

Волинський національний університет імені Лесі Українки

Факультет біології та лісового господарства

Кафедра зоології

О. П. Зінченко

Біометрія

*методичні рекомендації
до виконання практичних робіт*

Луцьк – 2023

УДК 57.087.1

З 63

*Рекомендовано до друку науково-методичною радою
Волинського національного університету імені Лесі Українки
(протокол № 7 від 16 березня 2023 р.)*

Рецензенти:

Степанюк Я. В. – завідувач кафедри завідувач кафедри гістології та медичної біології Волинського національного університету імені Лесі Українки, кандидат біологічних наук, доцент;
Григор'єва Н. В. – завідувач відділу природничих дисциплін Волинського інституту післядипломної педагогічної освіти.

Зінченко О. П.

З 63 Біометрія : Метод. рекомендації до викон. практичних робіт / Волинський національний університет імені Лесі Українки, факультет біології та лісового господарства, кафедра зоології. Луцьк, 2023. 56 с.

Видання вміщує методичні вказівки до виконання 12 практичних робіт із курсу «Біометрія», передбачених навчальним планом підготовки бакалавра галузі знань 20 «Аграрні науки та продовольство», спеціальності 205 «Лісове господарство», освітньо-професійної програми «Лісове господарство».

У роботах розглядаються теми, що висвітлюють основи комбінаторики та теорії множин, теорії ймовірностей та варіаційної статистики. До кожної практичної роботи наведені тема, мета, питання для контролю знань, хід виконання роботи. Наприкінці видання подано список рекомендованої літератури.

УДК 57.087.1

© Зінченко О. П., 2023

© Зінченко О. П. (обкладинка), 2023

Передмова

Методичні рекомендації з курсу «Біометрія» призначені для студентів освітнього рівня «Бакалавр» денної форми навчання спеціальності 205 «Лісове господарство», освітньо-професійної програми «Лісове господарство». Видання базується на знаннях з математики, що отриманні за попередній період навчання в середній та вищій школі.

Основу видання складають 12 практичних робіт, які віднесені до двох модулів: 1-го модуля – «Основи теорії ймовірностей» і до 2-го модуля – «Основи варіаційної статистики».

У 1-му модулі (практичні роботи №№1-4) розглядаються основи теорії множин, комбінаторики та теорії ймовірностей. 2-й модуль охоплює 8 практичних робіт (№№5–12) присвячених обробці та інтерпретації статистичних даних. Більшість завдань побудовано з використанням прикладів різних біологічних дисциплін: ботаніки, зоології, біотехнології, мікробіології, мікології, лісознавства, екології та ін. Вміння розв'язувати ці завдання необхідні фахівцю для грамотної інтерпретації та статистичного прогнозування експериментальних даних.

При виконанні робіт студенти мають широко використовувати калькулятори, комп'ютерну техніку і комп'ютерні програми Microsoft Office (текстовий редактор WORD, електронні таблиці EXCEL).

МОДУЛЬ 1. ОСНОВИ ТЕОРІЇ ЙМОВІРНОСТЕЙ

Практична робота №1

Тема: Вибіркові простори і простори рівних ймовірностей.

Мета: Ознайомитися з основними поняттями теорії ймовірностей і навчитися застосовувати їх для рішення прикладних задач. Закріпити знання з основ комбінаторики.

Обладнання: таблиці, гральні кубики, монети, калькулятори, комп'ютери.

Контрольні питання:

1. Основний принцип перелічення елементів.
2. Перестановки та сполучення елементів.
3. Біноміальний та поліноміальний коефіцієнти.
4. Поняття випробування.
5. Поняття вибіркового простору експерименту.
6. Поняття елементарної події, достовірної події і неможливої події.
7. Простір рівних ймовірностей.
8. Ймовірність подій в скінченному просторі рівних ймовірностей.
9. Додаткова подія.
10. Взаємно виключні події.

Хід роботи

Робота №1. Знаходження ймовірності подій в експерименті з підкидання монети.

Проведіть експеримент із підкидання монети. Підкиньте монету 100 разів і результат занесіть у таблицю.

Результат експерименту	Герб	Решка (аверс)	Разом всі результати
Кількість спроб			100

Порахуйте ймовірності випадання герба (G) і решки (P). Результати занесіть їх у зошит.

Робота №2. Знаходження ймовірності подій в експерименті з підкидання грального кубика.

Проведіть експеримент із підкидання грального кубика. Підкиньте кубик 100 разів і результат занесіть у таблицю.

Число на грані кубика	1	2	3	4	5	6	Парне число	Непарне число	Разом
Кількість вдалих спроб									100

Порахуйте ймовірності випадання на гральному кубіку парного (A) і непарного чисел (B). Результати запишіть у зошит.

В експерименті з підкидання грального кубика визначить і опишіть по одному прикладу: елементарної події, достовірної події; неможливої події; додаткових подій, взаємно виключних подій (Зразок. Елементарна подія – це подія, яка полягає в тому, що при підкиданні кубика на його грані випадає 1, або 2, і т. д.).

Робота №3. Розв'язок задач.

Розберіть приклади задач і виконайте завдання для самостійного розв'язку, які запропонує викладач.

Приклади розв'язання задач.

Приклад 1. Із 6 лабораторних щурів необхідно вибрати трьох і розмістити їх у 3-х клітках, що позначені А, В, С. Скільки існує способів це зробити?

Розв'язок

Існує 6 способів вибору одного щура для клітки А. Далі маємо 5 щурів, одного з яких відбираємо і розміщуємо в клітці В. І нарешті, одного з 4 щурів, що залишились, розміщуємо в клітці С. Згідно до основного принципу перелічення, для цього існує $6 \times 5 \times 4 = 120$ способів.

Приклад 2. Потрібно визначити перше, друге та третє місця на конкурсі, в якому бере участь 10 конкурсантів. Скільки теоретично існує способів розподілити ці місця?

Розв'язок

Необхідно визначити число перестановок з 10 об'єктів по три, тобто $10! / (10-3)! = 10! / 7! = 10 \times 9 \times 8 = 720$.

Приклад 3. В зоопарку утримують три лисиці та два єноти. Клітки розміщені в один ряд. Скільки існує способів розмістити клітки із тваринами?

Розв'язок

Один з шляхів вирішення цієї задачі полягає в тому, щоб просто переписати всі можливі варіанти. Позначимо лисиць літерою Л, а єнотів – Є. Можливі варіанти такі: ЛЛЛЄЄ, ЛЛЄЛЄ, ЛЄЛЛЄ, ЄЛЛЛЄ, ЄЄЛЛЛ, ЄЛЄЛЛ, ЄЛЛЄЛ, ЛЄЄЛЛ, ЛЄЛЄЛ, ЛЛЄЄЛ. Отримуємо 10 таких перестановок.

Існує і більш системний шлях вирішення задачі. Маємо $5! = 120$ перестановок з п'яти об'єктів, якщо всі об'єкти розрізняються. Якщо не можна розрізнити три з п'яти об'єктів, то для кожної перестановки з п'яти об'єктів знайдеться, всього по $3! = 6$ перестановок, які виглядають однаково. Щоб довести це, розглянемо розташування кліток з тваринами і позначимо лисиць за допомогою індексів L_1, L_2, L_3 . Тоді шість перестановок Будуть такими: $L_1L_2L_3ЄЄ, L_1L_2L_3ЄЄ, L_2L_1L_3ЄЄ, L_2L_3L_1ЄЄ, L_3L_1L_2ЄЄ, L_3L_2L_1ЄЄ$. Ці шість перестановок не відрізняються якщо цікавить тільки розміщення лисиць відносно єнотів. Таким чином, кількість перестановок, що відрізняються, дорівнює $5! / 3! = 120 / 6 = 20$. Але єноти в цьому випадку теж не розрізняються; наприклад, $LЛЛЄ_1Є_2$ – це те саме, що і $LЛЛЄ_2Є_1$. Отримуємо, що число різних розміщень є $20 / 2 = 5! / (3! \times 2!) = 10$. Це відповідає результатам, що отримані шляхом перелічення всіх варіантів.

Приклад 4. Три типи бактерій культивують в 9 пробірках. Три пробірки містять бактерій 1-го типу, чотири – бактерії 2-го типу, дві – бактерій 3-го типу. Скільки існує різних способів розмістити пробірки в один ряд на штативі, якщо для нас важливе розміщення лише типів бактерій?

Розв'язок

Множину з 9 пробірок розбивають на 3 підмножини, що містять відповідно 3, 4 і 2 об'єкти, що не розрізняються. Відповідне число різних

способів визначається поліноміальним коефіцієнтом: $\binom{9}{3,4,2} = \frac{9!}{3! \times 4! \times 2!} = 1260$.

Приклад 5. Для експерименту по визначенню швидкості росту необхідно вибрати 4 штами бактерій з 8. Скільки для цього існує способів вибору?

Розв'язок

Потрібно знайти число способів вибору 4 об'єктів з 8 не враховуючи порядок вибору. Це число є $\binom{8}{4} = \frac{8!}{4! \times (8-4)!} = 70$. Тобто, всього маємо 70 способів.

Приклад 6. У лабораторній клітці тримають 8 білих та 6 сірих мишей. Знайти число способів вибору п'яти мишей з клітки, якщо 3 них повинні бути білого, а дві сірого кольору.

Розв'язок

Існує $\binom{8}{3} = 56$ способів вибору трьох білих мишей та $\binom{6}{2} = 15$ способів

вибору двох сірих мишей.

Таким чином, маємо $56 \times 15 = 840$ способів вибору трьох білих та двох сірих мишей.

Приклад 7. При порівняльному дослідженні 16 людей, що страждають захворюванням щитоподібної залози, необхідно поділити їх на три групи по 12, 2 та 2 людини. Скільки існує способів це зробити?

Розв'язок

Необхідно підрахувати число способів розбиття множини з об'єктів на три підмножини, що містять 12, 2 та 2 об'єкти. Число визначається поліноміальним

коефіцієнтом $\binom{16}{12, 2, 2} = \frac{16!}{12! \times 2! \times 2!} = 10\,920$.

