

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Одеський національний університет імені І.І. Мечникова

П. Д. Варбанець, Н. А. Якімова

ЛІНГВОСТАТИСТИКА

Навчальний посібник

Одеса
ОНУ
2021

УДК 519.2:81'32/'33(075.8)

B18

Автори:

П. Д. Варбанець, доктор фізико-математичних наук, професор, зав. кафедри комп'ютерної алгебри та дискретної математики;

Н. А. Якімова, кандидат технічних наук, доцент, доцент кафедри комп'ютерної алгебри та дискретної математики.

*Рекомендовано до друку Науково-методичною радою
Одеського національного університету імені І.І. Мечникова
протокол № 6 від 15.10.2020 р.*

Рецензенти:

В. М. Пивоварчик, доктор фізико-математичних наук, професор, зав. кафедри вищої математики і статистики ДЗ ПНПУ імені К.Д.Ушинського;

Г. В. Ситар, доктор філологічних наук, доцент, професор кафедри загального та прикладного мовознавства Донецького національного університету імені Василя Стуса.

Варбанець П. Д., Якімова Н. А.

B18

Лінгвостатистика : навчальний посібник / П. Д. Варбанець, Н. А. Якімова ; Одес. нац. ун-т ім. І.І. Мечникова. – Одеса, 2021. – 182 с.

ISBN 978-617-689-428-5

У пропонованому навчальному посібнику розглядаються деякі основні поняття теорії ймовірностей та математичної статистики. У ньому розкритий зв'язок між строго математичною дисципліною та можливістю її застосування в моделюванні мови, що дозволяє студентам зрозуміти місце та роль їхньої спеціальності в рішенні проблем сучасної науки. В навчальному посібнику викладені правила обчислення ймовірності подій у різних умовах їх настання, методи математичної статистики при проведенні лінгвістичних експериментів, засоби аналізу отриманих результатів на предмет їх відповідності справжнім параметрам досліджуваних величин, а також засоби лінгвістичної інтерпретації отриманих математичних результатів.

Посібник складений для студентів першого (бакалаврського) рівня освіти спеціальності 035 «Філологія» спеціалізації 035.01 «Українська мова та література (прикладна лінгвістика)».

УДК 519.2:81'32/'33(075.8)

ISBN 978-617-689-428-5

© Варбанець П.Д., Якімова Н.А., 2021

© Одеський національний університет
імені І.І. Мечникова, 2021

Зміст

ПЕРЕДМОВА.....	6
1. ІМОВІРНІСНІ ПРОСТОРИ	
1.1. Класичне визначення ймовірності.....	9
1.2. Урнова схема.....	11
1.3. Скінченна схема з неоднаково можливими результатами	14
1.4. Вирахування подій	18
1.5. Аксиоматична побудова теорії ймовірностей	21
1.6. Контрольні запитання	24
1.7. Завдання для самостійного опрацювання	24
1.8. Контрольні завдання	30
2. УМОВНІ ЙМОВІРНОСТІ	
2.1. Повна ймовірність	31
2.2. Формула Байєса	33
2.3. Послідовності випробувань.....	35
2.4. Контрольні запитання	37
2.5. Задачі для самостійного опрацювання	37
2.6. Контрольні завдання	40
3. ВИПАДКОВА ВЕЛИЧИНА	
3.1. Визначення випадкової величини й її функція розподілу	41
3.2. Розподіл дискретних випадкових величин	44
3.2.1. Рівномірний розподіл на множині $\{1, 2, \dots, n\}$	48
3.2.2. Гіпергеометричний розподіл	48
3.2.3. Геометричний розподіл.....	48
3.2.4. Розподіл Пуассона	49

3.2.5. Біноміальний розподіл	50
3.3. Розподіл безперервних випадкових величин	50
3.3.1. Рівномірний розподіл	54
3.3.2. Експонентний розподіл	54
3.3.3. Нормальний розподіл	55
3.4. Контрольні запитання	56
4. ЧИСЛОВІ ХАРАКТЕРИСТИКИ ВИПАДКОВОЇ ВЕЛИЧИН	
4.1. Вибіркове середнє й вибірка дисперсія	57
4.2. Критерій нормальності розподілу	62
4.3. Критерій однорідності	69
4.4. Розподіл середнього арифметичного значення	72
4.5. Контрольні запитання	76
4.6. Задачі для самостійного опрацювання	77
4.7. Контрольні завдання	87
5. ЛІНГВІСТИЧНІ ГІПОТЕЗИ	
5.1. Постановка й перевірка лінгвістичних гіпотез	90
5.2. Критерій Стюдента	91
5.3. Критерій Ван дер Вардена	96
5.4. Контрольні запитання	99
5.5. Задачі для самостійного опрацювання	99
5.6. Контрольні завдання	104
6. ВИВЧЕННЯ ЗАЛЕЖНОСТІ ЛІНГВІСТИЧНИХ ОЗНАК	
6.1. Кореляційна залежність	105
6.2. Лінійна кореляція	110
6.3. Кореляційні відносини	117
6.4. Парціальна кореляція	118
6.5. Критерії кореляційного аналізу	119

6.5.1. Критерій вірогідності залежності ознак X і Y	119
6.5.2. Критерій значимості розходження кореляційних відносин.....	120
6.6. Регресійний аналіз.....	120
6.7. Контрольні запитання	125
6.8. Задачі для самостійного опрацювання	125
6.9. Контрольні завдання	130
7. ВАРІАНТИ ЗАВДАНЬ МОДУЛЬНОГО КОНТРОЛЮ	131
7.1. Модуль 1.....	131
7.2. Модуль 2.....	145
СПИСОК ЛІТЕРАТУРИ.....	174
Додаток 1	174
Додаток 2	175
Додаток 3	176
Додаток 4	177
Додаток 5	?

ПЕРЕДМОВА

Мова являє собою, як прийнято говорити в сучасному мовознавстві, деяке системно-структурне утворення. Окремі підсистеми мови називають *рівнями*, які представлені відповідними одиницями – фонемами, морфемами, лексемами, синтагмами (реченнями).

Оскільки одиниці кожного рівня мови перебувають в ієрархічній залежності від одиниць вищестоящего рівня, то зрозуміло, що, наприклад, число похідних слів у тій або іншій мові буде залежати від кількості афіксів з дериваційним значенням, а кількість морфем – від кількості фонем. У той же час кількість фонем у різних мовах не збігається. Ці прості приклади показують, що мова характеризується певними якісними й кількісними ознаками [6].

Якісний аналіз мови являє собою його категоризацію, тобто виділення в мові певних класів явищ, об'єднаних певними якісними ознаками. Цими явищами (категоріями) можуть бути одиниці мови (фонема, морфема, лексема), граматичні категорії, граматичні способи (афіксація, словоскладання, редуплікація тощо), типи слів (знаменні, службові; вульгаризми, діалектизми; архаїзми, неологізми тощо), типи речень (складні, прості; сурядні, підрядні тощо). Однак будь-яка категоризація, тобто якісний аналіз мови, нерозривно пов'язана із *квантифікацією* мови, тобто його кількісним аналізом. Таким чином, стає очевидним, що мова поряд з якісними ознаками володіє й кількісними. Ще більшою мірою має кількісні ознаки мова і її письмове втілення – текст.

У сучасній науці розрізняють так звані «добре організовані системи» і «погано організовані (дифузійні) системи» [9, стор.7]. До добре організованих систем належить, наприклад, рух планет. Завдяки чіткій упорядкованості цієї системи стає можливим точно обчислити й заздалегідь передбачити, наприклад, час сонячного затемнення. До погано організованих систем належить інтелектуальна діяльність людини, а разом з нею і мовне поведіння, тобто використання мови. Уважається, що

найбільш ефективними методами вивчення погано організованих систем є методи математичної статистики.

Таким чином, мова може бути дослідженою за допомогою якісних і кількісних методів. Залежно від цілей і задач, які ставить перед собою лінгвіст при вивченні явищ мови й мовлення, у здійснюваному дослідженні можуть застосовуватися або якісні, або кількісні методи аналізу, або й ті, й інші рівною мірою, або переважно перші або другі. Можуть виникнути також задачі, (особливо при аналізі тексту), які не можуть бути вирішені інакше, як за допомогою кількісних методів.

У самому математичному апараті, точніше, у сукупності математичних методів можна умовно розрізнити кількісні й не-кількісні методи [10]. За допомогою не кількісних методів (теорія множин, теорія алгоритмів, математична логіка) доцільно вивчати, насамперед, систему мови. Цей розділ науки називається *комбінаторна лінгвістика*. За допомогою кількісних методів (насамперед, теорія ймовірностей і математична статистика) доцільно досліджувати мову (текст). Цей другий напрямок називають *квантитативною лінгвістикою*. Якщо розглядати лінгвостатистику як одну зі складових частин квантитативної лінгвістики, стає очевидним, що між лінгвостатистикою і квантитативною (математичною) лінгвістикою існує помітна різниця, тому що кількість об'єктів і набір методів, за допомогою яких ці об'єкти вивчаються в лінгвостатистиці, значно вужче, ніж у математичній лінгвістиці в цілому.

За допомогою квантитативних методів у цей час досліджуються всі мовні підсистеми, тобто всі рівні мови й мовлення. Є спроби яким-небудь чином систематизувати лінгвостатистичні дослідження, тобто виділити ті галузі мовознавства, де квантитативні методи успішно застосовуються або можуть бути ефективно використані. Приведемо деякі сфери використання математичних методів у мовознавстві [6].

1. Дослідження фонетичного ладу мови (частота зустрічальності звуків, букв, букво- і словосполучень і т.д.).
2. Дослідження лексичного складу мови й частоти зустрічальності слів у тексті. Найважливішою задачею в цій галузі є створення частотних словників.

3. Дослідження авторського й функціонального стилів. Сюди ж можна віднести дослідження, пов'язані із установленням авторства твору. Можливості застосування статистичних методів для дослідження властивостей тексту безмежні, а потреби вивчення різних одиниць тексту надзвичайно великі. Властивості й відмінні риси текстів можуть бути досліджені з урахуванням різних параметрів і ознак: хронологічні, тематичні, жанрові, гендерні, соціально-статусні, вікові, лексичні, морфологічні, синтаксичні й інші кількісні характеристики тексту й уживаних у ньому одиниць.
4. Кількісна характеристика різних одиниць мови, вивчення довжини слова в різних текстах і різних мовах, вивчення довжини складів, морфем, речень.
5. Дослідження швидкості (темпів) і закономірностей розвитку й зміни мови - насамперед його лексичного складу.
6. Типологічне (порівняльне) вивчення різних мов і їхніх підсистем.
7. Квантитативне дослідження діалектології.
8. Квантитативний аналіз даних, отриманих за допомогою психолінгвістичних експериментів (асоціативний експеримент і семантичний диференціал Ч. Остуда).
9. Квантитативний аналіз семантичних і формальних відносин у реконструйованому за допомогою порівняльно-історичного методу лексичному складі прамови.
10. Дослідження семантики мови - парадигматичних і синтагматичних відносин у лексиці, синонімії, полісемії й інших явищах.
11. Перекладознавство. Є роботи, у яких кількісні методи використовуються при зіставленні двох мов при перекладі.
12. Методика викладання іноземних мов.

1. ІМОВІРНІСНІ ПРОСТОРИ

1.1. Класичне визначення ймовірності

Достовірною називається подія, що обов'язково відбудеться при здійсненні певного комплексу умов. Відповідно, *неможливою* називається подія, що при заданому комплексі умов не відбудеться ніколи. *Випадковою* називається така подія, що при заданому комплексі умов може як відбутися, так і не відбутися [1]. Міра можливості здійснення такої події і є її *ймовірність*. Достовірна й неможлива події можуть розглядатися як крайні окремі випадки випадкових подій.

Випадкові події будемо позначати великими латинськими буквами $A, B, C, \dots$. Достовірна подія позначається буквою Ω , неможлива - $\emptyset$. Уведемо тепер деякі відносини між подіями. Дві події A і B *несумісні*, якщо настання однієї з них виключає настання іншої. *Сума подій A і B* – це така третя подія $C=A+B$, що відбувається тоді, коли настає або подія A , або подія B , або вони обидві одночасно. *Добуток подій A і B* – це така третя подія $C=AB$, що настає тоді, коли відбуваються й подія A , і подія B . Подія $\bar{A}$ *протилежна* події A , якщо вона несумісна з подією A і разом з нею утворює достовірну подію, тобто $\bar{A} + A = \Omega$.

Однією з моделей зі скінченним числом результатів є класична імовірнісна схема. У цій схемі визначення ймовірності ґрунтується на рівній можливості кожного зі скінченного числа результатів. Таке визначення виникло на основі перших спроб вирахування шансів в азартних іграх. Так, у випадку із гральною кісткою при однократному киданні однакова можливість випадання кожної із шести граней, на які нанесені цифри 1, 2, 3, 4, 5, 6. Позначимо ці однаково можливі результати або *елементарні події* через $\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6$. Природно, що шанс здійснитися не одному результату, а одному із двох, наприклад, або ω_1 , або ω_2 , у два рази більше. Міркуючи таким чином, можна визначити шанси здійснення будь-якої складеної події, що складається з декількох елементарних.

У загальному випадку, коли є n однаково можливих елементарних подій $\omega_1, \dots, \omega_n$, ймовірність будь-якої складеної події A , що складається з m елементарних подій $\omega_{i_1}, \dots, \omega_{i_m}$, визначається як відношення кількості елементарних подій, що сприяють події A , до загальної кількості елементарних подій, тобто

$$P(A) = \frac{m}{n} \quad (1.1)$$

Наприклад, у випадку із гральною кісткою ймовірність події A , що полягає у випаданні парної кількості вічок (тобто $A = \{\omega_2, \omega_4, \omega_6, \dots\}$), дорівнює $P(A) = 3/6 = 1/2$, тому що в подію A входять три елементарних події, а загальна кількість елементарних подій дорівнює шести.

Із класичного визначення ймовірностей, зокрема, випливає, що ймовірність повної події Ω , що включає всі n елементарних подій, дорівнює $P(\Omega) = n/n = 1$. Але тоді повна подія Ω , що полягає в появі будь-якої з усього набору елементарних подій $\omega_1, \omega_2, \dots, \omega_n$, і є достовірною подією, тому що вона обов'язково відбувається. Тому ймовірність достовірної події дорівнює одиниці.

Якщо події розглядати як підмножини множини елементарних подій, то відносини між подіями, введені вище, можна інтерпретувати як співвідношення між множинами. Несумісні події – це такі події, які не містять спільних елементів. Сума $A+B$ і добуток AB – це відповідно їхнє об'єднання $A \cup B$ і перетинання $A \cap B$. Протилежна подія $\bar{A}$ – це доповнення A . Запис $A \subset B$ означає, що в B містяться всі елементарні події з A , і можуть міститися елементарні події, що не входять в A . Якщо $A \subset B$ і $B \subset A$, то $A=B$.

У випадку класичного визначення ймовірності справедлива наступна теорема додавання ймовірностей.

Теорема 1.1 (додавання ймовірностей). Формулювання.

Якщо дві складені події $A = \{\omega_{i_1}, \dots, \omega_{i_m}\}$ і $B = \{\omega_{j_1}, \dots, \omega_{j_k}\}$ є несумісними, то ймовірність об'єднаної події $C = A \cup B$ дорівнює сумі ймовірностей цих двох подій.

Доказ. Дійсно, імовірності подій A і B дорівнюють $\frac{m}{n}$ й $\frac{k}{n}$

відповідно. Подія $C=A \cup B = \{\omega_{i_1}, \dots, \omega_{i_m}, \omega_{j_1}, \dots, \omega_{j_k}\}$ містить $m+k$ елементарних подій, тому що за умовою теореми серед елементарних подій $\{\omega_{i_1}, \dots, \omega_{i_m}\}$ немає жодної, котра входила б у набір $\{\omega_{j_1}, \dots, \omega_{j_k}\}$. Тому, відповідно до класичного визначення, її ймовірність дорівнює

$$P(C) = \frac{m+k}{n} = \frac{m}{n} + \frac{k}{n} = P(A) + P(B).$$

Теорему доведено.

Подія $\bar{A}$ називається *протилежною* стосовно A , якщо до неї входять всі елементарні події, що не входять в A . Іншими словами, A і $\bar{A}$ - це такі несумісні події, які разом утворюють достовірну подію, тобто $A \cup \bar{A} = \Omega$. З теореми додавання випливає, що $P(\Omega) = P(A) + P(\bar{A}) = 1$, тому $P(\bar{A}) = 1 - P(A)$. Звідси, зокрема, випливає, що ймовірність неможливої події $\emptyset$, що є протилежною стосовно достовірної події Ω , дорівнює нулю.

1.2. Урнова схема

Класична схема, незважаючи на всю свою обмеженість, придатна для вирішення ряду суцього практичних задач. Розглянемо, наприклад, деяку сукупність елементів об'єму N . Це можуть бути вироби, кожний з яких є придатним або бракованим. Подібного роду ситуації описуються урнвою схемою: в урні є N куль, з них M білих, $(N - M)$ чорних.

Наприклад, уявимо собі, що є тільки руйнуючі засоби контролю кожного виробу на придатність (наприклад, сірника). У такому випадку не можна обстежити всю партію виробів, а тільки частину її. Отже, з урни, що містить N куль, у якій перебуває невідома кількість M білих куль, витягається вибірка об'єму n . Така процедура називається *вибіркою без повернення*. Необхідно

визначити ймовірність того, що у вибірці буде виявлено m білих куль. Це задача на застосування класичного визначення ймовірності. Справді, в описаній ситуації кожна вибірка не має переваги стосовно будь-якої іншої, тобто всі вони однаково можливі. Підрахуємо кількість всіх можливих вибірок об'єму n з N елементів. Як відомо з комбінаторики, кількість способів, за допомогою яких можна вибрати n елементів із загальної їхньої кількості N , дорівнює числу сполучень із N по n , тобто $C_N^n = \frac{N!}{n!(N-n)!}$,

де $N! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot N$. Таким чином, загальна кількість результатів дорівнює C_N^n . З'ясуємо, скільки результатів із загальної кількості елементарних результатів сприяє події A , тобто наявності у вибірці об'єму n білих куль у кількості m . Кількість способів, якими можна з M білих куль витягти m штук, дорівнює C_M^m , а кількість способів вибрати з $(N-M)$ чорних куль $(n-m)$ штук дорівнює C_{N-M}^{n-m} . Тому кількість результатів, сприятливих події A , дорівнює $C_M^m \cdot C_{N-M}^{n-m}$ і, отже, її ймовірність, що дорівнює відношенню кількості сприятливих результатів до їхньої загальної кількості, така:

$$P(A) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n} = \frac{C_n^m C_{N-n}^{M-m}}{C_N^M} = P_{M,N}(m, n). \quad (1.2)$$

Приклад 1.1. Нехай є партія, що складається з 500 виробів, у якій два бракованих. Яка ймовірність у вибірці з 5 виробів не виявити жодного бракованого?

Рішення. Скористаємося формулою (1.2). Маємо

$$P_{498,500}(5, 5) = \frac{C_{498}^5 C_2^0}{C_{500}^5} = \frac{498!}{493! 5!} \cdot \frac{2!}{0! 2!} = \frac{500!}{495! 5!} = 0.98$$

Розглянемо тепер, наприклад, урну з кулями, вибірка куль із якої відбувається послідовно по одній кулі, і при цьому щоразу фіксується номер кулі, а сама куля повертається знову до урни. Така процедура називається *вибіркою з поверненням* [2].

У цьому випадку ймовірність події A , обчислена аналогічним способом, дорівнює

$$P(A) = C_n^m \frac{M^m (N-M)^{n-m}}{N^n} = C_n^m \left(\frac{M}{N} \right)^m \left(1 - \frac{M}{N} \right)^{n-m}.$$

Говорити про ймовірності як про міри можливості здійснення випадкової події A має сенс тільки при здійсненні певного комплексу умов. При зміні умов зміниться й імовірність. Так, якщо до комплексу умов, при якому вивчалася ймовірність $P(A)$, додати нову умову, що полягає в появі події B , то одержимо інше значення ймовірності $P(A/B)$ – умовну ймовірність події A за умови, що відбулася подія B . Ймовірність $P(A)$, на відміну від умовної, називається *безумовною*.

Виведемо *формулу умовної ймовірності*. Нехай подіям A і B сприяють m і k елементарних результатів з n . Тоді, згідно (1.1),

їхні безумовні ймовірності дорівнюють $\frac{m}{n}$ й $\frac{k}{n}$ відповідно. Нехай події A за умови, що подія B відбулася, сприяє r елементарних результатів. Тоді, згідно (1.1), умовна ймовірність події A

дорівнює $P(A/B) = \frac{r}{k}$. Розділивши й чисельник, і знаменник на n ,

одержимо формулу умовної ймовірності:

$$P(A/B) = \frac{\frac{r}{n}}{\frac{k}{n}} = \frac{P(A \cap B)}{P(B)}, \quad (1.3)$$

оскільки події $A \cap B$ відповідає r результатів і, отже, $\frac{r}{n}$ - його безумовна ймовірність.

Подія A називається *незалежною від B* , якщо її умовна ймовірність дорівнює безумовній, тобто $P(A/B) = P(A)$. При цьому з формули (1.3) одержуємо

$$P(A \cap B) = P(A) \cdot P(B), \quad (1.4)$$

тобто властивість незалежності взаємна й для незалежних подій імовірність їхнього одночасного здійснення дорівнює добутку їхніх ймовірностей. Формула (1.3), записана у вигляді

$$P(A \cap B) = P(A) \cdot P(A/B), \quad (1.5)$$

називається *формулою множення для залежних подій*, а формула (1.4) – *теоремою множення для незалежних подій*.

Наприклад, у експерименті із гральною кісткою нехай подія A полягає у випаданні числа вічків, що ділиться на три, тобто $A = \{\omega_3, \omega_6, \dots\}$, а подія B – у випаданні парного числа вічків, тобто $B = \{\omega_2, \omega_4, \omega_6, \dots\}$. Тоді $A \cap B = \omega_6$ і по формулі умовної ймовірності (1.3) одержуємо, що

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{3/6} = \frac{1}{3}.$$

Але $P(A) = 2/6 = 1/3$. Тому $P(A/B) = P(A)$, тобто події A і B незалежні.

Взаємність незалежності подій означає, що якщо подія A не залежить від події B , то й подія B не залежить від події A . Тоді при $P(A) > 0$ і з огляду на формулу (1.5) маємо:

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A/B) \cdot P(B)}{P(A)} = \frac{P(A) \cdot P(B)}{P(A)} = P(B)$$

1.3. Скінченна схема з неоднаково можливими результатами

Обмеженість класичного визначення ймовірності, зокрема, закладена в однаковій можливості результатів. Дійсно, навіть невелике ускладнення практичної ситуації негайно ввійде в суперечність із однаковою можливістю, що може розглядатися, скоріше, як окремий випадок більш загальної ситуації.

Розглянемо, наприклад, стрілянину по круговій мішені. Елементарними результатами тут є влучення в те або інше кільце кругової мішені. Влучення в мале внутрішнє коло оцінюється в 10 вічок, у навколишнє його кільце – в 9 вічок, у наступне – в 8 вічок й т.д., у саме зовнішнє кільце – 1 вічко, невлучення в

кругову мішень – 0 вічка. Таким чином, є 11 елементарних подій $\omega_{10}, \omega_9, \dots, \omega_1, \omega_0$. Для кожного стрільця певного класу є свої певні стійкі шанси (імовірності) вибити за один постріл ту або іншу кількість вічок $p_{10}, p_9, \dots, p_1, p_0$. Ці події, загалом кажучи, неоднаково можливі. Наприклад, для майстрів спорту, очевидно, виключена подія ω_0 , тому $p_0=0$, тобто відразу виключається однакова можливість.

Скінченна схема з неоднаково можливими результатами визначається в такий спосіб. Є скінченний набір елементарних подій $\Omega=\{\omega_1, \omega_2, \dots, \omega_n\}$, і для кожної елементарної події ω_i задана її ймовірність p_i , $0 \leq p_i \leq 1$, причому $\sum_{i=1}^n p_i = 1$. Імовірність будь-якої складеної події $A=\{\omega_{i_1}, \dots, \omega_{i_m}\}$ визначається як сума ймовірностей входних у неї елементарних подій:

$$P(A) = \sum_{l=1}^m p_{i_l}. \quad (1.6)$$

Ця схема є узагальненням класичної схеми. Справді, якщо повернутися до випадку однакової можливості й приписати кожній елементарній події ймовірність $\frac{1}{n}$, то формула (1.6) приводить до класичного визначення ймовірності.

У випадку скінченної схеми також має місце теорема додавання.

Теорема 1.2. Формулювання. Для двох несумісних подій A і B , що є підмножинами Ω , $P(A \cup B) = P(A) + P(B)$.

Доказ. Нехай $A=\{\omega_{i_1}, \dots, \omega_{i_m}\}$, $B=\{\omega_{j_1}, \dots, \omega_{j_k}\}$. Згідно (1.6),

$$P(A) = \sum_{l=1}^m p_{i_l}, \quad P(B) = \sum_{q=1}^k p_{j_q}.$$

Оскільки A і B несумісні, вони не мають спільних елементарних подій і, отже, $C=A \cup B = \{\omega_{i_1}, \dots, \omega_{i_m}, \omega_{j_1}, \dots, \omega_{j_k}\}$. На підставі (1.6) маємо

$$P(C) = \sum_{l=1}^m p_{i_l} + \sum_{q=1}^k p_{j_q} = P(A) + P(B).$$

Теорему доведено.

Так само, як скінченна схема з неоднаково можливими результатами є узагальненням класичної скінченної схеми з однаково можливими результатами, дискретна схема з нескінченною кількістю неоднаково можливих подій, у свою чергу, є узагальненням скінченної схеми.

У дискретній схемі множина $\Omega = \{\omega_1, \omega_2, \dots, \omega_n, \dots\}$, загалом кажучи, містить зліченну кількість елементарних подій. Для кожної елементарної події задана її ймовірність $p_i = P(\omega_i)$, $0 \leq p_i \leq 1$,

причому $\sum_{i=1}^{\infty} p_i = 1$. Ймовірність будь-якої скінченної або злічен-

ної підмножини $A \subset \Omega$ множини елементарних подій Ω дорівнює сумі ймовірностей елементарних подій, що її складають, тобто

якщо $A = \bigcup_{l=1}^{\infty} \omega_{i_l}$, то $P(A) = \sum_{l=1}^{\infty} p_{i_l}$. Якщо ж $A = \bigcup_{l=1}^m \omega_{i_l}$, то має місце (1.6).

У скінченній схемі, як і в класичній, можна вивести формулу умовної ймовірності. Розглянемо події $A = \{\omega_{i_1}, \dots, \omega_{i_m}\}$ і $B = \{\omega_{j_1}, \dots, \omega_{j_l}, \dots, \omega_{j_k}\}$ такі, що $\omega_{i_l} = \omega_{j_l}, \dots, \omega_{i_l} = \omega_{j_l}, l \leq m, k$. Інакше кажучи, $A \cap B = \{\omega_{i_1}, \dots, \omega_{i_l}\}$. Тоді

$$P(A) = \sum_{\mu=1}^m p_{i_{\mu}}, P(B) = \sum_{q=1}^k p_{j_q} > 0, P(A \cap B) = \sum_{\mu=1}^l p_{i_{\mu}}.$$

Нехай подія B відбулася. Тому має місце нова скінченна схема з k результатами, $k \leq n$, отже, сума ймовірностей повного набору цих нових результатів повинна дорівнювати одиниці, а

вона, відповідно до первісної схеми, дорівнює $P(B) = \sum_{q=1}^k p_{j_q}$.

Щоб забезпечити рівність суми ймовірностей елементарних подій одиниці, введемо нові ймовірності результатів:

$$\tilde{p}_{j_q} = \frac{p_{j_q}}{P(B)}, \quad \sum_{q=1}^k \tilde{p}_{j_q} = \frac{\sum_{q=1}^k p_{j_q}}{P(B)} = 1.$$

У рамках нової схеми (тобто за умови, що відбулася подія B) визначаємо ймовірність події A :

$$P(A/B) = \sum_{q=1}^k \tilde{p}_{j_q} = \frac{\sum_{q=1}^l p_{j_q}}{P(B)} = \frac{P(A \cap B)}{P(B)}.$$

Таким чином, ми знову одержуємо ту ж формулу умовної ймовірності, що й у класичній схемі. Незалежність подій визначається аналогічно класичній схемі. Розглянемо найпростіші приклади схеми послідовних випробувань як ілюстрацію скінченної схеми з неоднаково можливими результатами.

Приклад 1.2. Система контролю виробів складається із двох незалежних перевірок, виконуваних одночасно. Виріб приймається, якщо він пройшло обидві перевірки. У результаті кожної перевірки бракований виріб приймається з ймовірностями α_1 , α_2 відповідно. Знайти ймовірність того, що буде прийнятий бракований виріб.

Рішення. Якщо на вхід системи контролю надійшов бракований виріб, то можливі наступні чотири елементарних результати: $\omega_1 = \{0, 0\}$, $\omega_2 = \{0, 1\}$, $\omega_3 = \{1, 0\}$, $\omega_4 = \{1, 1\}$, де 0 означає, що виріб визнаний бракованим, а 1 – що виріб визнаний придатним. Випробування незалежні, тому одержуємо наступні значення ймовірностей елементарних результатів $\omega = \{i_1, i_2\}$:

$$p_1 = p(\omega_1) = P\{i_1=0, i_2=0\} = P\{i_1=0\} \cdot P\{i_2=0\} = (1 - \alpha_1)(1 - \alpha_2),$$

$$p_2 = p(\omega_2) = P\{i_1=0, i_2=1\} = P\{i_1=0\} \cdot P\{i_2=1\} = (1 - \alpha_1)\alpha_2,$$

$$p_3 = p(\omega_3) = P\{i_1=1, i_2=0\} = P\{i_1=1\} \cdot P\{i_2=0\} = \alpha_1(1 - \alpha_2),$$

$$p_4 = p(\omega_4) = P\{i_1=1, i_2=1\} = P\{i_1=1\} \cdot P\{i_2=1\} = \alpha_1\alpha_2.$$

Сума ймовірностей елементарних подій повинна дорівнювати одиниці. Дійсно,

$$P(\Omega) = \sum_{l=1}^4 p(\omega_l) = (1 - \alpha_1)(1 - \alpha_2) + (1 - \alpha_1)\alpha_2 + \alpha_1(1 - \alpha_2) + \alpha_1\alpha_2 = 1.$$

Відповідно до умов задачі, у сформованій схемі ймовірність прийняти бракований виріб – це ймовірність елементарної події ω_4 , що полягає в тому, що й перша, і друга перевірки визнають бракований виріб придатним. Тому шукана ймовірність дорівнює $p_4 = \alpha_1 \alpha_2$.

Приклад 1.3. В умовах приклада 1.2 задані ймовірності β_1 , β_2 відбракувати придатний виріб у результаті першої й другої перевірок відповідно. Знайти ймовірність того, що буде відбракований придатний виріб.

Рішення. Якщо на вхід системи контролю надійшов придатний виріб, то можливі ті ж самі чотири елементарних результати, однак їхні ймовірності будуть іншими. Знову скористаємося незалежністю випробувань, тоді одержимо наступні ймовірності елементарних результатів:

$$\tilde{p}_1 = p(\omega_1) = P\{i_1=0, i_2=0\} = P\{i_1=0\} \cdot P\{i_2=0\} = \beta_1 \beta_2,$$

$$\tilde{p}_2 = p(\omega_2) = P\{i_1=0, i_2=1\} = P\{i_1=0\} \cdot P\{i_2=1\} = \beta_1 (1 - \beta_2),$$

$$\tilde{p}_3 = p(\omega_3) = P\{i_1=1, i_2=0\} = P\{i_1=1\} \cdot P\{i_2=0\} = (1 - \beta_1) \beta_2,$$

$$\tilde{p}_4 = p(\omega_4) = P\{i_1=1, i_2=1\} = P\{i_1=1\} \cdot P\{i_2=1\} = (1 - \beta_1)(1 - \beta_2).$$

Подія, що полягає в тому, що відбракованим є придатний виріб, містить у собі елементарні події ω_1 , ω_2 , ω_3 . Тому шукана ймовірність дорівнює

$$\tilde{p}_1 + \tilde{p}_2 + \tilde{p}_3 = \beta_1 \beta_2 + \beta_1 (1 - \beta_2) + (1 - \beta_1) \beta_2 = \beta_1 + \beta_2 - \beta_1 \beta_2.$$

1.4. Вирахування подій

Одним з основних понять теорії ймовірностей є простір елементарних подій Ω і події як деякої підмножини цього простору. У загальному випадку простір Ω може бути будь-якої природи: як скінченним, так і нескінченним, як дискретним, так і безперервним. Розглянемо, наприклад, стрілянину по мішені. Якщо нас цікавить тільки сам факт влучення в мішень, то елементарними результатами служать $\omega_1=1$ (влучення в мішень) і $\omega_0=0$ (невлучення в мішень). Якщо важливе влучення в окремі області мішені (області розрізняються з погляду уразливості ре-

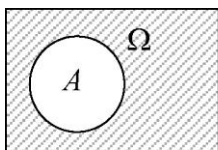
альної мети), то елементарними подіями можуть бути $\omega_{10}=10$, $\omega_9=9$, ..., $\omega_1=1$ (відповідають кількості вічків, приписаних влученню в певну область) і $\omega_0=0$ (невлучення в мішень). Нарешті, якщо істотно важливим є знання, у яку саме крапку щита, на якому зображена мішень, відбулося влучення, то довільний елементарний результат $\omega=\{x, y\}$ являє собою координати крапки влучення, а простір елементарних подій – це множина крапок щита.

Таким чином, нехай є простір елементарних подій Ω будь-якої природи. Будемо розглядати як події підмножини A, B, C цього простору. У такому випадку дії над подіями стають діями над підмножинами.

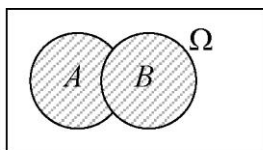
Подія A відбувається тоді й тільки тоді, коли відбувається одна з елементарних подій ω , з яких складається A . Несумісність подій A і B означає, що вони не мають жодної спільної елементарної події, тобто не перетинаються. Подія $A+B$ еквівалентна об'єднанню $A \cup B$, тобто такій множині елементарних подій, які входять або в A , або в B , а подія AB – перетинанню $A \cap B$, тобто множині елементарних подій, які є спільними для A і B . Подія, що включає всі елементарні події, тобто збігається із простором елементарних подій Ω , є достовірною. Звідси можна зробити висновок, що для будь-якої події $A \cap \Omega = A$, порожньою є подія, що не містить жодного елемента. Звідси випливає, що дві події A і B несумісні, якщо $A \cap B = \emptyset$. Очевидно, що $A \cup \emptyset = A$. *Теоретико-множинною різницею* двох подій A і B називається така подія $A \setminus B$, що містить ті елементарні події, що належать A , які не входять у B . Звідси випливає, що подія $\overline{A}$, протилежна події A , є $\Omega \setminus A$. Справді, $A \cap (\Omega \setminus A) = \emptyset$ і $A \cup (\Omega \setminus A) = \Omega$, тобто дійсно $\overline{A} = \Omega \setminus A$. Зокрема, $\overline{\Omega} = \emptyset$, $\overline{\emptyset} = \Omega$. Іноді використовується симетрична різниця подій $C = A \oplus B$, що представляє собою таку подію, у яку входять ті елементарні події, які входять в A або в B , але не входять у їхнє перетинання $A \cap B$. Таким чином, ця операція може бути представлена за допомогою вже введених операцій у такий спосіб:

$$C=A\oplus B=(A\setminus B)\cup(B\setminus A)=(A\cap \overline{B})\cup(\overline{A}\cap B).$$

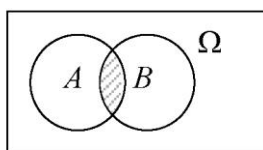
Для кращого розуміння операцій над подіями – підмножинами – звичай використовують умовні графічні зображення, представляючи достовірну подію Ω як прямокутник, а інші події – як кола. Тоді введені вище операції над подіями можуть бути представлені у вигляді діаграм Венна [1, стор. 19; 3. стор. 639], де результати операцій зображені у вигляді заштрихованих фігур. Операція доповнення зображена на мал. 1.1, об'єднання – на мал. 1.2, перетинання – на мал. 1.3, різниці – на мал. 1.4, симетричної різниці – на мал. 1.5.



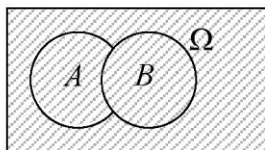
Мал. 1.1



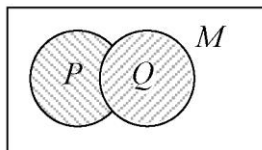
Мал. 1.2



Мал. 1.3



Мал. 1.4



Мал. 1.5

Дії над подіями, зокрема операції об'єднання (додавання) і перетинання (множення), аналогічні додаванню й множенню чисел. Ці операції

комутативні:

$$A\cup B=B\cup A, A\cap B=B\cap A;$$

асоціативні:

$$(A\cup B)\cup C=A\cup(B\cup C),$$

$$(A\cap B)\cap C=A\cap(B\cap C);$$

дистрибутивні:

$$(A\cup B)\cap C=(A\cap C)\cup(B\cap C).$$

Інша подібність із діями над числами полягає в тому, що для операції перетинання роль нуля й одиниці при множенні чисел виконують множини Ω і $\emptyset$, тому що $\Omega\cap A=A\cap\Omega=A$ і $\emptyset\cap A=A\cap\emptyset=\emptyset$. Разом з тим, теоретико-множинні рівності $A\cup A=A$ і $A\cap A=A$ і їм подібні показують, що повної аналогії немає.

Дії над подіями важливі не самі по собі, а як засіб визначення ймовірності одних подій через ймовірності інших подій. У випадку скінченної або зліченної теоретико-ймовірнісної схеми, розглянутої вище, як подія розглядалася будь-яка підмножина скінченного або зліченного простору елементарних подій Ω і ймовірність події визначалася як сума ймовірностей вхідних у нього елементарних подій. Якщо ж простір Ω неперервний, то має місце континуум елементарних результатів. Спроба вважати подією будь-яку підмножину безперервного простору Ω сполучена з великими труднощами. Тому в загальному випадку доводиться мати справу не з усіма підмножинами простору Ω , а лише з певним класом, замкнутим щодо операцій об'єднання й перетинання. Клас підмножин простору Ω , замкнутий щодо операцій доповнення, об'єднання й перетинання, в якому в той же час містяться множини $\emptyset$ і Ω , називається *полем*. Будемо позначати поле буквою S . Мінімальне поле S_0 складається з повної й порожньої множини $S_0 = \{\emptyset, \Omega\}$. Дійсно, $\emptyset$ і Ω входять у цей клас, а результатами операцій об'єднання, доповнення й перетинання над цими множинами знову служать дані множини: $\Omega \cup \emptyset = \Omega$, $\Omega \cap \emptyset = \emptyset$, $\overline{\Omega} = \emptyset$, $\overline{\emptyset} = \Omega$. Іншим, більш змістовним прикладом поля подій служить клас із чотирьох подій $S = \{\emptyset, A, \overline{A}, \Omega\}$. Дійсно,

$$\begin{aligned}\emptyset \cup A &= A, \emptyset \cup \overline{A} = \overline{A}, \Omega \cup \emptyset = \Omega, \emptyset \cap A = \emptyset, \emptyset \cap \overline{A} = \emptyset, \\ \Omega \cap \emptyset &= \emptyset, A \cup \overline{A} = \Omega, A \cup \Omega = \Omega, A \cap \overline{A} = \emptyset, A \cap \Omega = A, \\ \overline{A} \cup \Omega &= \Omega, \overline{A} \cap \Omega = \overline{A}, \overline{\emptyset} = \Omega, \overline{\overline{A}} = A, \overline{\Omega} = \emptyset.\end{aligned}$$

1.5. Аксиоматична побудова теорії ймовірностей

При аксіоматичній побудові теорії ймовірностей вихідним «матеріалом» служать простір елементарних подій Ω і виділений у ньому клас підмножин, що утворює поле подій S . Будова простору Ω і класу S визначається конкретною областю додатка.

