурс] : підручник / В. М. Горбачук, О. І. Кушлик-Дивульська. – Київ : КПІ ім. Ігоря Сікорського, 2023. – 351 с. – Режим доступу : <https://ela.kpi.ua/server/api/core/bitstreams/4b0ef359-532f-44b9-9436-4203204734db/content>.
3. Железнякова Е. Ю. Теорія ймовірностей та математична статистика [Електронний ресурс] : практикум / Е. Ю. Железнякова, Л. О. Норік. – Харків : ХНЕУ ім. С. Кузнеця, 2019. – 321 с. – Режим доступу : <http://repository.hneu.edu.ua/handle/123456789/21436>.
4. Жлуктенко В. І. Теорія ймовірностей і математична статистика : навч.-метод. посіб. У 2 ч. Ч. І. Теорія ймовірностей / В. І. Жлуктенко, С. І. Наконечний. – Київ : ХНЕУ, 2000. – 304 с.
5. Жлуктенко В. І. Теорія ймовірностей і математична статистика : навч.-метод. посіб. У 2 ч. Ч. ІІ. Математична статистика / В. І. Жлуктенко, С. І. Наконечний, С. С. Савіна. – Київ : ХНЕУ, 2001. – 336 с.
6. Збірник вправ з розділу "Теорія ймовірностей та математична статистика" навчальної дисципліни "Математика для економістів" для студентів галузі знань "Економіка і підприємництво" усіх форм навчання / уклад. Е. Ю. Железнякова, А. В. Ігначкова, З. Г. Попова та ін. – Харків : ХНЕУ, 2009. – 116 с.
7. Кігель В. Р. Теорія ймовірностей для економістів і менеджерів : навч. посіб. / В. Р. Кігель, О. І. Шаров. – Київ : Університет економіки та права "КРОК", 2018. – 144 с.
8. Лабораторний практикум з навчальної дисципліни "Теорія ймовірностей та математична статистика" : навч. посіб. / Е. Ю. Железнякова, І. Л. Лебедева, Л. О. Норік, К. В. Степанова. – Харків : ХНЕУ ім. С. Кузнеця, 2016. – 184 с.
9. Малярець Л. М. Математика для економістів. Теорія ймовірностей та математична статистика : навч. посіб. У 3 ч. Ч. 3 / Л. М. Малярець, І. Л. Лебедева, Л. Д. Широкоград. – Харків : ХНЕУ, 2011. – 568 с.

10. Малярець Л. М. Практикум з теорії ймовірностей та математичної статистики в Excel : навч.-практ. посіб. / Л. М. Малярець, І. Л. Лебедева, Е. Ю. Железнякова. – Харків : ХНЕУ, 2007. – 160 с.
11. Малярець Л. М. Теорія ймовірностей і математична статистика у вправах, прикладах та задачах : навч.-практ. посіб. / Л. М. Малярець, А. В. Ігначкова, Л. Д. Широкоград. – Харків : ХНЕУ, 2010. – 548 с.
12. Пістунів І. М. Теорія ймовірності та математична статистика для економістів з елементами електронних таблиць : навч. посіб. / І. М. Пістунів, І. Ю. Турчанінова. – Дніпро : «НТУ ДП», 2023. – 174 с.
13. Практикум з теорії ймовірностей та математичної статистики: навч. посіб. для студ. екон. спец. / А. М. Алілуйко, Н. В. Дзюбановська, В. О. Єрьоменко та ін. – Тернопіль : Підручники і посібники, 2023. – 352 с.
14. Радзієвська О. І. Спеціальні розділи математики для економістів (ч. I, II) : навч. посіб. / О. І. Радзієвська, Н. Л. Кузьмінська, Ю. О. Васютинська. – Київ : НУХТ, 2016. – 324 с.
15. Сайт персональних навчальних систем ХНЕУ ім. С. Кузнеця. – Режим доступу : <https://pns.hneu.edu.ua/course/index.php?categoryid=317>.
16. Теорія ймовірностей та математична статистика [Електронний ресурс] : зб. задач / С. М. Григулич, В. П. Лісовська, О. І. Макаренко та ін. – Київ : КНЕУ, 2014. – 277 с. – Режим доступу : <https://ir.kneu.edu.ua/server/api/core/bitstreams/e88570a8-60e2-49ef-9ce3-f00402ea0526/content>.
17. Теорія ймовірностей та математична статистика : навч. посіб. / Л. М. Малярець, І. Л. Лебедева, Е. Ю. Железнякова та ін. – Харків : ХНЕУ, 2010. – 404 с.