Приклад 8. З плазми крові, що містить червоні та білі кров'яні клітини, вибрати чотири клітини і відмітити, скільки з них буде червоними (еритроцитами).

Розв'язок

Можливі наступні варіанти: 0, 1, 2, 3, 4 (можливе число червоних клітин). Вибірковим простором буде $S = \{0, 1, 2, 3, 4\}$.

Приклад 9. Лабораторний щур знаходиться у лабіринті і повинен вибрати один з п'яти можливих шляхів. Лише один з них веде до заохочення у вигляді їжі. Припустимо, що щур з однаковою ймовірністю вибирає будь-який шлях. Яка ймовірність вибору шляху, що приводить до їжі?

Розв'язок

Вибірковий простір цього експерименту є $S = \{\text{шлях 1, шлях 2, шлях 3, шлях 4, шлях 5}\}$, і ймовірність вибору будь-якого шляху дорівнює $1/5$. Оскільки лише один шлях веде до їжі, то ймовірність p (щур знаходить їжу) $= 1/5$.

Приклад 10. Припустимо, що в групі з 10 людей є четверо чоловіків. Якщо випадковим чином вибирають двох людей, то яка ймовірність того, що вони: 1) чоловічої статі; 2) жіночої статі; 3) різної статі?

Розв'язок

Нехай A – подія, що полягає в тому, що обидві людини є чоловіками, B – обидві людини є жінками, C – що одна людина – чоловік, а інша – жінка. Простір вибірок S складається з можливих пар людей і містить $\binom{10}{2} = 45$ елементарних подій. Існує

$\binom{4}{2} = 6$ способів вибору двох чоловіків з чотирьох. Тому $p(A) = \frac{6}{45} = \frac{2}{15}$.

Аналогічно визначаємо кількість способів вибору двох жінок з шести $\binom{6}{2} = 15$.

Визначаємо подію, що обидві людини є жінками $p(B) = \frac{15}{45} = \frac{1}{3}$. І нарешті, існує 4 способи вибору одного чоловіка, та 6 способів, вибору однієї жінки, що дає $4 \times 6 = 24$ способи вибору по одній людині кожної статі. Таким чином, $p(C) = \frac{24}{45} = \frac{8}{15}$.

Практична робота №2

Тема: Скінченні простори ймовірностей.

Мета: Ознайомитися з застосуванням теорем додавання і множення ймовірностей елементарних подій для рішення прикладних задач.

Обладнання: таблиці, набори білих та чорних фішок, калькулятори, комп'ютери.

Контрольні питання:

1. Поняття множини.
2. Операції над множинами.
3. Діаграми Венна.
4. Визначення простору ймовірностей.
5. Теорема додавання і множення ймовірностей елементарних подій, що містяться в будь-якій події E .
6. Теорема для будь-яких подій A_1, A_2, A_3 у скінченному просторі ймовірностей.

Хід роботи

Робота №1. Знаходження ймовірності появи різних генотипів та фенотипів при моногібридному схрещуванні.

Припустимо, що в експерименті схрещуються гетерозиготні особини, які представлено в моделі 2 наборами фішок. В кожному наборі знаходяться по 50 чорних і 50 білих фішок, що символізують 2 алельні стани гену: A – домінуючу ознаку (чорне забарвлення) та a – рецесивну ознаку (відсутність чорного забарвлення) при моногібридному схрещуванні.

Схему такого моногібридного схрещування можна уявити так:

$$\begin{array}{l} P: \quad \quad \quad Aa \quad \times \quad Aa \\ \text{Гамети:} \quad \quad A, a \quad \quad \quad A, a \\ F_1: \quad \quad AA, \quad 2Aa, \quad aa \end{array}$$

З кожного набору випадково виберіть по одній фішці, відкладіть їх, а відповідний результат вибору (AA , Aa чи aa) зафіксуйте. Цю дію повторюйте до тих пір, поки у наборах не закінчатся фішки. Результати експерименту занесіть у таблицю.

Результат вибору фішок	AA (всі чорні)	Aa (неодноголірні)	aa (всі білі)	Разом
Кількість спроб				100

Нехай A означає подію вибору алелі A , а \bar{A} – вибору алелі a . Ймовірність потрапляння у гамету будь-якого з двох алелей дорівнює $p(A) = p(\bar{A}) = 0,5$ (у наборі чорних і білих фішок порівну).

Ймовірність появи будь-яких типів зигот дорівнює добуткам ймовірностей появи алелей, що потрапляють разом з гаметами, оскільки, ці події незалежні. Нехай подія E_1 – означає появу генотипу AA , E_2 – означає появу генотипу Aa та E_3 – означає появу генотипу aa . Поява забарвлених особин зумовлена появою у генотипі хоча б однієї домінуючої алелі (подія B), а поява білих (незабарвлених особин) відсутністю домінуючих алелей, тобто наявністю рецесивних алелей (подія \bar{B}).

Визначить ймовірності появи різних генотипів (AA), (Aa) і (aa) та фенотипів ($AA+Aa$) та (aa), що зумовлюють наявність або відсутність забарвлення. Результати у вигляді умовних позначень: $p(E_1)$, $p(E_2)$, $p(E_3)$, $p(B)$ та $p(\bar{B})$ занесіть у зошит.

Робота №2. Розв'язок задач.

Розберіть приклади знаходження ймовірності із застосуванням теорем додавання і множення ймовірностей елементарних подій і розв'яжіть задачі, які запропонує викладач.

Приклади розв'язання задач.

Приклад 1. У великій популяції дрозофіл 25% мух мають мутацію очей, 50% – мутацію крил, а 40% мух з мутацією очей мають ще й мутацію крил. Яка ймовірність того, що у випадково вибраної мухи буде хоча б одна з мутацій? Яка ймовірність того, що у випадково вибраної мухи, є мутація очей і відсутні мутації крил?

Розв'язок

Позначимо через E та W події, які полягають в тому, що випадково вибрана муха має мутації очей та мутацію крил.

Ймовірність того, що муха має одну чи обидві мутації, є $p(E \cup W) = p(E) + p(W) - p(E \cap W)$. Але за умовою задачі $p(E) = 0,25$; $p(W) = 0,5$; $p(E \cap W) = 0,25 \cdot 0,4 = 0,1$. Таким чином, $p(E \cup W) = 0,25 + 0,5 - 0,1 = 0,65$. Ймовірність того, що у випадково вибраної мухи, є мутація очей, але немає мутації крил – $p(E \setminus W) = p(E) - p(E \cap W) = 0,25 - 0,1 = 0,15$.

Приклад 2. В акваріумі, де тримають трьох риб A , B , та C час від часу поміщають шматочки їжі. Кожного разу, коли кидають шматочок їжі, риби конкурують за нього. Припустимо, що за тривалий період спостережень було встановлено, що риби A або B досягали успіху протягом $1/2$ часу, а риби A або C – протягом $3/4$ всього часу спостережень. Яка ймовірність того, що досягає успіху риба A ? Яка з риб нагодована краще за інших?

Розв'язок

Нехай A позначає подію, яка полягає в тому, що їжа достається рибі A , і т. п. Оскільки кожний шматочок їжі може достатися лише одній з риб, то $p(A \cap B) = p(A \cap C) = p(B \cap C) = 0$. Дано, що $p(A \cup B) = 1/2$ та $p(A \cup C) = 3/4$.

Застосуємо ствердження $p(A_1 \cup A_2) = p(A_1) + p(A_2) - p(A_1 \cap A_2)$ з теореми для будь-яких подій A_1 і A_2 в скінченному просторі ймовірностей. Звідси отримуємо, що $p(A \cup B) = p(A) + p(B) = 1/2$ і $p(A \cup C) = p(A) + p(C) = 3/4$. Отже, $p(B) = 1/2 - p(A)$ і $p(C) = 3/4 - p(A)$. Оскільки, $p(A) + p(B) + p(C) = 1$, то $1 = p(A) + [1/2 - p(A)] + [3/4 - p(A)] = 5/4 - p(A)$, що дає $p(A) = 1/4$. Звідси $p(B) = 1/4$ та $p(C) = 1/2$. Рибі C достається їжі вдвічі більше, ніж кожній з двох інших.

Практична робота №3

Тема: Умовна ймовірність.

Мета: Ознайомитися з застосуванням теореми умовної ймовірності для рішення прикладних задач.

Обладнання: таблиці, калькулятори, комп'ютери.

Контрольні питання:

1. Визначення умовної ймовірності.
2. Теорема умовної ймовірності.
3. Теорема множення умовних ймовірностей.
4. Визначення двох незалежних подій.
5. Визначення трьох незалежних подій.

Хід роботи

Робота №1. Розв'язок задач.

Розберіть приклади задач і виконайте завдання для самостійного розв'язку, що запропонує викладач.

Приклади розв'язання задач.

Приклад 1. У популяції дрозофіл 25 % особин мають мутацію крил, а 15 % – мутацію очей і 10 % – обидві мутації. Випадково вибирають одну мушку.

а) якщо у неї виявиться мутація крил, то яка ймовірність того, що в неї є і мутація очей?

б) якщо у неї виявиться мутація очей, то яка ймовірність того, що в неї є і мутація крил?

в) Яка ймовірність того, що в неї є хоча б одна з мутацій?

Розв'язок

Нехай W та E означають відповідно мутацію крил та очей. Тоді:

а) $p(E|W) = p(E \cap W) / p(W) = 0,1/0,25 = 2/5$;

б) $p(W|E) = p(W \cap E) / p(E) = 0,1/0,15 = 2/3$;

в) $p(W \cup E) = p(W) + p(E) - p(W \cap E) = 0,25 + 0,15 - 0,1 = 0,3$.

Приклад 2. На двох фермах А та В, де нараховують по 1000 голів великої худоби на кожній, відбувся спалах захворювання ящуром. Частки зараженої худоби складають відповідно $1/5$ та $1/4$. Випадково відбирають одну корову.

1) Яка ймовірність того, що відібрана корова належить фермі А і має захворювання?