Ймовірністю називається числова функція, визначена на полі подій S , що володіє наступними властивостями.

Аксіома 1. Для будь-якої події $A \in S$ $P(A) \geq 0$.

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

Аксіома 3. Імовірність об'єднання двох несумісних подій дорівнює сумі ймовірностей цих подій, тобто для будь-яких двох несумісних подій $A \in S, B \in S, A \cap B = \emptyset$ $P(A \cup B) = P(A) + P(B)$.

Скінченна схема задається скінченною множиною елементарних подій $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ і ймовірностями кожної з них

$0 \leq p_i \leq 1$, причому $\sum_{i=1}^n p_i = 1$. Імовірність будь-якої події A , що є

підмножиною Ω , тобто $A = \{\omega_{i_1}, \dots, \omega_{i_m}\}$, визначається як сума ймовірностей входних у неї елементарних подій, тобто по формулі (1.6).

Клас S всіх підмножин Ω утворює поле. Тепер переконаємося, що скінченна схема задовольняє вимозі аксіоми 1. Для цього виберемо довільну подію A , що є підмножиною Ω . Так як $A = \{\omega_{i_1}, \dots, \omega_{i_m}\}$, то, відповідно до скінченної схеми, $P(A) =$

$\sum_{l=1}^m p_{i_l}$, $0 \leq p_{i_l} \leq 1$, $l = \overline{1, m}$, тому $P(A) \geq 0$, тобто умова аксіоми 1

виконується. Умова аксіоми 2 виконується, тому що $\Omega = \{\omega_1, \omega_2,$

$\dots, \omega_n\}$ і на підставі (1.6) $P(\Omega) = \sum_{i=1}^n p_i = 1$. Умова аксіоми 3 також

виконується, тому що вона являє собою зміст теореми додавання для скінченної схеми. Таким чином, скінченна схема являє приклад об'єкта, для якого виконується система аксіом теорії ймовірностей.

Як приклад імовірнісної схеми з безперервним простором елементарних подій розглянемо схему з геометричними ймовірностями. Нехай простором елементарних подій служить множина крапок деякої області G , що має площу на площині. Як події будемо розглядати підмножини $A, B, C, \dots$ цієї області, що мають площу. Клас цих підмножин також утворює поле. При цьому ймовірність будь-якої події A (підмножини, що має площу $mes(A)$), можна задати в такий спосіб:

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

Описана схема також задовольняє аксіомам теорії ймовірностей. Аналогічно можна побудувати геометричні ймовірності в будь-якому скінченновимірному просторі.

У багатьох випадках виконання аксіоми 3 потрібно в розширеному варіанті. Аксіома 3 постулює додавання ймовірностей для скінченної кількості несумісних подій, тоді як у розширеному варіанті мова йде про зліченну кількість несумісних подій.

Аксіома 3'. Якщо $A_i \in S$, $i=1, 2, \dots$, $A_i \cap A_j = \emptyset$, $i \neq j$,

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

Однак, як відомо, не завжди події є несумісними. Вони також можуть бути залежними. Тому в загальному випадку для будь-яких (як несумісних, так і залежних) подій A і B має місце наступна формула додавання:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Аксіоми теорії ймовірностей лише постулюють існування ймовірностей для всіх подій, що утворюють поле S , і задають певні правила дії з ймовірностями. Експериментальне ж визначення ймовірності будь-якої події $A \in S$ може бути здійснене в результаті випробувань, виконуваних при певному тому ж самому комплексі умов. Із усього вищесказаного випливає, що певна теоретико-імовірнісна схема задається трійкою $\{\Omega, S, P\}$, тобто конкретним простором елементарних подій, конкретним набором підмножин Ω , що утворюють поле, а також конкретним завданням ймовірностей на множинах поля. Набір цих трьох компонентів називається *імовірнісним простором* [1, 4]. Імовірність P на $\{\Omega, S\}$ називається *розподілом ймовірностей на Ω* .

1.6. Контрольні запитання

1. Що таке випадкова подія?
2. Яка подія є достовірною, а яка неможливою?
3. Які події є несумісними?
4. Що таке добуток подій?
5. Як визначається ймовірність складеної події?
6. Сформулюйте теорему додавання ймовірностей.
7. В чому полягає урнова схема?
8. Які події є незалежними?
9. Наведіть формулу умовної ймовірності.
10. Сформулюйте теорему множення для незалежних подій.
11. У чому полягає скінченна схема з неоднаково можливими результатами?
12. Перелічіть основні дії над подіями та їх властивості.
13. Що таке поле подій?
14. Сформулюйте аксіоми теорії ймовірностей.
15. Що таке ймовірнісний простір?
16. Що таке розподіл ймовірностей?

1.7. Завдання для самостійного опрацювання

1. При киданні двох гральних кісток знайти ймовірність того, що на одній кістці випаде 4 очки, а на іншій – менше, ніж 4.

Розв'язання:

Розглянемо ситуацію, коли на першій кістці 4 очки. Грань з цифрою «4» у кістки лише одна. Тому ймовірність цієї події буде дорівнювати $1/6$ (одна подія з шести можливих). На другій кістці при цьому нас задовільняють грані з цифрами «1», «2» і «3», тобто три події з шести можливих. Тому ймовірність цієї події дорівнює $3/6$. Ці події повинні трапитися одночасно. Тому для обчислення ймовірності завданої ситуації їх треба помножити. Але нас задовільнить і дзеркальна ситуація, коли 4 очки випадають на другій кістці, а кількість менша за «4» - на першій. Тому ситуація, що описана в умові задачі, складається з двох однакових з математичної точки зору ситуацій. Нас задовіль-

нить будь-яка з них. В той же час це взаємовиключальні ситуації. Таким чином, ймовірність цих двох ситуацій треба додавати. Тобто маємо:

$$P(A) = P(\langle 4 \rangle) \cdot P(\langle \langle 4 \rangle \rangle) + P(\langle \langle 4 \rangle \rangle) \cdot P(\langle 4 \rangle) = \frac{1}{6} \cdot \frac{3}{6} + \frac{3}{6} \cdot \frac{1}{6} = 2 \cdot \frac{3}{36} = \frac{2}{12} = \frac{1}{6}.$$

Відповідь: $P(A) = 1/6$.

2. П'ять карток, що помічені цифрами від «1» до «5», ретельно перемішують. Знайти ймовірність того, що трьох-цифрове число, що утворилось із номерів трьох навмання вилучених карток, буде парним.

Розв'язання:

Число є парним, якщо закінчується парною цифрою. Із п'яти номерів карток є 3 непарні («1», «3» і «5») і два парні («2» і «4»). При вилученні трьох карток парне число можливе у випадку наступних ситуацій.

Ситуація А: 1-ша картка непарна, 2-га картка непарна, 3-тя картка парна.

Коли ми тягнемо першу картку, ймовірність витягти непарну дорівнює $3/5$ (три картки з можливих п'яти). Назад ми цю картку не повертаємо. Тому після цього в нас залишається 4 картки, з яких дві парні і дві непарні (одну непарну ми вже витягли). Тому ймовірність того, що при витягненні другої картки вона буде непарною дорівнює $2/4$ (дві картки з можливих чотирьох). Після цього в нас залишаються 3 картки, з яких одна непарна (дві з трьох непарних ми вже витягли) і дві парні. Тобто після кожного шагу у нас на одиницю зменшується загальна кількість карток, і також на одиницю зменшується кількість або парних, або непарних карток в залежності від того, яку з них ми витягли останньою. Таким чином, ймовірність витягти третю картку парну дорівнює $2/3$ (дві картки з трьох можливих). Тому ймовірність ситуації А буде наступною:

$$P(A) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{2}{3} = \frac{1}{5}.$$

Ситуація В: 1-ша картка непарна, 2-га картка парна, 3-тя картка парна.

Коли ми тягнемо першу картку, ймовірність витягти непарну дорівнює $3/5$ (три картки з можливих п'яти). Коли ми тягнемо другу картку, ймовірність витягти парну дорівнює $2/4$ (жодну парну ми ще не витягали, тому їх дві з можливих чотирьох). Коли ми тягнемо третю картку, ймовірність витягти парну дорівнює $1/3$ (одну парну ми вже витягли, тому одна картка з можливих трьох). Таким чином, ймовірність ситуації B буде наступною:

$$P(B) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{1}{3} = \frac{1}{10}.$$

Ситуація С: 1-ша картка парна, 2-га картка непарна, 3-тя картка парна.

Коли ми тягнемо першу картку, ймовірність витягти парну дорівнює $2/5$ (дві картки з можливих п'яти). Коли ми тягнемо другу картку, ймовірність витягти непарну дорівнює $3/4$ (жодну непарну ми ще не витягали, тому їх три з можливих чотирьох). Коли ми тягнемо третю картку, ймовірність витягти парну дорівнює $1/3$ (одну парну ми вже витягли, тому одна картка з можливих трьох). Таким чином, ймовірність ситуації C буде наступною:

$$P(C) = \frac{2}{5} \cdot \frac{3}{4} \cdot \frac{1}{3} = \frac{1}{10}.$$

Нас задовільнить будь яка з трьох описаних ситуацій. Тому для обчислення ймовірності події, яка описана в умові задачі, ці ймовірності треба додавати. Таким чином,

$$P = P(A) + P(B) + P(C) = \frac{1}{5} + \frac{1}{10} + \frac{1}{10} = \frac{2}{5}.$$

Відповідь: $P=2/5$.

3. При триразовому киданні двох гральних кісток знайти ймовірність сумарної кількості випадання рівно 8 очок.

Розв'язання:

У нас дві кістки. Тому за одне кидання не може випасти менше двох очків (по одному на кожній кістці) і більше 12 (по шість на кожній кістці). Тому за три кидання кількість 8 можна отримати наступними способами:

- 1) 2+2+4;
- 2) 2+4+2;
- 3) 4+2+2;
- 4) 2+3+3;
- 5) 3+2+3;
- 6) 3+3+2;

Усього можливі 6 способів. На перший погляд перші 3 способи здаються однаковими. Але це не так. Перший спосіб передбачає, що за перші два кидання отримано по 2 очка, а на останньому – 4 очка. При третьому способі, наприклад, 4 очка ми отримуємо при першому киданні, а по 2 очка – при другому і третьому. Тому ці три способи, хоч і складаються з однакових елементарних подій, насправді є різними ситуаціями. Нас задовільнить будь-яка з них. Аналогічно є різними і способи 4) – 6).

Тепер нам треба розглянути способи, якими ми можемо отримати 2, 3 і 4 очка на двох кістках за одне кидання (саме ці кількості беруть участь у створенні ситуацій 1) – 6)). Двійку можна отримати лише в один спосіб: коли на обох кістках випадає по одному очку. Тому

$$P(\langle\langle 2 \rangle\rangle) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{12}.$$

Трійку можна отримати двома способами. Коли на першій кістці випадає 2 очка, а на другій – одно очко. І навпаки, на першій 1 очко, а на другій – 2 очка. Тому

$$P(\langle\langle 3 \rangle\rangle) = \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} = \frac{2}{36} = \frac{1}{18}.$$

Четвірку можна отримати наступним чином:

- на обох кістках по 2 очка: 2+2;
- на одній кістці 1 очко, а на іншій – 3 очка: 1+3 або 3+1.

Ймовірність отримати кожен з цих цифр на кістці буде складати $1/6$. Тому ймовірність отримати сумарну четвірку на двох кістках за одне кидання обчислюється так:

$$P(\langle\langle 4 \rangle\rangle) = \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} = \frac{3}{36} = \frac{1}{12}.$$

Таким чином, ймовірність отримати «8» за триразове кидання двох кісток складає:

$$\begin{aligned} P(\langle\langle 8 \rangle\rangle) &= P(\langle\langle 2 \rangle\rangle) \cdot P(\langle\langle 2 \rangle\rangle) \cdot P(\langle\langle 4 \rangle\rangle) + P(\langle\langle 2 \rangle\rangle) \cdot P(\langle\langle 4 \rangle\rangle) \cdot P(\langle\langle 2 \rangle\rangle) + P(\langle\langle 4 \rangle\rangle) \cdot P(\langle\langle 2 \rangle\rangle) \cdot P(\langle\langle 2 \rangle\rangle) + \\ &+ P(\langle\langle 2 \rangle\rangle) \cdot P(\langle\langle 2 \rangle\rangle) + P(\langle\langle 2 \rangle\rangle) \cdot P(\langle\langle 3 \rangle\rangle) \cdot P(\langle\langle 3 \rangle\rangle) + P(\langle\langle 3 \rangle\rangle) \cdot P(\langle\langle 2 \rangle\rangle) \cdot P(\langle\langle 3 \rangle\rangle) + P(\langle\langle 3 \rangle\rangle) \cdot \\ &+ P(\langle\langle 3 \rangle\rangle) \cdot P(\langle\langle 2 \rangle\rangle) = 3 \cdot P(\langle\langle 2 \rangle\rangle) \cdot P(\langle\langle 2 \rangle\rangle) \cdot P(\langle\langle 4 \rangle\rangle) + 3 \cdot P(\langle\langle 2 \rangle\rangle) \cdot P(\langle\langle 3 \rangle\rangle) \cdot P(\langle\langle 3 \rangle\rangle) = \\ &= 3 \cdot \frac{1}{12} \cdot \frac{1}{12} \cdot \frac{1}{12} + 3 \cdot \frac{1}{12} \cdot \frac{1}{18} \cdot \frac{1}{18} = \frac{13}{5184} = 0,0025 = 0,25\%. \end{aligned}$$

Відповідь: $P(\langle\langle 8 \rangle\rangle) = 0,25\%$.

4. Два стрілка здійснюють по одному пострілу. Причому ймовірності влучення в ціль для кожного з них дорівнюють 0,8 і 0,9 відповідно. Знайти ймовірність влучення в ціль обома стрілками і ймовірність влучення хоча б одним з них.

Розв'язання:

Нехай $A = \langle\langle \text{В ціль влучили обидва стрілка} \rangle\rangle$ і $B = \langle\langle \text{В ціль влучив хоча б один із стрільців} \rangle\rangle$.

Подія A набуває чинності, коли жоден із стрільців не схибив. Тобто в ціль повинен влучити і один стрілок, і інший. Тому ймовірність влучення обома стрілками

$$P(A) = 0,8 \cdot 0,9 = 0,72.$$

Влучення в ціль стрілками не є взаємовиключними подіями. Тобто результат пострілу одного стрілка ніяк не впливає на результат іншого. Тому ці два постріли є подіями, що перетинаються. Нас влаштує ситуація, коли в ціль влучив або перший стрілок, або другий стрілок, або вони обидва. Не влаштує лише той випадок, коли схибили обидва. Таким чином, ймовірність влучення в ціль хоча б одним із стрільців

$$P(B) = 0,8 + 0,9 - 0,8 \cdot 0,9 = 1,7 - 0,72 = 0,98.$$

Відповідь: $P(A) = 0,72$; $P(B) = 0,98$.

5. Партія із 10 телевізорів приймається в магазині за умови, що випадково обрані два з них будуть справними. Знайти ймовірність того, що магазин прийме партію, що містить 4 поламаних телевізори.

Розв'язання:

Ця задача є задачею на застосування урнкової схеми. Загальна кількість телевізорів дорівнює $N=10$. З них обираються для перевірки 2. Тобто об'єм вибірки $n=2$. За M візьмемо справні телевізори. В партії, що перевіряється, згідно з умовою задачі, міститься 4 поламаних. Справних, відповідно, в ній $M=6$. Партію буде прийнято, якщо обидва перевірені телевізори виявляться справними. Тобто $m=2$. Тоді

$$P(A) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n} = \frac{6!}{2! \cdot (6-2)!} \cdot \frac{(10-6)!}{(2-2)! \cdot ((10-6)-(2-2))!} = \frac{10!}{2! \cdot (10-2)!} = \frac{6! \cdot 4!}{2! \cdot 4! \cdot 0! \cdot 4!} = \frac{1}{3}.$$

Але цю задачу можна розв'язати по-іншому. Нам відомо, що в пропонованій партії телевізорів 4 поламаних і $10 - 4 = 6$ справних. Нам потрібно, щоб два навмання вилучені телевізори виявилися справними. Аналогічно розв'язанню задачі 2), ми витягуємо один телевізор з 10 і він виявляється справним. Ймовірність цієї ситуації дорівнює $6/10$. Цей телевізор ми назад не повертаємо. Тому у нас залишається 9 телевізорів, з яких вже 5 справних (один ми вилучили). Тому ймовірність при другому витягуванні отримати справний телевізор дорівнює $5/9$. Виходячи з цього, ймовірність витягнути 2 справних телевізори із партії 10 телевізорів, серед яких 6 справних, буде складати

$$P(A) = \frac{6}{10} \cdot \frac{5}{9} = \frac{1}{3}.$$

Відповідь: $P(A)=1/3$.

1.8. Контрольні завдання

1. При киданні двох гральних кісток знайти ймовірність наступних подій:

- а) *на одній кістці кількість очок в два разі більша, ніж на іншій;*
- б) *сума кількості очок, що випали на двох кістках, менша за 5;*
- в) *сума кількості очок, що випали на двох кістках, більша за 8.*

2. П'ять карток, що помічені цифрами від «1» до «5», ретельно перемішують. Знайти ймовірність того, що:

- а) *чотирьохцифрове число, що утворилось із номерів трьох навмання вилучених карток, буде **непарним**;*
- б) *при випадковому розкладанні усіх карток на п'ять місць з номерами від «1» до «5» жодна картка не займе місця, відповідного до її номера;*
- в) *при почерговому перебиранні усіх карток їх номери будуть появлятися у порядку зростання.*

3. При триразовому киданні двох гральних кісток знайти ймовірність сумарної кількості випадання рівно 9 очок.

4. Три стрільця здійснюють по одному пострілу. Причому ймовірності влучення в ціль для кожного з них дорівнюють 0,6, 0,7 і 0,8 відповідно. Знайти ймовірність влучення в ціль двома стрільцями з трьох.

5. Зі 180 студентів, що складали іспити з англійської мови та математики, 15 не здали іспит з математики, 10 – з англійської мови, 5 студентів не здали обидва іспити. Знайти ймовірність того, що випадково обраний студент

- а) *не здав математику, але здав англійську мову;*
- б) *здав математику, але не здав англійську мову.*

6. В коробці знаходяться 13 кубиків. З них 5 червоних і 8 зелених. Знайти ймовірність того, що із трьох навмання вилучених кубиків усі будуть зеленими.

2. УМОВНІ ЙМОВІРНОСТІ

2.1. Повна ймовірність

Події $A_1, A_2, \dots, A_n$ утворюють *повну групу подій*, якщо вони попарно несумісні й разом утворюють достовірну подію, тобто

$$A_i \cap A_j = \emptyset, i \neq j, \bigcup_{i=1}^n A_i = \Omega.$$

Теорема 2.1 (про формулу повної ймовірності). Формулювання. Якщо події $A_1, A_2, \dots, A_n$, $P(A_i) > 0$, утворюють повну групу подій, то ймовірність події B може бути подана як сума добутків безумовних ймовірностей подій повної групи на умовні ймовірності події B :

$$P(B) = \sum_{i=1}^n P(A_i) \cdot P(B / A_i). \quad (2.1)$$

Доказ. Події повної групи $A_1, A_2, \dots, A_n$ попарно несумісні, тому попарно несумісні й їхні добутки (перетинання) з подією B , тобто події $B \cap A_i, B \cap A_j$ при $i \neq j$ несумісні. Так як подію B , з урахуванням (2.1), можна представити у вигляді $B = \bigcup_{i=1}^n (B \cap A_i)$,

то, застосувавши до цього розкладання події B аксіому додавання ймовірностей, маємо

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

Використовуючи формулу множення ймовірностей (1.5) для кожного доданка, остаточно одержуємо

$$P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(A_i) \cdot P(B / A_i).$$

Теорему доведено.

Вимога, що полягає в тому, що події A_i утворюють повну групу подій, може бути замінена більш слабкою: події A_i попар-

но не перетинаються, $B \in \bigcup_{i=1}^n A_i$. Крім того, на основі аксіоми 3' (аксіоми зліченної адитивності) теорему повної ймовірності можна поширити й на зліченну множину подій, що попарно не перетинаються A_i , $P(A_i) > 0$, $B \in \bigcup_{i=1}^n A_i$:

$$P(B) = \sum_{i=1}^{\infty} P(A_i) \cdot P(B / A_i).$$

Приклад 2.1. У трьох партіях деталей, що надійшли на склад, відсоток придатних становить відповідно 89%, 92% і 97%, а загальна кількість деталей у партіях співвідноситься як 1:2:3. Необхідно визначити ймовірність випадкового вибору непридатної деталі із всіх трьох партій.

Рішення. Позначимо через A_1, A_2, A_3 події, що полягають в тому, що обрана навмання деталь належить відповідно до першої, другої й третьої партій. Так як ці події утворюють повну групу подій, то $P(A_1) + P(A_2) + P(A_3) = 1$.

За умовою $P(A_1):P(A_2):P(A_3) = 1:2:3$, отже, $P(A_1) = \frac{1}{6}$,

$$P(A_2) = \frac{1}{3}, P(A_3) = \frac{1}{2}.$$

Розглядаючи вибір непридатної деталі як випадкову подію B , імовірності такої події за умови, що деталь вибирається з першої, другої й третьої партій, відповідно мають значення: $P(B/A_1) = 0.11$, $P(B/A_2) = 0.08$, $P(B/A_3) = 0.03$. Імовірність випадкового вибору непридатної деталі із всіх трьох партій визначається по формулі повної ймовірності:

$$\begin{aligned} P(B) &= P(A_1) \cdot P(B/A_1) + P(A_2) \cdot P(B/A_2) + P(A_3) \cdot P(B/A_3) = \\ &= \frac{1}{6} \cdot 0.11 + \frac{1}{3} \cdot 0.08 + \frac{1}{2} \cdot 0.03 = 0.06. \end{aligned}$$

2.2. Формула Байєса

На основі комутативності операції перетинання множин $A \cap B = B \cap A$ можна записати $P(A \cap B) = P(B \cap A)$ або $P(A) \cdot P(B/A) = P(B) \cdot P(A/B)$ [3]. Це співвідношення справедливо й для випадку, коли під A розуміється деяка подія A_k з повної групи подій $A_1, A_2, \dots, A_n$, тобто $P(A_k) \cdot P(B/A_k) = P(B) \cdot P(A_k/B)$, звідки

$$P(A_k/B) = P(A_k) \frac{P(B/A_k)}{P(B)}.$$

Підставляючи сюди $P(B)$ по формулі повної ймовірності (2.1) одержуємо *формулу Байєса*:

$$P(A_k/B) = \frac{P(A_k) \cdot P(B/A_k)}{\sum_{i=1}^n P(A_i) \cdot P(B/A_i)}. \quad (2.2)$$

По цій формулі можна обчислити ймовірності подій A_i , $i = \overline{1, n}$, за умови, що відбулася подія B , якщо відомі ймовірності $P(A_i)$ і $P(B/A_i)$. Очевидно, що одержувані при цьому умовні ймовірності $P(A_i/B)$ задовольняють співвідношенню
$$\sum_{i=1}^n P(A_i/B) = 1.$$

Ймовірності $P(A_i)$ подій A_i називають *апріорними ймовірностями*, тобто ймовірностями подій до виконання експерименту, а умовні ймовірності цих подій $P(A_i/B)$ – *апостеріорними*, тобто уточненими в результаті експерименту, результатом якого послужила поява події B [1, 2].

Приклад 2.2. На підприємстві виготовляють вироби певного виду на трьох поточкових лініях. На першій лінії виробляється 20% виробів від усього об'єму їхнього виробництва, на другій – 30%, на третій – 50%. Кожна з ліній характеризується відповідно наступними відсотками придатності виробів: 95%, 98% і 97%. Потрібно визначити ймовірність того, що навмання взятий виріб, випущений підприємством, виявиться бракованим, а також ймовірності того, що цей бракований виріб зроблений на першій, другій і третій лініях.

Рішення. Позначимо через A_1, A_2, A_3 події, що полягають в тому, що навання взятий виріб зроблений на першій, другій і третій лініях. Відповідно до умов задачі, $P(A_1)=0.2$, $P(A_2)=0.3$, $P(A_3)=0.5$ і ці події утворюють повну групу подій, тому що вони попарно несумісні й $P(A_1)+P(A_2)+P(A_3)=1$. Позначимо через B подію, що полягає в тому, що навання взятий виріб виявився бракованим. Відповідно до умов задачі, $P(B/A_1)=0.05$, $P(B/A_2)=0.02$, $P(B/A_3)=0.03$. Використовуючи формулу повної ймовірності, одержуємо

$$P(B)=P(A_1) \cdot P(B/A_1)+P(A_2) \cdot P(B/A_2)+P(A_3) \cdot P(B/A_3)= \\ = 0.05 \cdot 0.2+0.02 \cdot 0.3+0.03 \cdot 0.5=0.031,$$

тобто ймовірність того, що навання взятий виріб виявиться бракованим, дорівнює 3.1%.

Апріорні ймовірності того, що навання взятий виріб виготовлений відповідно на першій, другій і третій лініях, рівні 0.2, 0.3 і 0.5. Потім був виконаний експеримент, у результаті якого навання взятий виріб виявився бракованим. Визначимо тепер апостеріорні ймовірності того, що цей виріб виготовлений на першій, другій і третій лініях. По формулі Байєса маємо:

$$P(A_1/B)=\frac{P(A_1) \cdot P(B/A_1)}{\sum_{i=1}^3 P(A_i) \cdot P(B/A_i)}=\frac{0.05 \cdot 0.2}{0.031}=\frac{10}{31},$$

$$P(A_2/B)=\frac{P(A_2) \cdot P(B/A_2)}{\sum_{i=1}^3 P(A_i) \cdot P(B/A_i)}=\frac{0.02 \cdot 0.3}{0.031}=\frac{6}{31},$$

$$P(A_3/B)=\frac{P(A_3) \cdot P(B/A_3)}{\sum_{i=1}^3 P(A_i) \cdot P(B/A_i)}=\frac{0.03 \cdot 0.5}{0.031}=\frac{15}{31}.$$

Таким чином, ймовірності того, що навання взятий виріб, що виявився бракованим, виготовлено на першій, другій і третій лініях, відповідно рівні 0.322, 0.194 і 0.484.

Формула множення ймовірностей (1.5) може бути поширена на випадок довільної скінченної кількості подій $P(A_1 \cap A_2 \cap \dots \cap A_n)=P(A_1) \cdot P(A_1/A_2) \dots P(A_n/A_1 \cap A_2 \cap \dots \cap A_{n-1})$.

Події $A_1, A_2, \dots, A_n$ незалежні в сукупності, якщо для будь-якої їхньої підмножини

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \dots \cdot P(A_{i_k}).$$

Якщо ця умова виконується тільки для $k=2$, то такі події називаються *попарно незалежними*. З незалежності в сукупності випливає попарна незалежність, а з попарної незалежності не випливає незалежність у сукупності.

2.3. Послідовності випробувань

Нехай проводиться скінченне число n послідовних незалежних випробувань, у кожному з яких може відбутися певна подія – успіх – або наступить протилежна подія – невдача. Така послідовність випробувань називається *схемою Бернуллі*. У схемі Бернуллі одному випробуванню відповідає множина елементарних результатів, що складається із двох елементарних подій: $\{\omega_0, \omega_1\}$, ω_0 – невдача, ω_1 – успіх, при цьому $A = \{\omega_1\}$, $\bar{A} = \{\omega_0\}$. Множина елементарних результатів для n випробувань складається вже з 2^n елементарних подій $\omega_{i_1, i_2, \dots, i_n} = \{i_1, \dots, i_n\}$, кожна з яких відповідає конкретному результату випробувань, при цьому набір індексів $i_1, \dots, i_n$ являє собою конкретну послідовність нулів і одиниць, що відповідає результатам випробувань на кожному кроці.

Якщо задані ймовірності успіху й невдачі в окремому випробуванні $p_1=p$, $p_0=1-p=q$, то можна визначити ймовірність будь-якого елементарного результату в n випробуваннях. Дійсно, розглянемо будь-який елементарний результат $\omega_{i_1, i_2, \dots, i_n}$, при цьому $(i_1, \dots, i_n)$ – конкретна послідовність нулів і одиниць, що відповідає послідовності невдач або успіхів у кожному з n індивідуальних випробувань, наприклад $(A_{i_1}, A_{i_2}, \dots, A_{i_n})$. Тоді з незалежності друг від друга результатів окремих випробувань одержуємо

$$P\{\omega_{i_1, i_2, \dots, i_n}\} = P(\omega_{i_1})P(\omega_{i_2}) \dots P(\omega_{i_n}) = p_{i_1} p_{i_2} \dots p_{i_n} = pq \dots p.$$

Таким чином, якщо загальний елементарний результат включає m успіхів і $n - m$ невдач, то його ймовірність дорівнює

$$P(\omega_{i_1, i_2, \dots, i_n}) = p^m q^{n-m} \quad (2.3)$$

І, отже, по аксіомі додавання ймовірностей може бути визначена ймовірність будь-якої події, що складається з декількох елементарних подій. Зокрема, якщо нас цікавить ймовірність $P_n(m)$ того, що в n випробуваннях відбулося m успіхів, то її визначаємо як суму ймовірностей елементарних подій, що характеризуються m успіхами. Ймовірність такого елементарного результату, згідно (2.3), дорівнює $p^m q^{n-m}$. Отже, для знаходження ймовірності $P_n(m)$ треба визначити кількість елементарних подій, що характеризуються m успіхами, тобто встановити, скількома способами можуть бути на n місць розставлені m одиниць (інші $n - m$ місць займаються нулями). Але це аналогічно тому, що з n елементів треба вибрати (позначити) m елементів. Кількість таких вибірок, як відомо, дорівнює кількості сполучень із n по m , тобто C_n^m . Остаточню одержуємо

$$P_n(m) = C_n^m p^m q^{n-m}. \quad (2.4)$$

Сума біноміальних ймовірностей, що вийшли, дорівнює одиниці:

$$\sum_{m=0}^n P_n(m) = \sum_{m=0}^n C_n^m p^m q^{n-m} = (p+q)^n = 1^n = 1.$$

У випадку урнкової схеми можна уявити собі, що здійснюється вибірка об'єму n з урни не відразу, а послідовно куля за кулею. У результаті приходимо до схеми послідовних випробувань, однак на відміну від схеми Бернуллі тут результати наступних випробувань уже залежать від результатів попередніх. Так, якщо ймовірність на першому кроці витягти білу кулю дорівнює $\frac{M}{N}$, то умовна ймовірність витягти білу кулю на другому кроці

дорівнює $\frac{M-1}{N-1}$, якщо на першому кроці витягнута біла куля, і

$\frac{M}{N-1}$, якщо на першому кроці витягнута чорна куля. Але у ви-

падку, коли генеральна сукупність велика, тобто $N \rightarrow \infty$, урнову схему можна замінити схемою Бернуллі. На практиці це означає, що при об'ємі вибірки, істотно меншому об'єму генеральної сукупності, можна замість ймовірностей урнкової схеми приблизно використати відповідні ймовірності схеми Бернуллі., тобто при $n \ll N$

$$P_{M,N}(m, n) \approx C_n^m \left(\frac{M}{N} \right)^m \left(\frac{N-M}{N} \right)^{n-m}.$$

2.4. Контрольні запитання

1. Що таке повна група подій?
2. Сформулюйте теорему про формулу повної ймовірності.
3. Наведіть формулу Байєса.
4. Які ймовірності називають апіорними, а які апостеріорними?
5. Які події є незалежними у сукупності?
6. Яка послідовність випробувань називається схемою Бернуллі?

2.5. Задачі для самостійного опрацювання

1. У виробничому цеху три машини виготовляють однакові деталі. Перша виробляє 30% усіх деталей, друга – 25% і третя – 45%. Виробничий брак, якого припускається кожна машина, складає для них відповідно 1%, 1,2% і 2%. Знайти ймовірність того, що

а) навмання обрана деталь із вироблених протягом місяця виявиться бракованою.

Розв'язання

Нехай A_1 = «Деталь виготовлена на машині №1»,
 A_2 = «Деталь виготовлена на машині №2»,
 A_3 = «Деталь виготовлена на машині №3»,
 B = «Деталь є бракованою».

Тоді маємо наступні апіорні ймовірності:

$$P(A_1)=0,3; P(A_2)=0,25; P(A_3)=0,45.$$

Також можна визначити

$$P(B/A_1)=0,01; P(B/A_2)=0,012; P(B/A_3)=0,02.$$

Події A_1 , A_2 і A_3 попарно несумісні і $P(A_1)+P(A_2)+P(A_3)=1$. Тому ці події утворюють повну групу подій. Імовірність випадкового вибору бракованої деталі із всіх трьох партій визначається по формулі повної ймовірності:

$$\begin{aligned} P(B) &= P(A_1) \cdot P(B/A_1) + P(A_2) \cdot P(B/A_2) + P(A_3) \cdot P(B/A_3) = \\ &= 0,3 \cdot 0,01 + 0,25 \cdot 0,012 + 0,45 \cdot 0,02 = 0,015 = 1,5\%. \end{aligned}$$

Відповідь: $P(B)=1,5\%$.

б) бракована деталь виготовлена саме на 1-шій, на 2-гій і на 3-тій машинах.

Розв'язання

Для розв'язання цієї задачі треба застосувати формулу Байєса. При цьому треба врахувати, що $\sum_{i=1}^3 P(A_i) \cdot P(B/A_i) = P(B) = 0,015$. Тому маємо наступні апостеріорні ймовірності:

$$\begin{aligned} P(A_1/B) &= \frac{P(A_1) \cdot P(B/A_1)}{\sum_{i=1}^3 P(A_i) \cdot P(B/A_i)} = \frac{0,3 \cdot 0,01}{0,015} = 0,2; \\ P(A_2/B) &= \frac{P(A_2) \cdot P(B/A_2)}{\sum_{i=1}^3 P(A_i) \cdot P(B/A_i)} = \frac{0,25 \cdot 0,012}{0,015} = 0,2; \\ P(A_3/B) &= \frac{P(A_3) \cdot P(B/A_3)}{\sum_{i=1}^3 P(A_i) \cdot P(B/A_i)} = \frac{0,45 \cdot 0,02}{0,015} = 0,6. \end{aligned}$$

Відповідь: $P(A_1/B)=0,2$; $P(A_2/B)=0,2$; $P(A_3/B)=0,6$.

2. Деталі можуть виготовлятися із застосуванням двох технологій. Перша з них містить три технологічні операції. Ймовірності отримання браку при кожній з них дорівнюють відповідно 0,1, 0,2 і 0,3. Друга містить дві операції з однаковими ймовірностями отримання браку 0,3. Яка з технологій забезпечує більшу ймовірність отримання продукції 1-го сорту, якщо для небракованої деталі ймовірність приналежності до 1-го сорту по 1-шій технології дорівнює 0,9, а по другій технології – 0,8?

Розв'язання

Нехай A_{11} = «Деталь пройшла виготовлення 1-шою операцією по 1-шій технології»,

A_{12} = «Деталь пройшла виготовлення 2-гою операцією по 1-шій технології»,

A_{13} = «Деталь пройшла виготовлення 3-тєю операцією по 1-шій технології»,

A_{21} = «Деталь пройшла виготовлення 1-шою операцією по 2-гій технології»,

A_{22} = «Деталь пройшла виготовлення 2-гою операцією по 2-гій технології»,

B_1 = «Деталь, виготовлена за 1-шою технологією, є якісною»,

B_2 = «Деталь, виготовлена за 2-гою технологією, є якісною»,

C = «Деталь належить до 1-го сорту».

Будь-яка деталь може бути або бракованою, або якісною. Вона не може бути і бракованою, і якісною одночасно. Тому це взаємовиключальні події. З огляду на це можна обчислити наступні ймовірності недопущення браку на кожній з операцій:

$$P(B_1/A_{11})=1-0,1=0,9; P(B_1/A_{12})=1-0,2=0,8;$$

$$P(B_1/A_{13})=1-0,3=0,7;$$

$$P(B_2/A_{21})=P(B_2/A_{22})=1-0,3=0,7.$$

Деталь не є бракованою при виготовленні за деякою технологією, якщо при жодній операції цієї технології не було випущено браку. Тому маємо:

$$P(B_1)=P(B_1/A_{11}) \cdot P(B_1/A_{12}) \cdot P(B_1/A_{13})=0,9 \cdot 0,8 \cdot 0,7=0,504;$$

$$P(B_2)=P(B_2/A_{21}) \cdot P(B_2/A_{22})=0,7 \cdot 0,7=0,49.$$

За умовою задачі, ймовірності приналежності якісної деталі до 1-го сорту за умови виготовлення за тією чи іншою технологією:

$$P(C/B_1)=0,9; P(C/B_2)=0,8.$$

Події C і B_1 , а також події C і B_2 є залежними між собою. Тому для обчислення ймовірності добутку цих подій використається формула (1.5):

$$P(B_1 \cap C)=P(C/B_1) \cdot P(B_1)=0,9 \cdot 0,504=0,4536;$$

$$P(B_2 \cap C) = P(C/B_2) \cdot P(B_2) = 0,8 \cdot 0,49 = 0,392.$$

Очевидно, що $0,4536 > 0,392$.

Висновок: 1-ша технологія забезпечує більшу, ніж 2-га технологія, ймовірність отримання продукції 1-го сорту.

2.6. Контрольні завдання

1. У виробничому цеху чотири машини виготовляють однакові деталі. Перша виробляє 15% усіх деталей, друга – 35%, третя – 10%, четверта – 40%. Виробничий брак, якого припускається кожна машина, складає для них відповідно 1,1%, 1,3%, 2% і 1,5%. Знайти ймовірність того, що

а) навмання обрана деталь із вироблених протягом місяця виявиться бракованою;

б) бракована деталь виготовлена саме на 1-шій, на 2-гій, на 3-тій і на 4-тій машинах.

2. Деталі можуть виготовлятися із застосуванням двох технологій. Перша з них містить три технологічні операції. Ймовірності отримання браку при кожній з них дорівнюють відповідно 0,15, 0,12 і 0,32. Друга містить три операції з ймовірностями отримання браку 0,23, 0,11 і 0,09 відповідно. Яка з технологій забезпечує меншу ймовірність отримання продукції 1-го сорту, якщо для небракованої деталі ймовірність приналежності до 1-го сорту по 1-шій технології дорівнює 0,95, а по другій технології – 0,87?

3. ВИПАДКОВА ВЕЛИЧИНА

3.1. Визначення випадкової величини й її функція розподілу

Явища, які відбуваються в природі, є результатом дії внутрішніх і (або) зовнішніх факторів. Якщо зміна явища відбувається під впливом одного фактору, то цей вплив можна до деякої міри визначити, і результат зміни розглядається як детермінований цим фактором. Але часто кінцевий результат явища, що змінюється, викликаний дією багатьох причин, визначення яких сполучено з великими труднощами й не завжди є доцільним. Дія комплексу причин щораз викликає певні зміни в кількісних характеристиках розглянутого явища. Причому сутність явища не змінюється. Такі коливання в значеннях кількісної характеристики є випадковими. Вони викликані дією множини причин, кожна з яких може бути несуттєвою для природи даного явища в цілому або для окремих його якостей. При багаторазовому повторенні тих самих умов експерименту (при дії того самого комплексу причин) у змінах значень кількісної характеристики ознаки (явища) можна виявити деяку закономірність. Ця закономірність, що є присутньою у випадкових числах, буде відбивати ту якісну сторону (якісна ознака) явища, що найбільш яскраво проявляється у всіх експериментах (випробуваннях).

Кількісна характеристика розглянутої ознаки (явища), значення якої залежить від випадку, називається випадковою величиною [5]. Отже, *випадковою величиною* називається функція $X=X(\omega)$, визначена на множині елементарних подій Ω , $\omega \in \Omega$.