Додатки

Додаток А

$$\text{Значення функції } P(k, \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$\lambda \backslash k$	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	0,5458	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	0,0002	0,0007	0,0016	0,003	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
$\lambda \backslash k$	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	0,0572	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	0,0037	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	0,0126	0,0263	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

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	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3726	0,3712	0,3698
0,4	0,3683	0,3668	0,3652	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0395	0,0387	0,0379	0,0371	0,0363
2,2	0,0353	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001

Таблиця значень функції $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{x^2}{2}} dx$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2356	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3906	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4274	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4430	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4648	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4700	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4762	0,4757
2,0	0,4772	0,4778	0,4783	0,4788	0,4796	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4874	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4903	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4924	0,4927	0,4929	0,4930	0,1932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4958	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4973
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4980	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4986	0,4986	0,4987	0,4987	0,4988	0,4988	0,4988	0,4989	0,4989	0,4990
3,1	0,4990	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993
3,2	0,4993	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998	0,4998	0,4998	0,4998	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,5000	0,5000	0,5000	0,5000	0,5000
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

Таблиці значень $t = t(\gamma, n)$ і $q = q(\gamma, n)$

Таблиця Г.1

Таблиця значень $t = t(\gamma, n)$

n	γ			n	γ		
	0,95	0,99	0,999		0,95	0,99	0,999
5	2,78	4,60	8,61	20	2,093	2,861	3,883
6	2,57	4,03	6,86	25	2,064	2,797	3,745
7	2,45	3,71	5,96	30	2,045	2,756	3,659
8	2,37	3,50	5,41	35	2,032	2,720	3,600
9	2,31	3,36	5,04	40	2,023	2,708	3,558
10	2,26	3,25	4,78	45	2,016	2,692	3,527
11	2,23	3,17	4,59	50	2,009	2,679	3,502
12	2,20	3,11	4,44	60	2,001	2,662	3,464
13	2,18	3,06	4,32	70	1,996	2,649	3,439
14	2,16	3,01	4,22	80	1,991	2,640	3,418
15	2,15	2,98	4,14	90	1,987	2,633	3,403
16	2,13	2,95	4,07	100	1,984	2,627	3,392
17	2,12	2,92	4,02	120	1,980	2,617	3,374
18	2,11	2,90	3,97	∞	1,960	2,576	3,291
19	2,10	2,88	3,92				

Таблиця Г.2

Таблиця значень $q = q(\gamma, n)$

n	γ			n	γ		
	0,95	0,99	0,999		0,95	0,99	0,999
5	1,37	2,67	5,64	20	0,37	0,58	0,88
6	1,09	2,01	3,88	25	0,32	0,49	0,73
7	0,92	1,62	2,98	30	0,28	0,43	0,63
8	0,80	1,38	2,42	35	0,26	0,38	0,56
9	0,71	1,20	2,06	40	0,24	0,35	0,50
10	0,65	1,08	1,80	45	0,22	0,32	0,46
11	0,59	0,98	1,60	50	0,21	0,30	0,43
12	0,55	0,90	1,45	60	0,188	0,269	0,38
13	0,52	0,83	1,33	70	0,174	0,245	0,34
14	0,48	0,78	1,23	80	0,161	0,226	0,31
15	0,46	0,73	1,15	90	0,151	0,211	0,29
16	0,44	0,70	1,07	100	0,143	0,198	0,27
17	0,42	0,66	1,01	150	0,115	0,160	0,211
18	0,40	0,63	0,96	200	0,099	0,136	0,185
19	0,39	0,60	0,92	250	0,089	0,120	0,162