2) Якщо на кожній фермі 70 % зараженої худоби мають вік менш ніж один рік, то яка ймовірність того, що вибрана корова належить фермі В, має захворювання і вік понад рік?

Розв'язок

1) Нехай A – подія, яка полягає в тому, що корова взята з ферми А, B – з ферми В, а H – подія, яка полягає в тому, що корова заражена ящуром. Тоді шукана ймовірність є: $p(A \cap H) = p(A) \cdot p(H|A) = (1000/2000) \times (1/5) = 1/10$.

2) Визначимо C як подію, яка полягає в тому, що випадково відібрана корова має вік більше ніж один рік. Шукана ймовірність є:

$$p(B \cap H \cap C) = p(B)p(H|B)p(C|(B \cap H)) = (1/2) \cdot (1/4) \cdot (3/10) = 3/80 .$$

Приклад 3. В одному акваріумі знаходяться три білих, три червоних і три голубих рибки. Трьох випадково вибраних рибок переносять у інший акваріум. Яка ймовірність того, що всі ці три рибки білі?

Розв'язок

Розглянемо два методи рішення цієї задачі. Перший метод полягає в простому перерахуванні всіх можливостей.

Існує $\binom{9}{3} = 84$ способи вибору трьох рибок з дев'яти. Оскільки $\epsilon \binom{3}{3} = 1$

лише спосіб вибору трьох білих рибок, то шукана ймовірність дорівнює $1/84$.

У другому способі використовується теорема множення умовних ймовірностей. Уявимо, що ми вибираємо три рибки по черзі. Визначимо A , B і C як події, що полягають у тому, що перша, друга і третя вибрані рибки виявилися

білими. Ймовірність того, що всі відібрані рибки білі, є $p(A \cap B \cap C) = p(A) \cdot p(B|A) \cdot p(C|A \cap B)$. Зрозуміло, що $p(A) = 3/9 = 1/3$, оскільки є три білі рибки серед дев'яти. Якщо перша вибрана рыбка біла, що серед восьми рибок, які залишилися, опиняться дві білі. Це означає, що $p(B|A) = 2/8 = 1/4$. Аналогічно, $p(C|A \cap B) = 1/7$. Таким чином, $p(A \cap B \cap C) = (1/3) \cdot (1/4) \cdot (1/7) = 1/84$.

Приклад 4. Під час обстеження захворювань легень перевірялись 10 000 людей у віці понад 60 років. Виявилося, що 4 000 людей із цієї групи є постійними курцями. У 1 800 курців виявилися серйозні порушення у легенях. Серед тих, хто не курить, серйозні порушення в легенях мали 1 500 людей. Чи є паління і наявність порушень в легенях незалежними подіями?

Розв'язок

Визначимо A як подію, яка полягає в тому, що випадково вибрана людина є постійним курцем, і B – що у людини серйозні порушення в легенях.

Тоді, $p(A) = 4000/10000 = 0,4$ і $p(B) = 3300/10000 = 0,33$.

Умовна ймовірність паління за умови наявності порушень у легенях є $p(A|B) = p(A \cap B) / p(B) = (1800/10000) / (3300/10000) = 0,55$.

Оскільки $p(A|B) \neq p(A)$. Отже, A та B не є незалежними подіями.

Практична робота №4

Тема: Теорема Байєса. Повторні випробування.

Мета: Ознайомитися з застосуванням теореми Байєса для рішення задач та із застосуванням біноміального і поліноміального розподілів в експериментах з повторними випробуваннями.

Обладнання: Таблиці, калькулятори, комп'ютери.

Контрольні питання:

1. Поняття дискретної ймовірності.
2. Теорема для скінченного простора ймовірностей S , який розбито на підмножини, що не пересікаються.
3. Теорема Байєса.
4. Повторні випробування

5. Теорема про біноміальний розподіл.
6. Теорема про поліноміальний розподіл.

Хід роботи

Робота №1. Розв'язок задач.

Розберіть приклади задач і виконайте завдання для самостійного розв'язку, які запропонує викладач.

Приклади розв'язання задач.

Приклад 1. В популяції сірих мишей $2/3$ особин мають гетерозиготний генотип (Cc), а $1/3$ – гомозиготний генотип (CC). Уявимо, що альбінос (cc) та сіра миша, яка випадковим чином відібрана з популяції, схрещуються і дають потомство чисельністю 4 мишеняти. За умови, що ген альбінізму рецесивний, яка ймовірність того, що всі 4 мишеняти будуть сірими?

Розв'язок

Зрозуміло, що забарвлення потомства залежить від генотипу сірої миші – гомозиготного (CC), або гетерозиготного (Cc). В першому випадку все потомство буде сірим, а в другому – поява альбіносів та сірого потомства має рівну ймовірність. Таким чином,

$$p(\text{всі сірі}) = p(\text{всі сірі}|CC) \cdot p(CC) + p(\text{всі сірі}|Cc) \cdot p(Cc) = 1 \cdot (1/3) + (1/2)^4 \cdot (2/3) = 3/8.$$

Примітка: $p(\text{всі сірі}|CC) = 1$, оскільки все потомство буде мати гетерозиготний генотип Cc.

Приклад 2. Уявимо, що в певній великій популяції чоловіків та жінок порівну. В цій популяції 5% чоловіків і 0,25% жінок страждають дальтонізмом. Випадковим чином вибирають одного дальтоніка. Яка ймовірність того, що ця людина – чоловічої статі?

Розв'язок

Популяція поділена на дві підмножини, які не перетинаються – чоловіків (Ч) та жінок (Ж). Ми шукаємо $p(Ч|Д)$. За теоремою Байєса,

$$p(Ч|Д) = \frac{P(Ч) \cdot P(Д|Ч)}{P(Д|Ч) \cdot P(Ч) + P(Д|Ж) \cdot P(Ж)} = \frac{0,5 \times 0,05}{0,5 \times 0,05 + 0,5 \times 0,0025} = \frac{0,025}{0,02625} = 0,95.$$

Приклад 3. «Середня» людина з ймовірністю $3/5$ виконує певне завдання за 1 хвилину. Припустимо, що завдання виконувалося 10 людьми. Яка ймовірність рівно семи успішно виконаних завдань за 1 хвилину?

Розв'язок

Згідно теореми про біноміальний розподіл $f(n, k, p) = \binom{n}{k} p^k q^{n-k}$

Для даного випадку $n=10, k=7, p=3/5, q=1-p=1-3/5=2/5$.

$$\text{Отже, } f(10, 7, 3/5) = \binom{10}{7} (3/5)^7 (2/5)^3 = 0,215.$$

Приклад 4. В деякій великій популяції 20 % є шульги, 70 % людей, краще володіють правою рукою, ніж лівою, і 10 % – однаково вільно володіють обома руками. Якщо з популяції випадково вибирають 10 людей, то яка ймовірність того, що всі вони володіють правою рукою краще, ніж лівою? що семеро володіють правою рукою краще ніж лівою, двоє є шульгами і одна людина однаково вільно володіє обома руками?

Розв'язок

Використовуємо теорему про поліноміальний розподіл:

$$P(n_1, n_2, \dots, n_m) = \binom{n}{n_1, n_2, n_3} p_1^{n_1} p_2^{n_2} \dots p_m^{n_m}$$

Для даної задачі $n = 10$; $p_1 = 0,7$; $p_2 = 0,2$ та $p_3 = 0,1$;

$$\text{тобто: } P(10, 0, 0) = \binom{10}{10, 0, 0} 0,7^{10} 0,2^0 0,1^0 = 0,7^{10} = 0,028;$$

$$P(7, 2, 1) = \binom{10}{7, 2, 1} 0,7^7 0,2^2 0,1 = 0,119.$$

МОДУЛЬ 2. ОСНОВИ ВАРІАЦІЙНОЇ СТАТИСТИКИ

Практична робота №5

Тема: Фіксація і початкове впорядкування даних. Побудова графіків варіаційних рядів.

Мета: Навчитись фіксувати дані, розташовувати їх за ранжиром, складати ряди розподілу і будувати графіки розподілу частот..

Матеріал і обладнання: лінійки, міліметровий папір, роздатковий матеріал (набори з 50-100 листків берези, або тополі, клена та ін.; черепашки моллюсків, тощо), калькулятори, комп'ютери.

Контрольні питання:

1. Варіаційна крива.
2. Гістограма.
3. Полігон частот.
4. Кумулята і огіва.
5. Якісні дані.
6. Порядкові дані.
7. Кількісні дані.
8. Ранжирування.
9. Варіаційний ряд розподілу.

Хід роботи

Робота №1. Ранжирування варіаційного ряду

Розставте за ранжиром запропоновані дані і результати запишіть у зошит.

У кожному ряді, який ранжирують, визначить $x_1=x_{max}$, $x_n=x_{min}$ та розмах варіації R :

а) кількість личинок жуків-коваликів на 1 м^2 ґрунту:

3, 7, 1, 11, 7, 8, 7, 2, 7, 3, 6, 3, 1, 3;

б) діаметр клітин (у $\mu\text{м}$):

14, 19, 28, 18, 22, 23, 20, 21, 27, 17, 24, 21, 20, 21, 23, 19, 23, 22, 21, 22, 24, 20, 24, 20;

в) розміри тіла горала (у см): 115, 106, 113, 109, 110, 118, 117, 110.

Робота №2. Складання ряду розподілу варіант (на прикладі дискретних даних)

Розгляньте і оформіть в зошиті приклад складання ряду розподілу варіант.

Приклад. Кількість часточок вугілля у вакуолях інфузорії-парамеції:

0 1 1 4 1 3 1 1 2 0 3 1 3 2 1 0 2 2 2 3 2 3 3 4 1 0 2 3 3 3 1 3 2
1 1 3 3 2 3 3 1 2 1 4 2 3 1 2 3 1 1 3 2 4 1 4 0 1 1 1 2 2 3 3.