У випадку скінченної або зліченної Ω випадковою величиною є будь-яка функція, визначена на Ω . У загальному випадку функція $X(\omega)$ повинна бути такою, щоб для будь-яких x подія $A=\{\omega: X(\omega) < x\}$, що полягає в тому, що випадкова величина X попадає в інтервал $(-\infty, x)$, належало полю подій S і, таким чином, для будь-якої такої події була визначена ймовірність $P(A)=P\{X < x\}$.

Випадкові величини позначаються прописними латинськими буквами $X, Y, Z, \dots$, а значення випадкової величини на елементарній події – маленькими латинськими буквами $x, y, z, \dots$. Якщо множина елементарних подій скінченна, то випадкову величину можна задати, перелічивши її значення на всіх елементарних подіях. Однак у багатьох задачах немає необхідності розглядати випадкові величини як функції від елементарної події, а досить лише знати ймовірності будь-яких подій, пов'язаних з випадковою величиною, тобто закон розподілу випадкової величини. Говорять, що *закон розподілу випадкової величини X заданий*, якщо для будь-якої множини дійсних чисел B , що є об'єднанням або перетинанням скінченного або зліченного числа проміжків, задана ймовірність $P\{\omega: X(\omega) \in B\}$ події, що полягає в тому, що $X(\omega)$ прийме значення із цієї множини [1].

Випадкові величини діляться на безперервні й дискретні. Значення, прийняті безперервною випадковою величиною, повністю заповнюють один або кілька числових інтервалів. Значення дискретної випадкової величини розташовані дискретно, тобто суцільно не заповнюють ніякий числовий інтервал. Наприклад, нехай досліджується величина частотного інтервалу мелодійного завершення фрази (частотний інтервал виражається відношенням максимуму частоти основного тону до мінімуму частоти основного тону розглянутого мовного відрізка). Очевидно, що ця величина є випадковою і її значеннями можуть бути будь-які числа, більші одиниці. Таким чином, величина частотного інтервалу є безперервною випадковою величиною. Якщо ж вивчається кількість слів у різних фразах даної мови, то ця величина також буде випадковою, але її значеннями можуть бути тільки цілі позитивні числа, які, природно, не заповнюють ніякого числового інтервалу. Тому ця випадкова величина буде дискретною. Таким чином, випадкова величина дискретна, якщо існує скінченна або рахункова множина чисел $x_1, x_2, x_3, \dots$ таких, що $P\{X=x_n\}=p_n \geq 0, n=1, 2, 3, \dots, p_1+p_2+p_3+\dots=1$. Закон розподілу дискретної випадкової величини X визначений, якщо відомі всі x_n і ймовірності $p_n=P\{X=x_n\}$ такі, що $p_1+p_2+p_3+\dots=1$. Якщо скласти таблицю, у верхньому рядку якої помістити значення дискретної випадкової величини, а в нижньому – відповідні ймові-

рності, то одержимо *ряд розподілу випадкової величини*. Сума ймовірностей, записаних у другому рядку таблиці, повинна дорівнювати одиниці. Ряд розподілу задає закон розподілу дискретної випадкової величини.

Закон розподілу випадкової величини X заданий, якщо можна визначити ймовірність події $\{\omega: X(\omega) \in A\}$ для будь-якої множини A , що є об'єднанням або перетинанням довільних проміжків. У свою чергу, будь-які такі множини A можна одержати за допомогою операцій об'єднання, перетинання й доповнення із множини всіляких інтервалів виду $(-\infty, x)$. Отже, і події $\{\omega: X(\omega) \in A\}$ можна одержати за допомогою відповідних операцій над подіями із усіляких подій виду $\{\omega: X(\omega) < x\}$. По визначенню випадкової величини для будь-яких x подія $\{\omega: X(\omega) < x\}$ належить полю подій S . Отже, і події виду $\{\omega: X(\omega) \in A\}$ також належать S і для них визначена ймовірність. Таким чином, щоб задати закон розподілу довільної випадкової величини X , досить знати для будь-яких x імовірності подій $\{\omega: X(\omega) < x\}$.

Функція розподілу $F_X(x)$ випадкової величини X визначається формулою

$$F_X(x) = P\{\omega: X(\omega) < x\}.$$

Цю рівність зазвичай записують коротше у вигляді $F_X(x) = P\{X < x\}$. Для простоти в тих випадках, коли це не може привести до неточності, пишуть $F(x)$ замість $F_X(x)$.

Розглянемо властивості функції розподілу.

1. Функція розподілу приймає значення із проміжку $[0, 1]$: $0 \leq F(x) \leq 1$. Доказ. Дана властивість випливає з того, що функція розподілу – це ймовірність події $\{X < x\}$, а значення ймовірності будь-якої події не є від'ємним й не перевищує одиниці.

2. Імовірність того, що випадкова величина прийме значення з напівінтервалу $[x_1, x_2)$, дорівнює різниці $F(x_2) - F(x_1)$:

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

Доказ. Представимо подію, що полягає в тому, що випадкова величина прийме значення, менше x_2 , у вигляді суми неспільних подій $\{\omega: X(\omega) < x_2\} = \{\omega: X(\omega) < x_1\} \cup \{\omega: x_1 \leq X(\omega) < x_2\}$. За аксіомою додавання, імовірність суми неспільних подій дорівнює сумі ймовірностей цих подій:

$$F(x_2) = P\{\omega: X(\omega) < x_2\} = P\{\omega: X(\omega) < x_1\} + P\{\omega: x_1 \leq X(\omega) < x_2\} = \\ = F(x_1) + P\{\omega: x_1 \leq X(\omega) < x_2\},$$

звідки й випливає доказувана властивість.

3. Функція розподілу – неубутна функція, тобто $F(x_2) \geq F(x_1)$, якщо $x_2 > x_1$. Доказ. Нехай $x_2 > x_1$. Повторюючи наведені вище викладення й з огляду на те, що ймовірність будь-якої події не є від'ємною, одержуємо

$$F(x_2) = F(x_1) + P\{\omega: x_1 \leq X(\omega) < x_2\} \geq F(x_1).$$

4. $P\{X \geq x\} = 1 - F(x)$. Доказ. Подія $\{\omega: X(\omega) \geq x\}$ є протилежною події $\{\omega: X(\omega) < x\}$ і, отже, $P\{X \geq x\} = 1 - P\{X < x\} = 1 - F(x)$.

5. Якщо $x \rightarrow \infty$, то $F(x) \rightarrow 1$.

6. Якщо $x \rightarrow -\infty$, то $F(x) \rightarrow 0$.

7. Функція розподілу безперервна ліворуч, тобто $\lim_{\Delta \rightarrow +0} F(x - \Delta) = F(x)$.

Функцією розподілу може бути будь-яка функція, що задовольняє властивостям 1, 3, 5, 6 і 7, тобто безперервна ліворуч у кожній точці x монотонна неубутна функція, множиною значень якої є проміжок $(0, 1)$, що включає або не включає граничні точки 0 і 1.

3.2. Розподіл дискретних випадкових величин

Дискретні випадкові величини приймають скінченну або зліченну множину значень. Нехай X – дискретна випадкова величина, що приймає значення $x_1 < x_2 < x_3 < \dots$ з ймовірностями $p_1, p_2, p_3, \dots, p_1 + p_2 + p_3 + \dots = 1$. Якщо по осі абсцис відкласти $x_1, x_2, x_3, \dots$, а по осі ординат – відповідні ймовірності $p_1, p_2, p_3, \dots$ і з'єднати сусідні точки відрізками, то одержимо *багатокутник розподілу випадкової величини X*. Багатокутник розподілу – це графічне зображення ряду розподілу дискретної випадкової величини. Розглянемо функцію розподілу $F(x)$ дискретної випадкової величини X . Якщо $x \leq x_1$, то $F(x) = P\{X < x\} = 0$, тому що в цьому випадку подія $\{X < x\}$ є неможливою. Якщо $x_1 < x \leq x_2$, то подія $\{X < x\}$ наступить тоді й тільки тоді, коли наступить подія

$\{X=x_1\}$, тому $F(x)=P\{X<x\}=P\{X=x_1\}=p_1$. Якщо $x_2<X\leq x_3$, то подія $\{X<x\}$ дорівнює сумі подій $\{X=x_1\}$ і $\{X=x_2\}$ і

$$F(x)=P\{X<x\}=P\{X=x_1\}+P\{X=x_2\}=p_1+p_2.$$

Аналогічно, якщо $x_i<X<x_{i+1}$, то $F(x)=p_1+p_2+\dots+p_i$. Таким чином, функція розподілу дискретної випадкової величини дорівнює $F(x)=\sum_{x_i<x} p_i$, де $p_i=P\{X=x_i\}$ і підсумовування виробляється

по тим i , для яких $x_i<x$.

Функція розподілу дискретної випадкової величини постійна на проміжках $(-\infty, x_1]$, $(x_1, x_2]$, $(x_i, x_{i+1}]$, У точках $x_1, x_2, x_3, \dots$ функція розподілу має стрибки, що дорівнюють ймовірності того, що випадкова величина прийме відповідне значення.

Приклад 3.1. Знайти функцію розподілу випадкової величини, що дорівнює кількості випадків «герба» при киданні двох монет.

Рішення. Залежно від результату досвіду, ця кількість може виявитися рівною 0, 1 і 2. Виберемо в якості елементарних наступні події: $\omega_1=\Gamma\Gamma=\{\text{на першій монеті випав «герб», на другій монеті – теж «герб»}\}$, $\omega_2=\text{РГ}=\{\text{на першій монеті випала «решка», на другій монеті – «герб»}\}$, $\omega_3=\text{ГР}=\{\text{на першій монеті випав «герб», на другій монеті – «решка»}\}$, $\omega_4=\text{РР}=\{\text{на першій монеті випала «решка», на другій монеті – теж «решка»}\}$. Кількість X «гербів, що випали», є функцією від елементарної події $X(\omega_1)=X(\Gamma\Gamma)=2$, $X(\omega_2)=X(\text{РГ})=1$, $X(\omega_3)=X(\text{ГР})=1$, $X(\omega_4)=X(\text{РР})=0$. Результат експерименту є випадковим, тому значення X також є випадковим. Функція $X=X(\omega_i)$ – випадкова величина, визначена на множині $\Omega=\{\omega_1, \omega_2, \omega_3, \omega_4\}$.

Якщо монети симетричні, то елементарні події $\omega_1, \omega_2, \omega_3, \omega_4$ однаково можливі й тому природно в математичній моделі експерименту вважати ймовірності $P(\omega_i)$ рівними $1/4$, $i=1, 2, 3, 4$. Знаючи ймовірності $P(\omega_i)$, можна знайти ймовірність того, що випадкова величина X прийме те або інше значення. У наведеній таблиці подані елементарні події ω_i , ймовірності $P(\omega_i)$ і значення випадкової величини X , рівної кількості «гербів, що випали».

Таблиця 3.1

i	1	2	3	4
ω_i	ГГ	РГ	ГР	РР
$P(\omega_i)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
$X(\omega_i)$	21	1	1	0

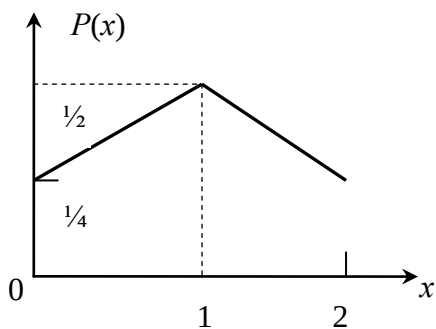
Слід зазначити, що випадкова величина X може приймати на різних елементарних подіях ті самі значення. Так, у розглянутому прикладі $X(\omega_2)=X(\omega_3)=1$. Подія, що полягає в тому, що випадкова величина X прийме значення, рівне нулю, дорівнює $\{PP\}$. Імовірність цієї події дорівнює $\frac{1}{4}$. Подія $\{\omega_i; X(\omega_i)=1\}$ дорівнює події $\{РГ, ГР\}$. Її ймовірність дорівнює $P\{X=1\}=P\{РГ, ГР\}=\frac{1}{2}$. Нарешті, $P\{X=2\}=P\{ГГ\}=\frac{1}{4}$. При побудові математичної моделі в цьому випадку можна було б об'єднати події РГ і ГР в одну й у якості елементарних розглядати події $\{0\}$, $\{1\}$ і $\{2\}$, що полягають в тому, що кількість «гербів, що випали», дорівнює нулю, одиниці або двійці відповідно: $\{0\}=\{PP\}$, $\{1\}=\{ГР, РГ\}$, $\{2\}=\{ГГ\}$. Щоб задати ймовірнісний простір і випадкову величину на «звуженому» просторі елементарних подій, у цьому випадку досить задати лише можливі значення випадкової величини й відповідні ймовірності, з якими ці значення приймаються. Вони наведені в таблиці 3.2, де x_i – значення випадкової величини, $p_i=P\{X=x_i\}$, $i=1, 2, 3$.

Таблиця 3.2

x_i	0	1	2
p_i	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

На мал. 3.1 наведений багатокутник розподілу розглянутої випадкової величини.

Знайдемо тепер функцію розподілу цієї випадкової величини. Функція розподілу при $x \leq 0$ дорівнює нулю. У точці $x=0$ вона має стрибок, рівний $\frac{1}{4}$, у точці $x=1$ – стрибок, рівний $\frac{1}{2}$, у точці $x=2$ – стрибок, рівний $\frac{1}{4}$. Між цими точками функція розподілу постійна. При $x > 2$ функція $F(x)=1$.

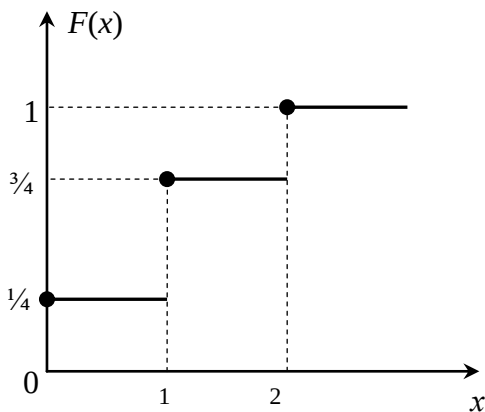


Мал.3.1

Таким чином,

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

Графік функції $F(x)$ зображений на мал. 3.2.



Мал. 3.2.

Приведемо тепер закони розподілу дискретних випадкових величин.

3.2.1. Рівномірний розподіл на множині $\{1, 2, \dots, n\}$

Випадкова величина X , що приймає цілочисельні значення від 1 до n , має рівномірний розподіл, якщо $P\{X=m\}=1/n$, $m=1, 2, \dots, n$. Багатокутник розподілу випадкової величини x являє собою відрізок прямої, паралельний осі абсцис. Кінці відрізка мають координати $(1, 1/n)$ і $(n, 1/n)$.

3.2.2. Гіпергеометричний розподіл

Випадкова величина X має гіпергеометричний розподіл, якщо

$$P\{X=m\} = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}, \quad m=0, 1, 2, \dots, \min(n, M).$$

Гіпергеометричний розподіл має місце, наприклад, у наступній задачі. Нехай у партії з N виробів M виробів 1-го сорту, а інші $N - M$ виробів – 2-го сорту. Із цієї партії витягають для контролю n виробів. Потрібно знайти закон розподілу випадкової величини X , рівної кількості виробів 1-го сорту серед обраних n виробів. Відповідно до (1.2) X має гіпергеометричний розподіл.

3.2.3. Геометричний розподіл

Випадкова величина X має геометричний розподіл, якщо $P_m = P\{X=m\} = q^m p$, $m=0, 1, 2, \dots$, $0 < p < 1$, $q=1-p$.

Підсумувавши нескінченно убутну геометричну прогресію, легко переконатися в тому, що

$$\sum_{m=0}^{\infty} P_m = 1: \sum_{m=0}^{\infty} P_m = \sum_{m=0}^{\infty} p q^m = p \frac{1}{1-q} = 1.$$

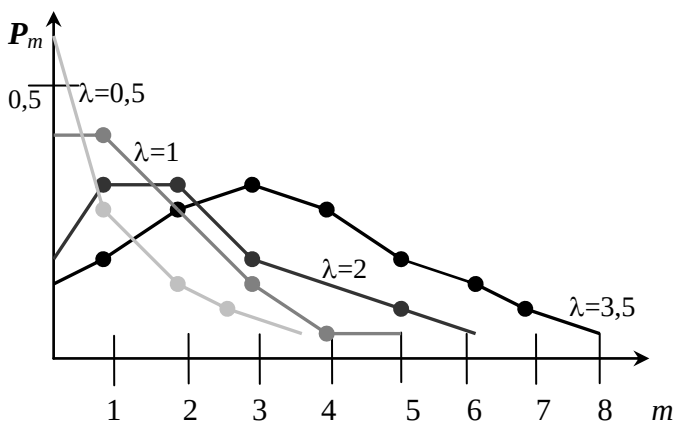
Геометричний розподіл має випадкова величина X , що дорівнює кількості випробувань Бернуллі до першого успіху з імовірністю успіху в одиничному випробуванні p .

3.2.4. Розподіл Пуассона

Випадкова величина X має розподіл Пуассона з параметром λ , якщо

$$p_m = P\{X=m\} = \frac{\lambda^m}{m!} e^{-\lambda}, \quad m=0, 1, 2, \dots, \lambda>0.$$

Скориставшись розкладанням e^λ у ряд Маклорена $e^\lambda = \sum_{m=0}^{\infty} \frac{\lambda^m}{m!}$, одержуємо



Мал. 3.3

$$\sum_{m=0}^{\infty} p_m = \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} e^{-\lambda} = e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = e^{-\lambda} e^{\lambda} = 1.$$

Багатокутники розподілу випадкових величин, що мають розподіл Пуассона з різними значеннями параметра λ , зображені на мал. 3.3.

По цьому типу розподіляються, наприклад, частоти слів з великою кількістю значень (у словнику) [6].

3.2.5. Біноміальний розподіл

Випадкова величина X , значення якої дорівнюють цілим числам від 0 до n , має біноміальний розподіл, якщо $P\{X=m\}$ задається формулою Бернуллі (2.4): $P\{X=m\}=P_n(m)=C_n^m p^m q^{n-m}$, $m=0, 1, 2, \dots, n$, $0 < p < 1$, $q=1-p$. Ряд розподілу випадкової величини X має такий вигляд:

m	0	1	2	...	k	...	n
$P_n(m)$	q^n	$C_n^1 p q^{n-1}$	$C_n^2 p^2 q^{n-2}$		$C_n^k p^k q^{n-k}$		p^n

По формулі Бернуллі можна знайти ймовірність m успіхів в n випробуваннях Бернуллі з імовірністю успіху p і невдачі $1-p=q$. Таким чином, випадкова величина, що дорівнює кількості успіхів в n випробуваннях Бернуллі, має біноміальний закон розподілу.

При великих значеннях n обчислення ймовірностей $P_n(m)$ по формулі Бернуллі стає незручним. Однак у ряді випадків вдається замінити формулу Бернуллі підходящою наближеною асимптотичною формулою. Зокрема, якщо n завелике, а ймовірність успіху p мала, то

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

Якщо при великих n імовірність p успіху у випробуванні Бернуллі близька до одиниці, то ймовірність невдачі q близька до нуля. У такому випадку пуассонове наближення можна використати для обчислення ймовірності кількості невдач.

3.3. Розподіл безперервних випадкових величин

Множина значень безперервної випадкової величини незліченна й звичайно являє собою деякий проміжок, скінченний або нескінченний. Безперервна випадкова величина визначена на незліченній множині елементарних подій Ω .

Нехай множина елементарних подій Ω збігається із множиною дійсних чисел $\Omega=\{x: x \in (-\infty, \infty)\}$ і нехай випадковими подіями з поля подій S є всілякі скінченні чи незліченні об'єднання. Імовірність на полі подій S задається за допомогою невід'ємної функції $\varphi(x)$, інтегрувальної на будь-якому скінченному або нескінченному проміжку й такий, що $\int_{-\infty}^{\infty} \varphi(x) dx = 1$. Якщо подія A

дорівнює об'єднанню проміжків, що не перетинаються, то ймовірність $P(A)$ дорівнює сумі інтегралів від $\varphi(x)$ по цих проміжках:

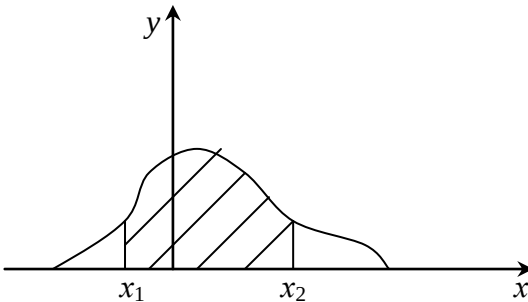
$$P(A) = \int_A \varphi(x) dx.$$

Визначена в такий спосіб на S імовірність задовольняє всім аксіомам теорії ймовірностей.

Описаний імовірнісний простір називається *безперервним імовірнісним простором*.

Не всяка випадкова величина, визначена на незліченній множині Ω , є безперервною. Випадкова величина X називається *безперервною*, якщо існує невід'ємна функція $p_X(x)$ така, що при будь-яких x функцію розподілу $F_X(x)$ можна представити у вигляді:

$$F_X(x) = \int_{-\infty}^x p_X(y) dy.$$



Мал. 3.4

Будемо розглядати тільки такі випадкові величини, для яких $p_X(x)$ безперервна скрізь, крім, можливо, скінченного числа точок.

Функція $p_X(x)$ називається *щільністю розподілу ймовірностей* або

щільністю розподілу. Якщо з контексту ясно, про яку випадкову величину мова йде, то пишуть $p(x)$ замість $p_X(x)$. З визначення випливає, що в точках безперервності щільність розподілу дорі-

внє похідній функції розподілу: $p(x)=F'(x)$. З визначення також випливає, що щільність розподілу $p_X(x)$ визначає закон розподілу випадкової величини X .

Щільність розподілу випадкової величини має наступні властивості:

$$1. \quad \text{Для будь-яких } x_1 < x_2 \quad P\{x_1 \leq X < x_2\} = \int_{x_1}^{x_2} p(x) dx.$$

Імовірність $P\{x_1 \leq X < x_2\}$ дорівнює площі фігури, обмеженої прямими $x=x_1$, $x=x_2$, $y=0$ і щільністю розподілу $y=p(x)$. На мал. 3.4 зображена щільність розподілу ймовірностей $p(x)$ і заштрихована площа фігури, що дорівнює ймовірності $P\{x_1 \leq X < x_2\}$.

$$2. \quad \text{Інтеграл по всій числовій прямій від щільності розподілу ймовірностей дорівнює одиниці, тобто} \quad \int_{-\infty}^{\infty} p(x) dx = 1.$$

3. Імовірність того, що безперервна випадкова величина прийме конкретне значення, дорівнює нулю, тобто $P\{X=a\}=0$.

Із властивості 3 випливає, що для безперервних випадкових величин

$$P\{x_1 \leq X < x_2\} = P\{x_1 \leq X \leq x_2\} = P\{x_1 < X < x_2\}.$$

Щільністю розподілу може бути будь-яка невід'ємна функція, інтеграл від якої по всій числовій осі дорівнює одиниці.

Приклад 3.2. Нехай простір елементарних подій Ω є відрізком, $\Omega = [-1, 1]$, подіями є всілякі об'єднання проміжків, що належать відріzkу $[-1, 1]$. Нехай далі ймовірність подій задається за допомогою функції $f(\omega) = 1/2$, побудованої на всьому відріzkу

$$[-1, 1]: \quad P\{A\} = P\{\omega; \omega \in A\} = \int_A \frac{1}{2} d\omega. \quad \text{Якщо } A = [a, b] \in [-1, 1], \text{ то}$$

$$P\{A\} = (b - a)/2.$$

Ми визначили всі елементи імовірнісного простору $\{\Omega, S, P\}$. Задамо тепер на множині елементарних подій Ω випадкову величину. Нехай $X(\omega) = 2 \cdot |\omega|$ (наприклад, при $\omega = -0.4$ $X(\omega) = 2 \cdot 0.4 = 0.8$). Знайдемо функцію розподілу й щільність розподілу випадкової величини X . Так як $0 \leq X(\omega) \leq 2$, то

$F(x)=0$ для $x \leq 0$ і $F(x)=1$ при $x \geq 2$. Знайдемо $F(x)$ для $x \in (0, 2)$.

Множина $A_x = \{\omega: X(\omega) < x\}$ являє собою інтервал $(-x/2, x/2)$:

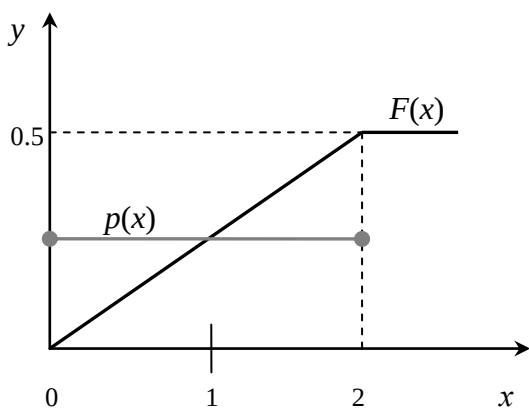
$$A_x = \{\omega: X(\omega) < x\} = \{\omega: 2 \cdot |\omega| < x\} = \{\omega: -x/2 < \omega < x/2\}.$$

Для $x \in (0, 2)$ функція $F(x)$ дорівнює

$$F(x) = P\{A_x\} = \int_{A_x} f(\omega) d\omega = \int_{-\frac{x}{2}}^{\frac{x}{2}} \frac{1}{2} d\omega = \frac{1}{2} x.$$

У результаті маємо

$$F(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ \frac{1}{2} x & \text{при } 0 < x \leq 2, \\ 1 & \text{при } x > 2. \end{cases}$$



Мал. 3.5

Графік функції $F(x)$ зображений на мал. 3.5. Функція розподілу $F(x)$ безперервна на всій числовій осі й диференцуєма скрізь, крім крапок $x=0$ і $x=2$. Отже, X – безперервна випадкова величина.

Знайдемо її щільність розподілу

$$p(x) = F'(x) = \begin{cases} 0 & \text{при } x < 0, \\ \frac{1}{2} & \text{при } 0 < x < 2, \\ 0 & \text{при } x > 2. \end{cases}$$

У крапках $x=0$ і $x=2$ можна покласти $p(x)$ рівною довільному числу, наприклад нулю. Графік щільності розподілу $p(x)$ також наведений на мал. 3.5.

Приведемо тепер закони розподілу безперервних випадкових величин.

3.3.1. Рівномірний розподіл

Безперервна випадкова величина X , що приймає значення на відрізку $[a, b]$, має рівномірний розподіл, якщо щільність розподілу $p(x)$ має вигляд

$$p(x) = \begin{cases} 0 & \text{при } x \notin [a, b], \\ \frac{1}{b-a} & \text{при } x \in [a, b]. \end{cases}$$

Для випадкової величини, що має рівномірний розподіл, імовірність того, що випадкова величина прийме значення із заданого інтервалу $(x, x+\Delta) \in [a, b]$, не залежить від розміщення цього інтервалу на числовій осі й пропорційна довжині цього інтервалу Δ :

$$P\{x < X < x + \Delta\} = \int_x^{x+\Delta} \frac{1}{b-a} dt = \frac{\Delta}{b-a}.$$

Функція розподілу має вигляд:

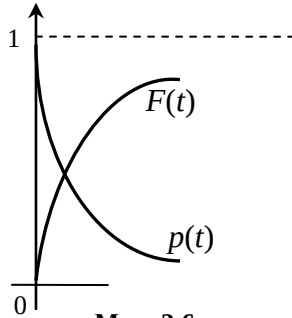
$$F(x) = \begin{cases} 0 & \text{при } x \leq a, \\ \frac{x-a}{b-a} & \text{при } a < x \leq b, \\ 1 & \text{при } x > b. \end{cases}$$

3.3.2. Експонентний розподіл

Безперервна випадкова величина X , що приймає ненегативні значення, має експонентний (показовий) розподіл з параметром λ , якщо щільність розподілу випадкової величини при $x \geq 0$ дорівнює $p(x) = \lambda e^{-\lambda x}$ і при $x < 0$ $p(x) = 0$. Функція розподілу випадкової величини X дорівнює

$$F(x) = \int_{-\infty}^x p(t) dt = \begin{cases} 0 & \text{при } x \leq 0, \\ 1 - e^{-\lambda x} & \text{при } x > 0. \end{cases}$$

Графіки функції розподілу й щільності розподілу наведені на мал. 3.6.



Мал. 3.6

3.3.3. Нормальний розподіл

Безперервна випадкова величина X має нормальний розподіл ймовірностей з параметрами a й $\sigma > 0$, якщо її щільність розподілу має вигляд [1, 4, 7]

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-a}{\sigma}\right)^2}.$$

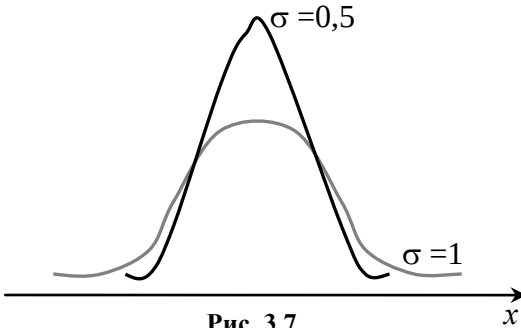


Рис. 3.7

Якщо X має нормальний розподіл, то коротко це записується у вигляді $X \sim N(a, \sigma)$. Графіки щільності розподілу при різних значеннях параметра σ і тому самому значенні a наведені на мал. 3.7.

Щільність розподілу $p(x)$ симетрична відносно прямої $x=a$ (тобто $p(a+y)=p(a-y)$). Якщо $x \rightarrow \pm\infty$, то $p(x) \rightarrow 0$. При зменшенні σ графік «стягається» до своєї осі симетрії $x=a$. Нормальний розподіл відіграє особливу роль у теорії ймовірностей і її застосуваннях. Це пов'язане з тим, що при виконанні певних умов сума великої кількості випадкових величин має «приблизно» нормальний розподіл.

Інтегральна функція цього розподілу виражається співвідношенням [4]

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

Зміст кривої нормального розподілу полягає в наступному: значення випадкової величини з нормальним законом розподілу розташовуються по обидві сторони від генерального середнього a , і чим далі ці значення віддалені від a , тим вони менш імовірні. Крім того, чим менше σ^2 , тим швидше гілки кривої наближаються до осі абсцис, тобто тим менш імовірні значення випадкової величини, що сильно ухиляються від a [5].

Незважаючи на те, що нормальний розподіл характеризує ряди з безперервними одиницями, при досить великій кількості спостережень воно добре описує й варіювання дискретних одиниць. До цього закону наближаються й інші типи розподілів. Прийнято вважати, що частотні одиниці мови (тексту) розподіляються за нормальним законом [6]. Навіть якщо на практиці отримані в ході лінгвостатистичних досліджень варіаційні ряди підкоряються якомусь іншому розподілу, вважається, що при низькій точності, запропонованої до лінгвостатистичним розрахунків, можна допустити, що розподіл є нормальним [8].

3.4. Контрольні запитання

1. Що таке випадкова величина?
2. Коли заданий закон розподілу випадкової величини?
3. Як одержати ряд розподілу випадкової величини?
4. Якою формулою визначається функція розподілу випадкової величини?
5. Перелічить властивості функції розподілу.
6. Які вам відомі закони розподілу дискретних випадкових величин.
7. Які лінгвістичні явища підкоряються цим законам.
8. Перелічить відомі вам закони розподілу безперервних випадкових величин.
9. Які лінгвістичні явища мають такий розподіл?

4. ЧИСЛОВІ ХАРАКТЕРИСТИКИ ВИПАДКОВОЇ ВЕЛИЧИНИ

4.1. Вибіркове середнє й вибіркова дисперсія

Для вивчення випадкової величини, що характеризує деяке явище (наприклад, мелодія фрази, наголос, темп проголошення тощо), проводиться експеримент. Кожне випробування (результат) експерименту доставляє дослідникові певне значення випадкової величини. Сукупність значень випадкової величини, отриманих у ході експерименту, називається *випадковою вибіркою*. Нехай випадкова величина X у ході експерименту, що складається з n випробувань, прийняла ряд значень $x_1, x_2, \dots, x_n$ (не обов'язково всі x різні між собою). Тоді говорять, що $(x_1, x_2, \dots, x_n)$ є випадковою вибіркою для X об'єму n . Кожна випадкова вибірка несе в собі деяку якісну й кількісну інформацію про досліджуване явище. Щоб зробити цю інформацію більш компактною й доступною для огляду, використовуються деякі числові характеристики випадкової вибірки, які, природно, розглядаються як наближення до відповідних характеристик самої випадкової величини X . До таких числових характеристик належать вибіркове середнє арифметичне значення й вибіркова дисперсія. Іноді слово «вибіркове» опускають.

Вибіркове середнє арифметичне значення – це число, навколо якого розташовуються (по обидві сторони) всі значення вибірки. У загальному виді для вибірки $(x_1, x_2, \dots, x_n)$ маємо:

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

Щоб визначити характер розташування вибірових значень щодо свого середнього арифметичного значення $\bar{x}$, обчислюється *вибіркова дисперсія*. Її зазвичай позначають через S^2 . Для її одержання необхідно від кожного вибіркового значення відняти середнє арифметичне значення, результати піднести до квадрата

та, а потім скласти й отриману суму розділити на $(n - 1)$, де n – об'єм вибірки, тобто

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

Зміст вибіркової дисперсії як міри розсіювання (розкиду) полягає в наступному: якщо є дві випадкові вибірки того самого об'єму n (або приблизно однакових об'ємів) і з тим самим середнім арифметичним значенням $\bar{x}$, то та вибірка, у якої S^2 менше, має значення, більш компактно розташовані відносно $\bar{x}$. Як правило, така вибірка виникає при вивченні більш однорідного матеріалу.

Корінь квадратний з вибіркової дисперсії, тобто S , називається *вибірковим середнім квадратичним відхиленням*.

Приклад 4.1. Нехай метою експерименту є вивчення тривалості наголошених голосних в англійській мові, обмірюваної в мсек. У результаті десяти випробувань була отримана випадкова вибірка: 206, 110, 136, 150, 200, 164, 170, 178, 140, 166. Очевидно, що об'єм вибірки дорівнює 10. Щоб одержати середнє арифметичне значення для цієї вибірки, необхідно всі знайдені числа скласти й суму розділити на 10. Таким чином, позначаючи вибіркове середнє через $\bar{x}$, будемо мати

$$\begin{aligned} \bar{x} &= \frac{206+110+136+150+200+164+170+178+140+166}{10} = \\ &= \frac{1620}{10} = 162 \text{ (мсек)}. \end{aligned}$$

$$\begin{aligned} S^2 &= \frac{1}{9} \times \\ &\times ((206-162)^2 + (110-162)^2 + (136-162)^2 + (150-162)^2 + (200-162)^2 + \\ &+ (164-162)^2 + (170-162)^2 + \\ &+ (178-162)^2 + (140-162)^2 + (166-162)^2) = \frac{7728}{9} = 858.67. \end{aligned}$$

$$S = \sqrt{858.67} = 29.3.$$

Розглянутий приклад показує, що визначення $\bar{x}$ й S^2 пов'язане з великою кількістю обчислень, що збільшуються разом з

об'ємом вибірки. Однак процес обчислень може бути істотно спрощений без особливої втрати точності. Для цього проводять *угрупкування* вибірових значень. Нехай є вибірка об'єму n ($x_1, x_2, \dots, x_n$). Можна вважати, що числа $x_1, x_2, \dots, x_n$ розташовані в порядку зростання (цього завжди можна домогтися). Вибираємо деяке ціле число k , таке, що $10 \leq k \leq 20$, складаємо різницю $x_n - x_1$ і визначаємо число h по формулі

$$h = \frac{x_n - x_1}{k}.$$

Тепер можна вказати інтервали групування: (x_1, x_1+h) , (x_1+h, x_1+2h) , $\dots$, $(x_1+(n-1)h, x_n)$. Кожне зі значень нашої випадкової вибірки $x_1, x_2, \dots, x_n$ потрапляє в один із зазначених вище інтервалів. Так, якщо деяке вибірове значення x_l задовольняє нерівності $x_1+lh \leq x_l < x_1+(l+1)h$, то воно належить до інтервалу $(x_1+lh, x_1+(l+1)h)$. Позначимо через n_i кількість значень вибірки, що потрапили в інтервал $(x_1+(i-1)h, x_1+ih)$. Якщо деяке вибірове значення потрапило на границю двох інтервалів, то воно належить до правого інтервалу. Числа n_i називаються *частотами відповідних інтервалів*. Тепер для визначення $\bar{x}$ й S^2 діють по наступному алгоритму:

1. Фіксується якийсь із побудованих інтервалів (зазвичай беруть інтервал з максимальним значенням частоти n_i).

2. Обчислюється середина зафіксованого інтервалу. Нехай це буде x_0 .

3. Всі інтервали нумеруються в зростаючому порядку від 1 до k (у нас рівно k інтервалів).

4. З номера кожного інтервалу віднімається номер зафіксованого інтервалу (ці різниці з відповідними знаками позначаються через $c_i = N_i - N_0$).

5. Далі користуються формулами:

$$\bar{x} = x_0 + \frac{1}{n} \left(\sum_{i=1}^k n_i c_i \right) \cdot h.$$

$$S^2 = \left[\sum_{i=1}^k n_i c_i^2 - \frac{1}{n} \left(\sum_{i=1}^k n_i c_i \right)^2 \right] \cdot h^2 \cdot \frac{1}{n-1}.$$

Тут $n_i c_i$ – добуток частоти n_i інтервалу на відповідне йому число c_i , n – об'єм вибірки.

Природно очікувати, що інтервальна розбивка приводить до деяких помилок в обчисленні вибіркового середнього й вибіркової дисперсії. Щоб нівелювати цю погрішність, можна внести виправлення на групування в $\bar{x}$ і S^2 . Виявляється, що виправлень потребує тільки величина S^2 . З урахуванням виправлення (її вираження зазначене Шеппардом і дорівнює $-\frac{h^2}{12}$) маємо вираження для S^2 :

$$S^2 = \left[\sum_{i=1}^k n_i c_i^2 - \frac{1}{n} \left(\sum_{i=1}^k n_i c_i \right)^2 \right] \cdot \frac{h^2}{n-1} - \frac{h^2}{12}.$$

Зазвичай виправлення Шеппарда становить менш 5% від виправленого S^2 , а тому її варто вносити тільки для вибірок досить великого об'єму, коли є впевненість, що вибіркова дисперсія добре наближає справжню дисперсію.

Приклад 4.2. При вивченні тривалості наголошених голосних в англійській мові була отримана вибірка, що містить 699 значень. Обчислення тривалості здійснювалося в мсек, і після групування отриманих даних були заповнені стовпці в таблиці 4.1.