Критичні точки розподілу Стьюдента

Рівень значущості α (двобічна критична область)				
Кількість ступенів свободи k	$\alpha = 0,01$	$\alpha = 0,02$	$\alpha = 0,05$	$\alpha = 0,1$
1	63,7	31,82	12,7	6,31
2	9,92	6,97	4,30	2,92
3	5,84	4,54	3,18	2,35
4	4,60	3,75	2,78	2,13
5	4,03	3,37	2,57	2,01
6	3,71	3,14	2,45	1,94
7	3,50	3,00	2,36	1,89
8	3,36	2,90	2,31	1,86
9	3,25	2,82	2,26	1,83
10	3,17	2,76	2,23	1,81
11	3,11	2,72	2,20	1,80
12	3,05	2,68	2,18	1,78
13	3,01	2,65	2,16	1,77
14	2,98	2,62	2,14	1,76
15	2,95	2,60	2,13	1,75
16	2,92	2,58	2,12	1,75
17	2,90	2,57	2,11	1,74
18	2,88	2,55	2,10	1,73
19	2,86	2,54	2,09	1,73
20	2,85	2,53	2,09	1,73
21	2,83	2,52	2,08	1,72
22	2,82	2,51	2,07	1,72
23	2,81	2,50	2,07	1,71
24	2,80	2,49	2,06	1,71
25	2,79	2,49	2,06	1,71
26	2,78	2,48	2,06	1,71
27	2,77	2,47	2,05	1,71
28	2,76	2,46	2,05	1,70
30	2,75	2,46	2,04	1,70
40	2,70	2,42	2,02	1,68
60	2,66	2,39	2,00	1,67
120	2,62	2,36	1,98	1,66
∞	2,58	2,33	1,96	1,64
k	$\alpha = 0,005$	$\alpha = 0,01$	$\alpha = 0,025$	$\alpha = 0,05$
Рівень значущості α (однобічна критична область)				

Критичні точки розподілу F Фішера – Снедекора

Таблиця Е.1

Критичні точки розподілу F Фішера – Снедекора
за рівнем 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	18,5	19,0	19,2	19,2	19,3	19,3	19,3	19,4	19,4	19,4	19,4	19,4
3	10,1	9,55	9,28	9,13	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	4,16	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	3,22	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
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,35	2,31	2,28
24	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	2,30	2,26	2,22	2,18
30	4,17	3,32	2,92	2,69	2,53	2,42	2,34	2,27	2,21	2,16	2,12	2,09
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,07	2,04	2,00
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,02	1,98	1,95
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,01	1,97	1,93	1,89
100	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	1,97	1,92	1,88	1,85
200	3,89	3,04	2,65	2,41	2,26	2,14	2,05	1,98	1,92	1,87	1,83	1,80
400	3,86	3,02	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,81	1,78
∞	3,84	2,99	2,60	2,37	2,21	1,09	2,01	1,94	1,88	1,83	1,79	1,75