Визначимо розмах мінливості, тобто максимальне та мінімальне значення варіант ($x_{max}=4$; $x_{min}=0$). Випишемо по порядку всі значення в цьому проміжку, а потім рухаючись по порядку запису у зошиті, робимо відмітки напроти відповідних значень у виписаному ряду. Отримуємо (табл. 1):

Таблиця 1

x_i :	0	1	2	3	4
f_i :	5	20	15	19	5
	•— ••	⊗ ⊗	⊗ •— ••	⊗ ⊗	•— ••

Сума частот дорівнює об'єму вибірки $\sum_{i=1}^n f_i = n = 64$.

Робота №3. Складання ряду розподілу варіант (на прикладі неперервних даних)

а) Розгляньте приклад складання ряду розподілу варіант і оформіть у зошиті алгоритм обробки варіаційного ряду.

Приклад. Довжина максимального стеблового листа (у мм) у 62 рослин королиці звичайної: 56; 61; 66; 67; 53; 84; 82; 80; 44; 68; 60; 61; 54; 60; 60; 62; 62; 55; 55; 54; 57; 47; 55; 60; 63; 38; 59; 45; 49; 57; 78; 68; 43; 62; 64; 48; 74; 43; 40; 53; 74; 72; 76; 77; 86; 50; 69; 69; 71; 76; 47; 71; 54; 63; 66; 56; 52; 87; 44; 69; 64; 64.

1) Визначаємо розмах мінливості $R = x_{max} - x_{min}$ від 38 до 87 і об'єм вибірки $n = 62$.

2) Визначаємо приблизне число класів за Старджесом: $k=7$.

3) Визначаємо класовий інтервал: $\lambda = \frac{(87 - 38)}{7} = 7$.

4) Класові границі можна вибрати наступні:

37-44-51-58-65-72-79-86-93. Але в такому випадку права границя останнього класу – 93. Весь ряд дещо зсувається вправо. Тому краще вибрати наступні границі класів: $x' = 35 - 42 - 49 - 56 - 63 - 70 - 77 - 84 - 91$.

Більш зручний запис для рознесення даних коли вказують початок і кінець кожного клас-інтервалу: 35-41 42-48 49-55 56-62 63-69 70-76 77-83 84-90.

5) Результат рознесення (табл. 2):

Таблиця 2

Розподіл листя королиці звичайної за довжиною максимального стеблового листа, мм (із застосуванням методу конвертів)

Клас-інтервали	35-41	42-48	49-55	56-62	63-69	70-76	77-83	84-90
f_i	• •	□	⊗	⊗	⊗	□	• •	• •
			•	• •	• •		•	•
				• •	•			

Після цього знайдемо x_{ci} – центральне, або середнє класове значення. Воно може бути отримане обчисленням, як середнє арифметичне границь класу. У першій колонці, наприклад, $(35+41) : 2=38$, в інших x_{ci} знаходимо аналогічно.

Частоти в кожному класі записуємо у цифровому вигляді. Отже, таблиця результатів набуває остаточного вигляду (табл. 3).

Таблиця 3

Розподіл листя королиці звичайної за довжиною максимального стеблового листа, мм

x_{ci}	38	45	52	59	66	73	80	87
f_i	2	8	11	14	13	8	3	3

б) Проведіть вимірювання зразків. Зафіксуйте результати і складіть ряд розподілу варіант аналогічно до попереднього прикладу.

в) На основі даних експерименту, які запропонує викладач, складіть ряд розподілу варіант. Результат оформіть у вигляді таблиці частот.

Робота №4. Побудова варіаційної кривої

Накресліть графік варіаційної кривої для ряду розподілу діаметра клітин (у мкм):

x_i	14	17	18	19	20	21	22	23	24	27	28
f_i	1	1	1	2	4	4	3	3	3	1	1

За допомогою програми Excel наберіть запропоновані дані на комп'ютері. Використайте майстер діаграм і побудуйте варіаційну криву.

Робота №5. Побудова гістограми

Накресліть стовпчасту гістограму для ряду розподілу кількості часточок вугілля у вакуолях інфузорії-туфельки, що представлений у табл. 1:

За допомогою програми MS Excel і майстра діаграм побудуйте стовпчасту (тип *Гістограма*) та кругову (тип *Кругова*) гістограми на комп'ютері.

Робота №6. Побудова полігону частот

Накресліть полігон частот для ряду розподілу варіант, представлений у табл. 1.

Побудуйте полігон частот на комп'ютері. Використайте майстер діаграм (тип *Графік*) у програмі MS Excel.

Робота №7. Побудова кривих накопичених частот

Для ряду розподілу варіант, представлений у табл. 1, побудуйте ряд накопичених частот.

Для цього треба перейти від середніх класових значень x_i до міжкласових границь x_i' , а потім підсумувати частоти (напрямо позначений стрілками на табл. 4) і отримати кумулятивний ряд, або ряд накопичених частот $\sum f_i$.

Додаючи частоти (напрямо позначений стрілками), отримуємо ряд, що складається з накопичених частот. Графік такого ряду з використанням накопичених частот називається кривою накопичених частот (кумулятою або огівною).

		Таблиця 4					
x_i		0	1	2	3	4	
x_i'	-0,5	0,5	1,5	2,5	3,5	4,5	
f_i		5	20	15	19	5	
$\sum f_i$	<u>0</u>	5	<u>25</u>	<u>40</u>	<u>59</u>	<u>64</u>	

Кумуляту будують на тій же координатній сітці, що і ламану розподілу, але по осі ординат відкладають накопичені частоти, а по осі абсцис відкладають значення класів із послідовним з'єднанням точок прямими лініями.

Для отримання огіви необхідно по осі абсцис відкладати накопичені частоти, а по осі ординат значення класів із послідовним з'єднанням геометричних точок прямими лініями.

Накресліть кумуляту і огіву за даними табл. 4.

Побудуйте криві накопичених частот на комп'ютері. Використайте майстер діаграм (тип *Графік*) у програмі Excel.

Практична робота №6

Тема: Показники положення і мінливості.

Мета: Навчитись підраховувати показники положення і мінливості способами умовної середньої, добутків, сум.

Матеріал і обладнання: калькулятори, комп'ютери, роздатковий матеріал для вимірювання ознак, лінійки.

Контрольні питання

1. Показники положення.
2. Показники мінливості.
3. Метод умовної середньої.
4. Метод добутків.
5. Метод доданків.

Хід роботи

Робота №1. Обробка варіаційного ряду способом добутків

Цей спосіб заснований на використанні відхилень варіант даної сукупності від їх середньої величини.

Розгляньте і оформіть в зошиті приклади обробки варіаційного ряду з варіантами, що не повторюються, і з варіантами, що повторюються.

Приклад 1 (варіанти у вибірці не повторюються). Обчислити середні величини та показники варіації визначень вітаміну С в паростках гороху (табл. 5).

Таблиця 5

№ п/п	Значення варіант x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1.	10,2	-1,5	2,25
2.	10,7	-1,0	1,0
3.	10,9	-0,8	0,64
4.	11,4	-0,3	0,09
5.	11,7	0	0
6.	11,8	0,1	0,01
7.	12,0	0,3	0,09
8.	12,1	0,4	0,16
9.	12,3	0,6	0,36
10.	12,6	0,9	0,81
11.	13,0	1,3	1,69

$$n = 11 \quad \sum x_i = 128,7 \quad \sum (x_i - \bar{x}) = 0 \quad \sum (x_i - \bar{x})^2 = 7,2$$

$$\bar{x} = \frac{1}{n} \sum x_i = 11,7. \quad S_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{7,2}{10}} = \pm 0,84.$$

Приклад 2 (варіанти у вибірці повторюються). Було проведено серію вимірювань діаметрів дерев певного виду в досліджуваному лісовому масиві. Було взято 50 об'єктів і одержано відповідну сукупність даних, яку згруповано по класах (табл. 6).

Таблиця 6

Класи	x_{u_i}	f_i	$f_i x_{u_i}$	$x_{u_i} - \bar{x}$	$(x_{u_i} - \bar{x})^2$	$f_i (x_{u_i} - \bar{x})^2$
41-45	43	2	86	-12,5	156	312
46-50	48	9	432	-7,5	56	504
51-55	53	14	742	-2,5	6	84
56-60	58	15	870	2,5	6	90
61-65	63	7	441	7,5	56	392
66-70	68	3	204	12,5	156	468
		$n = 50$	$\sum f_i \cdot x_{u_i} = 2775$			$\sum f_i (x_{u_i} - \bar{x})^2 = 1850$

Цей спосіб полягає в безпосередньому оперуванні частотами варіант f_i у класах, їх центральними значеннями x_{u_i} і відхиленнями останніх від середнього арифметичного значення \bar{x} . Відповідні формули мають такий вигляд:

$$\bar{x} = \frac{1}{n} \sum f_i x_{u_i}; \quad S_x = \sqrt{\frac{\sum f_i \cdot (x_{u_i} - \bar{x})^2}{n-1}}$$

У таблиці наведено розрахунки середнього арифметичного і середнього квадратичного відхилень методом добутків.

$$\bar{x} = \frac{1}{n} \sum f_i \cdot x_{u_i} = \frac{2775}{50} = 55,5 \text{ см}; \quad S_x = \sqrt{\frac{\sum f_i (x_{u_i} - \bar{x})^2}{n-1}} = \sqrt{\frac{1850}{49}} = \pm 6,15 \text{ см}.$$

Робота №2. Обробка варіаційного ряду способом умовної середньої (спосіб умовного нуля)

Візьмемо варіаційний ряд (ознака x – ширина листка у *Odontites litoralis*). Дані занесені у табл. 7.