З таблиці видно, що $k=15$, $h=20$. Зафіксуємо інтервал з № 7 і обчислимо числа c_i – стовпець № 3. Для одержання чисел стовпця № 4 перемножуються відповідні елементи другого й третього стовпців. Останній (п'ятий) стовпець отримується множенням елементів стовпця № 2 на квадрати чисел стовпця № 3 (піднесення у квадрат виражається зазвичай усно). Складаючи елементи другого, четвертого й п'ятого стовпців, відповідно одержуємо значення для $n = \sum n_i = 699$, $\sum_{i=1}^k n_i c_i = 363$, $\sum_{i=1}^k n_i c_i^2 = 5487$, звід-

ки

$$\bar{x} = 210 + \frac{363}{699} \cdot 20 = 210 + 10.4 \approx 220,$$

$$S^2 = \left[5487 - \left(\frac{363}{699} \right)^2 \right] \cdot \frac{20^2}{698} = 3000,$$

$$S = \sqrt{3000} = 54.77 \approx 55.$$

Таблиця 4.1

№ п/п	Інтервали	n_i	$c_i = N_i - N_0$	$n_i c_i$	$n_i c_i^2$
	1	2	3	4	5
1	80-100	2	-6	-12	72
2	100-120	5	-5	-25	125
3	120-140	32	-4	-128	512
4	140-160	59	-3	-177	531
5	160-180	71	-2	-142	284
6	180-200	92	-1	-92	92
7	200-220	127	-0	0	0
8	220-240	77	1	77	77
9	240-260	66	2	132	264
10	260-280	67	3	201	603
11	280-300	36	4	144	576
12	300-320	31	5	155	775
13	320-340	15	6	90	540
14	340-360	12	7	84	588
15	360-380	7	8	56	448
	Сума	$\sum n_i$		$\sum n_i c_i$	$\sum n_i c_i^2$

При рішенні було враховано, що серединою інтервалу № 7 є число 210. Так як $h=20$, то уточнене значення S^2 дорівнює

$$S^2 = 3000 - \frac{400}{12} \approx 2967,$$

$$S = \sqrt{2967} \approx 54.5.$$

4.2. Критерій нормальності розподілу

Одним з законів розподілу випадкової величини, що найбільш часто зустрічаються на практиці, є нормальний закон розподілу. Графік функції нормального розподілу симетричний відносно a . На практиці a знаходить своє вираження в середньому арифметичному значенні $\bar{x}$, а σ^2 – у вибірковому середньому S^2 . Як ми вже відзначали, значення випадкової величини з нормальним законом розподілу розташовуються по обидві сторони від генерального середнього a , і чим далі ці значення відділені від a , тим вони менш імовірні. Крім того, чим менше σ^2 , тим швидше гілки кривої наближаються до осі абсцис, тобто тим менш імовірні значення випадкової величини, що сильно ухиляються від a [5]. У силу того, що середнє арифметичне значення вибірки $\bar{x}$ так само, як і вибіркова дисперсія S^2 , є наближеннями до a і σ^2 , то можна очікувати, що характер розташування у вибірці значень випадкової величини (іноді значення називають варіантами), розподіленої за нормальним законом, буде мати зазначені властивості. Нормальна випадкова величина з генеральним середнім a й дисперсією σ^2 має наступну важливу властивість: з імовірністю β варіанти вилучені від a на відстань, не більше $d_\beta\sigma$, де d_β можна визначити з таблиці 4.2.

Таблиця 4.2

β	0.687	0.900	0.950	0.980	0.990	0.999
d_β	1.000	1.645	1.960	2.326	2.576	3.291

З таблиці видно, що практично (з імовірністю, більшою 0.99) всі значення нормальної випадкової величини вилучені від a на відстань, що не перевершує 3σ . Це твердження зветься *правилом трьох сигм*.

Нормально розподілені випадкові величини зазвичай відповідають явищам, що є результатом багатьох одночасно діючих факторів, дія кожного з яких не є переважаючою над іншими. Така ситуація виникає при вивченні явища, що протікає в однорідних умовах. Так, наприклад, акустичні характеристики наголошеного складу доцільно досліджувати на однорідному

лінгвістичному матеріалі за допомогою «однорідної» групи дикторів. Але умову однорідності не завжди можна витримати. Крім того, природа явища іноді буває мало вивчена, і тому важко визначити ступінь впливу окремих факторів на кількісні характеристики даного явища. До того ж мовне явище на рівні мовлення являє собою неоднорідний матеріал. Очевидно, не можна очікувати, що досліджуване явище обов'язково визначить нормально розподілену випадкову величину. Тому на практиці виникає задача визначення закону розподілу.

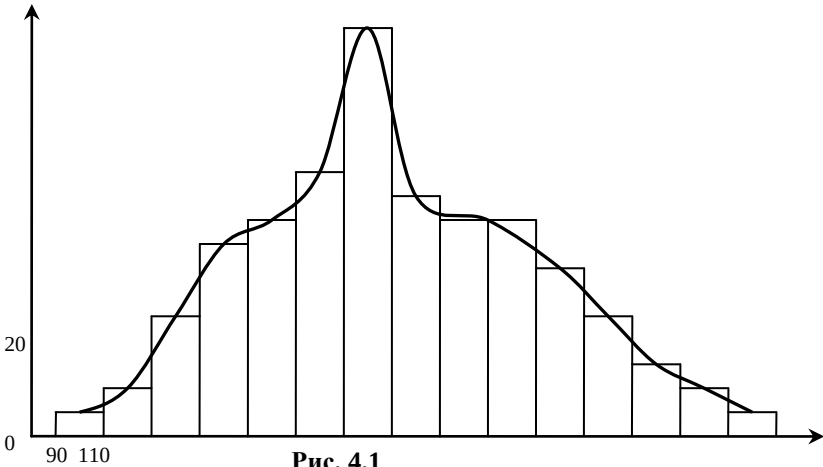
Визначення закону розподілу по вибірці розпадається на три етапи: формулювання виду закону розподілу, знаходження параметрів цього розподілу й перевірка згоди вихідної вибірки із прийнятим (сформульованим) законом розподілу. Однією з передумов для формулювання виду закону розподілу є емпіричний розподіл. Для побудови емпіричного розподілу в отриманій вибірці проводять інтервальну розбивку й складають таблицю й на її підставі будують гістограму емпіричного розподілу.

Приклад 4.3. Для умови приклада 4.2 маємо таблицю розподілу 4.3.

Таблиця 4.3

Інтервали	Частоти (n_i)	Відносні частоти (n/n)	Очікувані теоретичні ймовірності
80-100	2	2/699	
100-120	5	5/699	
120-140	32	32/699	
140-160	59	59/699	
160-180	71	71/699	
180-200	92	92/699	
200-220	127	127/699	
220-240	77	77/699	
240-260	66	66/699	
260-280	67	67/699	
280-300	36	36/699	
300-320	31	31/699	
320-340	15	15/699	
340-360	12	12/699	
360-380	7	7/699	
	$n = \sum n_i = 699$		

Цій таблиці відповідає гістограма



Приблизно через середини верхніх сторін прямокутників гістограми проводиться плавна крива. Отримана крива, хоч і віддалено, нагадує криву нормального розподілу. Тому природно припустити, що досліджувана вибірка взята з нормальної генеральної сукупності, параметри якої μ й σ^2 оцінюються на основі вибірки.

У загальному випадку, якщо із приводу гістограми було зроблене припущення про вид теоретичного розподілу й були оцінені його параметри, то залишається порівняти теоретичний закон розподілу з вибіркою. Для такого порівняння необхідно вибрати критерій перевірки гіпотези про погодженість емпіричного й теоретичного розподілу. Такі критерії називаються *критеріями згоди*. Всі критерії згоди складаються за однаковою схемою. Для цього вибирається деякий параметр, закон розподілу якого відомий, і який має досить велику «чутливість», тобто відмінність закону, що перевіряється, від дійсного істотно позначалося б на значеннях цього параметра. Найбільшою поширеністю користується *критерій Пірсона*, роль параметра в якому грає величина

$$\chi^2 = n \sum_{i=1}^k \frac{\left(\frac{n_i}{n} - p_i \right)^2}{p_i}, \quad (4.1)$$

де n – об'єм вибірки, k – кількість інтервалів розбивки, $\left(\frac{n_i}{n} \right)$ – відносні частоти інтервалів, p_i – теоретичні ймовірності цих інтервалів (у припущенні про вірність передбачуваного теоретичного розподілу).

Уведений параметр χ^2 використовується для перевірки (на підставі вибірки) гіпотези про те, що справжній закон розподілу збігається з передбачуванним. Цю гіпотезу позначають через H_0 і називають *нульовою гіпотезою*. Нульова гіпотеза H_0 відкидається, якщо ймовірність спостерігати дану вибірку в умовах гіпотези H_0 дуже мала.

На практиці прийнято вважати, що якщо ця ймовірність менше 0.05, то гіпотезу H_0 відкидають. Число 0.05 називається *рівнем значимості*. Іноді замість 0.05 беруть 0.01. Параметр χ^2 визначає критерій перевірки гіпотези H_0 . Його називають *критерієм «хі-квадрат»*. Застосування критерію засноване на тім, що для обраного рівня значимості визначають табличне χ_{05}^2 або χ_{01}^2 й порівнюють зі значенням χ^2 , обчисленим по формулі (4.1).

Щоб знайти табличне χ^2 , варто звернутися до таблиці розподілу χ^2 (див. Додаток 2) для значення $f = k - r - 1$, де k – кількість інтервалів групування, використаних при обчисленні χ^2 , а r – кількість параметрів теоретичного закону розподілу, оцінених на основі даної вибірки. Число f називається *числом ступенів волі*. Для випадку, коли передбачуваний теоретичний закон розподілу є нормальним, $f = k - 3$.

Для застосування критерію згоди χ^2 при перевірці нульової гіпотези H_0 : «дана вибірка взята з генеральної сукупності з нормальним законом розподілу» – проводиться обчислення в наступному порядку:

1. Нехай вирішено провести інтервальну розбивку спостережень вибірки на k інтервалів і нехай $y_1, y_2, \dots, y_{k+1}$ – границі всіх інтервалів.

2. Визначаються вибіркові $\bar{x}$ й S^2 .
3. Обчислюються різниці $y_1 - \bar{x}, y_2 - \bar{x}, \dots, y_{k+1} - \bar{x} \dots$
4. Ці різниці нормуються, тобто обчислюються величини

$$u_1 = \frac{y_1 - \bar{x}}{S}, u_2 = \frac{y_2 - \bar{x}}{S}, \dots, u_{k+1} = \frac{y_{k+1} - \bar{x}}{S}.$$

5. По таблиці знаходимо величини (див. Додаток 1) $\Phi(u_1), \Phi(u_2), \dots, \Phi(u_{k+1})$.
6. Складаємо різниці
7. $p_1 = |\Phi(u_1) - \Phi(u_2)|, \dots, p_k = |\Phi(u_k) - \Phi(u_{k+1})|$.
8. Обчислюємо $\tilde{n}_i = p_i n$.
9. Визначаємо значення χ^2 по формулі

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - \tilde{n}_i)^2}{\tilde{n}_i}.$$

Ця формула дає те ж значення χ^2 , що й формула (4.1), але вона більш зручна для обчислень. До того ж вона свідчить про правильність проведеної інтервальної розбивки. Вважається, що «очікувані частоти» $\tilde{n}_i$ не повинні бути менше п'яти. У протилежному випадку кілька середніх інтервалів поєднують в один, тобто тут уже не обов'язково піклуватися про рівність довжин інтервалів (таке об'єднання треба проводити вже після обчислення $\bar{x}$ й S^2).

Обчислене χ^2 порівнюємо з табличним χ_{05}^2 (або χ_{01}^2). Якщо виявиться, що $\chi^2 > \chi_{05}^2$, то гіпотеза H_0 відкидається з рівнем значимості 0.05. Рівень значимості α (зазвичай $\alpha=0.05$ або $\alpha=0.01$) означає, що ймовірність одержання даної вибірки в умовах гіпотези H_0 менше α , тобто гіпотеза H_0 слабо погоджується з даною вибіркою, а тому повинна бути відкинута. Якщо ж $\chi^2 < \chi_{05}^2$, то гіпотеза H_0 приймається. Точніше кажучи, ця гіпотеза не відкидається, і наступні експерименти можуть підтвердити її або спростувати. Особливо важливо продовжувати експеримент, якщо обчислене значення χ^2 задовольняє нерівності $\chi_{05}^2 < \chi^2 < \chi_{01}^2$.

, тому що дослідник (з обережності) може вибрати рівень значимості $\alpha=0.01$, а тому не відкидає гіпотезу H_0 , але наведена нерівність ставить під сумнів його висновок.

Приклад 4.4. Розглядаючи наведену вище гістограму, можна побачити, що в центрі її перебувають найбільш частотні варіанти, а ближче до країв – відносно рідкі варіанти. По цих ознаках гістограма нагадує криву щільності нормального розподілу. Використовуючи вибіркове середнє $\bar{x}=220$ і середнє квадратичне відхилення $S=55$, за допомогою критерію χ^2 можна перевірити гіпотезу H_0 : акустична характеристика – тривалість наголошеного складу в англійській мові – підкоряється нормальному закону розподілу з параметрами $a=220$ і $\sigma=55$. Емпіричне значення χ^2 (з його допомогою перевіряється гіпотеза H_0) обчислюється відповідно до зазначеної схеми. Результати проміжних обчислень наведені в таблиці 4.4.

Опишемо процес заповнення стовпців цієї таблиці. Було ухвалено рішення використати 15 інтервалів групування ($k=15$), причому початок першого інтервалу дорівнює 80, а довжина інтервалів дорівнює 20. У стовпці № 1 зазначені середини обраних інтервалів, у стовпці № 2 – кінці відповідних інтервалів, у стовпці № 3 – відхилення кінців інтервалів від вибіркового середнього $\bar{x}=220$. У стовпці № 4 записані числа u_i , що є нормованими відхиленнями кінців інтервалів. Ці числа отримуються діленням відповідних чисел стовпця № 3 на $S=55$ (з урахуванням знака). Стовпець № 5 містить значення $\Phi(u_i)$, які знаходять по таблиці значень функції Φ (див. Додаток 1). Так як для $u_1=-2.54$ маємо $\Phi(-2.54)=0.006$, а для $u_{10}=0.73$ маємо $\Phi(0.73)=0.767$. Числа стовпця № 6 відповідають серединам інтервалів і дорівнюють різницям значень функції Φ для кінців інтервалу (з більшого значення Φ віднімається менше значення). Числа стовпця № 6 і є теоретичні («очікувані») частоти $\tilde{n}_i$. Вони отримуються множенням чисел стовпця № 6 на об'єм вибірки n (у нашому випадку $n=699$). У стовпці № 8 зазначені емпіричні частоти n_i . Вони відбивають результат вибірки. Ці числа взяті зі стовпця № 2 таблиці 4.1.

Таблиця 4.4

Середини інтервалів	Границі інтервалів	Відхилення границь інтервалів від $\bar{x}$	Нормовані відхилення границь інтервалів u_i	$\Phi(u_i)$	$p_i = \Phi(u_i) - \Phi(u_{i+1})$	$\tilde{n}_i = n_i$	n_i
90	80	- 140	- 2.54	0.006	0.009	6.9	2
110	100	- 120	- 2.18	0.015	0.020	13.9	5
130	120	- 100	- 1.81	0.035	0.038	26.6	32
150	140	- 80	- 1.45	0.073	0.065	45.4	59
170	160	- 60	- 1.09	0.138	0.095	66.3	71
190	180	- 40	- 0.73	0.233	0.126	88.0	92
210	200	- 20	- 0.36	0.359	0.141	98.5	127
230	220	0	0	0.500	0.141	98.5	77
250	240	20	0.36	0.641	0.126	88.0	66
270	260	40	0.73	0.767	0.095	66.3	67
290	280	60	1.09	0.862	0.064	45.4	36
310	300	80	1.45	0.926	0.039	26.6	31
330	320	100	1.81	0.965	0.020	13.9	15
350	340	120	2.18	0.985	0.009	6.3	12
370	360	140	2.54	0.994	0.004	2.8	7
	380	160	2.91	0.998			

Тепер обчислюємо χ^2 . Для цього із чисел стовпця № 8 віднімаються числа стовпця № 7, отримані різниці зводяться у квадрат, а потім діляться на числа стовпця № 7. Результати підсумуються. Тому маємо:

$$\begin{aligned} \chi^2 = & \frac{(2-6.9)^2}{6.9} + \frac{(5-13.9)^2}{13.9} + \frac{(32-26.6)^2}{26.6} + \frac{(59-45.4)^2}{45.4} + \\ & + \frac{(71-66.3)^2}{66.3} + \frac{(92-88)^2}{88} + \frac{(127-98.5)^2}{98.5} + \frac{(77-98.5)^2}{98.5} + \frac{(66-88)^2}{88} + \\ & + \frac{(67-66.3)^2}{66.3} + \frac{(36-45.4)^2}{45.4} + \frac{(31-26.6)^2}{26.6} + \frac{(15-13.9)^2}{13.9} + \frac{(12-6.3)^2}{6.3} + \end{aligned}$$

$$+ \frac{(7-2.8)^2}{2.8} = \frac{24}{6.9} + \frac{79.3}{13.9} + \frac{29.2}{26.6} + \frac{207.4}{45.4} + \frac{22.1}{66.3} + \frac{16}{88} + \frac{756.2}{98.5} + \frac{462.2}{98.5} + \frac{484}{88} + \frac{0.5}{66.3} + \frac{88.4}{45.4} + \frac{18.4}{26.6} + \frac{1.2}{13.9} + \frac{32.5}{6.3} + \frac{17.6}{2.8} = 47.2$$

Тепер по таблиці розподілу χ^2 знаходимо χ_{05}^2 й χ_{01}^2 , що відповідають $f=15-3=12$. Маємо $\chi_{05}^2=21.0$, $\chi_{01}^2=26.2$. Оскільки $\chi^2=47.2>26.2$ (тобто χ_{01}^2), то гіпотеза H_0 відкидається. Таким чином, припущення про нормальний закон розподілу значень тривалості не підтвердилось.

4.3. Критерій однорідності

Цінність критерію χ^2 (для перевірки гіпотез) полягає в тому, що з його допомогою можна досліджувати якісні ознаки, які характеризуються якимись відтінками, але не мають кількісного вираження. Так, при вивченні мелодії інтонаційного типу необхідно враховувати комунікативне навантаження фраз, їхню емоційну насиченість, граматичний лад, ритмічну структуру, індивідуальні особливості голосу мовця. Якщо ні для одного із цих факторів не створено спеціальних умов, то досліджуване явище буде підкорятися нормальному закону.

На практиці часто доводиться мати справу з явищами, коли для супутніх цьому явищу факторів можуть, незалежно від бажання дослідника, створюватися сприятливі для цього фактору умови. Оскільки всяке явище вивчається в часі, то іноді ці умови можуть стати значущими. Тому перед дослідником може іноді виникати проблема підбора однорідного матеріалу для вивчення явища.

Наприклад, вивчалася частота основного тону голосного з головним наголосом в емоційних реченнях, що виражають подив. Аудиторський аналіз свідчив про чітке розходження позитивного й негативного відтінків емоції при прослуховуванні фраз у контексті. Однак коли експериментальні фрази були вичленовані з контексту, показання аудиторів виявилися плутаними, а іноді й суперечливими. Результати аудіро-

вання дозволяють висунути гіпотезу про те, що не існує акустичних ознак розходження позитивного й негативного відтінків емоції подиву, принаймні, такий параметр, як частота основного тону, не розрізняє цих відтінків. Інакше кажучи, матеріал про значення частоти основного тону, отриманий окремо для фраз із позитивними й негативними відтінками емоції подиву, є однорідним, тобто вибірки по цих відтінках емоції можна об'єднати. Висунуту гіпотезу можна перевірити за допомогою критерію χ^2 . Процес обчислення χ^2 у цьому випадку розглянемо на наступному прикладі.

Приклад 4.5. Значення частоти основного тону голосного з головним наголосом в емоційних реченнях, що виражають подив, зазначені в таблиці 4.5.

Таблиця 4.5.

Інтервали (частота основного тону)	Частоти для позитивних відтінків емоції (n_{1i})	Частоти для негативних відтінків емоції (n_{2i})	$n^{(i)}$	$\frac{n_1 n^{(j)}}{N}$	$\frac{n_2 n^{(j)}}{N}$
1	2	3	4	5	6
0.80 – 1.20	7	2	9	6.7	2.3
1.20 – 1.60	16	3	19	14.1	4.8
1.60 – 2.00	24	7	31	23.0	8.0
2.00 – 2.40	65	26	91	67.6	23.4
2.40 – 2.80	15	6	21	15.6	5.4
n_i	$n_1=127$	$n_2=44$	$N=171$		

Стовпці цієї таблиці заповнюються в такий спосіб: у результаті експерименту були отримані значення частоти основного тону (як для позитивних, так і для негативних відтінків емоції подиву), які були розподілені в п'ятьох інтервалах (0.80 – 1.20), (1.20 – 1.60), (1.60 – 2.00), (2.00 – 2.40), (2.40 – 2.80). Ці інтервали зазначені в стовпці № 1. У стовпці № 2 записано кількість фраз, що спостерігалися, з позитивним відтінком емоції подиву, склад з головним наголосом яких має значення частоти основного тону у відповідному інтервалі. Таким же чином заповнюється стовпець № 3, що відповідає негативним відтінкам емоції

подиву. Стовець № 4 – це сума відповідних елементів стовпців № 2 і № 3. Число N – сума всіх елементів стовпця № 4, n_1 – сума елементів стовпця № 2, n_2 – сума елементів стовпця № 3. Ясно, що $N=n_1+n_2$. У цьому прикладі $N=171$, $n_1=127$, $n_2=44$. Елементи стовпця № 5 (і відповідно стовпця № 6) отримуються як результат перемножування відповідних елементів стовпця № 4 на n_1 (відповідно, на n_2) і ділення результату на N .

Тепер можна приступитися до обчислення χ^2 . Для цього з кожного елемента стовпця № 2 віднімаємо відповідний елемент стовпця № 5, отримані різниці підносимо до квадрата, а результат ділимо на елемент стовпця № 5. У такий же спосіб вчиняємо з парами зі стовпців № 3 і № 6. Отримані числа підсумуємо. Сума і є значення χ^2 . Отже,

$$\begin{aligned}\chi^2 = & \frac{(7-6.7)^2}{6.7} + \frac{(2-2.3)^2}{2.3} + \frac{(16-14.1)^2}{14.1} + \frac{(3-4.8)^2}{4.8} + \frac{(24-23)^2}{23} + \\ & + \frac{(7-8)^2}{8} + \frac{(65-67.6)^2}{67.6} + \frac{(26-23.4)^2}{23.4} + \frac{(15-15.6)^2}{15.6} + \\ & + \frac{(6-5.4)^2}{5.4} = 1.63.\end{aligned}$$

Число ступенів волі $f=k-1$, де k – кількість інтервалів групування. У нашому прикладі $f=4$. З Додатку 2 знаходимо $\chi_{05}^2=9.488$ і $\chi_{01}^2=13.30$. Оскільки $\chi^2=1.61 < 9.488 = \chi_{05}^2$, то гіпотеза H_0 приймається, тобто позитивні й негативні відтінки емоції подиву однаково впливають на частоту основного тону складу з головним наголосом.

У загальному випадку, щоб побудувати критерій однорідності, припустимо, що досліджуване явище характеризується набором k взаємовиключних один одного якостей $P_1, \dots, P_k$ (у прикладі 4.5 якості характеризувалися інтервалами групування). Нехай є l вибірок відповідно об'ємів $n_1, \dots, n_l$ (у прикладі 4.5 $n=2$). Позначимо через n_{ij} кількість елементів в j -й вибірці, що володіють якістю P_j . Очевидно, що $n_i = n_{i,1} + \dots + n_{i,k}$. Позначимо ще $n^{(j)} = n_{1,j} + \dots + n_{l,j}$; $N = n_1 + \dots + n_l$.

Значення χ^2 обчислюється по формулі:

$$\chi^2 = \sum_{i,j} \frac{\left(n_{ij} - \frac{n_i n^{(j)}}{N} \right)^2}{\frac{n_i n^{(j)}}{N}}$$

Тут підсумовування ведеться по всіляких парах чисел (i, j) $i=1, \dots, k, j=1, \dots, k$. Це значення χ^2 порівнюється з табличним χ_{05}^2 або χ_{01}^2 , обчисленим для $f=(k-1)(n-1)$. Для випадку $n=2$ ми маємо випадок, уже розібраний у прикладі 4.5.

4.4. Розподіл середнього арифметичного значення

Вибіркове середнє $\bar{x}$ визначається на підставі деякої вибірки, отриманої в результаті експерименту. Тому як сама вибірка, так і вибіркове середнє $\bar{x}$, є випадковими величинами. Якщо зробити кілька вибірок, тобто кілька незалежних один від одного експериментів, то їх вибіркові середні $\bar{x}_1, \dots, \bar{x}_k$ можна розглядати як вибірку з деякої генеральної сукупності. Закон розподілу значень деякої лінгвістичної ознаки може бути відмінний від нормального, але вибіркові середні мають розподіл, близький до нормального. Це твердження стає практично достовірним, якщо об'єми вибірок, для яких визначалися $\bar{x}$, більші 30.

Як було відзначено вище, з великою ймовірністю (більшою 0.99) значення нормально розподіленої випадкової величини віддалені від її генерального середнього на відстань, меншу 3σ . Тому, визначивши вибіркове середнє $\bar{x}$, можна з імовірністю 0.99 указати інтервал (його серединою буде $\bar{x}$), у якому втримується генеральне середнє $\hat{x}$. Але для цього треба знати середнє квадратичне σ нормального розподілу, якому підкоряється $\bar{x}$.

Як оцінку для σ беруть величину $\frac{S}{\sqrt{n}}$, де S – середнє квад-

ратичне відхилення для вибірки, по якій знайдено $\bar{x}$, а n – об'єм цієї вибірки. Таким чином, з імовірністю 0.99 справжнє середнє $\hat{x}$ перебуває в інтервалі $(\bar{x} - 3 \frac{S}{\sqrt{n}}, \bar{x} + 3 \frac{S}{\sqrt{n}})$. Якщо можна за-

довольнятися меншою ймовірністю, то границі інтервалу, що містить $\hat{x}$, можна звузити. У загальному випадку ці границі мають вигляд $(\bar{x} - d_\beta \frac{S}{\sqrt{n}}, \bar{x} + d_\beta \frac{S}{\sqrt{n}})$, де d_β вибирається з таблиці 4.2 на підставі заданого значення ймовірності.

Приклад 4.6. У прикладі 4.2 було знайдено $\bar{x}=220$, $S=55$, причому $n=699$. Тому $\frac{S}{\sqrt{n}} = \frac{55}{\sqrt{699}} = 2.1$. Це значить, що з імовірністю, не меншою 0.99, справжнє середнє значення тривалості перебуває в межах 220 ± 6.3 , тобто між 213.7 і 226.3.

Іноді при порівнянні двох вибірок на предмет їхньої репрезентативності використовується параметр E , що обчислює по формулі

$$E = \frac{3 \cdot S}{\bar{x} \sqrt{n}} \cdot 100\%.$$

Уважається, що, якщо дві вибірки взяті з однієї й тієї ж генеральної сукупності (тобто отримані в результаті вивчення того самого явища, що протікає в однорідних умовах), то вибірка, для якої значення E менше, є більш представницькою. Інакше кажучи, у такій вибірці краще проявляється ознака досліджуваного явища.

Вибірка вважається *представницькою* серед подібних їй вибірок однієї й тієї ж генеральної сукупності, якщо її значення E менше 15%. У протидежному випадку до оцінок параметрів розподілу (μ і σ), що даються цією вибіркою, варто ставитися з обережністю.

Приклад 4.7. Розглянемо вибірки, вивчені в прикладах 4.1 і 4.2. Позначимо їхні середні значення відповідно через $\bar{x}_1$ і $\bar{x}_2$, а

середні квадратичні відхилення відповідно через S_1 і S_2 . Для вибірки із прикладу 4.1 $\bar{x}_1=162$, $S_1=29.3$, звідки випливає, що

$$\frac{S_1}{\sqrt{n_1}} = \frac{29.3}{\sqrt{10}} \approx 9.1,$$

$$(\bar{x}_1 \pm 3 \cdot S_1) = 162 \pm 29.3,$$

$$E_1 = \frac{3 \cdot 29.3}{162 \sqrt{10}} \cdot 100\% = 17\%.$$

Для вибірки із прикладу 4.2 $\bar{x}_2=220$, $S_2=55$. Отже,

$$\frac{S_2}{\sqrt{n_2}} = \frac{55}{\sqrt{699}} = 2.08,$$

$$(\bar{x}_2 \pm 3 \cdot S_2) = 220 \pm 165,$$

$$E_2 = \frac{3 \cdot 55}{220 \sqrt{699}} \cdot 100\% = 2.8\%.$$

Так як $E_1=17\%>15\%$, то відповідно до прийнятої границі обчислене $\bar{x}_1$ не можна прийняти, а відповідну вибірку вважати репрезентативною. У той же час, значення $\bar{x}_2$ цілком відповідає обраному критерію. Таким чином, можна прийняти, що середнє арифметичне значення тривалості голосних під наголосом в англійській мові укладене в межах 214 – 226 мсек. Границі визначені з імовірністю 0.99.

Погодженість розподілу $\bar{x}$ з нормальним законом дозволяє при будь-якому розподілі варіант із певною ймовірністю визначити границі для $\bar{x}$. А це означає, що кожна репрезентативна вибірка (тобто при $E < 15\%$) може мати надійну кількісну характеристику у вигляді $\bar{x} \pm 3S$. Дана характеристика добре відбиває якісну сторону досліджуваного явища, а це може бути використане при описі й порівнянні ознак різної природи. Крім того, оцінка $\bar{x}$ застосовується при визначенні помилки виміру. Припустимо, необхідно зняти з інтенограми значення частоти основного тону. Природно, що при вимірі частоти основного тону цілком може бути припущена якась помилка, величина якої визначається границями довірчого інтервалу X .

Приклад 4.8. Була обмірювана частота основного тону мелодійної кривої інтенोगрами фрази. Для визначення помилки виміру було взято десять однослівних фраз і зроблено по двадцять вимірів максимуму частоти основного тону голосного з головним наголосом кожного із цих слів. Були визначені середні арифметичні значення частоти основного тону по двадцяти вимірам. Таке усереднення робить закон розподілу середніх значень частоти основного тону розглянутих голосних близьким до нормального, що дозволяє користуватися правилом трьох сигм для визначеного інтервалу, у якому лежить справжнє середнє значення частоти основного тону (усереднення для основних акустичних параметрів рекомендується проводити по вибірках об'єму не менш 10). Результати обчислень представлені в таблиці 4.6.

Таблиця 4.6

№ досвіду	Слово	Голосний	Середнє значення частоти основного тону (по 20 вимірам)	Дисперсія в досвідах
1	2	3	4	5
1	'abstract	æ	300	1.29
2	'accent	»	272	6.18
3	'affix	»	271	1.32
4	'conflict	)	274	4.11
5	'contest	»	275	1.30
6	'contract	»	292	6.35
7	'decrease	[i:]	271	3.89
8	'detail	»	275	1.44
9	'discus	[I]	272	5.84
10	'increase	»	270	1.27
			2772	33.01

Суму елементів стовпця № 3 ділимо на 10 (на кількість випробувань) і одержуємо середнє значення частоти основного тону голосних під наголосом однослівних фраз. Отже, $\bar{x}=277.2$. Із суми елементів стовпця № 4 витягаємо квадратний корінь і ділимо на 10, що дає середнє квадратичне відхилення. У нашому випадку $S=0.57$. Тепер довірчий інтервал, побудований за пра-

вилом трьох сигм, має вигляд $277.2 \pm 3 \cdot 0.57 = 277.2 \pm 1.71$. Відповідне значення E дорівнює

$$E = \frac{3 \cdot 0.57}{277.2 \sqrt{10}} \cdot 100\% = 0.2\%.$$

Таким чином, наша вибірка є репрезентативною, а виходить, зазначений довірчий інтервал цілком придатний для вживання.

4.5. Контрольні запитання

1. Що таке випадкова вибірка?
2. Як обчислюється вибіркове середнє?
3. Як обчислюється вибіркова дисперсія?
4. В який спосіб проводиться угруповання вибіркових значень?
5. Що таке частота інтервалу?
6. Як і коли вноситься поправка Шеппарда?
7. Сформулюйте правило трьох сигм.
8. Що показує критерій згоди Пірсона?
9. Яка величина грає роль параметра в критерії Пірсона?
10. Як висувається нульова гіпотеза?
11. Як обчислюється число ступенів волі?
12. Що показує критерій однорідності?
13. Що таке репрезентативність вибірки?
14. Який закон розподілу має середнє арифметичне значення?
15. Як обираються границі довірчого інтервалу?
16. За яким критерієм досліджується репрезентативність вибірки?

4.6. Задачі для самостійного опрацювання

1. Дана випадкова вибірка

№ п/п	1	2	3	4	5	6	7	8	9	10
x_i	21	38	43	15	20	10	5	40	52	44

а) Обчислити середнє арифметичне значення, дисперсію та середньоквадратичне відхилення

Розв'язання:

Для розв'язання цієї задачі треба заповнити наступну таблицю.

№ п/п	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	2	3	4
1	21	$21 - 28,8 = -7,8$	$(-7,8)^2 = 60,84$
2	38	$38 - 28,8 = 9,2$	$9,2^2 = 84,64$
3	43	$43 - 28,8 = 14,2$	$14,2^2 = 201,64$
4	15	$15 - 28,8 = -13,8$	$(-13,8)^2 = 190,44$
5	20	$20 - 28,8 = -8,8$	$(-8,8)^2 = 77,44$
6	10	$10 - 28,8 = -18,8$	$(-18,8)^2 = 353,44$
7	5	$5 - 28,8 = -23,8$	$(-23,8)^2 = 566,44$
8	40	$40 - 28,8 = 11,2$	$11,2^2 = 125,44$
9	52	$52 - 28,8 = 23,2$	$23,2^2 = 538,24$
10	44	$44 - 28,8 = 15,2$	$15,2^2 = 231,04$
Σ	228		2429,6

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{228}{10} = 22,8$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{10-1} \cdot 2429,6 = \frac{2429,6}{9} = 269,9(5) = 269,96$$

$$S = \sqrt{S^2} = \sqrt{269,96} = 16,43.$$

Для обчислення $\bar{x}$ використовується сума елементів стовпця № 2, а для обчислення дисперсії – сума елементів стовпця № 4. Значення елементів стовпця № 3 є проміжними. Вони підносяться до другого ступеня (до квадрату), щоб отримати елементи стовпця № 4. Тому суму елементів стовпця № 3 обчислювати не треба. Округлення проводити до сотих.

Відповідь: $\bar{x}=28,8$; $S^2=269,96$; $S=16,43$.

б) Побудувати довірчий інтервал

Розв'язання:

Згідно правила «трьох сігм», будь яке значення випадкової величини, що розподілена за нормальним законом, з ймовірністю більшою за 99% знаходиться по обидві сторони від середнього арифметичного на відстані, що не перевищує 3σ . При розрахунках параметр нормального розподілу σ апроксимується значенням середньоквадратичного відхилення S . Тому:

$$\hat{x} \in \left(\bar{x} - 3 \frac{S}{\sqrt{n}}, \bar{x} + 3 \frac{S}{\sqrt{n}} \right) = \left(28,8 - 3 \cdot \frac{16,43}{\sqrt{10}}; 28,8 + 3 \cdot \frac{16,43}{\sqrt{10}} \right) = (28,8 - 15,59; 28,8 + 15,59) = (13,21; 44,39).$$

Відповідь: $\hat{x} \in (13,21; 44,39)$.

в) Обчислити коефіцієнт репрезентативності вибірки та зробити висновок

Розв'язання:

$$E = \frac{3 \cdot S}{\bar{x} \sqrt{n}} \cdot 100\% = \frac{3 \cdot 16,43}{28,8 \cdot \sqrt{10}} \cdot 100\% = 0,5413 \cdot 100\% = 54,13\% > 15\%.$$

Висновок: вибірка не є репрезентативною.

2. Дана випадкова вибірка

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
x_i	58	5	50	11	15	28	31	40	10	19	34	65	53	61	45	39	51	27	31	18

а) Провести групування для $k=5$. Вказати шаг групування, отримані інтервали та їх частоти.

Розв'язання:

№ п/п	Інтер-вали	n_i	$c_i = N_i - N_0$	$n_i c_i$	$n_i c_i^2$
	1	2	3	4	5
1	5 - 17	4	$1 - 3 = -2$	$4 \cdot (-2) = -8$	$4 \cdot (-2)^2 = (-2) \cdot (-8) = 16$
2	17 - 29	4	$2 - 3 = -1$	$4 \cdot (-1) = -4$	$4 \cdot (-1)^2 = (-1) \cdot (-4) = 4$
3	29 - 41	5	$3 - 3 = 0$	$5 \cdot 0 = 0$	$5 \cdot 0^2 = 0 \cdot 0 = 0$
4	41 - 53	3	$4 - 3 = 1$	$3 \cdot 1 = 3$	$3 \cdot 1^2 = 1 \cdot 3 = 3$
5	53 - 65	4	$5 - 3 = 2$	$4 \cdot 2 = 8$	$4 \cdot 2^2 = 2 \cdot 8 = 16$
	Σ			- 1	39

Щоб побудувати інтервали, треба обчислити їх довжину. Для цього треба виявити найбільше та найменше значення елементів вибірки. Найменшим є $x_2=5$, а найбільшим $x_{12}=65$. Тому довжина інтервалів дорівнює

$$h = \frac{x_{\max} - x_{\min}}{k} = \frac{65 - 5}{5} = \frac{60}{5} = 12.$$

Починаючи відлік від найменшого значення, отримуємо 5 інтервалів довжиною 12 кожний. В нашій задачі довжина інтервалів є цілим числом. Округлення не було потреби проводити. Тому останній інтервал закінчується рівно в найбільшому значенні. Для заповнення стовпця №2 треба підрахувати, скільки значень вибірки потрапило до кожного інтервалу. Наприклад, до 1-го інтервалу (5 – 17) потрапили значення $x_2=5$, $x_4=11$, $x_5=15$, $x_9=10$. Усього їх 4. Тому частота 1-го інтервалу дорівнює $n_1=4$. До 2-го інтервалу (17 – 29) потрапили значення $x_6=28$, $x_{10}=19$, $x_{18}=27$, $x_{20}=18$. Їх також 4. Тому частота 2-го інтервалу також дорівнює $n_2=4$. До 3-го інтервалу (29 - 41) потрапили 5 значень: $x_7=31$, $x_8=40$, $x_{11}=34$, $x_{16}=39$, $x_{19}=31$. Тому $n_3=5$. До 4-го інтервалу

(41 – 53) потрапили 3 значення: $x_3=50$, $x_{15}=45$, $x_{17}=51$. Слід звернути увагу, що $x_{13}=53$ (права границя 4-го інтервалу) потрапляє до наступного, 5-го, інтервалу, який з цього числа починається. Тому $n_4=3$. До останнього 5-го інтервалу (53 – 65) потрапило 4 значення: $x_1=58$, $x_{12}=65$, $x_{13}=53$, $x_{14}=61$. Слід зазначити, що $x_{12}=65$, що співпадає з правою границею цього інтервалу, належить саме до цього інтервалу. Останній інтервал – єдиний інтервал, що містить свою праву границю. Тому $n_5=4$.

Після того, як обчислені частоти усіх інтервалів, треба підрахувати їх суму, незважаючи на те, що об'єм вибірки відомий. Це треба робити для самоконтролю. Якщо сума усіх частот співпадає з об'ємом вибірки, то частоти обчислені, скоріш за все, правильно. Якщо не співпадає – треба шукати помилку. Найрозповсюдженішою помилкою при обчисленні частот є саме хибне віднесення границі інтервалу до лівого інтервалу, або не віднесення до жодного, або віднесення і до лівого, і до правого інтервалів. Якщо припуститися помилки на етапі проведення групування, то усі подальші обчислення будуть неправильними.

б) Обчислити середнє арифметичне значення, дисперсію та середньоквадратичне відхилення

Розв'язання:

Для виконання цього завдання треба продовжити заповнювати таблицю. При розв'язанні попередньої задачі були заповнені стовпці №1 і №2. Для обчислення елементів стовпця №3 треба визначити найчастотніший інтервал. Ним є інтервал №3. Його частота $n_3=5$ є найбільшою серед отриманих інтервалів. Тому фіксується номер і середина саме цього інтервалу:

$$N_0 = 3; \quad x_0 = 35.$$

Для обчислення елементів стовпця №3 із номера поточного інтервалу віднімається число 3 (номер зафіксованого інтервалу). Елементи стовпця №4 є добутком відповідних елементів стовпця №2 і стовпця №3. Для отримання елементів стовпця №5 треба підвести до другого ступеня елементи стовпця №3 і помножити цей результат на елемент стовпця №2. Але можна зробити простіше: помножити елементи стовпця №3 на відповідні еле-

менти стовпця № 4. Тобто для отримання елементів стовпців № 4 і № 5 треба помножити відповідні елементи попередніх двох стовпців (стовпців № 2 і № 3 для отримання стовпця № 4, а також стовпців № 3 і № 4 для отримання стовпця № 5). Для подальших підрахунків нам будуть потрібні суми по стовпцям № 4 і № 5. Таким чином, маємо:

$$\begin{aligned}\bar{x} &= x_0 + \frac{1}{n} \left(\sum_{i=1}^k n_i c_i \right) \cdot h = 35 + \frac{(-1)}{20} \cdot 12 = 35 - \frac{12}{20} = 34,4; \\ S^2 &= \left[\sum_{i=1}^k n_i c_i^2 - \frac{1}{n} \left(\sum_{i=1}^k n_i c_i \right)^2 \right] \cdot h^2 \cdot \frac{1}{n-1} = \left(39 - \frac{(-1)^2}{20} \right) \cdot \frac{12^2}{19} = \\ &= 295,2; \\ S &= \sqrt{S^2} = \sqrt{295,2} = 17,18.\end{aligned}$$

Відповідь: $\bar{x}=34,4$; $S^2=295,2$; $S=17,18$.