**Критичні точки розподілу F Фішера – Снедекора
за рівнем значущості $\alpha = 0,01$**

k_2	k_1									
	1	2	3	4	5	6	8	12	24	∞
1	4 052	4 999	5 403	5 625	5 764	5 859	5 981	6 106	6 234	6 366
2	98,5	99,0	99,2	99,3	99,3	99,4	99,3	99,4	99,5	99,5
3	34,1	30,8	29,5	28,7	28,2	27,9	27,5	27,1	26,6	26,1
4	21,2	18,0	16,7	16,0	15,5	15,2	14,8	14,4	13,9	13,5
5	16,3	13,3	12,1	11,4	11,0	10,7	10,3	9,9	9,5	9,6
6	13,7	10,9	9,8	9,2	8,8	8,5	8,1	7,7	7,3	6,9
7	12,3	9,6	8,5	7,9	7,5	7,2	6,8	6,5	6,1	5,7
8	11,3	8,7	7,6	7,0	6,6	6,5	6,0	5,7	5,3	4,9
9	10,6	8,0	7,0	6,4	6,1	5,8	5,5	5,1	4,7	4,3
10	10,0	7,6	6,6	6,0	5,6	5,4	5,1	4,7	4,3	3,9
11	9,7	7,2	6,2	5,7	5,3	5,1	4,7	4,4	4,0	3,6
12	9,3	6,9	6,0	5,4	5,1	4,8	4,5	4,2	3,8	3,4
13	9,1	6,7	5,7	5,2	4,9	4,6	4,3	4,0	3,6	3,2
14	8,9	6,5	5,6	5,0	4,7	4,5	4,1	3,8	3,4	3,0
15	8,7	6,4	5,4	4,9	4,6	4,3	4,0	3,7	3,3	2,9
16	8,5	6,2	5,3	4,8	4,4	4,2	3,9	3,6	3,2	2,8
17	8,4	6,1	5,2	4,7	4,3	4,1	3,8	3,5	3,1	2,7
18	8,3	6,0	5,1	4,6	4,3	4,0	3,7	3,4	3,0	2,6
19	8,2	5,9	5,0	4,5	4,2	3,9	3,6	3,3	2,9	2,4
20	8,1	5,9	4,9	4,4	4,1	3,9	3,6	3,2	2,9	2,4
22	7,9	5,7	4,8	4,3	4,0	3,8	3,5	3,1	2,8	2,3
24	7,8	5,6	4,7	4,2	3,9	3,7	3,3	3,0	2,7	2,2
26	7,7	5,5	4,6	4,1	3,8	3,6	3,3	3,0	2,6	2,1
28	7,6	5,5	4,6	4,1	3,8	3,5	3,2	2,9	2,5	2,1
30	7,6	5,4	4,5	4,0	3,7	3,5	3,2	2,8	2,5	2,0
40	7,3	5,2	4,3	3,8	3,5	3,3	3,0	2,7	2,3	1,8
60	7,1	5,0	4,1	3,7	3,3	3,1	2,8	2,5	2,1	1,6
120	6,9	4,8	4,0	3,5	3,2	3,0	2,7	2,3	2,0	1,4
∞	6,6	4,6	3,8	3,3	3,0	2,8	2,5	2,2	1,8	1,0

Критичні точки розподілу χ^2

Кількість ступенів свободи k	Рівень значущості α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,362	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,7	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,76	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

Предметний покажчик

Б

Баєса формули 30
Бернуллі схема 45
– формула 45

В

Варіанта 198
Варіаційний ряд 197
– – дискретний 198
– – інтервальний 198
Вибірка (вибіркова сукупність) 197
– репрезентативна 197
Вибіркова дисперсія 215
– середня 214
Вибірковий метод 197
Випадкова величина 64
– – двовимірна 134
– – закон розподілу 64
– – дискретна 64
– – способи задання 65
– – неперервна 64
– – одновимірна 64
Випадкові події 11
– – незалежні 29
– – несумісні 30
– – сумісні 30
– – протилежні 29
Виправлена дисперсія 215
Випробування 11
Відбір 197
– безповторний 197
– повторний 197

Г

Генеральна сукупність 196
Гіпотеза статистична 229

Гістограма частот 201
– відносних частот 201

Д

Дисперсія випадкової величини 66
– властивості 67
Дисперсійний аналіз 284
– – двофакторний 287
– – однофакторний 284
Діаграми Венна – Ейлера 11
Довірчий інтервал 216
– – прогнозу 308

З

Закон великих чисел 181
Закон розподілу 64
– – біноміальний 87
– – геометричний 88
– – гіпергеометричний 89
– – двовимірної випадкової величини 135
– – нормальний 105
– – показниковий 103
– – Пуассона 88
– – рівномірний 102

Й

Ймовірність події 10

К

Коефіцієнт асиметрії 69
– варіації 68, 216
– ексцесу 69
– коваріації 143, 261
– кореляції 144, 261
– рангової кореляції Спірмена 265
– регресії 304
Композиція 148