Таблиця 7

$x, \text{ мм}$	x_{δ_3}	f_i	a_i	$f_i a_i$	$f_i a_i^2$
1	2	3	4	5	6
1,0–1,9	1,45	7	–2	–14	28
2,0–2,9	2,45	42	–1	–42	42
3,0–3,9	$A=3,45$	52	0	0	0
4,0–4,9	4,45	36	+1	+36	36
5,0–5,9	5,45	11	+2	+22	44
6,0–6,9	6,45	2	+3	+6	18
		$n=150$		$\sum f_i a_i = +8$	$\sum f_i a_i^2 = 168$

Спочатку в окремому стовпці виписують центральні значення класів x_{δ_3} та відповідні до них частоти f_i . Потім вибирають умовну середню A – варіанту, або центральне класове значення, що відповідає двом вимогам: умовна середня повинна по можливості відповідати максимальній частоті і в той же час знаходитися приблизно посередині ряду. Якщо ряд асиметричний, то вона може лежати у сусідньому з модальним класі, ближче до середини ряду. В нашому прикладі умовну середню краще вибрати в модальному класі (3,0–3,9), а її значення буде відповідно $A=3,45$. Потім підраховують відхилення варіант (a_i) від умовної середньої через $a_i = (x_{u_i} - A)$.

$$\bar{x} = A + \frac{\sum f_i \cdot a_i}{n}, \text{ або } \bar{x} = A + \frac{\sum f_i \cdot (x_{u_i} - A)}{n} \quad S_x = \sqrt{\left(\frac{n}{n-1}\right) \left[\frac{\sum f_i a_i^2}{n} - \left(\frac{\sum f_i a_i}{n}\right)^2 \right]}.$$

Можна спростити обчислення, якщо відхилення класових варіант від умовної середньої A віднести до величини класового інтервалу, тобто замість $a_i = (x_{u_i} - A)$, брати $a_i = \frac{(x_{u_i} - A)}{\lambda}$. Тоді для всіх випадків (для рівноінтервальних рядів) відхилення варіант від умовної середньої A , де $a=0$, перетвориться в ряд 1, 2, 3 і т. д., які розглядають в сторону менших, ніж A , значень варіант з «–», а в сторону більших – з «+». Тоді формули мають набувають вигляду:

$$\bar{x} = A + \lambda \cdot \frac{\sum f_i \cdot a_i}{n}, \quad S_x = \lambda \cdot \sqrt{\left(\frac{n}{n-1}\right) \left[\frac{\sum f_i a_i^2}{n} - \left(\frac{\sum f_i a_i}{n}\right)^2 \right]}.$$

У стовпці 4 розрахункової таблиці положення умовної середньої A позначається нулем, а верх і вниз від нуля беремо відлік в натуральних ($-1, -2, \dots; +1, +2, \dots$) Ці відліки характеризують відхилення центральних класових інтервалів $x_{\dot{o}_3}$ від умовної середньої, замінюючи менш зручні натуральні відхилення.

Стовпчик 5 містить в собі добутки частот на відповідні відхилення класів від умовної середньої. В останньому стовпці ці добутки ще раз множимо на величину a_i , і отримуємо добуток частот на квадрат відхилень від умовної середньої.

Стовпчик 5 містить дані, які необхідні для підрахунку середньої арифметичної \bar{x} , а стовпчик 6 – дані для підрахунку середнього квадратичного відхилення S_x . Цій меті служать суми $\sum f_i a_i$ та $\sum f_i a_i^2$, що розміщені у нижньому рядку розрахункової таблиці.

Для розрахунку використовують робочі формули, за умови, що $A=3,45$, а класовий інтервал $\lambda = 1$. Спочатку знаходимо середнє арифметичне

$$\bar{x} = A + \lambda \cdot \frac{\sum f_i \cdot a_i}{n} = 3,45 + 1 \cdot \frac{8}{150} = 3,50 \text{ мм.}$$

Потім підраховуємо середнє квадратичне відхилення

$$S_x = \lambda \cdot \sqrt{\left(\frac{n}{n-1}\right) \left[\frac{\sum f_i a_i^2}{n} - \left(\frac{\sum f_i a_i}{n}\right)^2 \right]} = 1 \cdot \sqrt{\left(\frac{150}{150-1}\right) \left[\frac{168}{150} - \left(\frac{8}{150}\right)^2 \right]} = \pm 1,06 \text{ мм.}$$

Коефіцієнт варіації визначаємо за формулою $C_v = \frac{S_x}{\bar{x}} \cdot 100\% = \frac{1,06}{3,50} \% = 30,5\%$.

Підраховуємо моду та медіану:

$$M_o = x_{\text{мод}} + \lambda \cdot \left(\frac{f_2 - f_1}{2f_2 - f_1 - f_3} \right) = 3,0 + 1 \cdot \left(\frac{52 - 42}{2 \cdot 52 - 42 - 36} \right) = 3,38 \text{ мм,}$$

$$M_d = x_{\text{мод}} + \lambda \cdot \left(\frac{\frac{n}{2} - \sum f_i}{f_{M_d}} \right) = 3,0 + 1 \cdot \left(\frac{0,5 \cdot 150 - 49}{52} \right) = 3,50 \text{ мм.}$$

Робота №3. Обробка варіаційного ряду способом сум

Спосіб сум відрізняється тим, що в розрахунковій таблиці дія множення замінюється дією додавання.

Цей спосіб також заснований на використанні умовної середньої. Візьмемо той же варіаційний ряд (ознака x – ширина листка у *Odontites litoralis*). Переконаємось чи співпадають результати за цим (табл. 8) і попереднім способами визначення середнього арифметичного.

Спочатку обираємо умовну середню $A=3,45$. В чотирьох наступних стовпцях проти умовної середньої робимо прочерки, виключаючи відповідні клітини розрахункової таблиці від подальшої обробки. Число таких прочерків послідовно збільшують у такій послідовності: у 4-му стовпці – 1 прочерк, у 5-му – 3, у 6-му – 5 і в 7-му – 7 прочерків. Якщо ряд вкорочений, то не всі прочерки можуть поміститися у таблиці, а крім того, один, або два останніх стовпця можуть залишитися без цифр.

Таблиця 8

x , мм	$x_{\bar{o}_s}$	f_i	$a_1=56$	$a_2=7$	$a_3=0$	$a_4=0$
1	2	3	4	5	6	7
1,0–1,9	1,45	7	7	7	—	—
2,0–2,9	2,45	42	49	—	—	—
3,0–3,9	$A=3,45$	52	—	—	—	—
4,0–4,9	4,45	36	49	—	—	—
5,0–5,9	5,45	11	13	15	—	—
6,0–6,9	6,45	2	2	2	2	—
Пояснення:		$n=150$	$b_1=64$	$b_2=17$	$b_3=2$	$b_4=0$
$s=a+b$			$s_1=120$	$s_2=24$	$s_3=2$	$s_4=0$
$d=b-a$			$d_1=+8$	$d_2=+10$	$d_3=+2$	—

Наступний етап роботи полягає в написанні так званих начотних рядів окремо у верхній та нижній половині кожного стовпця. Ці ряди складаються шляхом підсумовування чисел у напрямку показаному стрілками. Потім в кожній половині стовпців підрахувати суми a_1, a_2, a_3, a_4 та b_1, b_2, b_3, b_4 , які і записуються у відповідних графах. Після того, у стовпцях підраховують загальні суми за формулою $s=a+b$ та різниці за формулою $d=b-a$. Різниця d_4 для звичайної мети не потрібна.

Для розрахунку параметрів ряду попередньо підраховують три статистичних моменти: умовний момент першого порядку (\bar{v}_1), умовний момент другого порядку (\bar{v}_2), центральний момент другого порядку (μ_2):

$$\bar{v}_1 = \frac{d_1}{n} = \frac{8}{150} = 0,0533; \quad \bar{v}_2 = \frac{s_1 + 2s_2}{n} = \frac{120 + 2 \cdot 24}{150} = \frac{168}{150} = 1,1200;$$

$$\mu_2 = \bar{v}_2 - \bar{v}_1^2 = 1,1200 - 0,0533^2 = 1,1172.$$

Підставляємо знайдені значення у робочі формули, отримуємо середнє арифметичне і середнє квадратичне відхилення:

$$\bar{x} = A + \lambda \cdot \bar{v}_1 = 3,45 + 1 \cdot 0,0533 = 3,50 \text{ мм}; \quad S_x = \lambda \cdot \sqrt{\mu_2} = 1 \cdot \sqrt{1,1172} = \pm 1,06 \text{ мм}.$$

Робота №4. Статистична обробка варіаційного ряду на комп'ютері

На основі даних експерименту, які запропонує викладач, проведіть їх статистичну обробку. Використайте для роботи майстер функцій (категорія *Статистичні*) програми MS Excel. Внесіть у комп'ютер дані і отримайте мінімальне та максимальне значення варіант, кількість варіант у вибірці, середнє арифметичне значення і дисперсію, стандартне відхилення, моду, медіану. Використовуючи отримані значення напишіть необхідну формулу для коефіцієнта варіації у програмі MS Excel і порахуйте його значення. та їх результати

Оформіть у текстовому редакторі MS Word позначення всіх отриманих статистичних величин, підставте їх значення і роздрукуйте.

Практична робота №7

Тема: t -критерій Ст'юдента. F -критерій Фішера.

Мета: Навчитись застосовувати t -критерій Ст'юдента і F -критерій Фішера для статистична оцінки істотності різниці середніх та мінливості.

Матеріал і обладнання: калькулятори, комп'ютери, розрахункові таблиці.

Контрольні питання

1. Критерій Ст'юдента.
2. Коефіцієнт Ст'юдента.

3. Статистична оцінки істотності різниці середніх арифметичних значень для будь-яких великих або малих однакових вибірок.
4. Статистична оцінки істотності різниці середніх арифметичних значень для малих неоднакових вибірок
5. Статистична оцінки істотності різниці середніх арифметичних значень для парних контрольних та експериментальних даних.
6. Критерій Фішера.

Хід роботи

Робота 1. Порівняння двох однакових і малих за обсягом груп експериментальних даних

Розгляньте приклад і запишіть його у зошит.

Приклад. Досліджували вміст вітаміну В₁₂ у суспензії дріжджів в умовах експериментального мутагенезу. Було проведено по три аналізи для контрольного (вихідного) і мутантного штабів і дістали такі дані (в умовних одиницях): контрольний штаб – 5,7; 6,4; 7,1; мутантний штаб – 8,2; 6,6; 6,9. Треба зробити висновок про продуктивність досліджуваного мутанта дріжджів щодо вітаміну В₁₂ порівняно з вихідним штабом. Запишемо обидві групи даних в одну таблицю (табл. 9).