в) Побудувати довірчий інтервал

Розв'язання:

Як і при розв'язанні задачі 1б), для побудови довірчого інтервалу скористаємось правилом «трьох сігм»:

$$\begin{aligned}\hat{x} \in \left(\bar{x} - 3 \frac{S}{\sqrt{n}}, \bar{x} + 3 \frac{S}{\sqrt{n}} \right) &= \left(34,4 - 3 \cdot \frac{17,18}{\sqrt{20}}; 34,4 + 3 \cdot \frac{17,18}{\sqrt{20}} \right) = \\ &= (34,4 - 11,52; 34,4 + 11,52) = (22,88; 45,92).\end{aligned}$$

Відповідь: $\hat{x} \in (22,88; 45,92)$.

г) Обчислити коефіцієнт репрезентативності вибірки

Розв'язання:

$$E = \frac{3 \cdot S}{x \sqrt{n}} \cdot 100\% = \frac{3 \cdot 17,18}{34,4 \cdot \sqrt{20}} \cdot 100\% = 0,3349 \cdot 100\% = 33,49\% >$$

15%.

Висновок: вибірка не є репрезентативною.

3. Порівняти дві наведені вибірки щодо їх репрезентативності

Розв'язання:

З двох вибірок більш репрезентативною вважається та, у якій коефіцієнт репрезентативності E менший. Для вибірки із задачі 1) $E=54,13\%$. Для вибірки із задачі 2) $E=33,49\%$. Для 2-ї вибірки значення цього коефіцієнту є кращим.

Відповідь: 2-га вибірка є більш репрезентативною, ніж 1-ша.

4. В умовах задачі 2) перевірити гіпотезу щодо нормального закону розподілу випадкової величини.

Розв'язання:

Для вибірки із задачі 2) були отримані наступні результати: $\bar{x}=34,4$; $S=17,18$. З урахуванням цього, можна заповнити наступну таблицю. Елементами стовпця №2 є усі границі інтервалів. Слід звернути увагу, що границь завжди на одну більше, ніж інтервалів. Тобто це усі ліві границі та одна права границя останнього інтервалу, яка вже не є лівою границею для жодного іншого інтервалу.

Для отримання елементів стовпця №3 із усіх елементів стовпця №2 віднімаємо одне й те ж саме число: $\bar{x}=34,4$. Для отримання елементів стовпця №4 усі елементи стовпця №3 ділимо на одне й те ж саме число: $S=17,18$. Елементи стовпця №5 знаходимо за таблицею Додатка 1. Для обчислення елементів стовпця №6 із елемента наступного рядка стовпця №5 віднімаємо елемент поточного рядка стовпця №5 і записуємо результат у поточний рядок стовпця №6. Для отримання елементів стовпця №7 усі елементи стовпця №6 множимо на одне й те ж саме число: об'єм вибірки $n=20$. Елементи стовпця №8 переписуємо із стовпця №2 попередньої таблиці (задача 2). Для обчислення елементів стовпця №9 різницю відповідних елементів стовпців №7 і №8 підводимо до другого ступеня і результат ділимо на відповідний елемент стовпця №7. Сума елементів стовпця №9 і є потрібна нам величина критерію Пірсона.

Середини інтервалів	Границі інтервалів	Відхилення границь інтервалів від $\bar{x}$	Нормовані відхилення границь інтервалів u_i	$\Phi(u_i)$	$p_i = \Phi(u_i) - \Phi(u_{i+1}) $	$\tilde{n}_i = np_i$	n_i	χ^2
1	5	5 – 34,4 = -29,4	(-29,4)/17,18 = -1,71	0,0436	0,1562 – 0,0436 = 0,1126	0,1126·20 = 2,25	4	$\frac{(2,25-4)^2}{2,25} = 1,36$
2	17	17 – 34,4 = -17,4	(-17,4)/17,18 = -1,01	0,1562	0,3783 – 0,1562 = 0,2221	0,2221·20 = 4,44	4	$\frac{(4,44-4)^2}{4,44} = 0,04$
3	29	29 – 34,4 = -5,4	(-5,4)/17,18 = -0,31	0,3783	0,6480 – 0,3783 = 0,2697	0,2697·20 = 5,39	5	$\frac{(5,39-5)^2}{5,39} = 0,03$
4	41	41 – 34,4 = 6,6	6,6/17,18 = 0,38	0,6480	0,8599 – 0,6480 = 0,2119	0,2119·20 = 4,24	3	$\frac{(4,24-3)^2}{4,24} = 0,36$
5	53	53 – 34,4 = 18,6	18,6/17,18 = 1,08	0,8599	0,9625 – 0,8599 = 0,1026	0,1026·20 = 2,05	4	$\frac{(2,05-4)^2}{2,05} = 1,85$
Σ		65 – 34,4 = 30,6	30,6/17,18 = 1,78	0,9625				3,64

$$\chi^2=3,64.$$

Обчислюємо тепер число ступенів свободи для нашої вибірки. Взагалі при обчисленні критерію Пірсона приймається $f=k-r-1$. Якщо перевіряється відповідність вибірки випадковій величині з нормальним законом розподілу, то $r=2$, тому що на основі вибірки оцінюється 2 параметри розподілу: середнє і дисперсія. Тому для нормальної випадкової величини при обчисленні критерію Пірсона завжди $f=k-3$. При проведенні групування для досліджуваної вибірки було обрано 5 інтервалів. Тому для даної задачі

$$f=5-3=2.$$

Далі за таблицею Додатка 2 для $f=2$ знаходимо $\chi_{05}^2=6,0$ і $\chi_{01}^2=9,21$. Їх слід порівняти із обчисленим за допомогою таблиці $\chi^2=3,64$. Очевидно, що отриманий для розглянутої вибірки $\chi^2 < \chi_{05}^2$. Це означає, що ймовірність спостерігати дану вибірку в умовах нормального закону розподілу складає 95%. Тобто з ймовірністю 95% випадкова величина, з якої отримана дана вибірка, розподілена за нормальним законом.

Висновок: гіпотеза щодо нормального закону розподілу досліджуваної випадкової величини підтверджується.

5. Перевірити гіпотезу щодо однорідності матеріалу, що міститься у двох вибірках.

Інтервали	1 – 1.5	1.5 – 2	2 – 2.5	2.5 – 3	3 – 3.5
Частоти для вибірки I (n_{1i})	1	4	15	10	8
Частоти для вибірки II (n_{2i})	2	2	8	7	2

Розв'язання:

Для розв'язання даної задачі треба заповнити наступну таблицю. Стовпці № 1 - № 3 є умовою задачі.

Інтервали	Частоти для вибірки I (n_{1i})	Частоти для вибірки II (n_{2i})	$n^{(i)}$	$\frac{n_1 n^{(i)}}{N}$	$\frac{n_2 n^{(i)}}{N}$	χ_{1i}^2	χ_{2i}^2
1 – 1.5	1	2	1+2=3	$\frac{3 \cdot 38}{59} = 1,93$	$\frac{3 \cdot 21}{59} = 1,07$	$\frac{(1,93-1)^2}{1,93} = 0,45$	$\frac{(1,07-2)^2}{1,07} = 0,81$
1.5 – 2	4	2	4+2=6	$\frac{6 \cdot 38}{59} = 3,86$	$\frac{6 \cdot 21}{59} = 2,14$	$\frac{(3,86-4)^2}{3,86} = 0,01$	$\frac{(2,14-2)^2}{2,14} = 0,01$
2 – 2.5	15	8	15+8=23	$\frac{23 \cdot 38}{59} = 14,81$	$\frac{23 \cdot 21}{59} = 8,19$	$\frac{(14,81-15)^2}{14,81} = 0,002$	$\frac{(8,19-8)^2}{8,19} = 0,004$
2.5 – 3	10	7	10+7=17	$\frac{17 \cdot 38}{59} = 10,95$	$\frac{17 \cdot 21}{59} = 6,05$	$\frac{(10,95-10)^2}{10,95} = 0,08$	$\frac{(6,05-7)^2}{6,05} = 0,15$
3 – 3.5	8	2	8+2=10	$\frac{10 \cdot 38}{59} = 6,44$	$\frac{10 \cdot 21}{59} = 3,56$	$\frac{(6,44-8)^2}{6,44} = 0,38$	$\frac{(3,56-2)^2}{3,56} = 0,68$
Σ	$n_1=38$	$n_2=21$	$N=59$			$\Sigma = \chi_1^2 = 0,92$	$\Sigma = \chi_2^2 = 1,65$

Елементи стовпця № 4 є сумою відповідних елементів стовпців №2 і №3. Числа n_1 і n_2 є сумами елементів стовпців №2 і №3 відповідно. Число N є сумою елементів стовпця №4 і одночасно сумою чисел n_1+n_2 . Для самоперевірки рекомендується обчислювати число N обома способами. Якщо результати відрізняються, то десь на цьому етапі підрахунків було припущено помилки. Якщо результати співпадають, то, скоріш за все, елементи стовпця №4 і числа n_1 і n_2 обчислені правильно. В цьому випадку можна продовжити процес розв'язання задачі. Щоб отримати елементи стовпця №5, треба відповідний елемент стовпця №4 помножити на число n_1 і розділити на число N . Елементи стовпця №6 розраховуються аналогічно, але замість числа n_1 в обчисленнях бере участь число n_2 . Щоб отримати елементи стовпця №7, треба різницю відповідних елементів стовпців №5 і №2 підвести до другого ступеня і результат розділити на елемент стовпця №5. Щоб отримати елементи стовпця №8, треба різницю відповідних елементів стовпців №6 і №3 підвести до

другого ступеня і результат розділити на елемент стовпця №6. Суми по стовпцях №7 і №8 є значеннями критерію χ^2 для 1-ї й 2-ї вибірок відповідно. Таким чином, стовпці №2, №5 і №7 відповідають 1-й вибірці, а стовпці №3, №6 і №8 – 2-й вибірці. В нашій задачі є тільки дві вибірки. Тому у нас є тільки 2 групи стовпців. Якщо б за умовою задачі досліджувалися на однорідність, наприклад, три вибірки, то була б третя група стовпців, яка б обчислювалась аналогічним чином. Для прийняття рішення щодо висунутої гіпотези про однорідність матеріалу у всіх вибірках нам потрібний сумарний критерій по всіх вибірках. У нашій задачі досліджуються дві вибірки. Тому для обчислення цього критерію були використані лише два доданки:

$$\chi^2 = \chi_1^2 + \chi_2^2 = 0,92 + 1,65 = 2,57.$$

Далі також треба обчислити число ступенів свободи. Для перевірки гіпотези щодо однорідності матеріалу він дорівнює $f=(k-1)(m-1)$. В нашій задачі для двох вибірок ($m=2$) було використано п'ять інтервалів ($k=5$). Тому

$$f=(5-1)(2-1)=4.$$

Для цього значення $f=4$ по таблиці Додатку 2 знаходимо $\chi_{05}^2=9,5$ і $\chi_{01}^2=13,3$. Порівнюючи обчислене в ході розв'язання значення критерію з табличним, бачимо, що $2,57 < 9,5$. Тобто $\chi^2 < \chi_{05}^2$. Це означає, що ознака, за якою матеріал був поділений на декілька вибірок, не є суттєвою для значення досліджуваної випадкової величини і матеріал можна об'єднати в одну вибірку і в подальшому досліджувати як одне ціле.

Висновок: гіпотеза щодо однорідності матеріалу приймається. Вибірki можна об'єднати в одну.

4.7. Контрольні завдання

1. Дана випадкова вибірка

№ п/п	1	2	3	4	5	6	7	8	9	10
x_i	19	22	13	18	33	15	25	40	15	43

а) обчислити середнє арифметичне значення, дисперсію та середньоквадратичне відхилення;

б) побудувати довірчий інтервал;

в) обчислити коефіцієнт репрезентативності вибірки та зробити висновок.

2. Дана випадкова вибірка

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
x_i	41	19	5	23	9	11	31	59	47	16	24	13	15	33	12	10	46	50	7	65

а) Провести групування для $k=5$. Вказати шаг групування, отримані інтервали та їх частоти.

б) Обчислити середнє арифметичне значення, дисперсію та середньоквадратичне відхилення.

в) Побудувати довірчий інтервал.

г) Обчислити коефіцієнт репрезентативності вибірки.

г) Перевірити гіпотезу щодо нормального закону розподілу випадкової величини.

3. Дана випадкова вибірка

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
x_i	8	5	30	19	39	26	24	15	32	23	11	17	38	27	19	31	40	6	12	28

а) Провести групування для $k=5$. Вказати шаг групування, отримані інтервали та їх частоти.

б) Обчислити середнє арифметичне значення, дисперсію та середньоквадратичне відхилення.

в) Побудувати довірчий інтервал.

г) Обчислити коефіцієнт репрезентативності вибірки.

г) Перевірити гіпотезу щодо нормального закону розподілу випадкової величини.

4. Дана випадкова вибірка

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
x_i	6	13	23	5	18	48	31	39	26	14	49	55	30	52	10	17	43	24	19	9

а) Провести групування для $k=5$. Вказати шаг групування, отримані інтервали та їх частоти.

б) Обчислити середнє арифметичне значення, дисперсію та середньоквадратичне відхилення.

в) Побудувати довірчий інтервал.

г) Обчислити коефіцієнт репрезентативності вибірки.

г) Перевірити гіпотезу щодо нормального закону розподілу випадкової величини.

5. Дана випадкова вибірка

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
x_i	9	13	23	5	18	44	38	40	21	36	8	19	22	35	41	45	16	27	29	43

а) Провести групування для $k=5$. Вказати шаг групування, отримані інтервали та їх частоти.

б) Обчислити середнє арифметичне значення, дисперсію та середньоквадратичне відхилення.

в) Побудувати довірчий інтервал.

г) Обчислити коефіцієнт репрезентативності вибірки.

г) Перевірити гіпотезу щодо нормального закону розподілу випадкової величини.

6. Порівняти усі наведені вибірки щодо їх репрезентативності. Вказати найбільш і найменш репрезентативні вибірки серед досліджуваних.

7. Перевірити гіпотезу щодо однорідності матеріалу, що міститься у двох вибірках.

Інтервали	1,5 – 2	2 – 2,5	2,5 – 3	3 – 3,5	3,5 – 4
Частоти для вибірки I (n_{1i})	3	6	7	5	1
Частоти для вибірки II (n_{2i})	3	7	12	10	4

8. Перевірити гіпотезу щодо однорідності матеріалу, що міститься у трьох вибірках.

Інтервали	1,5 – 2,5	2,5 – 3,5	3,5 – 4,5	4,5 – 5,5	5,5 – 6,5
Частоти для вибірки I (n_{1i})	3	5	9	8	2
Частоти для вибірки II (n_{2i})	1	2	6	4	3
Частоти для вибірки III (n_{3i})	4	4	10	9	5

5. ЛІНГВІСТИЧНІ ГІПОТЕЗИ

5.1. Постановка й перевірка лінгвістичних гіпотез

Постановка лінгвістичної задачі визначається метою дослідження. Залежно від задачі планується й проводиться експеримент. Чітко поставлена лінгвістична задача є необхідною умовою ефективності статистичного аналізу.

Припустимо, досліджуються інтонаційні одиниці певної мови. Фрази, що представляють ту або іншу інтонаційну одиницю, відбираються дослідником у результаті лінгвістичного аналізу із залученням носіїв даної мови. Над отриманими групами фраз проводиться електроакустичний аналіз, у результаті якого кожна фраза подається фізичними характеристиками - частотою основного тону, інтенсивністю й тривалістю. Далі дослідник, спираючись на вже наявні результати експерименту, визначає окремі ознаки, гіпотетично істотні для даної інтонаційної одиниці. У підсумку будуть отримані інтонаційні ознаки, подані певним рядом значень тієї або іншої фізичної характеристики. До цих вихідних фізичних значень і застосовується ймовірно-статистичний аналіз.

Звичайно виходять із припущення, що ряд значень фізичної величини являє собою розподіл деякої випадкової величини. З досліджень по лінгвістиці відомо, що мовним явищам властиво статичне коливання, потенційність. Однак границі варіювання даних характеристик не довільні, існують «області розсіювання» кожної лінгвістичної одиниці. Виходячи із властивості потенційності мовної одиниці й існування певних границь варіювання її, можна зробити висновок про ймовірно-статистичний характер прояву даної одиниці. Розподіл значень фізичних характеристик розпізнавальних ознак двох лінгвістичних одиниць, незважаючи на деяке перетинання їхніх інтервалів, має досить чітке протиставлення, що виражається в статистичних параметрах. Це підтверджується даними багатьох експериментально-лінгвістичних досліджень. Отже, кожна лінгвістична одиниця характеризується певним імовірнісним розподілом (або розподілами), представленими у своїх фізичних характеристиках.

На початковому етапі статистичного аналізу йде *опис* у статистичних термінах того різноманіття значень, у яких представлений кожний з гіпотетично істотних лінгвістичних ознак. Без такої обробки фізичних даних не можна виділити із усього різноманіття величин, що змінюються безупинно, самі характерні прояви, типові для досліджуваного явища. Цей етап припускає обчислення середнього арифметичного значення й дисперсії, трьохсігмених границь для середньоарифметичного, побудову розподілу (гістограми) варіант тощо.

Наступний етап застосування статистичного аналізу є саме перевіркою певних лінгвістичних гіпотез. Це може бути вирішення питання істотності тієї або іншої ознаки для лінгвістичної одиниці в результаті застосування критеріїв розходження (критерію Стьюдента, Ван дер Вардена).

Перебираючи в такий спосіб ознаки, установлюється набір розпізнавальних ознак для кожної лінгвістичної одиниці (тут виходять із припущення, що розходження на лінгвістичному рівні обов'язково знаходить своє відбиття у фізичних характеристиках). Ймовірно-статистичний аналіз дозволяє також розглянути взаємодію різних ознак усередині лінгвістичної одиниці. При цьому рекомендується застосовувати кореляційний аналіз або аналіз кореляційних відносин (якщо розподіл варіант відрізняється від нормального). Визначити ступінь залежності тих або інших ознак може допомогти регресійний аналіз. Якщо лінгвіст припускає дію на розподіл варіант якогось другорядного фактору, то перевірити цю гіпотезу можна за допомогою дисперсійного аналізу [5].

5.2. Критерій Стьюдента

При вивченні показників лінгвістичних характеристик велике значення має наступна постановка питання: деякий лінгвістичний показник визначений по двох різних вибірках. Як правило, середнє значення цього показника в одній вибірці відрізняється від середнього значення його в іншій вибірці. Чи істотно це розходження?

У тих випадках, коли закон розподілу цього показника близький до нормального, відповідь на поставлене питання дає критерій Стюдента. Вимогу нормальності можна опустити, якщо об'єми порівнюваних вибірок досить великі (більше 30). Сутність критерію Стюдента проілюструємо на наступному прикладі.

Приклад 5.1. Для двох груп дикторів (група **A** – чоловіки, група **B** – жінки) зафіксовані значення сумарної енергії голосного у двоскладових словах англійської мови з наголосом на початковому складі. Результати були піддані інтервальному угрупованню й були побудовані дві таблиці.

Таблиця 5.1

Група А

Інтервали	Частоти (n_i)	x_i	$n_i x_i$	$n_i x_i^2$
180-210	2	-4	-8	32
210-240	4	-3	-12	36
240-270	5	-2	-10	20
270-300	6	-1	-6	6
300-330	8	0	0	0
330-360	6	1	6	6
360-390	4	2	8	16
390-420	3	3	9	27
420-450	2	4	8	32
	40		-7	175

Група В

Інтервали	Частоти (n_j)	y_j	$n_j y_j$	$n_j y_j^2$
120-140	1	-5	-5	25
140-160	1	-4	-4	16
160-180	3	-3	-9	27
180-200	3	-2	-6	12
200-220	4	-1	-4	4
220-240	5	0	0	0
240-260	5	1	5	5
260-280	3	2	6	12
280-300	3	3	9	27
300-320	2	4	8	32
320-340	2	5	10	50
	32		+10	210

З таблиці маємо:

$$\bar{a} = 315 + \frac{(-7)}{40} \cdot 30 = 310; \quad \bar{b} = 230 + \frac{10}{32} \cdot 20 = 236,$$

де $\bar{a}$ - середнє значення сумарної енергії для групи **A**, а $\bar{b}$ - середнє значення для групи **B**.

Розходження вибірових середніх (310 і 236) ще не доводить, що групи **A** і **B** істотно різні з точки зору впливу на значення сумарної енергії. Висувається гіпотеза H_0 : розходження вибірових середніх $\bar{a}$ і $\bar{b}$ є несуттєвим, випадковим, тобто справжні середні значення сумарної енергії для груп **A** і **B** однакові. Для перевірки цієї гіпотези обчислюємо $S_{\bar{a}-\bar{b}}$ по формулі:

$$S_{\bar{a}-\bar{b}} = \sqrt{\frac{\sum (a_i - \bar{a})^2 + \sum (b_j - \bar{b})^2}{n_A + n_B - 2} \cdot \frac{n_A + n_B}{n_A n_B}},$$

де n_1 і n_2 - об'єми відповідно вибірок у групах **A** і **B**, крім того суми $\sum (a_i - \bar{a})^2$ і $\sum (b_j - \bar{b})^2$ мають той же зміст, що й при обчисленні дисперсій вибірок відповідно для груп **A** і **B**. Без проведення групування дисперсія дорівнює

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

В той же час після проведення групування дисперсія дорівнює

$$S^2 = \left[\sum_{i=1}^k n_i c_i^2 - \frac{1}{n} \left(\sum_{i=1}^k n_i c_i \right)^2 \right] \cdot h^2 \cdot \frac{1}{n-1}.$$

Враховуючи, що це одна й та ж сама величина незалежно від того, було проведено групування чи ні, маємо:

$$\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \left[\sum_{i=1}^k n_i c_i^2 - \frac{1}{n} \left(\sum_{i=1}^k n_i c_i \right)^2 \right] \cdot h^2 \cdot \frac{1}{n-1}.$$

Помножуючи обидві частини цієї рівності на $n-1$, отримуємо:

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \left[\sum_{i=1}^k n_i c_i^2 - \frac{1}{n} \left(\sum_{i=1}^k n_i c_i \right)^2 \right] \cdot h^2.$$

Тому

$$\Sigma(a_i - \bar{a})^2 = \left(175 - \frac{(-7)^2}{40} \right) \cdot 30^2 = 156420,$$

$$\Sigma(b_j - \bar{b})^2 = \left(210 - \frac{(10)^2}{32} \right) \cdot 20^2 = 82800.$$

Тому

$$S_{\bar{a}-\bar{b}} = \sqrt{\frac{156420 + 82800}{40 + 32 - 2} \cdot \frac{40 + 32}{40 \cdot 32}} = 13.9.$$

Тепер обчислюємо

$$t = \frac{|\bar{a} - \bar{b}|}{S_{\bar{a}-\bar{b}}} = \frac{310 - 236}{13.9} = 5.33.$$

По таблиці розподілу Стьюдента (див. Додаток 3) визначаємо t_{05} для числа ступенів волі $f = n_1 + n_2 - 2$. У розглянутому прикладі $t_{05} = 2.00$. Далі необхідно порівняти обчислене значення t з табличним t_{05} . Критерій Стьюдента затверджує: якщо $t > t_{05}$, то гіпотеза H_0 відкидається. Причому ймовірність того, що вона помилкова, не менша 0.95, тобто в 95% вона не вірна. Якщо ж $t \leq t_{05}$, то гіпотеза H_0 приймається. У розглянутому прикладі $t = 5.33 > 2.00 = t_{05}$, а тому припущення про відсутність розходжень у впливі дикторів груп А і В на сумарну енергію помилкове.

У розглянутому прикладі вибірки для груп **А** і **В** були незалежні одна від одної, і для перевірки гіпотези H_0 (у загальному випадку гіпотеза H_0 формулюється так: розходження в середніх значеннях двох вибірок несуттєве) була використана величина

$$t = \frac{|\bar{a} - \bar{b}|}{S_{\bar{a}-\bar{b}}}.$$

Критерій Стьюдента в лінгвістиці часто використовується для аналізу лексики, наприклад для вирішення питання про розходження частот уживання слів [11, стор.68-69]. Критерій Стьюдента може бути використаний не тільки як показник ступеня розходження між частотами слів, але і як показник ступеня їх-

ньої близькості. З його допомогою можна виміряти відстань між словами (попарно). Він також може бути використаний і при дослідженні функціональних стилів [12]. Критерій Стюдента в сполученні із правилом трьох сигм може також використатися для дослідження середньої довжини слова й речення, лексичних, морфологічних і синтаксичних властивостей тексту [6, стор.134]. Дослідження морфологічного рівня тексту подано, зокрема, у роботі [13], у якій вивчена частота зустрічальності різних частин мови української мови в трьох функціональних стилях. Показано, що за допомогою критерію Стюдента можна встановити ступінь розходження між художнім, публіцистичним і науковим стилями. В більшості випадків середні частоти зустрічальності різних частин мови істотно відрізняються одна від одної й, таким чином, середня частота вживання частин мови може бути одним з диференціальних ознак кожного стилю.

Дуже важливим є один окремий випадок застосування критерію Стюдента. Нехай, наприклад, вивчається енергетична неповноцінність переднаголосних і післянаголосних складів трискладових слів української мови. Висловлено припущення, що в межах того самого слова розходження немає (це гіпотеза H_0). Для перевірки цієї гіпотези не слід окремо вивчати переднаголосні й післянаголосні склади, а досить розглянути пари: переднаголосний-післянаголосний склади. Знайти різницю їх енергетичних наповненостей і перевірити, чи значно середнє значення цих різниць відрізняється від нуля [5]. Такий метод перевірки гіпотези H_0 іноді називають *методом порівнянь*. Він зручний для математичної обробки (замість $2n$ спостережуваних значень обробляється тільки n чисел). Але застосовуючи метод парних порівнянь, слід дотримуватися певних умов – пари утворюються в процесі дослідження як результат дії двох постійних факторів, розходження в дії яких цікавить дослідника. Не можна довільно зіставляти пари тільки на тій підставі, що є дві вибірки однакових об'єктів.

Використання критерію Стюдента має велике значення у фонетиці при відборі однорідного матеріалу для експериментів. Так, наприклад, досліджуючи енергетичні характеристики наголошених складів на відміну від ненаголошених, можна зтикну-

тися з тим фактом, що група ненаголошених включає дві підгрупи – переднаголосних і післянаголосних. Причому позиційні умови можуть впливати на акустичні характеристики цих складів. Отримані середні значення для переднаголосних і післянаголосних складів трохи відрізняються, але чи настільки істотно це розходження, що можна говорити про дві різні сукупності, затверджувати неможливо без відповідного статистичного аналізу. Експериментатор висуває гіпотезу H_0 про неістотність розходження вибірових середніх переднаголосних і післянаголосних складів і перевіряє її, використовуючи критерій Стюдента. Підтвердження гіпотези H_0 дає можливість надалі розглядати переднаголосні й післянаголосні голосні як одну сукупність. Критерій Стюдента також часто використовується в експериментальній фонетиці, якщо за даними аудиторського аналізу немає істотних розходжень у сприйнятті окремих підгруп явища, і експериментатор хоче об'єднати їх у більші групи.

5.3. Критерій Ван дер Вардена

Критерій Ван дер Вардена, або критерій X , належить до непараметричних критеріїв, при застосуванні яких немає необхідності обчислювати статистичні параметри вибірки (середнє, дисперсію). Подібно критерію Стюдента, критерій X рекомендується застосовувати при вирішенні питання про істотність розбіжностей порівнюваних вибірок варіант, тобто про значимості тієї або іншої ознаки досліджуваної лінгвістичної одиниці. Незважаючи на схожість розв'язуваних задач, обидва критерії розрізняються умовами застосування. Застосування критерію X , як і будь-якого непараметричного критерію, не вимагає знання функції розподілу варіант. Цей критерій може застосовуватися при малій кількості варіант, тобто в тих випадках, коли не можна вирішити питання про функції розподілу вихідних даних. При великій кількості варіант ($n \geq 30$) критерій Ван дер Вардена діє аналогічно критерію Стюдента й тому в цих випадках застосовується саме критерій Стюдента.

Приклад 5.2. Досліджується вплив ритміки фрази на значення її фізичних характеристик, зокрема, на величину частот-

ного діапазону. Досліджуються фрази наступних ритмічних структур: з початковим головним наголошеним складом і з кінцевим головним наголошеним складом однієї й тієї ж інтонаційної одиниці. Значення частотного діапазону першої групи фраз позначимо як прояв випадкової величини x , другої групи – через y . У результаті експерименту (були записані німецькі емоційні фрази) отримані наступні значення частотного діапазону:

x : 1.55, 1.61, 1.73, 1.80, 1.92, 2.09, 2.33, 2.44; $n_1=8$;

y : 1.33, 1.51, 1.66, 1.70, 1.82, 1.88, 2.20; $n_2=7$.

Розглянемо сукупність чисел, що складається зі значень x і y , і розташуємо числа цієї сукупності в зростаючому порядку. Тим самим кожному розглянутому числу приписується порядковий номер. Для спрощення розрахунків припустимо, що серед всіх значень, x і y , немає однакових. У розглянутому прикладі саме така ситуація. Розташування чисел і приписування їм порядкових номерів r зручно вести за допомогою таблиці 5.2:

Таблиця 5.2

x			1.55	1.64			1.73	1.80			1.92	2.09		2.33	2.44
y	1.33	1.61			1.66	1.70			1.82	1.88			2.20		
r	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Таким чином, числам першої групи, тобто числам, що відповідають випадковій величині x , приписані номери 3, 4, 7, 8, 11, 12, 14, 15. Кількість чисел у сукупності x (відповідно y) позначене через n_1 (відповідно n_2) і покладається $n=n_1+n_2$. У розглянутому прикладі $n_1=8$, $n_2=7$, $n=15$. Складемо послідовність дробів

$\frac{r}{n+1}$, де r пробігає всі номери, приписані числам із сукупності x . У розглянутому прикладі це дробі $\frac{3}{16}$, $\frac{4}{16}$, $\frac{7}{16}$, $\frac{8}{16}$, $\frac{11}{16}$, $\frac{12}{16}$, $\frac{14}{16}$, $\frac{15}{16}$. Для кожної дробі по таблиці функції Ψ (Додаток 4) зна-

ходимо $\Psi\left(\frac{r}{n+1}\right)$, а потім складаємо суму

$X = \sum_r \Psi\left(\frac{r}{n+1}\right)$. У розглянутому прикладі

$$\begin{aligned}
 X &= \Psi\left(\frac{3}{16}\right) + \Psi\left(\frac{4}{16}\right) + \Psi\left(\frac{7}{16}\right) + \Psi\left(\frac{8}{16}\right) + \Psi\left(\frac{11}{16}\right) + \Psi\left(\frac{12}{16}\right) + \Psi\left(\frac{14}{16}\right) + \Psi\left(\frac{15}{16}\right) = \\
 &= (-0.89) + (-0.67) + (-0.16) + (0.00) + (0.49) + (0.67) + (1.15) + (1.53) = (-1.72) + (3.84) = 2.12.
 \end{aligned}$$

Тепер варто звернутися до таблиці критерію X (Додаток 5). По числах n і $n_1 - n_2$ (якщо $n_1 < n_2$, то по числах n і $n_2 - n_1$) знаходимо X_{05} . Якщо обчислене значення X перевершує X_{05} , то робиться висновок про існування розходження середніх значень сукупностей x і y (точніше, робиться висновок про приналежність двох вибірок до різних генеральних сукупностей). У протилежному випадку, тобто коли $X \leq X_{05}$, приймається гіпотеза про відсутність розходжень у сукупностях x і y . Іноді, з метою обережності, при виконанні нерівності $X > X_{05}$ порівнюють X з X_{01} . Якщо $X > X_{01}$, то, безумовно, робиться висновок про розходження генеральних сукупностей, з яких витягнуті вибірки. Якщо ж $X_{05} < X < X_{01}$, то ситуація вважається сумнівною, і для остаточного висновку варто продовжувати експеримент.

Для розглянутого приклада $X_{05} = 3.24$, і значить $X < X_{05}$. Тому гіпотеза H_0 приймається. У такий спосіб, розходження в значеннях частотного діапазону двох ритмічних структур є несуттєвим і, отже, можна з великою ймовірністю припустити, що ритміка фрази в подібних реченнях не впливає на величину їхнього частотного діапазону. Природно, щоб даний висновок мав більш загальний характер, необхідно досліджувати різні ритмічні структури на фразах з різним інтонаційним оформленням.

Практика застосування критеріїв розходження показує, що критерій Ван дер Вардена має більшу чутливість до розрізнення, і отримані висновки завжди погодяться або можуть бути пояснені лінгвістичним аналізом. Умовою для застосування критерію X є вимога, щоб різниця кількостей порівнюваних варіант не перевищувала 5 (тобто $|n_1 - n_2| \leq 5$).

5.4. Контрольні запитання

1. Чим визначається постановка лінгвістичної задачі?
2. Коли використовується критерій Стюдента?
3. В чому полягає метод порівнянь?
4. До якої групи критеріїв належить критерій Ван дер Вардена?
5. Коли може бути застосований критерій Ван дер Вардена?

5.5. Задачі для самостійного опрацювання

1. Перевірити гіпотезу щодо випадковості розбіжностей середніх для двох вибірок

Група А

№ з/п	Інтервали	Частоти (n_i)
1	150-200	18
2	200-250	23
3	250-300	39
4	300-350	30
5	350-400	25

Група В

№ з/п	Інтервали	Частоти (n_i)
1	100-110	2
2	110-120	4
3	120-130	8
4	130-140	9
5	140-150	15
6	150-160	13
7	160-170	7

Розв'язання:

Для розв'язання поставленої задачі треба для кожної вибірки (для кожної групи) заповнити таблицю, подібну таблиці 4.1. Спираючись на результати обчислень в цих таблицях, можна буде отримати усі необхідні для подальших розрахунків величини.

З умови задачі видно, що довжина усіх інтервалів для групи А дорівнює 50. Найбільш частотним є третій інтервал (250 – 300) з частотою 39. Тому доцільно зафіксувати саме його. Його середина припадає на 275. Виходячи з цього, для групи А маємо:

$$x_0^{(A)}=275; \quad h_A = 50; \quad N^{(A)} = 3.$$

Група А

№ з/п	Інтервали	Частоти (n_i)	$y_i=c_i^{(A)}$	$n_i y_i$	$n_i y_i^2$
1	150-200	18	$1 - 3 = -2$	$18 \cdot (-2) = -36$	$(-36) \cdot (-2) = 72$
2	200-250	23	$2 - 3 = -1$	$23 \cdot (-1) = -23$	$(-23) \cdot (-1) = 23$
3	250-300	39	$3 - 3 = 0$	$39 \cdot 0 = 0$	$0 \cdot 0 = 0$
4	300-350	30	$4 - 3 = 1$	$30 \cdot 1 = 30$	$30 \cdot 1 = 30$
5	350-400	25	$5 - 3 = 2$	$25 \cdot 2 = 50$	$50 \cdot 2 = 100$
Σ		$n_A=135$		21	225

Об'єм вибірки для групи А (сума частот усіх інтервалів цієї групи) дорівнює $n_A=135$. Враховуючи все це, можна обчислити середнє арифметичне значення $\bar{a}$ для цієї групи, а також показник $\Sigma(a_i - \bar{a})^2$, необхідний для подальших розрахунків.

$$\bar{a} = x_0^{(A)} + \frac{1}{n_A} \left(\sum_{i=1}^5 n_i y_i \right) \cdot h_A = 275 + \frac{21}{135} \cdot 50 = 282,78;$$

$$\begin{aligned} \Sigma(a_i - \bar{a})^2 &= \left[\sum_{i=1}^5 n_i y_i^2 - \frac{1}{n_A} \left(\sum_{i=1}^5 n_i y_i \right)^2 \right] h_A^2 = \\ &= \left(225 - \frac{21^2}{135} \right) \cdot 50^2 = 554333,33. \end{aligned}$$

Група В

№ з/п	Інтервали	Частоти (n_j)	$z_j=c_j^{(B)}$	$n_j z_j$	$n_j z_j^2$
1	100-110	2	$1 - 5 = -4$	$2 \cdot (-4) = -8$	$(-8) \cdot (-4) = 32$
2	110-120	4	$2 - 5 = -3$	$4 \cdot (-3) = -12$	$(-12) \cdot (-3) = 36$
3	120-130	8	$3 - 5 = -2$	$8 \cdot (-2) = -16$	$(-16) \cdot (-2) = 32$
4	130-140	9	$4 - 5 = -1$	$9 \cdot (-1) = -9$	$(-9) \cdot (-1) = 9$
5	140-150	15	$5 - 5 = 0$	$15 \cdot 0 = 0$	$0 \cdot 0 = 0$
6	150-160	13	$6 - 5 = 1$	$13 \cdot 1 = 13$	$13 \cdot 1 = 13$
7	160-170	7	$7 - 5 = 2$	$7 \cdot 2 = 14$	$14 \cdot 2 = 28$
Σ		$n_B=58$		- 18	150

Аналогічно для групи В з умови задачі видно, що довжина усіх інтервалів дорівнює 10. Найбільш частотним є п'ятий інтервал (140 – 150) з частотою 15. Тому доцільно зафіксувати саме його. Його середина припадає на 145. Виходячи з цього, для групи В маємо:

$$x_0^{(B)}=145; \quad h_B = 10; \quad N_0^{(B)} = 5.$$

Об'єм вибірки для групи **B** (сума частот усіх інтервалів цієї групи) дорівнює $n_B=58$. Враховуючи все це, можна обчислити середнє арифметичне значення $\bar{b}$ для цієї групи, а також показник $\Sigma(b_j - \bar{b})^2$, необхідний для подальших розрахунків.

$$\bar{b}=x_0^{(B)}+\frac{1}{n_B}\left(\sum_{j=1}^7 n_j z_j\right) \cdot h_B = 145 - \frac{18}{58} \cdot 10 = 141,90;$$

$$\begin{aligned} \Sigma(b_j - \bar{b})^2 &= \left[\sum_{j=1}^7 n_j z_j^2 - \frac{1}{n_B} \left(\sum_{j=1}^7 n_j z_j \right)^2 \right] \cdot h_B^2 = \\ &= \left(150 - \frac{18^2}{58} \right) \cdot 10^2 = 14441,38. \end{aligned}$$

Тепер можна обчислити $S_{\bar{a}-\bar{b}}$.

$$\begin{aligned} S_{\bar{a}-\bar{b}} &= \sqrt{\frac{\sum (a_i - \bar{a})^2 + \sum (b_j - \bar{b})^2}{n_A + n_B - 2} \cdot \frac{n_A + n_B}{n_A n_B}} = \\ &= \sqrt{\frac{554333,33 + 14441,38}{135 + 58 - 2} \cdot \frac{135 + 58}{135 \cdot 58}} = \sqrt{73,40} = 8,57. \end{aligned}$$

Тепер, маючи $\bar{a}$, $\bar{b}$ і $S_{\bar{a}-\bar{b}}$, можна обчислити безпосередньо критерій Стюдента:

$$t = \frac{|\bar{a} - \bar{b}|}{S_{\bar{a}-\bar{b}}} = \frac{282,78 - 141,9}{8,57} = 16,44.$$

Як і для попередніх критеріїв перевірки гіпотез, обчислюється число ступенів свободи. У разі використання критерію Стюдента воно дорівнює $f=n_A+n_B-2$. Для досліджуваних в даній задачі вибірок маємо

$$f=135+58-2=191.$$

Для цього числа із таблиці Додатка 3 знаходимо для порівняння t_{05} і t_{01} . В таблиці Додатка 3 немає даних для $f=191$. Тому обираємо найближче до нього значення по таблиці. Таким значенням є $f=200$. Для нього знаходимо:

$$t_{05}=1,97; \quad t_{01}=2,6.$$

Очевидно, що $16,44 > 2,6$. Тому порівняння табличних значень критерію з обчисленим у ході розв'язання задачі призводить до висновку, що розбіжності в значенні середніх для цих двох вибірок, що розділені за якоюсь ознакою, є суттєвим, не випадковим.