Кореляційна таблиця 258
– залежність 259
Кореляційне відношення 263
– поле 260
Кореляційний аналіз 258
– момент 142
Критична область 230
Крок 199
Кумулята 200

Л

Лінія регресії 303
– – емпірична 303
– – довірча смуга 308

М

Математичне сподівання 66
– – властивості 66
– – дискретної випадкової
величини 66
– – умовне 141
Матриця коваріаційна 143
Медіана 69, 215
Многокутник розподілу 65
Мода 69, 215
Момент початковий 68, 142, 216
– центральний 68, 142, 216
– кореляційний 142

Н

Найімовірніше число появи події 46
Нерівність Маркова 181
– Чебишова 181

О

Обсяг вибіркової сукупності 197
Оцінка статистична 213
– – ефективна 213

– – зсунута 213
– – інтервальна 214
– – незсунута 213
– – спроможна 214
– – точкова 214

П

Перестановки 9
Повна група випадкових подій 30
Полігон 200
Потужність критерію 230
Помилка 230
– другого роду 230
– першого роду 230
Правило добутку 10
– сум 10
– трьох сигм 108
Простір елементарних подій 11

Р

Ранг 265
Регресія 303
– від'ємна 304
– додатна 304
– лінійна 304
– множинна 304
– нелінійна 304
– парна 304
Рівень значущості 216
Рівняння регресії 303
Розмах варіаційного ряду 199
Розміщення 9
Розподіл біноміальний 87
– варіаційний 197
– умовний 138

Ряд розподілу 64

– – дискретної випадкової величини 64

– – статистичний 200

С

Середнє квадратичне відхилення 67, 215

– – – вибіркове 215

– – – виправлене 215

Сполучення 10

Статистична гіпотеза 229

– – альтернативна (конкурентна) 229

– – непараметрична 230

– – неспрямована 229

– – основна (нульова) 229

– – параметрична 230

– – спрямована 229

Статистичний критерій 230

– – Колмогорова 239

– – Пірсона 236

– – Романовського 238

– – Стьюдента 217

– – Фішера – Снедекора 231

Сукупність вибіркова 197

– генеральна 196

Сума квадратів відхилень загальна (SST) 285

– – помилок моделі (SSE) 285

– – пов'язані з регресією (SSR) 285

Схема Бернуллі 45

Т

Теорема Бернуллі 182

– Лапласа інтегральна 47

– Ляпунова 184

– множення ймовірностей 28, 29

– Муавра – Лапласа локальна 46

– Хінчина 183

– центральна гранична 183

– Чебишова 182

У

Умовне середнє 259

Ф

Формула повної ймовірності 30

– Пуассона 46

– Стерджеса 198

Функція випадкового аргументу 144

– Гаусса 47, 107

– двох випадкових аргументів 148

– диференціальна 100

– інтегральна 101

– Лапласа 47

– умовна диференціальна 139

Функція розподілу ймовірностей 65

– – – випадкової величини 65

– – – двовимірної випадкової величини 136

– – емпірична 200

Ч

Частота 198

– відносна 198

– емпірична 237

НАВЧАЛЬНЕ ВИДАННЯ

Железнякова Еліна Юріївна
Норік Лариса Олексіївна

ТЕОРІЯ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНА СТАТИСТИКА

Навчальний посібник

Самостійне електронне текстове мережеве видання

Відповідальний за видання *Л. М. Малярець*

Відповідальний редактор *О. С. Вяткіна*

Редактор *О. Г. Доценко*

Коректор *О. Г. Доценко*

План 2025 р. Поз. № 20-ЕНП. Обсяг 338 с.

Видавець і виготовлювач – ХНЕУ ім. С. Кузнеця, 61165, м. Харків, просп. Науки, 9-А

*Свідоцтво про внесення суб'єкта видавничої справи до Державного реєстру
ДК № 4853 від 20.02.2015 р.*