Таблиця 9

Вихідний штаб			Мутантний штаб		
Значення варіант	$(x_{i_1} - \bar{x}_1)$	$(x_{i_1} - \bar{x}_1)^2$	Значення варіант	$(x_{i_2} - \bar{x}_2)$	$(x_{i_2} - \bar{x}_2)^2$
5,7	-0,7	0,49	8,2	1,0	1,00
6,4	0	0	6,6	-0,6	0,36
7,1	0,7	0,49	6,9	-0,3	0,09

Отже, для вихідного (контрольного) штабу $n_1 = 3$; $\sum x_{i_1} = 19,2$; $\bar{x}_1 = 6,4$;

$$m_{\bar{x}_1}^2 = \frac{\sum (x_{i_1} - \bar{x}_1)^2}{n_1 \cdot (n_1 - 1)} = \frac{0,98}{6} = 0,16;$$

Для мутантного штабу $n_2 = 3$; $\sum x_{i_2} = 21,7$; $\bar{x}_2 = 7,2$;

$$m_{x_2}^2 = \frac{\sum (x_{i_2} - \bar{x}_2)^2}{n_2 \cdot (n_2 - 1)} = \frac{1,45}{6} = 0,24;$$

$$d = |\bar{x}_1 - \bar{x}_2| = 7,2 - 6,4 = 0,8;$$

$$m_d = \sqrt{m_{\bar{x}_1}^2 + m_{\bar{x}_2}^2} = \sqrt{0,16 + 0,24} = \sqrt{0,40} = 0,63;$$

$$t = \frac{d}{m_d} = \frac{0,80}{0,63} = 1,3.$$

В обох групах було проведено всього 6 аналізів, а отже, $\nu = 3 + 3 - 2 = 4$.

Згідно табл. 8 [4] значенню $t = 1,3$ для графі $\nu = 4$ відповідає ймовірність $p = 0,716$.

Ця ймовірність є занадто малою, щоб на основі проведених дослідів стверджувати, що мутантний штам статистично істотно продукує більше вітаміну B₁₂, ніж вихідний штам. Отже, щоб оцінити промислове значення нового штаму, треба продовжити його лабораторне вивчення.

Робота 2. Порівняння двох неоднакових і малих за обсягом груп експериментальних даних

Розгляньте приклад і запишіть його у зошит.

Приклад. Визначали електропровідність м'язової тканини в умовах опромінення потоком нейтронів. Було проведено 13 дослідів при опроміненні і 9 – у нормальних умовах (контроль). Дослідника цікавить питання, чи впливає опромінення нейтронами (при даній дозі) на електропровідність м'язової тканини. Результати вимірювань подали в умовних одиницях і звели в дві таблиці за аналогічними з попереднім прикладом схемами. У табл. 10 наведено дані контрольної серії аналізів, а в табл. 11 – дослідної.

Таблиця 10

Порядковий номер	Значення варіант (x_{i_1})	$(x_{i_1} - \bar{x}_1)$	$(x_{i_1} - \bar{x}_1)^2$
1	2,6	-1,0	1,00
2	2,9	-0,7	0,49
3	3,3	-0,3	0,09
4	4,8	1,2	1,44
5	3,7	0,1	0,01
6	3,0	-0,6	0,36
7	4,2	0,6	0,36
8	4,2	0,6	0,36
9	3,8	0,2	0,04
$n_1 = 9$	$\sum x_{i_1} = 32,5$ $\bar{x}_1 = 3,6$		$\sum (x_{i_1} - \bar{x}_1)^2 = 4,96$

Таблиця 11

Порядковий номер	Значення варіант (x_{i_2})	$(x_{i_2} - \bar{x}_2)$	$(x_{i_2} - \bar{x}_2)^2$
1	4,3	-1,3	1,69
2	7,2	1,6	2,56
3	5,4	-0,2	0,04
4	4,1	-1,5	2,25
5	4,6	-1,0	1,00
6	6,6	1,0	1,00
7	3,9	-1,7	2,89
8	7,0	1,4	1,96
9	6,3	0,7	0,49
10	5,7	0,1	0,01
11	5,4	-0,2	0,04
12	6,8	1,2	1,44
13	4,9	-0,7	0,49
$n_2 = 13$	$\sum x_{i_2} = 72,2$ $\bar{x}_2 = 5,6$		$\sum (x_{i_2} - \bar{x}_2)^2 = 15,86$

Далі визначаємо:

$$d = |5,6 - 3,6| = 2,0;$$

$$m_d = \sqrt{\left(\frac{4,96 + 15,86}{9 + 13 - 2}\right)\left(\frac{9 + 13}{9 \cdot 13}\right)} = 0,443;$$

$$t = \frac{2,0}{0,443} = 4,5.$$

Величину ймовірності, яка відповідає коефіцієнту $t = 4,5$, визначаємо за таблицею Ст'юдента. Загальна кількість проведених дослідів була $n_1 + n_2 = 9 + 13 = 22$, отже, $\nu = 20$.

З табл. 8 [4] видно, що вірогідність підвищення електропровідності м'язової тканини під впливом опромінення нейтронами доведено з імовірністю більшою ніж 0,999.

Робота 3. Порівняння парних експериментів: контрольних і дослідних аналізів

Розгляньте приклад і запишіть його у зошит.

Приклад. Дослідника цікавить питання про вплив γ -випромінювання на вміст дезоксирибонуклеїнової кислоти (ДНК) в печінці. Досліди були поставлені на морських свинках. Щоб забезпечити однорідність матеріалу, однакових за

віком і статтю тварин відбирали попарно з генетично однорідної групи, і з кожної пари тварин одна була контролем, а другу опромінювали променями. Аналізи вмісту ДНК в печінці проводили за однаковою методикою, причому контрольну і дослідну серію аналізів ставили паралельно в однакових умовах.

Добуті дані (було поставлено 8 пар аналізів) подано в табл. 12 у відповідних одиницях визначень.

Таблиця 12

Порядковий номер	Контрольна серія	Серія з опроміненням	Попарна різниця (x_{d_i})	$x_{d_i} - \bar{x}_d$	$(x_{d_i} - \bar{x}_d)^2$
1	12,9	10,3	2,6	1,2	1,40
2	9,0	7,9	1,1	-0,3	0,09
3	11,6	10,1	1,5	0,1	0,01
4	10,5	8,2	2,3	0,9	0,81
5	9,8	9,9	-0,1	-1,5	2,25
6	9,2	10,1	-0,9	-2,3	5,29
7	11,1	9,7	1,4	0	0
8	12,2	8,7	3,5	2,1	4,41
$n = 8$			$\sum x_{d_i} = 11,4$		$\sum (x_{d_i} - \bar{x}_d)^2 = 14,26$

Отже, середня різниця $\bar{x}_d = \frac{\sum x_{d_i}}{n} = 1,4$,

а її стандартна похибка $m_d = \sqrt{\frac{\sum (x_{d_i} - \bar{x}_d)^2}{n \cdot (n-1)}} = \sqrt{\frac{14,26}{56}} = \sqrt{0,25} = 0,5$.

Коефіцієнт Ст'юдента у цьому випадку дорівнює:

$$t = \frac{\bar{x}_d}{m_d} = \frac{1,4}{0,5} = 2,8; \quad \nu = n - 1 = 8 - 1 = 7. \quad \text{Значенню } t = 2,8 \text{ для } \nu = 7 \text{ в табл. 8 [4]}$$

відповідає ймовірність $p = 0,969$. Отже, можна стверджувати, що γ -опромінення в досліджуваній дозі приводить до зменшення вмісту ДНК в печінці морських свинок.

Робота 4. Статистичне доведення істотності різниці в мінливості двох експериментальних груп

Розгляньте приклад і запишіть його у зошит.

Приклад. При дослідженні нового методу зберігання м'яса було зібрано такі дані: а) контрольний спосіб (1) – 120 ± 6 днів ($n_1 = 20$); б) досліджуваний спосіб

(2) – 128 ± 9 днів ($n_2 = 50$). Абсолютна різниця в часі зберігання ($d = 8$ днів) між обома методами статистично неістотна, оскільки ймовірність її істотності мала.

Справді, $t = \frac{8}{\sqrt{6^2 + 9^2}} = \frac{8,0}{10,8} = 0,74$, що відповідає ймовірності $p = 0,54$. Однак

різниця в мінливості часу зберігання між обома методами є статистично істотною. Щоб пересвідчитися в цьому, обчислимо спочатку величини дисперсій для обох методів і на їх основі – коефіцієнт Фішера F .

$$S^2_{x_1} = (m_{\bar{x}_1} \cdot \sqrt{n_1})^2 = m_{\bar{x}_1}^2 \cdot n_1 = 36 \cdot 20 = 720.$$

$$S^2_{x_2} = (m_{\bar{x}_2} \cdot \sqrt{n_2})^2 = m_{\bar{x}_2}^2 \cdot n_2 = 81 \cdot 50 = 4050.$$

$$F_{експ} = \frac{4050}{720} = 5,62.$$

Табличне значення визначаємо на перетині горизонтальної графі з $\nu_1 = 19$ і вертикальної графі $\nu_2 = 49$ (50 є найближче значення) з табл. 9 [4]. Дістаємо $F_{табл} = 2,00$ при ймовірності $p = 0,95$. У зв'язку з тим, що в нашому прикладі $F_{експ} > F_{табл}$ можна стверджувати, що різниця між обома методами за їх мінливістю статистично істотна з високим рівнем ймовірності ($p = 0,95$).

Практична робота №8

Тема: Оцінка характеру розподілу. Розв'язання прикладів на нормальний закон розподілу та за характером розподілу. Оцінки істотності належності експериментального розподілу частот у класах до теоретично відомого стандарту.

Мета: Навчитись оцінювати характер розподілу за даними вибірки. Оволодіти методикою розв'язання прикладів для оцінки істотності належності експериментального розподілу частот у класах до теоретично відомого стандарту, а також до нормального розподілу.

Матеріал і обладнання: калькулятори, розрахункові таблиці.