Висновок: гіпотеза відкидається, ознака, за якою вибірки були розділені, дійсно суттєво впливає на середнє значення випадкової величини.

2. Невелика кількість (16) опитуваних респондентів розділена на дві групи за якісною ознакою. Семеро з них володіють цією ознакою, а дев'ятеро – ні. Їм було запропоновано кількісно оцінити певну величину. Отримані наступні відповіді:

I) 3; 5; 1; 4,4; 7; 6,2; 6; 10; 12.

II) 3,2; 4,6; 2; 8; 9; 5,8; 10,5.

Визначити, чи суттєво впливає зазначена якісна ознака на кількісне сприйняття досліджуваного явища.

Розв'язання:

В зв'язку з невеликою кількістю респондентів групування проводити недоцільно. Різниця між їх кількістю в двох групах $n_1 - n_2 = 9 - 7 = 2 < 5$. Тому можна скористатися критерієм Ван дер Вардена. Для цього ранжуємо і перенумеруємо отримані оцінки в порядку зростання. Результат запишемо у наступну таблицю:

№з/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
I (x)	1		3		4,4		5		6	6,2	7			10		12
II (y)		2		3,2		4,6		5,8				8	9		10,5	

Першій групі присвоєно номери 1, 3, 5, 7, 9, 10, 11, 14 і 16. Кількість чисел у сукупності x (відповідно y) позначене через n_1 (відповідно n_2) і покладається $n = n_1 + n_2$. У розглянутому прикладі $n_1 = 9$, $n_2 = 7$, $n = 16$. Складемо послідовність дробів $\frac{r}{n+1}$, де r пробігає всі номери, приписані числам із сукупності x . У розгляну-

тому прикладі це дробі $\frac{1}{17}, \frac{3}{17}, \frac{5}{17}, \frac{7}{17}, \frac{9}{17}, \frac{10}{17}, \frac{11}{17}, \frac{14}{17}, \frac{16}{17}$. Для кожної дробі по таблиці функції Ψ (Додаток, табл. 4) знаходимо $\Psi(\frac{r}{n+1})$. Результат можна записати у вигляді таблиці:

№з/п	Дріб	$\Psi(\frac{r}{n+1})$
1	$\frac{1}{17}=0,059$	- 1,56
3	$\frac{3}{17}=0,176$	- 0,93
5	$\frac{5}{17}=0,294$	- 0,54
7	$\frac{7}{17}=0,412$	- 0,22
9	$\frac{9}{17}=0,529$	0,07
10	$\frac{10}{17}=0,588$	0,22
11	$\frac{11}{17}=0,647$	0,38
14	$\frac{14}{17}=0,824$	0,93
16	$\frac{16}{17}=0,941$	1,56
		- 0,09

Сума по останньому стовпцю таблиці і є наш обчислений критерій. Тобто в запропонованій задачі критерій $X = - 0,09$. Для чисел $n_1 - n_2 = 9 - 7 = 2$ і $n = 16$ по таблиці Додатку 5 знаходимо $X_{05} = 3,36$ і $X_{01} = 4,25$. Обчислений $X < X_{05}$. Тому можна стверджувати, що наявність зазначеної якісної ознаки не впливає на кількісну оцінку досліджуваного явища.

Висновок: гіпотеза приймається, якісна ознака не має впливу на вимірювану кількісну характеристику.

5.6. Контрольні завдання

1. Перевірити гіпотезу щодо випадковості розбіжностей середніх для двох вибірок

Група А

№ з/п	Інтервали	Частоти (n_i)
1	130-200	15
2	200-270	15
3	270-340	22
4	340-410	38
5	410-480	30
6	480-550	29
7	550-620	14
8	620-690	10

Група В

№ з/п	Інтервали	Частоти (n_i)
1	100-120	9
2	120-140	11
3	140-160	23
4	160-180	36
5	180-200	19
6	200-220	17

2. Експеримент було проведено двічі в різних умовах. Перший раз було зафіксовано 7 спостережень: 12, 22, 13, 14, 25, 48, 30. Друге випробування дало 10 спостережень: 10, 42, 11, 15, 29, 33, 18, 45, 20, 47. Визначити, чи належать результати обох випробувань до однієї генеральної сукупності або різниця в умовах суттєво впливає на результат.

6. ВИВЧЕННЯ ЗАЛЕЖНОСТІ ЛІНГВІСТИЧНИХ ОЗНАК

6.1. Кореляційна залежність

Для встановлення загальних законів, по яких протікають лінгвістичні явища, необхідно не тільки проаналізувати кожний з компонентів якого-небудь явища, але представити детальну кількісну і якісну характеристику різних зв'язків, що існують між ознаками, що цікавлять дослідника процесу, і описати результати впливу даних зв'язків на досліджувані процеси [5].

Прояв однієї ознаки перебуває в тісному зв'язку з багатьма іншими ознаками досліджуваного явища. Внутрішня структура явища характеризується багатопричинним зв'язком. Зведення складної системи відносин до більш простих її видів, виділення тих зв'язків, які є основними в досліджуваній функції мовного явища – це ті задачі, які повинні бути вирішені експериментатором на початковому етапі дослідження.

Теоретично всі зв'язки, з узагальнення яких виникають закони науки, можна розділити на два види – функціональна залежність і кореляційна залежність. При *функціональній залежності* будь-якому фіксованому значенню однієї ознаки відповідає строго визначене, завжди те саме, значення іншої ознаки. При *кореляційній залежності* фіксованому значенню однієї ознаки можуть відповідати кілька значень іншої ознаки, причому до проведення експерименту це відповідне значення другої ознаки дослідникові невідомо. Причина цього полягає в тому, що в жодному експерименті, загалом кажучи, не можна повністю врахувати вплив другорядних факторів. У лінгвістичних явищах всі зв'язки, як правило, кореляційні. І в тих випадках, коли лінгвістичне явище описується рядом ознак (факторів), що мають якісне вираження, характер і ступінь залежності ознак можна вивчити за допомогою кореляційного аналізу.

Кореляційний зв'язок може розглядатися з погляду його «тісноти» і «форми». Під *тісністю* кореляційного зв'язку розуміється сила впливу досліджуваних ознак однієї на одну. По

тісноті кореляція може бути *слабкою*, *середньою* й *сильною*. *Форма* кореляційного зв'язку показує, як у середньому змінюються значення однієї ознаки при зміні іншої. За формою кореляція може бути прямолінійною (*лінійною*) або *криволінійною*. Прямо-лінійна форма кореляційного зв'язку виникає тоді, коли рівним змінам першої ознаки відповідають рівні (у середньому) за величиною й за знаком зміни другої ознаки.

Приклад 6.1. Нехай є дві ознаки, X і Y , що характеризують одне й теж лінгвістичне явище. У результаті експерименту були отримані наступні значення:

x	4	6	9	10	12
y	20	28	30	35	37

**Повна пряма
кореляція**

З отриманих даних видно, що зі збільшенням кожного значення ознаки X збільшується кожне значення ознаки Y . Залежність, при якій зі збільшенням (або зменшенням) кожного значення однієї ознаки збільшується (або зменшується) кожне значення іншої ознаки, називається *повною*.

В іншому випадку ознаки X і Y одержали такі значення:

x	4	6	9	10	12
y	30	20	35	28	37

**Часткова пряма
кореляція**

Наведені дані показують, що зі збільшенням значень ознаки X збільшуються не кожне значення ознаки Y . Залежність називається *частковою*, коли в середньому зі збільшенням (або зменшенням) значень однієї ознаки збільшуються (зменшуються) значення іншої ознаки. Повна й часткова кореляція за своєю формою є *прямою*, якщо збільшення (зменшення) значень однієї ознаки спричиняє збільшення (зменшення) значень іншої ознаки. У тому випадку, коли зі збільшенням (зменшенням) значень однієї ознаки зменшуються (збільшуються) значення іншої ознаки, кореляція називається *зворотною*.

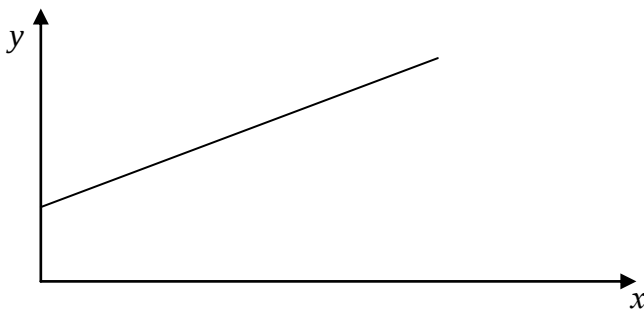
x	4	6	9	10	12
y	37	35	30	28	20

**Повна зворотна
кореляція**

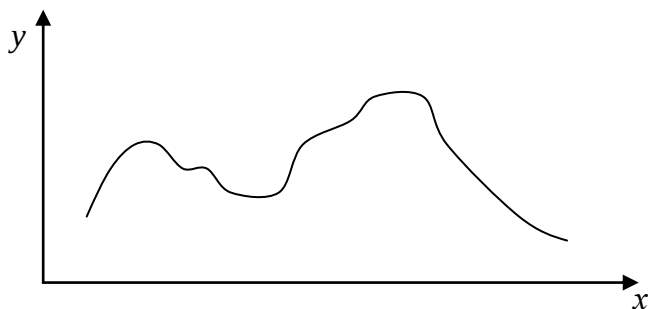
x	4	6	9	10	12
y	37	28	35	20	30

Часткова зворотна кореляція

При лінійній залежності розташування значень ознак X і Y можна подати графічно уздовж однієї прямої. Ця пряма називається *лінією регресії*. Залежність вважається *лінійною*, якщо рівним збільшенням значень однієї ознаки відповідають більш-менш рівні зміни того самого знака іншої ознаки. Якщо рівним збільшенням значень однієї ознаки відповідають різні зміни як за знаком, так і за величиною, іншої ознаки, то така залежність називається *криволінійною*.



Мал. 6.1 – Лінійна залежність двох ознак



Мал. 6.2 – Криволінійна залежність двох
ознак

Кореляційному аналізу завжди повинен передувати лінгвістичний аналіз досліджуваних ознак. Якщо в результаті лінгвістичного аналізу виявлено, що між двома ознаками якого-небудь лінгвістичного явища існує зв'язок, то застосування кореляційного аналізу дає можливість виявити ступінь і характер даного зв'язку, або відсутність зв'язку між даними ознаками. При цьому необхідно враховувати вплив інших ознак на досліджувані. У цьому випадку, коли ці додаткові ознаки дуже впливають (цей висновок робиться тільки на основі лінгвістичного аналізу), то дані кореляційного аналізу не будуть відбивати реальну картину залежності досліджуваних двох ознак. У таких випадках застосовується парціальна кореляція.

Кореляційна залежність двох ознак не має двостороннього характеру, тобто некомутативна. Точніше кажучи, якщо вплив 1-ї ознаки на 2-гу має певну міру й форму зв'язку, то тіснота й форма кореляційного зв'язку між 2-ю і 1-ю ознаками може бути іншою. Але в тих випадках, коли кореляційний зв'язок між двома ознаками є лінійним, ступені впливу однієї ознаки на іншу однакові. Цим пояснюється те, що надалі велику увагу буде приділено саме лінійному кореляційному зв'язку.

У лінгвістичних дослідженнях кореляційний аналіз застосовується досить широко. Він служить для встановлення подібності й зв'язку між одиницями мови й мовлення, якщо відомо частоту зустрічальності цих одиниць. Зокрема, використання кореляційного аналізу для встановлення парадигматичних зв'язків між словами засновано на припущенні, що чим більшою мірою збігається лексична сполучуваність двох слів, тим більшою мірою проявляється семантична подібність між цими словами. Досвід лінгвістичних досліджень показує, що при використанні кореляційного аналізу для встановлення парадигматичних відносин у лексиці необхідно розрізняти два основних процедурних варіанти [6]:

- **формальна процедура**, коли аналіз здійснюється на основі частоти сполучуваності досліджуваних одиниць без якої-небудь додаткової семантичної обробки й угруповання останніх;

- **напівформальна процедура**, коли аналіз проводиться на основі спільної зустрічальності (сполучуваності) досліджуваної

одиниці із семантичними підкласами слів, угруповання й вичленовування яких здійснюється дослідником неформальним способом.

Експериментальним шляхом встановлено, що застосування чисто формальної процедури підвищує ступінь об'єктивності одержуваних результатів, але знижує ступінь їхньої повноти й відповідності інтуїтивним уявленням про зв'язки слів, відбитих у тлумачних і синонімічних словниках. Застосування напівформальної процедури знижує ступінь об'єктивності (тому що дослідник інтуїтивно робить якісне угруповання контекстуальних партнерів по подібності їхньої семантики) одержуваних результатів, але більш повно відбиває сполучуваності властивості всіх лексико-семантичних варіантів, а разом з тим і парадигматичні зв'язки аналізованих слів. Ще одна особливість використання кореляційного аналізу полягає в тому, що часто вихідна таблиця, на основі якої здійснюється кореляційний аналіз, залежно від цілей і задач дослідження використовується двічі: при знаходженні коефіцієнтів кореляції таблиця розвертається таким чином, щоб рядки й стовпці в ній мінялися місцями. Наприклад, були досліджені парадигматичні відносини між прикметниками зі значенням «міцний» в англійській мові на основі їхньої сполучуваності з підкласами іменників. У результаті були отримані дані про парадигматичні зв'язки цих прикметників і на цій основі побудоване «семантичне поле». Однак, строго говорячи, дослідження констатують у таких ситуаціях не семантичне, а лексичне поле, тому що встановлюється зв'язок між словами, а не між значеннєвими одиницями. Розгорнувши вихідну таблицю, де була зафіксована частота зустрічальності прикметників з підкласами іменників («значеннєвими одиницями», вираженими й обмеженими певним набором слів), були знайдені величини коефіцієнтів кореляції між частотами семантичних підкласів і на цій основі побудоване семантичне поле, що покриває лексемами зі значенням «міцний» в англійській мові. Потім був здійснений наступний крок: було проведене порівняння семантичного поля «міцний» в англійській мові з відповідним семантичним полем у французькій мові [14, 15]. В інших аналогічних випадках можна

досліджувати зв'язки між іншими одиницями мови на фонетичному, морфологічному або синтаксичному рівнях.

6.2. Лінійна кореляція

Припустимо, що є значення двох ознак X і Y якогось лінгвістичного явища. У кожному прояві явища обидві ці ознаки приймають певне числове значення. У результаті численних спостережень було помічено, що ознаки X і Y взаємозалежні. Зв'язок може прийняти різний вид, але ми вправі припустити найпростішу залежність між ними – прямолінійну (лінійну). Лінійний зв'язок є найпростішим й досить часто зустрічається. Мірою тісноти лінійного зв'язку служить *коефіцієнт кореляції* r , обумовлений формулою

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}.$$

Тут, як звичайно, $\bar{x}$ (відповідно $\bar{y}$) позначає середнє арифметичне значення ознаки x за спостереженнями $x_1, x_2, \dots, x_n$ (відповідно для y).

Приклад 6.2. Лінгвістичний аналіз показав, що між компонентами словесного наголосу – інтенсивністю, тривалістю й сумарною енергією існують складні зв'язки, і що роль цих компонентів у створенні ефекту наголошеності складу не однакова. Виникає задача про об'єктивне вивчення ступеня й характеру зв'язку між інтенсивністю, тривалістю й сумарною енергією методами математичної статистики. Для визначення кореляційного зв'язку між тривалістю й інтенсивністю наголошеного голосного кінцевого складу був проведений експеримент із 50 спостережень. Експериментальний матеріал складався зі слів того самого звукового складу й однорідної акцентної структури. Були обчислені показники основних акустичних характеристик – інтенсивності, тривалості й сумарної енергії наголошених і ненаголошених складів. Позначимо тривалість через X й інтенсив-

ність через Y . Якщо за результатами експерименту ми в змозі визначити значення цих ознак у кожному експерименті (спостереженні), то ми одержимо сукупність пар $(x_1, y_1), \dots, (x_n, y_n)$. Тут n – кількість спостережень в експерименті, а пари (x_i, y_i) означають, що в i -тому спостереженні ознака X прийняла значення x_i , а ознака Y – значення y_i . Розрахунки для обчислення коефіцієнта кореляції наведені в таблиці 6.1, де X – тривалість наголошеного голосного кінцевого складу, Y – інтенсивність наголошеного голосного кінцевого складу.

Таблиця 6.1

№	1		2		3		4
	Значення ознаки		Відхилен. від серед. арифм.		Квадрати відхилень від серед. арифм		Добутки відхилень
	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	2	3	4	5	6	7	8
1	350	17	+110	+3.5	12100	12.25	+385
2	320	17	+80	+3.5	6400	12.25	+280
3	360	16	+120	+2.5	14400	6.25	+300
4	360	15	+120	+1.5	1440	2.25	+180
5	340	17	+100	+3.5	10000	12.25	+350
6	240	10.5	0	- 3.0	0	9.0	0
7	200	9.5	- 40	+4.0	160	16.0	+160
8	260	14.5	+20	- 1.0	400	1.0	+20
9	200	10.5	- 40	- 3.0	1600	9.0	+120
10	240	11	0	- 2.5	0	6.25	0
11	290	16.5	- 50	+3.0	2500	9.0	- 150
12	220	19.5	- 20	+6.0	400	36.0	- 120
13	200	16	- 40	+2.5	1600	6.25	- 100
14	220	10	- 20	- 3.5	400	12.25	+70
15	300	13	+60	- 0.5	3600	0.25	- 30
16	230	16	- 10	+2.5	100	6.25	- 25
17	270	15	+30	+1.5	900	2.25	+45
18	300	15.5	+60	- 2.0	3600	4.0	+120
19	300	19.5	+60	+6.0	3600	36.0	+360
20	270	19.5	+30	+6.0	900	36.0	+180
21	250	20	+10	+6.5	100	42.25	+65
22	250	15	+10	+1.5	100	2.25	- 15

Продовження таблиці 6.1

1	2	3	4	5	6	7	8
23	310	16	+70	+2.5	4900	6.25	+175
24	320	14.5	+80	+1.0	6400	1.0	+80
25	280	11	+40	- 2.5	1600	6.25	- 100
26	230	15	- 10	+1.5	100	2.25	- 15
27	250	14	+10	+0.5	100	0.25	+5
28	280	19	+40	+5.5	1600	30.25	+220
29	230	13	- 10	- 0.5	100	0.25	+5
30	270	17	+30	+3.5	900	12.25	+105
31	250	16	+10	+2.5	100	6.25	+2.5
32	200	16	- 40	+2.5	1600	6.25	- 100
33	180	14	- 60	+0.5	3600	0.25	- 30
34	240	10	0	- 3.5	0	12.25	0
35	200	14	- 40	+0.5	1600	0.25	- 20
36	270	13	+30	- 0.5	900	0.25	- 15
37	170	11	- 70	- 2.5	4900	6.25	+175
38	190	13	- 50	- 0.5	2500	0.25	+25
39	130	3.5	- 110	- 10.0	12100	100.0	+1100
40	130	4	- 110	- 9.0	12100	90.25	+1054
41	150	7	- 90	- 6.5	8100	42.25	+585
42	150	3	- 90	- 10.5	8100	110.25	+945
43	220	9.5	- 20	- 4.0	400	16.0	+80
44	220	17.5	- 20	- 4.0	400	16.0	+80
45	240	14	0	+0.5	0	0.25	0
46	240	16	0	+2.5	0	6.25	0
47	2400	18	0	+4.5	0	20.25	0
48	180	7	- 60	- 6.5	3600	42.25	+390
49	180	9	- 60	- 4.5	3600	20.25	+270
50	180	6	- 60	- 7.5	3600	56.25	+450
Ра- зом	12000	674.5			161600	890.25	7545

$$r = \frac{7545}{\sqrt{161600 \cdot 890}} = \frac{7545}{11928} = 0.63.$$

З наведеного приклада видно, що безпосередній підрахунок r являє собою досить громіздку процедуру, особливо при великих n . Значення r також можна визначити, використовуючи ме-

тод групування даних у кореляційних ґратах, що істотно скорочує кількість необхідних обчислень. До того ж, кореляційні ґрати дозволяють проводити попередній статистичний аналіз характеру кореляційної залежності. Для побудови кореляційних ґрат діємо в такий спосіб: по спостережуваним значенням ознаки X (відповідно, Y), тобто по числах $x_1, x_2, \dots, x_n$ (відповідно, $y_1, y_2, \dots, y_n$) визначаємо інтервали групування. Рекомендується, щоб кількість інтервалів групування по кожній ознаці перебувала в межах від 8 до 15 (для кожної ознаки X і Y своя кількість інтервалів). Не рекомендується розбивати отримані дані на занадто малу кількість інтервалів, тому що це може привести до зменшення точності вимірів тісноти зв'язку між ознаками.

Приклад 6.3. Для умови приклада 6.2 обчислимо коефіцієнт кореляції з використанням кореляційних ґрат. Для цього всі отримані виміри ознаки X (тривалості) розбиваємо на 8 інтервалів, а всі виміри ознаки Y (інтенсивності) – на 9 інтервалів. Отримані дані наведені в таблиці 6.2.

З наведеної таблиці видно, що крім кореляційних ґрат, вона містить ще ряд додаткових рядків і стовпців. Але насамперед розглянемо самі кореляційні ґрати (на числа кліток, що розташовані у дужках, поки не звертаємо уваги). Видно, що хоча заповнені клітки й не лежать строго на діагоналі, що з'єднує лівий верхній і правий нижній кути таблиці, все-таки зосереджені біля цієї діагоналі. Тому природно очікувати, що кореляційний зв'язок між тривалістю й інтенсивністю близький до лінійного (причому зі збільшенням тривалості збільшується й інтенсивність). Розглянемо тепер зміст додаткових рядків і стовпців таблиці 6.2.

Перший стовпець (позначений n_x) складається із чисел, що дорівнюють сумі чисел, зазначених у заповнених клітках відповідного рядка кореляційних ґрат. Підраховуємо, скільки разів значення ознаки X у кожному з інтервалів зустрічається в кожному з інтервалів ознаки Y . Так, перше число стовпця n_x , тобто число 4, означає, що значенню тривалості з інтервалу 130 – 160 відповідає 4 спостереження в експерименті, а саме, три зі значенням інтенсивності в інтервалі 3 – 5 і одне в інтервалі 7 – 9.

Стовпець і рядок №2 (позначені відповідно c_x і c_y) являють собою штучні шкали. Для одержання елементів другого стовпця вибирають інтервал ознаки X , приписують йому значення 0, а іншим послідовно значення -1 , -2 і т.д. (якщо ці інтервали передують зафіксованому) і значення 1 , 2 , 3 і т.д. (для наступних за зафіксованим інтервалів). Зазвичай за нуль приймають одну із центральних частот (переважно максимальну). У розглянутому прикладі значення 0 приписане інтервалу $250 - 280$.

Третій стовпець (позначений як $n_x c_x$) складається з попарних добутків елементів 1-го й 2-го стовпців (з урахуванням знаків).

Четвертий стовпець (позначений як $n_x c_x^2$) складається з попарних добутків елементів 1-го стовпця на квадрати елементів 2-го стовпця. Всі добутки беруться зі знаком «+».

Аналогічним чином будуються елементи чотирьох додаткових рядків.

Щоб одержати елементи 5-го стовпця (рядка), спочатку визначають числа в круглих дужках усередині кореляційних грат. Ці числа є добутками відповідних елементів 2-го стовпця й 2-го рядка (з урахуванням знаків). Так, в 1-му рядку кореляційних грат під числом 3 у дужках розташоване число 12, що є результатом добутку числа 3 на число 4. Тепер для одержання елементів стовпця № 5 (позначеного як Σ) складаються попарні добутки чисел із кліток відповідного рядка кореляційних грат на відповідні числа в круглих дужках (з урахуванням знака). Результати добутків відповідного рядка складаються. Так, число 42 у стовпці № 5 дорівнює $3 \cdot 12 + 6 \cdot 1$. Аналогічно знаходять елементи рядка № 5.

Тепер для визначення коефіцієнта кореляції обчислюємо:

а) число, що дорівнює сумі елементів стовпця № 1 (перевіркою потрібно переконаватися, що таке ж значення n дає сума елементів рядка № 1 (тобто рядка n_y). У розглянутому прикладі $n=50$;

б) суму елементів стовпця № 3 позначимо через X_1 , а суму елементів рядка № 3 – через Y_1 . У нас $X_1=15$, $Y_1=48$;

в) суму елементів стовпця № 4 позначимо через X_2 , а суму елементів рядка № 4 – як Y_2 . У нас $X_2=183$, $Y_2=288$;

Таблица 6.2

X	Y										n _x	c _x	n _x c _x	n _x C _x ²	Σ ₁
	3-5	5-7	7-9	9-11	11-13	13-15	15-17	17-19	19-21						
130-160	3 (12)		1 (6)							4	-3	-12	36	42	
160-190		1 (6)	1 (4)	1 (2)	1 (10)	1 (-2)				5	-2	-10	20	10	
190-220				2 (1)		2 (-2)	3 (-2)			7	-1	-7	7	-6	
220-250				4 (0)	1 (0)	2 (0)	3 (0)	2 (0)	1 (0)	13	0	0	0	0	
250-280						3 (1)	3 (2)	1 (3)	2 (4)	9	1	9	9	20	
280-310					1 (0)	1 (2)	1 (4)		2 (8)	5	2	10	20	22	
310-340						1 (3)	1 (6)	1 (9)		3	3	9	27	18	
340-370							2 (8)	2 (12)		4	4	16	64	40	
n _y (1)	3	1	2	7	3	10	13	6	5	50		15	183	146	
c _y (2)	-4	-3	-2	-1	0	1	2	3	4						
n _y c _y (3)	-12	-3	-4	-7	0	10	26	18	20	48					
n _y C _y ² (4)	48	9	8	7	0	10	52	54	80	268					
Σ ₂	36	6	10	4	0	4	26	36	24	146					

г) сума елементів стовпця № 5 позначається через Z , у нас $Z=146$. Таке ж значення Z повинне вийти (і це повинне служити як перевірка) як сума елементів рядка № 5.

$$\text{Обчислимо тепер } \Sigma_{xy}=Z - \frac{X_1 Y_1}{n}, \Sigma_{xx}=X_2 - \frac{X_1^2}{n}, \Sigma_{yy}=Y_2 - \frac{Y_1^2}{n}.$$

Звідси

$$r = \frac{\Sigma_{xy}}{\sqrt{\Sigma_{xx} - \Sigma_{yy}}}.$$

Для розглянутого приклада

$$\Sigma_{xy}=146 - \frac{15 \cdot 48}{50} = 146 - 14.4 = 131.6;$$

$$\Sigma_{xx}=183 - \frac{15^2}{50} = 183 - 4.5 = 178.5;$$

$$\Sigma_{yy}=268 - \frac{48^2}{50} = 268 - 46 = 222.$$

Отже,

$$r = \frac{131.6}{\sqrt{178.5 \cdot 222}} = 0.66.$$

Для того, щоб зробити висновок про кореляційну залежність ознак X і Y варто врахувати, що коефіцієнт кореляції r може приймати значення від -1 до $+1$, причому негативне значення для r говорить про те, що залежність між X і Y зворотна (тобто з ростом однієї ознаки інша зменшується), а при позитивних значеннях r зв'язок прямий (з ростом однієї ознаки збільшується й інша). Крім того, чим ближче r до ± 1 , тим сильніше лінійний зв'язок між X і Y , а коли r близько до нуля, то між X і Y лінійний зв'язок відсутній. Загалом за силою зв'язку величину коефіцієнта лінійної кореляції можна інтерпретувати наступним чином: , якщо $|r| \in [0.7; 1]$ – є сильний лінійний зв'язок, якщо $|r| \in [0.3; 0.7)$ – є лінійний зв'язок середньої сили, якщо $|r| \in [0; 0.3)$ – лінійний зв'язок слабкий, майже відсутній. Однак при цьому цілком можливо, що між X і Y має місце кореляційний зв'язок, і навіть сильний, але криволінійний.

6.3. Кореляційні відносини

Будь-який вид зв'язку двох ознак можна звести до прямолінійної або криволінійної форми залежності. Тіснота зв'язку ознак як при прямолінійній, так і при криволінійній кореляції, визначається так званими кореляційними відносинами. На відміну від коефіцієнта кореляції, кореляційні відносини виражають односторонній зв'язок ознак як у випадку лінійної, так і у випадку криволінійної кореляції, тобто залежність X від Y , позначувану через $e_{x/y}$, і залежність Y від X , позначувану через $e_{y/x}$. Величина e завжди позитивна. Вона змінюється від 0 до 1. Якщо вона приймає значення, близьке або рівне одиниці, то говорять про тісний зв'язок (або повну кореляцію) досліджуваних ознак, і, навпаки, значення e , близьке або рівне нулю, свідчить про відсутність якої-небудь залежності між досліджуваними ознаками.

Для визначення $e_{x/y}$ і $e_{y/x}$ також використовують кореляційну таблицю, але при цьому для зручності обчислень додають ще два стовпці № 6 і № 7 і два рядки № 6 і № 7. Щоб одержати елементи 6-го стовпця (позначеного як y_x), варто помножити числа відповідного рядка кореляційних ґрат на відповідні елементи 2-го рядка (позначеної як c_y). Сума цих попарних добутків і дає значення відповідного елемента стовпця № 6. Елементи 7-го стовпця (7-го рядка) дорівнюють квадратам елементів 6-го стовпця (6-го рядка), діленим на відповідні елементи 1-го стовпця (1-го рядка). Позначимо через S_1 (відповідно S_2) суму всіх елементів 7-го стовпця (відповідно, 7-го рядка) і обчислимо $S_x = X_2 - S_2$, $S_y = Y_2 - S_1$. Значення X_2 і Y_2 вже були знайдені при обчисленні коефіцієнта кореляції r . З урахуванням того, що кількість інтервалів розбивки ознаки X дорівнює k , а кількість інтервалів розбивки ознаки Y дорівнює l , одержуємо формули для обчислення кореляційних відносин:

$$e_{x/y} = \sqrt{1 - \frac{n-1}{n-k} \cdot \frac{S_x}{\sum_{xx}}}, \quad e_{y/x} = \sqrt{1 - \frac{n-1}{n-l} \cdot \frac{S_y}{\sum_{yy}}}.$$

Якщо об'єм вибірки, тобто n , невеликий (наприклад, менше 30), то можна користуватися спрощеними формулами для $e_{x/y}$ і $e_{y/x}$:

$$e_{x/y} = \sqrt{1 - \frac{S_x}{\Sigma_{xx}}}, \quad e_{y/x} = \sqrt{1 - \frac{S_y}{\Sigma_{yy}}}.$$

Приклад 6.4. Обчислимо кореляційні відносини для ознак X і Y в умовах приклада 6.2. Для цього необхідно таблицю 6.2 доповнити ще двома стовпцями й ще двома рядками, значення елементів яких обчислюється наведеним способом. Так, наприклад, перший елемент 6-го стовпця дорівнює $3 \cdot (-4) + 1 \cdot (-2) = -14$, другий елемент дорівнює $1 \cdot (-3) + 1 \cdot (-2) + 1 \cdot (-1) + 1 \cdot 0 + 1 \cdot 1 = -5$, і т.д. Аналогічно визначаються елементи 6-го рядка. Так, його 4-й елемент дорівнює $1 \cdot (-2) + 2 \cdot (-1) + 4 \cdot 0 = -4$. Отже, у цілому таблиця 6.2 доповнюється такими стовпцями й строками:

6-й стовпець: - 14; - 5; 6; 14; 20; 11; 6; 10.

7-й стовпець: 49; 5; 5.1; 44.5; 24.2; 12; 6.

6-й рядок: 9; - 2; - 5; - 4; 0; 4; 13; 12; 6.

7-й рядок: 27; 4; 12.5; 2.3; 0; 1.6; 13; 24; 7.2.

Далі знаходимо:

$$S_1 = 179.8, S_2 = 91.6, S_x = 183 - 91.6 = 91.4, S_y = 268 - 179.8 = 88.2,$$

$$e_{x/y} = \sqrt{1 - \frac{50-1}{50-8} \cdot \frac{91.6}{178.5}} = 0.63, \quad e_{y/x} = \sqrt{1 - \frac{50-1}{50-9} \cdot \frac{88.2}{222}} = 0.72.$$

Отримані значення кореляційних відносин дозволяють затверджувати, що кожна з ознак X і Y (тобто тривалість і інтенсивність) досить сильно впливають одна на одну.

6.4. Парціальна кореляція

У більшості випадків при лінгвістичних дослідженнях характер кореляції між двома ознаками повністю або частково залежить від того, що дві досліджуваних ознаки або одна з них залежать, у свою чергу, від якої-небудь третьої ознаки. Для того, щоб точно оцінити зв'язок між двома ознаками, необхідно виключити залежність від третьої ознаки. Це можна зробити, зокрема, при постійному значенні третьої ознаки. Але на практиці це здійснити дуже важко, практично неможливо. Однак можна спробувати виключити залежність від третьої ознаки, зіставив-

ши її з кожною з основних ознак. Коефіцієнт кореляції між двома ознаками X і Y при постійних значеннях третьої ознаки Z називається *парціальним коефіцієнтом кореляції*. Парціальний коефіцієнт кореляції може бути обчислений тільки у випадку строго лінійного зв'язку всіх трьох ознак за допомогою наступної формули:

$$r_{xy/z} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}.$$

6.5. Критерії кореляційного аналізу

6.5.1. Критерій вірогідності залежності ознак X і Y

Часто виникає необхідність визначити, чи дійсно залежить X від Y , або отримане по одній вибірці значення є чисто випадковим. Інакше кажучи, чи можна на підставі одного лише значення $e_{x/y}$, близького до одиниці, затверджувати, що в тій же генеральній сукупності ознака X у тому ж ступені залежить від ознаки Y . Щоб можна було поширити висновок за межі вибірки, перевіряють отримане значення $e_{x/y}$ за допомогою формули

$$K = \frac{e_{x/y}}{m(e_{x/y})},$$

де $m(e_{x/y}) = \frac{1 - e_{x/y}^2}{\sqrt{n}}$ - середня помилка обчислення. Нульова гіпотеза затверджує незалежність ознак X і Y . Відповідно до критерію, при $K > 3$ H_0 відкидається, при $K \leq 3$ H_0 приймається. Подібним же чином перевіряється вірогідність висновку щодо залежності ознаки Y від X .

Приклад 6.5. В умовах прикладів 6.2 і 6.3 перевіримо вірогідність залежності ознаки X від Y і ознаки Y від X .

$$m(e_{x/y}) = \frac{1 - e_{x/y}^2}{\sqrt{n}} = \frac{0.596}{7.07} = 0.08, K = \frac{e_{x/y}}{m(e_{x/y})} = \frac{0.63}{0.08} = 8 > 3.$$

Отже, гіпотеза H_0 відкидається з імовірністю 0.99, тобто ознака X залежить від Y .

$$m(e_{y/x}) = \frac{1 - e_{y/x}^2}{\sqrt{n}} = \frac{0.475}{7.07} = 0.07, K = \frac{e_{y/x}}{m(e_{y/x})} = \frac{0.72}{0.07} = 10.3 > 3.$$

Отже, і в цьому випадку гіпотеза H_0 відкидається з імовірністю 0.99, тобто варто визнати, що ознака Y залежить від X .

6.5.2. Критерій значимості розходження кореляційних відносин

Часто виникає необхідність знати, чи можна вважати значення $e_{x/y}$ і $e_{y/x}$ рівними (тобто їхні розходження є чисто випадковим фактором) або необхідно встановити, що їхні розходження значимі, а виходить, можна затверджувати, що одне з них завжди більше іншого в генеральній сукупності. Отже, нульова гіпотеза затверджує, що $e_{x/y} = e_{y/x}$. Для перевірки H_0 використовується критерій

$$K = \frac{(e_{x/y} - e_{y/x})^2}{m^2(e_{x/y}) + m^2(e_{y/x})}.$$

Якщо $K > 9$, то гіпотеза H_0 відкидається з імовірністю 0.99. Якщо $K \leq 9$, то з тією же ймовірністю гіпотеза H_0 приймається.

Приклад 6.6. Для розглянутих умов $K = \frac{(0.72 - 0.63)^2}{0.08^2 + 0.07^2} =$

$$\frac{0.09^2}{0.0113} = 0.71 < 9. \text{ Отже, гіпотеза } H_0 \text{ приймається, тобто } e_{x/y} = e_{y/x}.$$

6.6. Регресійний аналіз

З кореляційного аналізу відомо, як перевірити наявність залежності між ознаками й установити тісноту цього зв'язку. Кореляційні відносини дають нам ступінь однобічного зв'язку ознак, але не вказують, наскільки змінюється одна ознака зі зміною іншої. Часто буває необхідно довідатися, як поводить

ознака X із збільшенням (або зменшенням) значення ознаки Y . На це питання дає відповідь регресійний аналіз. Залежно від форми зв'язку ознак розрізняють: лінійну й криволінійну регресію. Найпростішою формою зв'язку є лінійна. При строго функціональній лінійній залежності двох ознак X і Y їхня форма зв'язку геометрично виразиться однією прямою лінією. При кореляційній лінійній залежності (коли коефіцієнт кореляції, загалом кажучи, відмінний від ± 1) однією прямою лінією цей зв'язок виразити не вдається. Однак за допомогою двох прямих можна показати, як кількісні зміни однієї ознаки впливають на кількісні зміни іншої ознаки. Ці дві прямі називаються *лініями регресії*. У загальному випадку, тобто при криволінійному кореляційному зв'язку, можна за допомогою двох кривих описати вплив однієї ознаки на іншу. Будемо розглядати тільки той випадок, коли регресійний зв'язок виражається за допомогою прямих ліній регресії. Перша з них називається *лінією регресії Y по X* и показує, як змінюється Y при зміні X . Її рівняння має вигляд

$$y - \bar{y} = b_{x/y}(x - \bar{x}).$$

Щоб визначити x , y , $b_{x/y}$ варто звернутися до кореляційної таблиці 6.2. Тоді

$$\bar{x} = c_x + \frac{\sum n_x c_x}{n} d_x, \quad \bar{y} = c_y + \frac{\sum n_y c_y}{n} d_y, \quad b_{x/y} = \frac{\sum_{xy}}{\sum_{xx}} \cdot \frac{d_y}{d_x}$$

Тут c_x (відповідно c_y) – середина інтервалу для ознаки X (відповідно Y), якому приписане значення 0, а d_x (відповідно d_y) – довжини інтервалів групування ознаки X (відповідно Y). Для розглянутого приклада 6.3 маємо:

$$\bar{x} = 235 + \frac{15}{50} \cdot 30 = 244; \quad \bar{y} = 12 + \frac{48}{50} \cdot 2 \approx 14; \quad b_{x/y} = \frac{131.6}{178.5} \cdot \frac{2}{30} = 0.05.$$

Таким чином, рівняння прямої лінії регресії Y по X має вигляд:

$$y - 14 = 0.05(x - 244).$$

Коефіцієнт 0.05 показує, що при збільшенні (зменшенні) значення ознаки X на одиницю ознака Y збільшується (зменшується) на 0.05.