Контрольні питання:

1. Критерій Пірсона хі-квадрат.

2. Оцінка істотності належності експериментального розподілу частот у класах до теоретично відомого стандарту.
3. Закон нормального розподілу.

Хід роботи

Робота 1. Оцінка істотності належності експериментального розподілу частот у класах до теоретично відомого стандарту

Розгляньте приклад і запишіть його у зошит.

Приклад. При схрещуванні двох сортів помідорів, які мали жовті й червоні плоди, в другому поколінні спостерігали розщеплення ознак у такій пропорції – 312 червоних і 88 жовтих. Всього було обстежено 400 гібридів. Чи відповідає це розщеплення класичному закону Менделя, згідно з яким ознаки в другому поколінні розщеплюються за пропорцією 3 : 1?

Складемо таблицю з графами, необхідними для обчислення критерію хі-квадрат (табл. 13). Кількість ступенів вільності $\nu = 1$, бо кількість класів k у даному випадку 2, а $\nu = k - 1$.

Таблиця 13

Гібриди	Фактична кількість (f)	Теоретично обчислена кількість (f') для співвідношення 3 : 1	$ f - f' $	$(f - f')^2$	$\frac{(f - f')^2}{f'}$
Червоні	312	300	12	144	0,48
Жовті	88	100	12	144	1,44
	$n = 400$	$n = 400$			$\Sigma = 1,92$

У табл. 10 [4] знаходимо, що для $\nu = 1$ граничне значення $\chi^2_{табл, p=0,95}$, яке свідчить про істотну різницю в розподілах фактичних (f) і теоретично очікуваних (f') даних, дорівнює 3,84. У нашому випадку $\chi^2_{експ.} = 1,92$. Гіпотеза про належність характеру розщеплення гібридів в експерименті до теоретично передбаченого законом Менделя підтвердилась, бо $\chi^2_{табл, p=0,95} > \chi^2_{експ.}$, а отже різниця в розподілах неістотна.

Робота 2. Оцінка істотності належності експериментального розподілу до нормального розподілу

Розгляньте приклад і запишіть його у зошит.

Приклад. Досліджували діаметри стовбурів сосни (у см) одного віку на експериментальній ділянці. Проведено 200 обстежень і після вибору класового інтервалу $\lambda = 2$ дані згрупували у 8 класів ($k = 8$) з частотами f_j , як це показано в гр. 1–3 (табл. 14).

У гр. 4–7 наведено елементи формул для обчислення середнього арифметичного \bar{x} і середнього квадратичного відхилення S_x . Обчислення елементів для визначення теоретичних частот f'_j у класах подано у гр. 8–10. Для кожного класу обчислюють нормоване відхилення t_j (гр. 8), за табл. 2 [4] для кожного t_j визначають ординати функції густини нормального розподілу $f(t_j)$ (гр. 9) і за формулою обчислюють для кожного класу теоретичні частоти f'_j (гр. 10). Для зручності підрахунків емпіричні частоти наводяться у гр. 11 (гр. 11 дублює гр. 3). В тих класах, де кількість частот менша за 5, частоти об'єднують з частотами сусідніх класів (гр.10* і 11*), кількість класів за рахунок об'єднання при цьому зменшується.

Таблиця 14

Класи	Центральні значення класів (x_{u_i})	Фактично добуті частоти (f_i)	$f_i \cdot x_{u_i}$	$x_{u_i} - \bar{x}$	$(x_{u_i} - \bar{x})^2$	$f_i(x_{u_i} - \bar{x})^2$
1	2	3	4	5	6	7
8,1–10,0	9	3	27	–6,76	45,7	137,1
10,1–12,0	11	19	209	–4,76	22,7	431,3
12,1–14,0	13	31	403	–2,76	7,6	235,6
14,1–16,0	15	49	735	–0,76	0,58	28,4
16,1–18,0	17	57	969	1,24	1,54	87,8
18,1–20,0	19	27	513	3,24	10,5	283,5
20,1–22,0	21	13	273	5,24	27,5	357,5
22,1–24,0	23	1	23	7,24	52,4	52,4
$k = 8$ $\lambda = 2$		$n = 200$	$\sum f_i \cdot x_{u_i} = 3152$ $\bar{x} = 15,76$		$\sum f_i(x_{u_i} - \bar{x})^2 = 1613,6$ $S_x = 2,85$	

Класифікація експериментальних даних				Обчислення \bar{x} та S_x				
$t = \left \frac{x_{u_j} - \bar{x}}{S_x} \right $	Функція нормальної густини ймовірності $f(t_j)$	Теоретичні частоти f_j'	f_j'	Емпіричні частоти f_j	f_j	$ f_j - f_j' $	$(f_j - f_j')^2$	$\frac{(f_j - f_j')^2}{f_j'}$
8	9	10	10*	11	11*	12	13	14
2,37	0,024	3	17	3	22	5	25	1,47
1,67	0,099	14	35	19	31	4	16	0,46
0,97	0,249	35	55	49	49	6	36	0,65
0,27	0,385	55	52	57	57	5	25	0,47
0,44	0,362	52	29	27	27	2	4	0,14
1,14	0,208	29	12	13	14	2	4	0,33
1,84	0,073	10		1				
2,54	0,016	2						
		$n = 200$ $k = 8.$	$k = 6$	$n = 200$ $k = 8.$	$k = 6$	$k = 6$	$\chi^2_{експ.} = 3,52$ $\nu = k - 3 = 3$	
Обчислення теоретичних частот у класах (f')				Обчислення χ^2				

У гр. 12–14 остаточно обчислюють елементи формули хі-квадрат. Сума величин гр. 14 є значенням $\chi^2_{експ.}$.

Таким чином, у розглянутому прикладі маємо: $\chi^2_{експ.} = 3,52$; $\nu = k - 3 = 3$. Згідно табл. 10 [4] $\chi^2_{табл., p=0,95} = 7,81$, $\chi^2_{експ.} < \chi^2_{табл., p=0,95}$. Отже, гіпотезу про нормальний розподіл експериментальних вимірів сосни на досліджуваній ділянці вважаємо доведеною. Відхилити її немає достатніх підстав.

Практична робота №9

Тема: Оцінка характеру розподілу. Розв'язання прикладів на біноміальний закон розподілу, закон розподілу Пуассона. Порівняння за характером розподілу.

Мета: Навчитись оцінювати характер розподілу за даними вибірки. Оволодіти методикою розв'язання прикладів на біноміальний закон розподілу, закон розподілу Пуассона та за характером розподілу.

Матеріал і обладнання: калькулятори, комп'ютери, розрахункові таблиці.

Контрольні питання:

1. Біноміальний розподіл.
2. Розподіл Пуассона
3. Порівняльна оцінка істотності різниці в характері двох експериментальних розподілів.

Хід роботи

Робота №1. Оцінка істотності належності експериментального розподілу частот до біноміального розподілу

Розгляньте приклад і запишіть його у зошит.

Приклад. Було обстежено 100 приплодів свиней породи А (табл. 15). Кількість поросят чоловічої статі у них (фактичні частоти) розподілилася так, як це показано в гр. 1 і 2. Чи можна віднести цей розподіл до біноміального типу?

Таблиця 15

Класи з кількістю подій (v)	Фактичні частоти (f_i)	f_i	$\psi_{\text{бін.}}$	$f'_i = n \cdot \psi_{\text{бін.}}$	f'_i	$ f_i - f'_i $	$(f_i - f'_i)^2$	$\frac{(f_i - f'_i)^2}{f'_i}$
1	2	2*	3	4	4*	5	6	7
0	1	9	0,016	2	11	2	4	0,36
1	8		0,094	9				
2	15	15	0,234	23	23	8	64	2,78
3	45	45	0,312	31	31	14	196	6,32
4	19	19	0,234	23	23	4	16	0,70
5	10	12	0,094	9	11	1	1	0,09
6	2		0,016	2				
$k = 7$	$n = 100$			$n = 100$		$k = 5$		$\chi^2 = 10,25$

Обчислимо теоретично очікувані частоти f' (гр. 3 і 4), для чого виписуємо з табл. 4 [4] для $k = 7$ відповідні коефіцієнти $\psi_{\text{бін.}}$ (гр. 3) і множимо кожен з них на кількість даних $n = 100$ з заокругленням до цілих чисел (гр. 4). (Слід простежити, щоб після заокруглення $n_{\text{експ.}} = n_{\text{теор.}}$).

Далі робимо стандартні розрахунки для обчислення величини χ^2 за формулою

$$\chi^2 = \sum \frac{(f_i - f_i')^2}{f_i'}. \text{ Вони зроблені у гр. 5–7. Дістаємо } \chi^2_{\text{експ.}} = 10,25, \nu = 5 - 3 = 2;$$

згідно табл. 10 [4] $\chi^2_{\text{табл.,}p=0,95} = 5,99, \chi^2_{\text{експ.}} > \chi^2_{\text{табл.,}p=0,95}.$

Отже, одержаний в експерименті розподіл не може бути віднесений до біноміального закону варіювання (коли прийняти рівень гарантії $p = 0,99$, то достатніх підстав відхилити гіпотезу про біноміальний розподіл не буде).

Робота №2. Оцінка істотності належності експериментального розподілу частот до розподілу Пуассона

Розгляньте приклад і запишіть його у зошит.

Приклад. Протягом тижня провели обстеженняпологових будинків щодо частоти народження близнюків (табл. 16).