Аналогічно лінія регресії ознаки X по Y задається рівнянням

$$x - \bar{x} = b_{y/x}(y - \bar{y}).$$

Значення $\bar{x}$ й $\bar{y}$ визначаються по наведеним вище формулах, а значення $b_{y/x}$ задається співвідношенням

$$b_{y/x} = \frac{\sum_{xy}}{\sum_{yy}} \cdot \frac{d_x}{d_y}.$$

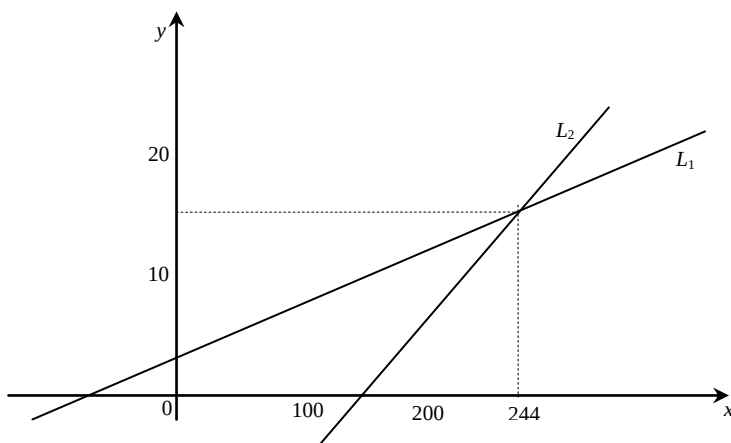
Для прикладу 6.3 маємо:

$$b_{y/x} = \frac{131.6}{222} \cdot \frac{30}{2} = 8.9.$$

Тоді рівняння прямої лінії регресії X по Y приймає вид:

$$x - 244 = 8.9(y - 14).$$

На графіку ці прямі мають вигляд:



Мал. 6.3 – L_1 – лінія регресії Y по X ; L_2 – лінія регресії X по Y

Практичне застосування цих ліній регресії таке, що якщо дослідник задається деяким значенням ознаки X (наприклад, у розглянутому випадку задамося значенням тривалості наголошеного голосного 200 мсек), то в середньому варто очікувати, що ознака Y прийме значення ординати відповідної точки пря-

мої L_1 (у розглянутому прикладі $Y \approx 11.8$ мм). І навпаки, якщо задаватися значенням ознаки Y , то розглядаючи пряму L_2 , одержуємо середнє очікуване значення ознаки X . Так, при інтенсивності 7 мм тривалість у середньому дорівнює 181.7 мсек.

Однак варто помітити, що користуватися графіком ліній регресії в зазначеному вище змісті можна тільки в певних межах. Іншими словами, задаватися значенням однієї з ознак можна тільки в межах найменшого - найбільшого зі значень цієї ознаки, отриманих у даній вибірці. Так, лінією регресії L_1 із приклада 6.2 можна користуватися, задаючись значеннями ознаки тривалості тільки з інтервалу 130 - 370.

При обчисленні коефіцієнтів $b_{y/x}$ і $b_{x/y}$ можуть вийти й негативні значення, і знак мінус у відповідного коефіцієнта говорить лише про те, що зі збільшенням однієї ознаки інша зменшується. У всьому іншому зміст ліній регресії аналогічний уже описаному. Величина кута між лініями регресії L_1 і L_2 характеризує тісноту зв'язку ознак X і Y – чим сильніше залежність ознак, тим менше кут між прямими. Коли цей кут близький до 90° , то практично лінійний зв'язок між ознаками відсутній. У цьому випадку доцільно побудувати емпіричну криву регресії Y по X (і аналогічно X по Y). Процес побудови емпіричної лінії регресії виглядає в такий спосіб. Нехай $(x_1, y_1), \dots, (x_n, y_n)$ – n пар отриманих значень ознак X і Y , і нехай $X_1, \dots, X_k$ – всі різні значення X серед чисел $x_1, \dots, x_n$ (деякі з x_i можуть збігатися, так що $k \leq n$). Якщо значення X_1 зустрілося серед чисел $x_1, \dots, x_n$ рівно n_1 раз, то йому відповідає n_1 значень Y серед чисел $y_1, \dots, y_n$. Позначимо через $\bar{Y}_1$ середнє значення цих значень Y . Одержуємо пари $(X_1, \bar{Y}_1)$. Аналогічно знаходять пари $(X_2, \bar{Y}_2), \dots, (X_k, \bar{Y}_k)$. Пари точок $(X_1, \bar{Y}_1), \dots, (X_k, \bar{Y}_k)$ зображуємо на малюнку, а потім побудовані точки з'єднуємо прямими відрізками. Отримана ламана лінія і є емпіричною лінією регресії Y по X .

Якщо в прикладі 6.2 середини інтервалів прийняти за справжні значення, які приймають ознаки X і Y у вибірці, то одержимо ряд значень:

$$X_1=145, X_2=175, X_3=205, X_4=235, X_5=265, X_6=295, X_7=325, \\ X_8=355.$$

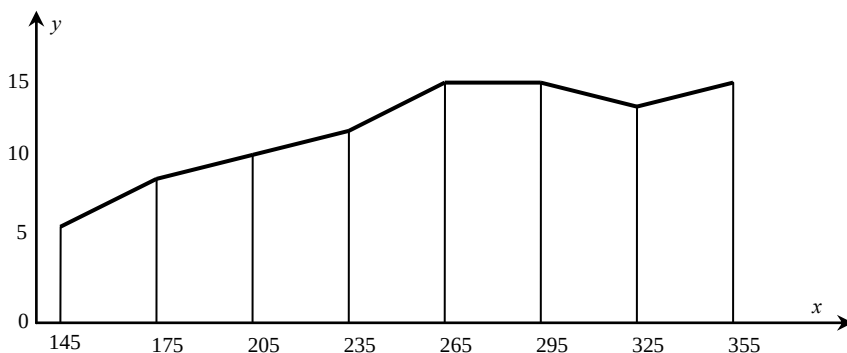
Числа $n_1, \dots, n_8$ по суті є числами з першого стовпця кореляційної таблиці 6.2. Щоб знайти $\bar{Y}_1$, урахуємо, що $X_1=145$ зустрілося три рази в парі з $Y=4$ (середина інтервалу 3 – 5) і один раз у парі з $Y=8$ (середина інтервалу 7 – 9). Звідси

$$\bar{Y}_1 = \frac{3 \cdot 4 + 1 \cdot 8}{4} = 5.$$

Аналогічно знаходимо:

$$\bar{Y}_2 = 10, \bar{Y}_3 = 13.7, \bar{Y}_4 = 14.1, \bar{Y}_5 = 16.4, \bar{Y}_6 = 16.4, \bar{Y}_7 = 16, \\ \bar{Y}_8 = 17.$$

Тепер зображуємо крапки $(X_i, \bar{Y}_i)$ і будуємо ламану.



Мал. 6.4 - Емпірична лінія регресії

Коли кількість побудованих точок досить велика, їх вдається з'єднати плавною кривою, що дає уявлення про теоретичну лінію регресії.

Варто помітити, що навіть коли залежність ознак X і Y не є лінійною, можна з певним ступенем наближення (хоча б з метою якісного аналізу) використати прямі лінії регресії.

6.7. Контрольні запитання

1. Чим кореляційна залежність відрізняється від функціональної?
2. Якою може бути кореляція за тісністю?
3. Якою може бути кореляція за формою?
4. Яка кореляція є повною, а яка частковою?
5. Яка кореляція є прямою, а яка зворотною?
6. За якою формулою обчислюється коефіцієнт лінійної кореляції?
7. Яким чином інтерпретується числове значення коефіцієнту кореляції?
8. Що таке кореляційні грати?
9. Що виражають кореляційні відносини?
10. Яка кореляція називається парціальною?
11. Які вам відомі критерії кореляційного аналізу?
12. Як побудувати лінії регресії?
13. Що можна дослідити за допомогою ліній регресії?

6.8. Задачі для самостійного опрацювання

1. Обчислити коефіцієнт лінійної кореляції між двома показниками, значення яких подано у таблиці

№з/п	1	2	3	4	5	6	7	8	9	10
x_i	125	119	98	87	100	102	97	89	104	109
y_i	48	38	34	33	45	43	40	36	41	42

Розв'язання:

Щоб розв'язати поставлену задачу, треба заповнити наступну таблицю. Перші три стовпці цієї таблиці є умовою задачі.

Підрахувавши суми по стовпцях №2 і №3, можна обчислити середні арифметичні значення обох показників:

$$\bar{x} = \frac{1030}{10} = 103; \quad \bar{y} = \frac{400}{10} = 40.$$

№	1		2		3		4
	Значення ознаки		Відхилення від середнього арифметичного		Квадрати відхилень від середнього арифметичного		Добутки відхилень $(x_i - \bar{x})(y_i - \bar{y})$
	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	
1	2	3	4	5	6	7	8
1	125	48	$125 - 103 = 22$	$48 - 40 = 8$	$(22)^2 = (125 - 103)^2 = 484$	$(8)^2 = (48 - 40)^2 = 64$	$22 \times 8 = 176$
2	119	38	$119 - 103 = 16$	$38 - 40 = -2$	$(16)^2 = (119 - 103)^2 = 256$	$(-2)^2 = (38 - 40)^2 = 4$	$16 \times (-2) = -32$
3	98	34	$98 - 103 = -5$	$34 - 40 = -6$	$(-5)^2 = (98 - 103)^2 = 25$	$(-6)^2 = (34 - 40)^2 = 36$	$(-5) \times (-6) = 30$
4	87	33	$87 - 103 = -16$	$33 - 40 = -7$	$(-16)^2 = (87 - 103)^2 = 256$	$(-7)^2 = (33 - 40)^2 = 49$	$(-16) \times (-7) = 112$
5	100	45	$100 - 103 = -3$	$45 - 40 = 5$	$(-3)^2 = (100 - 103)^2 = 9$	$(5)^2 = (45 - 40)^2 = 25$	$(-3) \times 5 = -15$
6	102	43	$102 - 103 = -1$	$43 - 40 = 3$	$(-1)^2 = (102 - 103)^2 = 1$	$(3)^2 = (43 - 40)^2 = 9$	$(-1) \times 3 = -3$
7	97	40	$97 - 103 = -6$	$40 - 40 = 0$	$(-6)^2 = (97 - 103)^2 = 36$	$(0)^2 = (40 - 40)^2 = 0$	$(-6) \times 0 = 0$
8	89	36	$89 - 103 = -14$	$36 - 40 = -4$	$(-14)^2 = (89 - 103)^2 = 196$	$(-4)^2 = (36 - 40)^2 = 16$	$(-14) \times (-4) = 56$
9	104	41	$104 - 103 = 1$	$41 - 40 = 1$	$(1)^2 = (104 - 103)^2 = 1$	$(1)^2 = (41 - 40)^2 = 1$	$1 \times 1 = 1$
10	109	42	$109 - 103 = 6$	$42 - 40 = 2$	$(6)^2 = (109 - 103)^2 = 36$	$(2)^2 = (42 - 40)^2 = 4$	$6 \times 2 = 12$
Σ	1030	400			1300	208	337

Елементи стовпця №6 є квадратами елементів стовпця №4. Тому ці стовпці можна заповнювати одночасно. Щоб отримати елементи стовпця №4, треба із відповідних елементів стовпця №2 відняти одне й те ж саме число $\bar{x}=103$. Підводячи ці результати до другого ступеня, можна відразу отримати елементи стовпця №6.

Елементи стовпця №7 є квадратами елементів стовпця №5. Тому ці стовпці також можна заповнювати одночасно. Щоб отримати елементи стовпця №5, треба із відповідних елементів стовпця №3 відняти одне й те ж саме число $\bar{y}=40$. Підводячи ці результати до другого ступеня, можна відразу отримати елементи стовпця №7.

Елементи стовпця №8 є добутками відповідних елементів стовпців №4 і №5.

Елементи стовпців №4 і №5 є проміжними результатами. Вони потрібні лише для подальших обчислень. Тому суми елементів по цих стовпцях обчислювати не треба. Остаточними результатами є елементи стовпців №6, №7 і №8. Суми саме по цих стовпцях потрібні для обчислення коефіцієнта лінійної кореляції. Таким чином, маємо:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{337}{\sqrt{1300 \cdot 208}} = \frac{337}{520} = 0,65.$$

Слід зазначити, що напрямок лінійного кореляційного зв'язку між досліджуваними показниками визначає знак суми саме по стовпцю №8. Елементи стовпців №6 і №7 є парними ступенями елементів стовпців №4 і №5. Тому незалежно від того, який знак мають елементи цих проміжних стовпців, елементи стовпців №6 і №7 завжди є додатними числами. Тому і суми елементів по цих стовпцях теж завжди є додатними числами. На це слід звернути увагу, щоб при обчисленнях без застосування ЕОМ не припуститися механічної помилки. З проведених обчислень бачимо, що $0,65 > 0$; $0,65 \in [0,3; 0,7)$.

Висновок: між досліджуваними ознаками є прямий лінійний зв'язок середньої сили.

2. Обчислити коефіцієнт лінійної кореляції між двома показниками, значення яких подано у таблиці

№з/п	1	2	3	4	5
x_i	13	15	10	23	25
y_i	101	108	93	80	87

Розв'язання:

Щоб розв'язати поставлену задачу, треба заповнити наступну таблицю. Перші три стовпці цієї таблиці є умовою задачі.

№	1		2		3		4
	Значення ознаки		Відхилення від середнього ариф- метичного		Квадрати відхилень від серед- нього арифметичного		Добутки відхилень
	x_i	y_i	$\bar{x} - x_i - \bar{X}$	$\bar{y} - y_i - \bar{Y}$	$(\bar{x} - x_i - \bar{X})^2$	$(\bar{y} - y_i - \bar{Y})^2$	$(\bar{x} - x_i - \bar{X})(\bar{y} - y_i - \bar{Y})$
1	2	3	4	5	6	7	8
1	13	101	$13 - 17 = -4$	$101 - 94 = 7$	$(-4)^2 = (13 - 17)^2 = 16$	$(7)^2 = (101 - 94)^2 = 49$	$(-4) \times 7 = -28$
2	14	108	$14 - 17 = -3$	$108 - 94 = 14$	$(-3)^2 = (14 - 17)^2 = 9$	$(14)^2 = (108 - 94)^2 = 196$	$(-3) \times 14 = -42$
3	10	93	$10 - 17 = -7$	$93 - 94 = -1$	$(-7)^2 = (10 - 17)^2 = 49$	$(-1)^2 = (93 - 94)^2 = 1$	$(-7) \times (-1) = 7$
4	23	81	$23 - 17 = 6$	$81 - 94 = -13$	$(6)^2 = (23 - 17)^2 = 36$	$(-13)^2 = (81 - 94)^2 = 169$	$6 \times (-13) = -78$
5	25	87	$25 - 17 = 8$	$87 - 94 = -7$	$(8)^2 = (25 - 17)^2 = 64$	$(-7)^2 = (87 - 94)^2 = 49$	$8 \times (-7) = -56$
Σ	85	470			174	464	-197

Підрахувавши суми по стовпцях №2 і №3, можна обчислити середні арифметичні значення обох показників:

$$\bar{x} = \frac{85}{5} = 17; \quad \bar{y} = \frac{470}{5} = 94.$$

Елементи стовпця №6 є квадратами елементів стовпця №4. Тому ці стовпці можна заповнювати одночасно. Щоб отримати елементи стовпця №4, треба із відповідних елементів стовпця

№2 відняти одне й те ж саме число $\bar{x}=17$. Підводячи ці результати до другого ступеня, можна відразу отримати елементи стовпця №6.

Елементи стовпця №7 є квадратами елементів стовпця №5. Тому ці стовпці також можна заповнювати одночасно. Щоб отримати елементи стовпця №5, треба із відповідних елементів стовпця №3 відняти одне й те ж саме число $\bar{y}=94$. Підводячи ці результати до другого ступеня, можна відразу отримати елементи стовпця №7.

Елементи стовпця №8 є добутками відповідних елементів стовпців №4 і №5.

Остаточними результатами є елементи стовпців №6, №7 і №8. Суми саме по цих стовпцях використаємо для обчислення коефіцієнта лінійної кореляції. Таким чином, маємо:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{-197}{\sqrt{174 \cdot 464}} = \frac{-197}{284,14} = -0,69.$$

З проведених обчислень бачимо, що $(-0,69) < 0$;
 $0,69 \in [0,3; 0,7)$.

Висновок: між досліджуваними ознаками є зворотній лінійний зв'язок середньої сили.

6.9. Контрольні завдання

1. Обчислити коефіцієнт лінійної кореляції між двома показниками, значення яких подано у таблиці

№з/п	1	2	3	4	5	6	7	8	9	10
x_i	1	15	4	38	25	12	3	6	25	10
y_i	110	52	87	24	60	43	98	95	71	82

2. Обчислити коефіцієнт лінійної кореляції між двома показниками, значення яких подано у таблиці

№з/п	1	2	3	4	5	6	7	8	9	10
x_i	14	2	39	42	18	33	26	42	18	30
y_i	- 28	- 6	- 53	- 31	- 50	- 52	- 39	- 60	- 21	- 48

3. Обчислити коефіцієнт лінійної кореляції між двома показниками, значення яких подано у таблиці

№з/п	1	2	3	4	5	6	7	8	9	10
x_i	- 45	- 18	- 33	- 55	- 67	- 27	- 40	- 41	- 29	- 19
y_i	6,2	12,5	8,1	5,1	3,9	10,1	7,6	7,4	8,9	10,4

7. ВАРІАНТИ ЗАВДАНЬ МОДУЛЬНОГО КОНТРОЛЮ

7.1. Модуль 1

Варіант № 1

1. При дворазовому киданні трьох гральних кісток знайти ймовірність випадання рівно 8 очок.

2. В урні міститься 10 шариків. З них 5 білих, решта чорні. Із цієї урни виймаються 7 шариків. Знайти ймовірність того, що 4 з них будуть білими.

3. В ході експерименту було отримано наступні значення елементів вибірки:

№ п/п	x_i		№ п/п	x_i
1	5		16	64
2	12		17	22
3	46		18	15
4	48		19	69
5	31		20	26
6	30		21	34
7	56		22	58
8	80		23	71
9	74		24	9
10	63		25	10
11	25		26	13
12	85		27	19
13	41		28	79
14	32		29	16
15	17		30	48

а) провести групування при $k=10$;

б) обчислити $\bar{x}$ і S^2 ;

в) перевірити гіпотезу H_0 =«Дану вибірку взято із генеральної сукупності, що розподілена за нормальним законом».

Варіант № 2

1. При дворазовому киданні трьох гральних кісток знайти ймовірність випадання рівно 9 очок.
2. В урні міститься 15 шариків. 3 них 7 білих, решта чорні. Із цієї урни виймаються 10 шариків. Знайти ймовірність того, що 3 з них будуть білими.
3. В ході експерименту було отримано наступні значення елементів вибірки:

№ п/п	x_i		№ п/п	x_i
1	10		16	44
2	84		17	32
3	62		18	13
4	31		19	18
5	58		20	20
6	54		21	41
7	16		22	39
8	5		23	7
9	29		24	11
10	85		25	23
11	36		26	28
12	74		27	37
13	33		28	73
14	42		29	82
15	16		30	60

- а) провести групування при $k=10$;
- б) обчислити $\bar{x}$ і S^2 ;
- в) перевірити гіпотезу H_0 =«Дану вибірку взято із генеральної сукупності, що розподілена за нормальним законом».

Варіант № 3

1. При дворазовому киданні двох гральних кісток знайти ймовірність випадання рівно 6 очок.

2. В урні міститься 13 шариків. З них 9 білих, решта чорні. Із цієї урни виймаються 8 шариків. Знайти ймовірність того, що 5 з них будуть білими.

3. В ході експерименту було отримано наступні значення елементів вибірки:

№ п/п	x_i		№ п/п	x_i
1	13		16	6
2	22		17	11
3	25		18	34
4	64		19	79
5	68		20	66
6	76		21	58
7	81		22	27
8	54		23	19
9	12		24	14
10	5		25	37
11	33		26	46
12	23		27	41
13	21		28	62
14	84		29	17
15	85		30	7

а) провести групування при $k=10$;

б) обчислити $\bar{x}$ і S^2 ;

в) перевірити гіпотезу H_0 =«Дану вибірку взято із генеральної сукупності, що розподілена за нормальним законом».

Варіант №4

1. При дворазовому киданні двох гральних кісток знайти ймовірність випадання рівно 7 очок.

2. В урні міститься 9 шариків. З них 5 білих, решта чорні. Із цієї урни виймаються 7 шариків. Знайти ймовірність того, що 4 з них будуть білими.

3. В ході експерименту було отримано наступні значення елементів вибірки:

№ п/п	x_i		№ п/п	x_i
1	8		16	49
2	41		17	55
3	46		18	54
4	21		19	70
5	35		20	12
6	39		21	64
7	27		22	59
8	19		23	20
9	47		24	9
10	54		25	40
11	11		26	23
12	5		27	18
13	83		28	30
14	80		29	33
15	44		30	85

- а) провести групування при $k=10$;
- б) обчислити $\bar{x}$ і S^2 ;
- в) перевірити гіпотезу H_0 =«Дану вибірку взято із генеральної сукупності, що розподілена за нормальним законом».

Варіант № 5

1. При дворазовому киданні двох гральних кісток знайти ймовірність випадання рівно 8 очок.
2. В урні міститься 19 шариків. З них 10 білих, решта чорні. Із цієї урни виймаються 15 шариків. Знайти ймовірність того, що 8 з них будуть білими.
3. В ході експерименту було отримано наступні значення елементів вибірки:

№ п/п	x_i		№ п/п	x_i
1	40		16	29
2	5		17	33
3	13		18	34
4	12		19	52
5	16		20	50

№ п/п	x_i		№ п/п	x_i
6	19		21	60
7	85		22	64
8	83		23	67
9	64		24	35
10	21		25	72
11	17		26	8
12	25		27	22
13	34		28	15
14	67		29	9
15	60		30	80

- а) провести групування при $k=10$;
- б) обчислити $\bar{x}$ і S^2 ;
- в) перевірити гіпотезу H_0 =«Дану вибірку взято із генеральної сукупності, що розподілена за нормальним законом».

Варіант №6

1. При дворазовому киданні чотирьох гральних кісток знайти ймовірність випадання рівно 9 очок.
2. В урні міститься 14 шариків. З них 10 білих, решта чорні. Із цієї урни виймаються 7 шариків. Знайти ймовірність того, що 4 з них будуть білими.
3. В ході експерименту було отримано наступні значення елементів вибірки:

№ п/п	x_i		№ п/п	x_i
1	11		16	49
2	13		17	63
3	64		18	25
4	69		19	38
5	85		20	44
6	20		21	52
7	47		22	58
8	62		23	61
9	19		24	84
10	36		25	33

№ п/п	x_i		№ п/п	x_i
11	5		26	65
12	21		27	44
13	66		28	18
14	9		29	21
15	41		30	50

- а) провести групування при $k=10$;
 б) обчислити $\bar{x}$ і S^2 ;
 в) перевірити гіпотезу H_0 =«Дану вибірку взято із генеральної сукупності, що розподілена за нормальним законом».

Варіант № 7

1. При дворазовому киданні чотирьох гральних кісток знайти ймовірність випадання рівно 10 очок.
2. В урні міститься 16 шариків. З них 9 білих, решта чорні. Із цієї урни виймаються 5 шариків. Знайти ймовірність того, що 2 з них будуть білими.
3. В ході експерименту було отримано наступні значення елементів вибірки:

№ п/п	x_i		№ п/п	x_i
1	12		16	65
2	51		17	21
3	81		18	78
4	64		19	72
5	77		20	31
6	45		21	28
7	36		22	55
8	34		23	51
9	85		24	42
10	26		25	39
11	19		26	81
12	5		27	56
13	44		28	53
14	82		29	21
15	33		30	28

- а) провести групування при $k=10$;
 б) обчислити $\bar{x}$ і S^2 ;
 в) перевірити гіпотезу H_0 =«Дану вибірку взято із генеральної сукупності, що розподілена за нормальним законом».

Варіант № 8

1. При триразовому киданні двох гральних кісток знайти ймовірність випадання рівно 7 очок.
 2. В урні міститься 8 шариків. З них 3 білих, решта чорні. Із цієї урни виймаються 6 шариків. Знайти ймовірність того, що 2 з них будуть білими.
 3. В ході експерименту було отримано наступні значення елементів вибірки:

№ п/п	x_i		№ п/п	x_i
1	33		16	83
2	23		17	66
3	26		18	51
4	84		19	37
5	72		20	12
6	69		21	66
7	36		22	52
8	22		23	14
9	15		24	17
10	85		25	19
11	5		26	21
12	22		27	20
13	17		28	74
14	84		29	81
15	63		30	16

- а) провести групування при $k=10$;
 б) обчислити $\bar{x}$ і S^2 ;
 в) перевірити гіпотезу H_0 =«Дану вибірку взято із генеральної сукупності, що розподілена за нормальним законом».

Варіант № 9

1. При чотириразовому киданні двох гральних кісток знайти ймовірність випадання рівно 9 очок.

2. В урні міститься 12 шариків. З них 7 білих, решта чорні. Із цієї урни виймаються 8 шариків. Знайти ймовірність того, що 6 з них будуть білими.

3. В ході експерименту було отримано наступні значення елементів вибірки:

№ п/п	x_i		№ п/п	x_i
1	14		16	19
2	12		17	82
3	55		18	41
4	84		19	66
5	85		20	32
6	23		21	35
7	39		22	57
8	73		23	74
9	64		24	75
10	25		25	65
11	28		26	41
12	85		27	49
13	16		28	82
14	5		29	33
15	54		30	47

а) провести групування при $k=10$;

б) обчислити $\bar{x}$ і S^2 ;

в) перевірити гіпотезу $H_0 = \text{«Дану вибірку взято із генеральної сукупності, що розподілена за нормальним законом»}$.

Варіант №10

1. При чотириразовому киданні двох гральних кісток знайти ймовірність випадання рівно 10 очок.

2. В урні міститься 11 шариків. З них 5 білих, решта чорні. Із цієї урни виймаються 8 шариків. Знайти ймовірність того, що 3 з них будуть білими.

3. В ході експерименту було отримано наступні значення елементів вибірки:

№ п/п	x_i		№ п/п	x_i
1	84		16	85
2	23		17	24
3	8		18	13
4	10		19	35
5	36		20	25
6	5		21	84
7	14		22	69
8	79		23	66
9	82		24	21
10	63		25	25
11	66		26	23
12	54		27	22
13	52		28	45
14	18		29	44
15	9		30	32

а) провести групування при $k=10$;

б) обчислити $\bar{x}$ і S^2 ;

в) перевірити гіпотезу H_0 =«Дану вибірку взято із генеральної сукупності, що розподілена за нормальним законом».

Варіант № 11

1. При чотириразовому киданні трьох гральних кісток знайти ймовірність випадання рівно 13 очок.

2. В урні міститься 7 шариків. З них 4 білих, решта чорні. Із цієї урни виймаються 5 шариків. Знайти ймовірність того, що 3 з них будуть білими.

3. В ході експерименту було отримано наступні значення елементів вибірки:

№ п/п	x_i		№ п/п	x_i
1	47		16	24
2	69		17	10
3	82		18	28
4	85		19	37
5	54		20	43
6	55		21	12
7	36		22	69
8	31		23	84
9	25		24	55
10	19		25	19
11	82		26	22
12	5		27	7
13	8		28	18
14	33		29	34
15	11		30	12

- а) провести групування при $k=10$;
- б) обчислити $\bar{x}$ і S^2 ;
- в) перевірити гіпотезу H_0 =«Дану вибірку взято із генеральної сукупності, що розподілена за нормальним законом».

Варіант №12

1. При чотириразовому киданні трьох гральних кісток знайти ймовірність випадання рівно 14 очок.
2. В урні міститься 12 шариків. З них 6 білих, решта чорні. Із цієї урни виймаються 5 шариків. Знайти ймовірність того, що 2 з них будуть білими.
3. В ході експерименту було отримано наступні значення елементів вибірки:

№ п/п	x_i		№ п/п	x_i
1	14		16	5
2	12		17	17
3	58		18	64
4	31		19	21
5	85		20	16

№ п/п	x_i		№ п/п	x_i
6	52		21	73
7	82		22	77
8	9		23	54
9	41		24	13
10	20		25	25
11	34		26	9
12	13		27	10
13	59		28	11
14	82		29	13
15	43		30	17

- а) провести групування при $k=10$;
б) обчислити $\bar{x}$ і S^2 ;
в) перевірити гіпотезу H_0 =«Дану вибірку взято із генеральної сукупності, що розподілена за нормальним законом».

Варіант №13

- При триразовому киданні трьох гральних кісток знайти ймовірність випадання рівно 10 очок.
- В урні міститься 14 шариків. З них 7 білих, решта чорні. Із цієї урни виймаються 9 шариків. Знайти ймовірність того, що 5 з них будуть білими.
- В ході експерименту було отримано наступні значення елементів вибірки:

№ п/п	x_i		№ п/п	x_i
1	22		16	41
2	31		17	65
3	56		18	27
4	58		19	41
5	9		20	12
6	14		21	10
7	73		22	8
8	5		23	7
9	24		24	55
10	68		25	64

№ п/п	x_i		№ п/п	x_i
11	71		26	32
12	80		27	48
13	85		28	77
14	2		29	69
15	79		30	50

- а) провести групування при $k=10$;
 б) обчислити $\bar{x}$ і S^2 ;
 в) перевірити гіпотезу H_0 =«Дану вибірку взято із генеральної сукупності, що розподілена за нормальним законом».

Варіант № 14

1. При триразовому киданні трьох гральних кісток знайти ймовірність випадання рівно 11 очок.
2. В урні міститься 19 шариків. З них 11 білих, решта чорні. Із цієї урни виймаються 13 шариків. Знайти ймовірність того, що 8 з них будуть білими.
3. В ході експерименту було отримано наступні значення елементів вибірки:

№ п/п	x_i		№ п/п	x_i
1	80		16	38
2	33		17	31
3	24		18	24
4	26		19	29
5	19		20	85
6	8		21	17
7	7		22	33
8	5		23	10
9	10		24	6
10	16		25	14
11	83		26	52
12	77		27	58
13	62		28	61
14	51		29	37
15	46		30	81

- а) провести групування при $k=10$;
 б) обчислити $\bar{x}$ і S^2 ;
 в) перевірити гіпотезу $H_0 = \text{«Дану вибірку взято із генеральної сукупності, що розподілена за нормальним законом»}$.

Варіант № 15

1. При триразовому киданні чотирьох гральних кісток знайти ймовірність випадання рівно 13 очок.

2. В урні міститься 17 шариків. З них 10 білих, решта чорні. Із цієї урни виймаються 11 шариків. Знайти ймовірність того, що 6 з них будуть білими.

3. В ході експерименту було отримано наступні значення елементів вибірки:

№ п/п	x_i		№ п/п	x_i
1	16		16	42
2	22		17	48
3	14		18	35
4	84		19	41
5	36		20	81
6	34		21	63
7	50		22	26
8	68		23	18
9	79		24	44
10	85		25	71
11	11		26	22
12	9		27	34
13	5		28	56
14	17		29	7
15	64		30	6

- а) провести групування при $k=10$;
 б) обчислити $\bar{x}$ і S^2 ;
 в) перевірити гіпотезу $H_0 = \text{«Дану вибірку взято із генеральної сукупності, що розподілена за нормальним законом»}$.

Варіант № 16

1. При триразовому киданні чотирьох гральних кісток знайти ймовірність випадання рівно 14 очок.
2. В урні міститься 12 шариків. З них 7 білих, решта чорні. Із цієї урни виймаються 8 шариків. Знайти ймовірність того, що 5 з них будуть білими.
3. В ході експерименту було отримано наступні значення елементів вибірки:

№ п/п	x_i		№ п/п	x_i
1	64		16	85
2	31		17	73
3	29		18	42
4	5		19	54
5	49		20	64
6	83		21	84
7	17		22	22
8	16		23	16
9	48		24	17
10	55		25	31
11	63		26	14
12	31		27	16
13	22		28	22
14	18		29	17
15	44		30	19

- а) провести групування при $k=10$;
- б) обчислити $\bar{x}$ і S^2 ;
- в) перевірити гіпотезу H_0 =«Дану вибірку взято із генеральної сукупності, що розподілена за нормальним законом».

7.2. Модуль 2

Варіант № 1

1. Визначити, чи містять дані вибірки однорідний матеріал і чи можна їх об'єднати в одну.

Інтервали	Частоти для 1-ї вибірки	Частоти для 2-ї вибірки
5-10	8	4
10-15	14	12
15-20	16	18
20-25	24	21
25-30	25	32
30-35	36	35
35-40	20	33
40-45	14	18
45-50	10	17
50-55	3	4

2. В ході експерименту було отримано наступні значення випадкової величини:

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	12	13	55	34	23	54	67	22	45	75	33	31	18	23	28

а) побудувати довірчий інтервал для генерального середнього $\bar{x}$;

б) оцінити репрезентативність вибірки.

3. Значення одного й того ж показника було визначено за двома різними вибірками і згруповано в інтервали. По результатах групування було побудовано дві таблиці.

Вибірка А

Інтервали	Частоти (n_i)
5-15	2
15-25	4
25-35	15
35-45	17
45-55	23
55-65	31
65-75	45
75-85	43
85-95	40
95-105	32
105-115	28
115-125	26
125-135	16
135-145	9
145-155	7

Вибірка В

Інтервали	Частоти (n_j)
10-30	3
30-50	8
50-70	18
70-90	25
90-110	30
110-130	29
130-150	19
150-170	6
170-190	4
190-210	1

Чи суттєва різниця в середніх значеннях цих двох вибірок?

4. Вимірювання двох показників подано в таблиці.

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	11	23	43	45	13	24	9	16	32	14	5	16	22	13	3
Y	21	26	54	67	22	28	15	20	40	25	10	22	18	32	7

Визначити ступінь і напрямок лінійної кореляційної залежності між цими двома показниками.

Варіант № 2

1. Визначити, чи містять дані вибірки однорідний матеріал і чи можна їх об'єднати в одну.

Інтервали	Частоти для 1-ї вибірки	Частоти для 2-ї вибірки
5-10	23	18
10-15	32	21
15-20	43	27
20-25	44	35
25-30	42	38
30-35	49	40
35-40	39	35
40-45	29	34
45-50	20	25
50-55	16	15

2. В ході експерименту було отримано наступні значення випадкової величини:

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	11	13	43	23	47	68	54	33	76	13	42	54	11	7	40

а) побудувати довірчий інтервал для генерального середнього $\bar{x}$;

б) оцінити репрезентативність вибірки.

3. Значення одного й того ж показника було визначено за двома різними вибірками і згруповано в інтервали. По результатах групування було побудовано дві таблиці.

Вибірка А

Інтервали	Частоти (n_i)
5-15	12
15-25	14
25-35	17
35-45	22
45-55	34
55-65	37
65-75	46
75-85	44

Вибірка В

Інтервали	Частоти (n_i)
10-30	10
30-50	13
50-70	15
70-90	22
90-110	29
110-130	31
130-150	28
150-170	18

Вибірка А (продовження)	
85-95	40
95-105	31
105-115	25
115-125	22
125-135	18
135-145	13
145-155	6

Вибірка В (продовження)	
170-190	10
190-210	5

Чи суттєва різниця в середніх значеннях цих двох вибірок?

4. Вимірювання двох показників подано в таблиці.

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
<i>X</i>	5	8	19	45	76	34	56	23	10	4	38	41	30	22	65
<i>Y</i>	12	17	25	51	89	42	65	30	4	10	45	57	47	36	80

Визначити ступінь і напрямок лінійної кореляційної залежності між цими двома показниками.

Варіант № 3

1. Визначити, чи містять дані вибірки однорідний матеріал і чи можна їх об'єднати в одну.

Інтервали	Частоти для 1-ї вибірки	Частоти для 2-ї вибірки
5-10	2	3
10-15	5	4
15-20	7	5
20-25	12	10
25-30	15	14
30-35	13	12
35-40	10	11
40-45	8	7
45-50	6	5
50-55	1	2

2. В ході експерименту було отримано наступні значення випадкової величини:

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	2	15	32	66	3	45	81	23	54	12	9	17	37	34	20

а) побудувати довірчий інтервал для генерального середнього $\bar{x}$;

б) оцінити репрезентативність вибірки.

3. Значення одного й того ж показника було визначено за двома різними вибірками і згруповано в інтервали. По результатах групування було побудовано дві таблиці.

Вибірка А

Інтервали	Частоти (n_i)
5-15	3
15-25	4
25-35	6
35-45	8
45-55	10
55-65	11
65-75	11
75-85	15
85-95	18
95-105	16
105-115	15
115-125	9
125-135	7
135-145	5
145-155	1

Вибірка В

Інтервали	Частоти (n_j)
10-30	2
30-50	5
50-70	8
70-90	10
90-110	19
110-130	16
130-150	13
150-170	11
170-190	4
190-210	1

Чи суттєва різниця в середніх значеннях цих двох вибірок?

4. Вимірювання двох показників подано в таблиці.

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	3	65	44	13	87	54	19	23	9	14	32	66	15	71	43
Y	90	12	52	71	5	9	67	56	87	68	40	31	56	80	33

Визначити ступінь і напрямок лінійної кореляційної залежності між цими двома показниками.

Варіант № 4

1. Визначити, чи містять дані вибірки однорідний матеріал і чи можна їх об'єднати в одну.

Інтервали	Частоти для 1-ї вибірки	Частоти для 2-ї вибірки
5-10	35	43
10-15	37	47
15-20	48	54
20-25	48	56
25-30	76	60
30-35	59	59
35-40	55	57
40-45	41	37
45-50	31	29
50-55	26	15

2. В ході експерименту було отримано наступні значення випадкової величини:

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	12	36	52	14	79	551	33	21	16	4	8	19	65	33	87

а) побудувати довірчий інтервал для генерального середнього $\bar{x}$;

б) оцінити репрезентативність вибірки.

3. Значення одного й того ж показника було визначено за двома різними вибірками і згруповано в інтервали. По результатах групування було побудовано дві таблиці.

Вибірка А

Інтервали	Частоти (n_i)
5-15	44
15-25	49
25-35	56
35-45	59
45-55	76
55-65	89
65-75	91
75-85	93
85-95	85
95-105	71
105-115	63
115-125	55
125-135	47
135-145	36
145-155	28

Вибірка В

Інтервали	Частоти (n_i)
10-30	58
30-50	63
50-70	74
70-90	85
90-110	97
110-130	88
130-150	71
150-170	52
170-190	47
190-210	34

Чи суттєва різниця в середніх значеннях цих двох вибірок?

4. Вимірювання двох показників подано в таблиці.

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	15	21	36	95	14	26	32	47	64	82	41	39	77	82	9
Y	21	26	54	67	22	28	15	20	40	25	10	22	18	32	7

Визначити ступінь і напрямок лінійної кореляційної залежності між цими двома показниками.

Варіант № 5

1. Визначити, чи містять дані вибірки однорідний матеріал і чи можна їх об'єднати в одну.

Інтервали	Частоти для 1-ї вибірки	Частоти для 2-ї вибірки
5-10	6	4
10-15	15	12
15-20	16	18
20-25	19	21
25-30	29	32
30-35	31	35
35-40	30	33
40-45	30	18
45-50	18	17
50-55	7	4

2. В ході експерименту було отримано наступні значення випадкової величини:

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
xi	14	18	52	64	87	51	24	55	34	78	95	94	68	87	36

а) побудувати довірчий інтервал для генерального середнього $\bar{x}$;

б) оцінити репрезентативність вибірки.

3. Значення одного й того ж показника було визначено за двома різними вибірками і згруповано в інтервали. По результатах групування було побудовано дві таблиці.

Вибірка А

Інтервали	Частоти (n_i)
5-15	1
15-25	1
25-35	2
35-45	3
45-55	3
55-65	3
65-75	8
75-85	15
85-95	35
95-105	33
105-115	27
115-125	14
125-135	10
135-145	9
145-155	8

Вибірка В

Інтервали	Частоти (n_i)
10-30	2
30-50	3
50-70	9
70-90	15
90-110	34
110-130	21
130-150	19
150-170	11
170-190	6
190-210	3

Чи суттєва різниця в середніх значеннях цих двох вибірок?

4. Вимірювання двох показників подано в таблиці.

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	11	23	43	45	13	24	9	16	32	14	5	16	22	13	3
Y	18	36	35	14	78	91	25	34	71	12	8	9	3	14	75

Визначити ступінь і напрямок лінійної кореляційної залежності між цими двома показниками.

Варіант № 6

1. Визначити, чи містять дані вибірки однорідний матеріал і чи можна їх об'єднати в одну.

Інтервали	Частоти для 1-ї вибірки	Частоти для 2-ї вибірки
5-10	112	98
10-15	126	114
15-20	132	146
20-25	148	167
25-30	189	210
30-35	17	159
35-40	165	133
40-45	110	128
45-50	91	89
50-55	60	71

2. В ході експерименту було отримано наступні значення випадкової величини:

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	125	138	98	71	159	180	87	54	68	122	147	36	154	64	79

а) побудувати довірчий інтервал для генерального середнього $\bar{x}$;

б) оцінити репрезентативність вибірки.

3. Значення одного й того ж показника було визначено за двома різними вибірками і згруповано в інтервали. По результатах групування було побудовано дві таблиці.

Вибірка А	
Інтервали	Частоти (n_i)
5-15	210
15-25	215
25-35	246
35-45	258
45-55	259
55-65	293
65-75	250
75-85	243

Вибірка В	
Інтервали	Частоти (n_i)
10-30	197
30-50	198
50-70	204
70-90	209
90-110	219
110-130	243
130-150	235
150-170	208

Вибірка А (продовження)	
85-95	236
95-105	208
105-115	198
115-125	197
125-135	183
135-145	179
145-155	168

Вибірка В (продовження)	
170-190	189
190-210	177

Чи суттєва різниця в середніх значеннях цих двох вибірок?

4. Вимірювання двох показників подано в таблиці.

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	48	89	65	18	47	32	54	11	89	17	65	98	14	37	27
Y	156	187	113	146	112	110	98	159	102	119	146	100	147	154	99

Визначити ступінь і напрямок лінійної кореляційної залежності між цими двома показниками.

Варіант № 7

1. Визначити, чи містять дані вибірки однорідний матеріал і чи можна їх об'єднати в одну.

Інтервали	Частоти для 1-ї вибірки	Частоти для 2-ї вибірки
5-10	23	39
10-15	29	41
15-20	45	45
20-25	68	65
25-30	99	77
30-35	87	69
35-40	84	54
40-45	76	37
45-50	61	36
50-55	48	29

2. В ході експерименту було отримано наступні значення випадкової величини:

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	34	12	89	74	64	13	11	8	95	41	53	64	71	25	84

а) побудувати довірчий інтервал для генерального середнього $\bar{x}$;

б) оцінити репрезентативність вибірки.

3. Значення одного й того ж показника було визначено за двома різними вибірками і згруповано в інтервали. По результатах групування було побудовано дві таблиці.

Вибірка А

Інтервали	Частоти (n_i)
5-15	38
15-25	39
25-35	46
35-45	49
45-55	51
55-65	53
65-75	68
75-85	79
85-95	74
95-105	62
105-115	35
115-125	29
125-135	11
135-145	10
145-155	10

Вибірка В

Інтервали	Частоти (n_i)
10-30	25
30-50	29
50-70	34
70-90	48
90-110	56
110-130	57
130-150	56
150-170	41
170-190	32
190-210	19

Чи суттєва різниця в середніх значеннях цих двох вибірок?

4. Вимірювання двох показників подано в таблиці.

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	123	156	159	161	68	179	182	186	188	193	197	204	219	221	241
Y	98	97	84	83	64	61	52	48	44	37	33	28	21	14	9

Визначити ступінь і напрямок лінійної кореляційної залежності між цими двома показниками.

Варіант № 8

1. Визначити, чи містять дані вибірки однорідний матеріал і чи можна їх об'єднати в одну.

Інтервали	Частоти для 1-ї вибірки	Частоти для 2-ї вибірки
5-10	22	21
10-15	23	26
15-20	23	29
20-25	23	31
25-30	28	35
30-35	23	33
35-40	22	33
40-45	20	33
45-50	16	21
50-55	13	17

2. В ході експерименту було отримано наступні значення випадкової величини:

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	47	46	45	48	64	52	51	89	32	18	53	64	69	19	22

а) побудувати довірчий інтервал для генерального середнього $\bar{x}$;

б) оцінити репрезентативність вибірки.

3. Значення одного й того ж показника було визначено за двома різними вибірками і згруповано в інтервали. По результатах групування було побудовано дві таблиці.

Вибірка А

Інтервали	Частоти (n_i)
5-15	11
15-25	11
25-35	11
35-45	19
45-55	24
55-65	26
65-75	38
75-85	41
85-95	40
95-105	40
105-115	32
115-125	27
125-135	24
135-145	24
145-155	10

Вибірка В

Інтервали	Частоти (n_i)
10-30	15
30-50	19
50-70	22
70-90	24
90-110	27
110-130	45
130-150	40
150-170	37
170-190	14
190-210	13

Чи суттєва різниця в середніх значеннях цих двох вибірок?

4. Вимірювання двох показників подано в таблиці.

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	119	118	110	103	98	97	65	61	52	48	44	32	25	12	8
Y	12	15	19	28	33	46	49	52	69	71	77	84	96	99	117

Визначити ступінь і напрямок лінійної кореляційної залежності між цими двома показниками.

Варіант № 9

1. Визначити, чи містять дані вибірки однорідний матеріал і чи можна їх об'єднати в одну.

Інтервали	Частоти для 1-ї вибірки	Частоти для 2-ї вибірки
5-10	12	11
10-15	13	19
15-20	16	25
20-25	24	27
25-30	26	32
30-35	29	39
35-40	13	21
40-45	10	21
45-50	9	20
50-55	7	9

2. В ході експерименту було отримано наступні значення випадкової величини:

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	115	124	136	187	112	98	97	134	147	162	103	87	158	166	131

а) побудувати довірчий інтервал для генерального середнього $\bar{x}$;

б) оцінити репрезентативність вибірки.

3. Значення одного й того ж показника було визначено за двома різними вибірками і згруповано в інтервали. По результатах групування було побудовано дві таблиці.

Вибірка А	
Інтервали	Частоти (n_i)
5-15	41
15-25	45
25-35	46
35-45	49
45-55	56
55-65	65
65-75	69
75-85	71
85-95	79

Вибірка В	
Інтервали	Частоти (n_i)
10-30	32
30-50	33
50-70	38
70-90	39
90-110	41
110-130	44
130-150	37
150-170	35
170-190	33

Вибірка А (продовження)	
95-105	69
105-115	51
115-125	44
125-135	31
135-145	28
145-155	17

Вибірка В (продовження)	
190-210	29

Чи суттєва різниця в середніх значеннях цих двох вибірок?

4. Вимірювання двох показників подано в таблиці.

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	148	12	123	108	146	179	54	62	38	37	1	16	69	72	89
Y	11	13	16	25	34	10	8	26	31	17	10	21	23	34	9

Визначити ступінь і напрямок лінійної кореляційної залежності між цими двома показниками.

Варіант № 10

1. Визначити, чи містять дані вибірки однорідний матеріал і чи можна їх об'єднати в одну.

Інтервали	Частоти для 1-ї вибірки	Частоти для 2-ї вибірки
5-10	1	2
10-15	2	3
15-20	2	4
20-25	2	6
25-30	8	6
30-35	10	15
35-40	9	14
40-45	5	13
45-50	4	8
50-55	3	7

2. В ході експерименту було отримано наступні значення випадкової величини:

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	21	32	33	12	15	14	8	7	36	41	20	30	4	24	3

а) побудувати довірчий інтервал для генерального середнього $\bar{x}$;

б) оцінити репрезентативність вибірки.

3. Значення одного й того ж показника було визначено за двома різними вибірками і згруповано в інтервали. По результатах групування було побудовано дві таблиці.

Вибірка А

Інтервали	Частоти (n_i)
5-15	3
15-25	3
25-35	3
35-45	3
45-55	5
55-65	5
65-75	8
75-85	10
85-95	9
95-105	9
105-115	9
115-125	4
125-135	2
135-145	2
145-155	2

Вибірка В

Інтервали	Частоти (n_j)
10-30	1
30-50	9
50-70	6
70-90	5
90-110	4
110-130	3
130-150	3
150-170	3
170-190	3
190-210	2

Чи суттєва різниця в середніх значеннях цих двох вибірок?

4. Вимірювання двох показників подано в таблиці.

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	89	85	86	47	65	62	74	67	80	90	99	45	63	50	81
Y	123	128	129	132	138	145	105	109	124	157	133	163	112	117	100

Визначити ступінь і напрямок лінійної кореляційної залежності між цими двома показниками.

Варіант № 11

1. Визначити, чи містять дані вибірки однорідний матеріал і чи можна їх об'єднати в одну.

Інтервали	Частоти для 1-ї вибірки	Частоти для 2-ї вибірки
5-10	111	97
10-15	135	96
15-20	132	94
20-25	130	92
25-30	119	100
30-35	117	106
35-40	110	108
40-45	101	109
45-50	101	114
50-55	99	117

2. В ході експерименту було отримано наступні значення випадкової величини:

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	145	123	145	169	98	79	100	96	83	71	60	110	116	103	87

а) побудувати довірчий інтервал для генерального середнього $\bar{x}$;

б) оцінити репрезентативність вибірки.

3. Значення одного й того ж показника було визначено за двома різними вибірками і згруповано в інтервали. По результатах групування було побудовано дві таблиці.

Вибірка А

Інтервали	Частоти (n_i)
5-15	100
15-25	103
25-35	106
35-45	109
45-55	125
55-65	139
65-75	147
75-85	169
85-95	154
95-105	144
105-115	139
115-125	128
125-135	124
135-145	123
145-155	101

Вибірка В

Інтервали	Частоти (n_i)
10-30	108
30-50	109
50-70	116
70-90	123
90-110	127
110-130	129
130-150	115
150-170	114
170-190	110
190-210	103

Чи суттєва різниця в середніх значеннях цих двох вибірок?

4. Вимірювання двох показників подано в таблиці.

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	10	13	18	24	26	22	24	13	10	69	25	21	20	19	8
Y	105	109	117	113	119	124	129	114	103	151	129	128	117	121	103

Визначити ступінь і напрямок лінійної кореляційної залежності між цими двома показниками.

Варіант № 12

1. Визначити, чи містять дані вибірки однорідний матеріал і чи можна їх об'єднати в одну.

Інтервали	Частоти для 1-ї вибірки	Частоти для 2-ї вибірки
5-10	2	1
10-15	3	8
15-20	5	9
20-25	8	11
25-30	8	16
30-35	9	19
35-40	7	15
40-45	4	13
45-50	3	7
50-55	1	3

2. В ході експерименту було отримано наступні значення випадкової величини:

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	10	54	87	52	31	29	28	54	63	27	9	62	46	50	60

а) побудувати довірчий інтервал для генерального середнього $\bar{x}$;

б) оцінити репрезентативність вибірки.

3. Значення одного й того ж показника було визначено за двома різними вибірками і згруповано в інтервали. По результатах групування було побудовано дві таблиці.

Вибірка А

Інтервали	Частоти (n_i)
5-15	13
15-25	16
25-35	18
35-45	21
45-55	29
55-65	38
65-75	45
75-85	45

Вибірка В

Інтервали	Частоти (n_i)
10-30	9
30-50	9
50-70	11
70-90	16
90-110	28
110-130	46
130-150	31
150-170	30

Вибірка А (продовження)	
85-95	49
95-105	32
105-115	21
115-125	18
125-135	17
135-145	12
145-155	4

Вибірка В (продовження)	
170-190	27
190-210	5

Чи суттєва різниця в середніх значеннях цих двох вибірок?

4. Вимірювання двох показників подано в таблиці.

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	13	15	18	19	19	25	26	34	39	67	78	96	99	102	103
Y	215	210	190	165	168	130	120	119	104	103	102	92	81	64	31

Визначити ступінь і напрямок лінійної кореляційної залежності між цими двома показниками.

Варіант № 13

1. Визначити, чи містять дані вибірки однорідний матеріал і чи можна їх об'єднати в одну.

Інтервали	Частоти для 1-ї вибірки	Частоти для 2-ї вибірки
5-10	23	13
10-15	25	15
15-20	29	19
20-25	31	21
25-30	39	25
30-35	42	28
35-40	38	26
40-45	21	23
45-50	19	20
50-55	17	17

2. В ході експерименту було отримано наступні значення випадкової величини:

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	98	74	54	89	32	54	69	82	71	34	46	62	21	73	36

а) побудувати довірчий інтервал для генерального середнього $\bar{x}$;

б) оцінити репрезентативність вибірки.

3. Значення одного й того ж показника було визначено за двома різними вибірками і згруповано в інтервали. По результатах групування було побудовано дві таблиці.

Вибірка А

Інтервали	Частоти (n_i)
5-15	23
15-25	28
25-35	29
35-45	33
45-55	33
55-65	33
65-75	35
75-85	35
85-95	39
95-105	41
105-115	44
115-125	43
125-135	40
135-145	36
145-155	27

Вибірка В

Інтервали	Частоти (n_j)
10-30	20
30-50	20
50-70	23
70-90	25
90-110	26
110-130	24
130-150	23
150-170	22
170-190	21
190-210	16

Чи суттєва різниця в середніх значеннях цих двох вибірок?

4. Вимірювання двох показників подано в таблиці.

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	115	119	100	103	95	87	63	108	79	81	101	128	131	104	106
Y	10	13	18	9	7	21	14	17	8	5	16	8	4	9	10

Визначити ступінь і напрямок лінійної кореляційної залежності між цими двома показниками.

Варіант № 14

1. Визначити, чи містять дані вибірки однорідний матеріал і чи можна їх об'єднати в одну.

Інтервали	Частоти для 1-й виборки	Частоти для 2-й виборки
5-10	15	16
10-15	18	23
15-20	26	29
20-25	27	34
25-30	35	39
30-35	47	51
35-40	44	49
40-45	30	33
45-50	21	20
50-55	8	17

2. В ході експерименту було отримано наступні значення випадкової величини:

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	47	89	54	91	62	80	54	77	71	60	91	88	70	50	63

а) побудувати довірчий інтервал для генерального середнього $\bar{x}$;

б) оцінити репрезентативність вибірки.

3. Значення одного й того ж показника було визначено за двома різними вибірками і згруповано в інтервали. По результатах групування було побудовано дві таблиці.

Выборка А	
Інтервали	Частоти (n_i)
5-15	5
15-25	9
25-35	18
35-45	19
45-55	22
55-65	26
65-75	38
75-85	43
85-95	20
95-105	19
105-115	17
115-125	10
125-135	5
135-145	3
145-155	3

Выборка В	
Інтервали	Частоти (n_i)
10-30	11
30-50	13
50-70	18
70-90	24
90-110	27
110-130	33
130-150	16
150-170	15
170-190	9
190-210	7

Чи суттєва різниця в середніх значеннях цих двох вибірок?

4. Вимірювання двох показників подано в таблиці.

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	125	139	114	100	109	118	134	127	116	108	123	114	120	128	131
Y	10	13	8	5	7	9	12	11	9	6	9	7	8	14	15

Визначити ступінь і напрямок лінійної кореляційної залежності між цими двома показниками.

Варіант № 15

1. Визначити, чи містять дані вибірки однорідний матеріал і чи можна їх об'єднати в одну.

Інтервали	Частоти для 1-й виборки	Частоти для 2-й виборки
5-10	2	16
10-15	3	15
15-20	5	13
20-25	9	8
25-30	14	7
30-35	15	2
35-40	13	9
40-45	10	10
45-50	8	12
50-55	6	19

2. В ході експерименту було отримано наступні значення випадкової величини:

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	44	46	32	51	24	27	30	14	39	48	23	17	26	41	50

а) побудувати довірчий інтервал для генерального середнього $\bar{x}$;

б) оцінити репрезентативність вибірки.

3. Значення одного й того ж показника було визначено за двома різними вибірками і згруповано в інтервали. По результатах групування було побудовано дві таблиці.

Выборка А

Интервалы	Частоты (n_i)
5-15	48
15-25	49
25-35	51
35-45	55
45-55	62
55-65	78
65-75	91
75-85	90
85-95	88
95-105	70
105-115	63
115-125	54
125-135	40
135-145	39
145-155	38

Выборка В

Интервалы	Частоты (n_i)
10-30	31
30-50	38
50-70	40
70-90	56
90-110	59
110-130	63
130-150	61
150-170	60
170-190	41
190-210	30

Чи суттєва різниця в середніх значеннях цих двох вибірок?

4. Вимірювання двох показників подано в таблиці.

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	148	153	137	179	184	121	164	128	137	139	152	163	167	158	159
Y	33	43	27	61	73	20	50	24	29	30	46	48	51	45	45

Визначити ступінь і напрямок лінійної кореляційної залежності між цими двома показниками.

Варіант № 16

1. Визначити, чи містять дані вибірки однорідний матеріал і чи можна їх об'єднати в одну.

Інтервали	Частоти для 1-й виборки	Частоти для 2-й виборки
5-10	121	115
10-15	132	118
15-20	136	129
20-25	138	143
25-30	146	149
30-35	151	160
35-40	150	158
40-45	147	131
45-50	130	121
50-55	119	110

2. В ході експерименту було отримано наступні значення випадкової величини:

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	10	18	26	34	46	21	51	32	85	17	63	61	48	56	27

- а) побудувати довірчий інтервал для генерального середнього $\bar{x}$;
б) оцінити репрезентативність вибірки.

3. Значення одного й того ж показника було визначено за двома різними вибірками і згруповано в інтервали. По результатах групування було побудовано дві таблиці.

Выборка А

Интервалы	Частоты (n_i)
5-15	17
15-25	18
25-35	22
35-45	29
45-55	31
55-65	39
65-75	48
75-85	55
85-95	59
95-105	61
105-115	67
115-125	80
125-135	70
135-145	50
145-155	20

Выборка В

Интервалы	Частоты (n_i)
10-30	20
30-50	22
50-70	29
70-90	24
90-110	23
110-130	22
130-150	18
150-170	17
170-190	10
190-210	4

Чи суттєва різниця в середніх значеннях цих двох вибірок?

4. Вимірювання двох показників подано в таблиці.

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	119	138	117	154	169	132	187	147	158	112	110	164	153	150	120
Y	21	14	30	3	2	15	1	4	3	35	40	1	4	6	18

Визначити ступінь і напрямок лінійної кореляційної залежності між цими двома показниками.

СПИСОК ЛІТЕРАТУРИ

1. Колемаев В.А., Староверов О.В., Турундаевский В.Б. Теория вероятностей и математическая статистика. – М.: Высшая школа, 1991. – 400с.
2. Севастьянов Б.А. Курс теории вероятностей и математической статистики. – М.: Наука, 1982. – 256с.
3. Сигорский В.П. Математический аппарат инженера. – Киев: Техника, 1975. – 768с.
4. Уилкс С. Математическая статистика. – М.: Наука, 1967. – 632 с.
5. Бровченко Т.А., Варбанец П.Д., Таранец В.Г. Метод статистического анализа в фонетических исследованиях. /учебное пособие. – Одесса, 1976. – 100с.
6. Левицкий В.В. Квантитативные методы в лингвистике. – Черновцы: Рута, 2004. – 190с.
7. Коршунов Ю.М. Математические основы кибернетики. – М.: Энергия, 1980. – 424с.
8. Носенко И.А. Начала статистики для лингвистов. – М.: Высшая школа, 1981.
9. Налимов В.В. Теория эксперимента. – М.: Наука, 1971.
10. Пиотровский Р.Г., Бектаев К.Б., Пиотровская А.А. Математическая лингвистика. – М.: Высшая школа, 1997.
11. Перебийніс В.С., Муравицька М.П., Дарчук Н.П. Частотні словники та їх використання. – Київ: Наукова думка, 1985.
12. Статистичні параметри стилів. – Київ: Наукова думка, 1967.
13. Тищенко В. Частота частин мови в різних функціональних стилях сучасної української мови// Перебийніс В.С., Муравицька М.П. Питання структурної лексикології. – Київ, 1970. – С. 215-224.
14. Мусурівська О.В. Прикметники, об'єднані семантикою фірму сучасній англійській мові. - Одеса: АКД, 1993.
15. Ковтанюк В.Р. Зіставлення лексико-семантичних мікросистем із значенням «міцний» у французькій та англійській мовах. – Донецьк: АКД, 2001.

Додаток 1

Значення функції нормального розподілу $\Phi(u)$

u	0	1	2	3	4	5	6	7	8	9
-0.0	5000	4960	4920	4880	4840	4801	4761	4721	4681	4641
-0.1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
-0.2	4207	4168	4129	4090	4052	4013	3974	3936	3897	3859
-0.3	3821	3783	3745	3707	3669	3632	3594	3557	3520	3483
-0.4	3446	3409	3372	3336	3300	3264	3228	3192	3156	3121
-0.5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
-0.6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
-0.7	2420	2389	2358	2327	2297	2266	2236	2206	2177	2148
-0.8	2119	2090	2061	2033	2005	1977	1949	1922	1894	1867
-0.9	1841	1814	1788	1762	1736	1711	1685	1660	1635	1611
-1.0	1587	1562	1539	1515	1492	1469	1446	1423	1401	1379
-1.1	1357	1335	1314	1292	1271	1251	1230	1210	1190	1170
-1.2	1151	1131	1112	1093	1075	1056	1038	1020	1003	0985
-1.3	0968	0951	0934	0918	0901	0885	0869	0853	0838	0823
-1.4	0808	0793	0778	0764	0749	0735	0721	0708	0694	0681
-1.5	0668	0655	0643	0630	0618	0606	0594	0582	0571	0559
-1.6	0548	0537	0526	0516	0505	0495	0485	0475	0465	0455
-1.7	0446	0436	0427	0418	0409	0401	0392	0384	0375	0367
-1.8	0359	0351	0344	0336	0329	0322	0314	0307	0301	0294
-1.9	0288	0281	0274	0268	0262	0256	0250	0244	0239	0233
-2.0	0228	0222	0217	0212	0207	0202	0197	0192	0188	0183
-2.1	0179	0174	0170	0166	0162	0158	0154	0150	0146	0143
-2.2	0139	0136	0132	0129	0125	0122	0119	0116	0113	0110
-2.3	0107	0104	0102	0099	0096	0094	0091	0089	0087	0084
-2.4	0082	0080	0078	0075	0073	0071	0069	0068	0066	0064
-2.5	0062	0060	0059	0057	0055	0054	0052	0051	0049	0048
-2.6	0047	0045	0044	0043	0041	0040	0039	0038	0037	0036
-2.7	0035	0034	0033	0032	0031	0030	0029	0028	0027	0026
-2.8	0026	0025	0024	0023	0023	0022	0021	0021	0020	0019
-2.9	0019	0018	0017	0016	0016	0015	0015	0014	0014	0014
	-3.0	-3.1	-3.2	-3.3	-3.4	-3.5	-3.6	-3.7	-3.8	-3.9
	0013	0010	0007	0005	0003	0002	0002	0001	0001	0000

Для визначення значень $\Phi(u)$ при $u \geq 0$ використовують співвідношення $\Phi(u) = 1 - \Phi(-u)$. Так, наприклад, $\Phi(1.65) = 1 - \Phi(-1.65) = 1 - 0.0495 = 0.9505$.

Додаток 2

Критичні значення χ^2

f	5%	15%	f	5%	1%	f	5%	1%
1	3.8	6.6	18	28.9	34.8	35	49.8	57.3
2	6.0	9.21	19	30.1	36.2	36	51.0	58.6
3	7.8	11.3	20	31.4	37.6	37	52.2	59.9
4	9.5	13.3	21	32.7	38.9	38	53.4	61.2
5	11.1	15.1	22	33.9	40.3	39	54.6	62.4
6	12.6	16.8	23	35.2	41.6	40	55.8	63.7
7	14.1	18.5	24	36.4	43.0	41	56.9	65.0
8	15.5	20.1	25	37.7	44.3	42	58.1	66.2
9	16.9	21.7	26	38.9	45.6	43	59.3	67.5
10	18.3	23.1	27	40.1	47.0	44	60.5	68.7
11	19.7	24.7	28	41.3	48.3	45	61.7	70.0
12	21.0	26.2	29	42.6	49.6	46	62.8	71.2
13	22.4	27.7	30	43.8	50.9	47	64.0	72.4
14	23.7	29.1	31	45.0	52.2	48	65.2	73.7
15	25.0	30.6	32	46.2	53.5	49	66.3	74.9
16	26.3	32.0	33	47.4	54.8	50	67.5	76.2
17	27.6	33.4	34	48.6	56.1			

Нульова гіпотеза приймається при $\chi^2 \leq \chi_{05}^2$ й відкидається при $\chi^2 > \chi_{01}^2$.

Додаток 3

Критичні значення t (критерію Стьюдента)

f	Рівень довіри			f	Рівень довіри		
	95%	99%	99.9%		95%	99%	99.9%
1	12.71	63.60		21	2.08	2.83	3.82
2	4.30	9.93	31.60	22	2.07	2.82	3.79
3	3.18	5.84	12.94	23	2.07	2.81	3.77
4	2.78	4.60	8.61	24	2.06	2.80	3.75
5	2.57	4.03	6.86	25	2.06	2.79	3.73
6	2.45	3.71	5.96	26	2.06	2.78	3.71
7	2.37	3.50	5.41	27	2.05	2.77	3.69
8	2.31	3.36	5.04	28	2.05	2.76	3.67
9	2.26	3.25	4.78	29	2.04	2.76	3.66
10	2.23	3.17	4.59	30	2.04	2.75	3.65
11	2.20	3.11	4.44	40	2.02	2.70	3.55
12	2.18	3.06	4.32	50	2.01	2.68	3.50
13	2.16	3.01	4.22	60	2.00	2.66	3.46
14	2.15	2.98	4.14	80	1.99	2.64	3.42
15	2.13	2.95	4.07	100	1.98	2.63	3.39
16	2.12	2.92	4.02	120	1.98	2.62	3.37
17	2.11	2.90	3.97	200	1.97	2.60	3.34
18	2.10	2.88	3.92	500	1.96	2.59	3.31
19	2.09	2.86	3.88	∞	1.96	2.58	3.29
20	2.09	2.85	3.85				
f	5%	1%	0.1%	f	5%	1%	0.1%
	Рівень значимості				Рівень значимості		

Нульова гіпотеза приймається при $t \leq t_{05}$ й відкидається при $t > t_{01}$.

Додаток 4

Значення функції Ψ

$p \rightarrow$ $\downarrow$	0	1	2	3	4	5	6	7	8	9
0.00	- ∞	- 3.09	- 2.88	- 2.75	- 2.65	- 2.58	- 2.51	- 2.46	- 2.41	- 2.37
0.01	- 2.33	- 2.29	- 2.26	- 2.23	- 2.20	- 2.17	- 2.14	- 2.12	- 2.10	- 2.07
0.02	- 2.05	- 2.03	- 2.01	- 2.00	- 1.98	- 1.96	- 1.94	- 1.93	- 1.91	- 1.90
0.03	- 1.88	- 1.87	- 1.85	- 1.84	- 1.83	- 1.81	- 1.80	- 1.79	- 1.77	- 1.76
0.04	- 1.75	- 1.74	- 1.73	- 1.72	- 1.71	- 1.70	- 1.68	- 1.67	- 1.66	- 1.65
0.05	- 1.64	- 1.64	- 1.63	- 1.62	- 1.61	- 1.60	- 1.59	- 1.58	- 1.57	- 1.56
0.06	- 1.55	- 1.55	- 1.54	- 1.53	- 1.52	- 1.51	- 1.51	- 1.50	- 1.49	- 1.48
0.07	- 1.48	- 1.47	- 1.46	- 1.45	- 1.45	- 1.44	- 1.43	- 1.43	- 1.42	- 1.41
0.08	- 1.41	- 1.40	- 1.39	- 1.39	- 1.38	- 1.37	- 1.37	- 1.36	- 1.35	- 1.35
0.09	- 1.34	- 1.33	- 1.33	- 1.32	- 1.32	- 1.31	- 1.30	- 1.30	- 1.29	- 1.29
0.10	- 1.28	- 1.28	- 1.27	- 1.26	- 1.26	- 1.25	- 1.25	- 1.24	- 1.24	- 1.23
0.11	- 1.23	- 1.22	- 1.22	- 1.21	- 1.21	- 1.20	- 1.20	- 1.19	- 1.19	- 1.18
0.12	- 1.18	- 1.17	- 1.17	- 1.16	- 1.16	- 1.15	- 1.15	- 1.14	- 1.14	- 1.13
0.13	- 1.13	- 1.12	- 1.12	- 1.11	- 1.11	- 1.10	- 1.10	- 1.09	- 1.09	- 1.09
0.14	- 1.08	- 1.08	- 1.07	- 1.07	- 1.06	- 1.06	- 1.05	- 1.05	- 1.05	- 1.04
0.15	- 1.04	- 1.03	- 1.03	- 1.02	- 1.02	- 1.02	- 1.01	- 1.01	- 1.00	- 1.00
0.16	- 0.99	- 0.99	- 0.99	- 0.98	- 0.98	- 0.97	- 0.97	- 0.97	- 0.96	- 0.96
0.17	- 0.95	- 0.95	- 0.95	- 0.94	- 0.94	- 0.93	- 0.93	- 0.93	- 0.92	- 0.92
0.18	- 0.92	- 0.91	- 0.91	- 0.90	- 0.90	- 0.90	- 0.89	- 0.89	- 0.89	- 0.88
0.19	- 0.88	- 0.87	- 0.87	- 0.87	- 0.86	- 0.86	- 0.86	- 0.85	- 0.85	- 0.85
0.20	- 0.84	- 0.84	- 0.83	- 0.83	- 0.83	- 0.82	- 0.82	- 0.82	- 0.81	- 0.81
0.21	- 0.81	- 0.80	- 0.80	- 0.80	- 0.79	- 0.79	- 0.79	- 0.78	- 0.78	- 0.78
0.22	- 0.77	- 0.77	- 0.77	- 0.76	- 0.76	- 0.76	- 0.75	- 0.75	- 0.75	- 0.74
0.23	- 0.74	- 0.74	- 0.73	- 0.73	- 0.73	- 0.72	- 0.72	- 0.72	- 0.71	- 0.71
0.24	- 0.71	- 0.70	- 0.70	- 0.70	- 0.69	- 0.69	- 0.69	- 0.68	- 0.68	- 0.68
0.25	- 0.67	- 0.67	- 0.67	- 0.67	- 0.66	- 0.66	- 0.66	- 0.65	- 0.65	- 0.65
0.26	- 0.64	- 0.64	- 0.64	- 0.63	- 0.63	- 0.63	- 0.63	- 0.62	- 0.62	- 0.62

$p \rightarrow$ $\downarrow$	0	1	2	3	4	5	6	7	8	9
0.27	- 0.61	- 0.61	- 0.61	- 0.60	- 0.60	- 0.60	- 0.59	- 0.59	- 0.59	- 0.59
0.28	- 0.58	- 0.58	- 0.58	- 0.57	- 0.57	- 0.57	- 0.57	- 0.56	- 0.56	- 0.56
0.29	- 0.55	- 0.55	- 0.55	- 0.54	- 0.54	- 0.54	- 0.54	- 0.53	- 0.53	- 0.53
0.30	- 0.52	- 0.52	- 0.52	- 0.52	- 0.51	- 0.51	- 0.51	- 0.50	- 0.50	- 0.50
0.31	- 0.50	- 0.49	- 0.49	- 0.49	- 0.48	- 0.48	- 0.48	- 0.48	- 0.47	- 0.47
0.32	- 0.47	- 0.46	- 0.46	- 0.46	- 0.46	- 0.45	- 0.45	- 0.45	- 0.45	- 0.44
0.33	- 0.44	- 0.44	- 0.43	- 0.43	- 0.43	- 0.43	- 0.42	- 0.42	- 0.42	- 0.42
0.34	- 0.41	- 0.41	- 0.41	- 0.40	- 0.40	- 0.40	- 0.40	- 0.39	- 0.39	- 0.39
0.35	- 0.39	- 0.38	- 0.38	- 0.38	- 0.37	- 0.37	- 0.37	- 0.37	- 0.36	- 0.36
0.36	- 0.36	- 0.36	- 0.35	- 0.35	- 0.35	- 0.35	- 0.34	- 0.34	- 0.34	- 0.33
0.37	- 0.33	- 0.33	- 0.33	- 0.32	- 0.32	- 0.32	- 0.32	- 0.31	- 0.31	- 0.31
0.38	- 0.31	- 0.30	- 0.30	- 0.30	- 0.30	- 0.29	- 0.29	- 0.29	- 0.28	- 0.28
0.39	- 0.28	- 0.28	- 0.27	- 0.27	- 0.27	- 0.27	- 0.26	- 0.26	- 0.26	- 0.26
0.40	- 0.25	- 0.25	- 0.25	- 0.25	- 0.24	- 0.24	- 0.24	- 0.24	- 0.23	- 0.23
0.41	- 0.23	- 0.23	- 0.22	- 0.22	- 0.22	- 0.21	- 0.21	- 0.21	- 0.21	- 0.20
0.42	- 0.20	- 0.20	- 0.20	- 0.19	- 0.19	- 0.19	- 0.19	- 0.18	- 0.18	- 0.18
0.43	- 0.18	- 0.17	- 0.17	- 0.17	- 0.17	- 0.16	- 0.16	- 0.16	- 0.16	- 0.15
0.44	- 0.15	- 0.15	- 0.15	- 0.14	- 0.14	- 0.14	- 0.14	- 0.13	- 0.13	- 0.13
0.45	- 0.13	- 0.12	- 0.12	- 0.12	- 0.12	- 0.11	- 0.11	- 0.11	- 0.11	- 0.10
0.46	- 0.10	- 0.10	- 0.10	- 0.09	- 0.09	- 0.09	- 0.09	- 0.08	- 0.08	- 0.08
0.47	- 0.08	- 0.07	- 0.07	- 0.07	- 0.07	- 0.06	- 0.06	- 0.06	- 0.06	- 0.05
0.48	- 0.05	- 0.05	- 0.05	- 0.04	- 0.04	- 0.04	- 0.04	- 0.03	- 0.03	- 0.03
0.49	- 0.03	- 0.02	- 0.02	- 0.02	- 0.02	- 0.01	- 0.01	- 0.01	- 0.01	- 0.00
0.50	0.00	0.00	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02
0.51	0.03	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.05	0.05
0.52	0.05	0.05	0.06	0.06	0.06	0.06	0.07	0.07	0.07	0.07
0.53	0.08	0.08	0.08	0.08	0.09	0.09	0.09	0.09	0.10	0.10
0.54	0.10	0.10	0.11	0.11	0.11	0.11	0.12	0.12	0.12	0.12
0.55	0.13	0.13	0.13	0.13	0.14	0.14	0.14	0.14	0.15	0.15
0.56	0.15	0.15	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17

$p \rightarrow$ $\downarrow$	0	1	2	3	4	5	6	7	8	9
0.57	0.18	0.18	0.18	0.18	0.19	0.19	0.19	0.19	0.20	0.20
0.58	0.20	0.20	0.21	0.21	0.21	0.21	0.22	0.22	0.22	0.23
0.59	0.23	0.23	0.23	0.24	0.24	0.24	0.24	0.25	0.25	0.25
0.56	0.25	0.26	0.26	0.26	0.26	0.27	0.27	0.27	0.27	0.28
0.61	0.28	0.28	0.28	0.29	0.29	0.29	0.30	0.30	0.30	0.30
0.62	0.31	0.31	0.31	0.31	0.32	0.32	0.32	0.32	0.33	0.33
0.63	0.33	0.33	0.34	0.34	0.34	0.35	0.35	0.35	0.35	0.36
0.64	0.36	0.36	0.36	0.37	0.37	0.37	0.37	0.38	0.38	0.38
0.65	0.39	0.39	0.39	0.39	0.40	0.40	0.40	0.40	0.41	0.41
0.66	0.41	0.42	0.42	0.42	0.42	0.43	0.43	0.43	0.43	0.44
0.67	0.44	0.44	0.45	0.45	0.45	0.45	0.46	0.46	0.46	0.46
0.68	0.47	0.47	0.47	0.48	0.48	0.48	0.49	0.49	0.49	0.49
0.69	0.50	0.50	0.50	0.50	0.51	0.51	0.52	0.52	0.52	0.52
0.70	0.52	0.53	0.53	0.53	0.54	0.54	0.54	0.54	0.55	0.55
0.71	0.55	0.56	0.56	0.56	0.57	0.57	0.57	0.57	0.58	0.58
0.72	0.58	0.59	0.59	0.59	0.59	0.60	0.60	0.60	0.61	0.61
0.73	0.61	0.62	0.62	0.62	0.63	0.63	0.63	0.63	0.64	0.64
0.74	0.64	0.65	0.65	0.65	0.66	0.66	0.67	0.67	0.67	0.67
0.75	0.67	0.68	0.68	0.68	0.69	0.69	0.70	0.70	0.70	0.70
0.76	0.71	0.71	0.71	0.72	0.72	0.72	0.73	0.73	0.73	0.74
0.77	0.74	0.74	0.75	0.75	0.75	0.76	0.76	0.76	0.77	0.77
0.78	0.77	0.78	0.78	0.78	0.79	0.79	0.80	0.80	0.80	0.80
0.79	0.81	0.81	0.81	0.82	0.82	0.82	0.83	0.83	0.83	0.84
0.80	0.84	0.85	0.85	0.85	0.86	0.86	0.87	0.87	0.87	0.87
0.81	0.88	0.88	0.89	0.89	0.89	0.90	0.90	0.90	0.91	0.91
0.82	0.92	0.92	0.92	0.93	0.93	0.93	0.94	0.94	0.95	0.95
0.83	0.95	0.96	0.97	0.97	0.97	0.97	0.98	0.98	0.99	0.99
0.84	0.99	1.00	1.00	1.01	1.01	1.02	1.02	1.02	1.03	1.03
0.85	1.04	1.04	1.05	1.05	1.05	1.06	1.07	1.07	1.07	1.08
0.86	1.08	1.09	1.09	1.09	1.10	1.10	1.11	1.11	1.12	1.12

$p \rightarrow$ $\downarrow$	0	1	2	3	4	5	6	7	8	9
0.87	1.13	1.13	1.14	1.14	1.15	1.15	1.16	1.16	1.17	1.17
0.88	1.18	1.18	1.19	1.19	1.20	1.20	1.21	1.21	1.22	1.22
0.89	1.23	1.23	1.24	1.24	1.25	1.25	1.26	1.26	1.27	1.28
0.90	1.28	1.29	1.29	1.30	1.30	1.31	1.32	1.32	1.33	1.33
0.91	1.34	1.35	1.35	1.36	1.37	1.37	1.38	1.39	1.39	1.40
0.92	1.41	1.41	1.42	1.43	1.43	1.44	1.45	1.45	1.46	1.47
0.93	1.48	1.48	1.49	1.50	1.51	1.51	1.52	1.53	1.54	1.55
0.94	1.55	1.56	1.57	1.58	1.59	1.60	1.61	1.62	1.63	1.64
0.95	1.64	1.65	1.66	1.67	1.68	1.70	1.71	1.72	1.73	1.74
0.96	1.75	1.76	1.77	1.79	1.80	1.81	1.83	1.84	1.85	1.87
0.97	1.88	1.90	1.91	1.93	1.94	1.96	1.98	2.00	2.01	2.03
0.98	2.05	2.07	2.10	2.12	2.14	2.17	2.20	2.23	2.26	2.29
0.99	2.33	2.37	2.41	2.46	2.51	2.58	2.65	2.75	2.88	3.09

Навчальне видання

**Варбанець Павло Дмитрович,
Якімова Наталія Анатоліївна**

ЛІНГВОСТАТИСТИКА

Навчальний посібник

Видано за авторською редакцією

Підписано до друку 26.05.2021 р. Формат 84x108/32.

Ум.-друк. арк. 10,34. Зам. № 2288. Тираж 18 прим.

Видавець і виготовлювач

Одеський національний університет імені І. І. Мечникова
(свідцтво суб'єкта видавничої справи серія ДК № 4215 від 22.11.2011 р.)

65082, м. Одеса, вул. Єлісаветинська, 12, Україна

Тел.: (048) 723 28 39