Таблиця 16

Класи (v_i)	Фактичні частоти (f_i)	f_i	$f_i \cdot v_i$	$v_i - \bar{x}$	$(v_i - \bar{x})^2$	$f_i \cdot (v_i - \bar{x})^2$
1	2	2*	3	4	5	6
0	108	108	0	-0,64	0,41	44,3
1	62	62	62	0,36	0,13	8,1
2	25	25	50	1,36	1,85	46,2
3	4	5	12	2,36	5,57	22,3
4	1		4	3,36	11,29	11,3

$$k = 5 \quad n = 200$$

$$\sum f_i \cdot v_i = 128$$

$$\sum f_i \cdot (v_i - \bar{x})^2 = \sum f_i \cdot (v_i - \bar{x})^2 = 132$$

$$\bar{x} = \frac{\sum f_i \cdot v_i}{n} = 0,64$$

$$= 132$$

$$S_x^2 = \frac{\sum f_i \cdot (v_i - \bar{x})^2}{n - 1} = 0,66$$

$$S_x = \sqrt{0,66} = 0,81$$

Позначимо кількість фактів народження близнюків за тиждень (рідкісна подія) літерою v ($v = 0,1,2,3,4$) – це класи досліджуваної нами події; їх кількість $k = 5$ (гр.1). У гр. 2 показано, скількипологових будинків зареєстрували за цей час народження близнюків у кількостях, зазначених у кожному з класів. Це графа одержаних частот подій у класах; позначимо її літерою f_i . Очевидно, що $\sum f_i = n = 200$.

Середнє арифметичне \bar{x} обчислюємо за формулою, тобто значення кожного класу перемножуємо на його частоту (гр. 3), і суму добутків по всіх класах ділимо на кількість даних.

Середнє квадратичне відхилення і дисперсію обчислюємо за формулою, для чого використовуємо гр. 4–6 (табл. 16).

Зміст обчислених статистичних показників залишається в розподілі Пуассона таким самим, як і у випадку нормального або біноміального розподілу даних.

В результаті проведених обчислень дістаємо для нашого прикладу $\bar{x} = 0,64$ і $S_x^2 = 0,66$. Середнє арифметичне практично дорівнює дисперсії, а отже, є всі підстави припустити, що ми маємо справу з розподілом Пуассона.

Значення функції Пуассона $\Psi_{\text{Пуас.}}$ наведено в табл. 5 [4]. Слід звернути увагу, що ця функція залежить від середнього арифметичного \bar{x} і показника класу ν .

Оскільки досліджувана подія рідкісна, то найчастіше трапляються в сукупності події нульового класу, тобто класу, в якому досліджуваної події взагалі немає. З певною частотою трапляються події з малими показниками класів – 1, 2, тобто класів, в яких події здійснюються рідко. Класи з високими частотами подій трапляються в розподілі Пуассона настільки рідко, що ними можна практично нехтувати. Звідси очевидно, що і середня кількість появи рідкісної за своєю природою події \bar{x} для розподілу Пуассона – величина мала.

Коли кількість класів зростає і середнє арифметичне також приймає більші значення, розподіл Пуассона може зберігатись (за умови, що $\bar{x} = S_x^2$), але він тоді втрачає свою характерну асиметричність і наближається за формою до біноміального розподілу. Це легко помітити, порівнявши, наприклад, дані граф табл. 5 [4] для $\bar{x} = 5$ і для $\bar{x} = 10$. На практиці у тих випадках, коли $\bar{x} > 10$, біноміальний розподіл точніше описує експериментальний матеріал, ніж розподіл Пуассона.

Продовження таблиці 16

$\Psi_{\text{Пуас.}}$ ($\bar{x} = 0,6$)	$f'_i = n \cdot \Psi_{\text{Пуас.}}$	f'_i	f_i	$ f_i - f'_i $	$(f_i - f'_i)^2$	$\frac{(f_i - f'_i)^2}{f'_i}$
7	8	8*	2*	9	10	11
0,549	109	109	108	1	1	0,01
0,329	66	66	62	4	16	0,24
0,099	20	20	25	5	25	1,25
0,020	4	} 5	5	0	0	0
0,003	1					
	$n = 200$			$k = 4$		$\chi^2 = 1,50 \quad \nu = k - 2 = 2$

Повернемося тепер до продовження табл. 16 і проведемо обробку даних

експерименту. Підрахунок теоретичних частот проводимо в гр. 7 і 8. У гр. 7 вписуємо значення функції Пуассона $\psi_{\text{табл.}}$ з табл. 5 [4] для $\bar{x} = 0,6$ – перші 5 коефіцієнтів, оскільки у нас кількість класів $k = 5$. У гр. 8 перемножуємо $\psi_{\text{табл.}}$ у кожному класі на загальну кількість даних n і заокруглюємо до цілих чисел. Наступні гр. 9–11 належать до критерію хі-квадрат. Сума чисел гр. 11 – величина $\chi^2_{\text{експ.}}$. Слід нагадати, що кількість ступенів вільності ν для розподілу Пуассона дорівнює $k - 2$. Отже, для нашого прикладу

$$\chi^2_{\text{експ.}} = 1,50, \quad \nu = k - 2 = 4 - 2 = 2, \quad \chi^2_{\text{табл., } p=0,95} = 5,99; \quad \chi^2_{\text{експ.}} < \chi^2_{\text{табл., } p=0,95}$$

Гіпотезу про належність експериментальних даних до розподілу Пуассона підтверджено.

Робота №3. Оцінка істотності різниці в характері двох експериментальних розподілів

Розгляньте приклад і запишіть його у зошит.

Приклад. Досліджували реакцію чоловіків (група 1) і жінок (група 2) на спинномозкову анестезію. Реакція полягала в зниженні кров'яного тиску у відповідь на введення анестезуючої речовини (гіпотензивний ефект) і досліджувалася для різних вікових категорій. У кожній віковій категорії обстежували по 30 чоловіків і 30 жінок і реєстрували частоту гіпотензивного ефекту (табл. 17).

Таблиця 17

Вік (в роках)	Частота гіпотензивного ефекту		$f_1 + f_2$	$n_1 f_2$	$n_2 f_1$	$n_1 f_2 - n_2 f_1$	$(n_1 f_2 - n_2 f_1)^2$	$\frac{(n_1 f_2 - n_2 f_1)^2}{f_1 + f_2}$
	в групі f_1	в групі f_2						
1	2	3	4	5	6	7	8	9
10–19	6	5	11	400	420	20	400	36
20–29	27	18	45	1440	1890	450	202500	4500
30–39	24	27	51	2160	1680	480	230400	4518
40–49	15	15	30	1200	1050	150	22500	750
50–59	8	5	13	400	560	160	25600	1969
$k = 5$	$n_1 = 80$	$n_2 = 70$	$n = 150$					$\Sigma = 11773$

Заповнюючи таблицю, ми дістаємо всі елементи, необхідні для обчислення

хі-квадрата. Остаточо, $\chi^2 = \frac{1}{n_1 \cdot n_2} \sum \frac{(n_1 f_2 - n_2 f_1)^2}{f_1 + f_2} = 2,10$.

Кількість ступенів вільності $\nu = (k - 1)(l - 1)$, де k – кількість класів ($k = 5$), а l кількість груп, які порівнюють ($l = 2$). Отже, $\nu = (5 - 1)(2 - 1) = 4$.

Теоретичне граничне значення $\chi^2_{табл., p=0,95}$, вище якого гіпотезу про ідентичність розподілів відхиляють (різниця в розподілах стає істотною), знаходимо у табл. 10 [4], воно дорівнює 9,49.

Оскільки $\chi^2_{експ} < \chi^2_{табл., p=0,95}$, то різниця в характері вікових розподілів гіпотензивного ефекту при анестезії між чоловіками і жінками статистично неістотна. Ми можемо сказати точніше, що ймовірність істотності цієї різниці менша за 0,50, як це видно з табл. 10 [4].

Практична робота №10

Тема: Знаходження коефіцієнту кореляції (на конкретних біологічних прикладах).

Мета: Навчитись обчислювати коефіцієнт кореляції для сукупності малого і великого обсягу.

Матеріал і обладнання: калькулятори, комп'ютери, розрахункові таблиці.

Контрольні питання:

1. Кореляція та причинна залежність.
2. Коваріація.
3. Коефіцієнт кореляції.
4. Властивості коефіцієнта кореляції.
5. Інтерпретація коефіцієнту кореляції.

Хід роботи

Робота №1. Обчислення коефіцієнта кореляції для малого обсягу сукупності

Розгляньте приклад і запишіть його у зошит.

Приклад. У 12 аналізах визначали взаємозалежність між величиною амплітуди пікового біоелектричного потенціалу і вмістом заліза в м'язі при подразненні нервово-м'язового препарату. Результати аналізів наведено в табл. 18 у відповідних одиницях вимірювання.

Таблиця 18

№ п/п	Величина амплітуди пікового біоелектричного потенціалу (x_i)	Вміст заліза в м'язі (y_i)	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$
1	73	30	-7	49	-3	9	21
2	77	32	-3	9	-1	1	3
3	87	37	7	49	4	16	28
4	75	34	-5	25	1	1	-5
5	97	40	17	289	7	49	119
6	63	26	-17	289	-7	49	119
7	82	30	2	4	-3	9	-6
8	82	34	2	4	1	1	2
9	80	33	0	0	0	0	0
10	79	34	-1	1	1	1	-1
11	90	38	10	100	5	25	50
12	75	28	-5	25	-5	25	25
$n=12$	$\Sigma = 960$	$\Sigma = 396$		$\Sigma = 844$		$\Sigma = 186$	$\Sigma = 355$

Величину взаємозв'язку між цими двома показниками дослідимо за допомогою обчислень коефіцієнта кореляції за формулою

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

Для цього в табл. 19 введено графі, що дає змогу

обчислити відповідні суми, які входять у дану формулу. Отже, $n = 12$; $\bar{x} = 80$; $S_x = 8,8$; $\bar{y} = 33$; $S_y = 4,1$.

Тоді $r = \frac{355}{\sqrt{844 \cdot 186}} = +0,896$. У нашому випадку кореляція позитивна. Це

означає, що в середньому із збільшенням вмісту заліза в м'язі збільшується також його піковий потенціал.

Для нашого прикладу $r = 0,896$. Йому відповідає значення $z = 1,45$ (у табл. 11 [4]), а $t = 1,45 \cdot \sqrt{12 - 3} = 4,35$.

Значенню $t = 4,35$ у табл. 7 [4] відповідає ймовірність $p > 0,999$. Отже, наявність тісного взаємозв'язку між двома досліджуваними показниками доведено з високою вірогідністю.

Робота №2. Обчислення коефіцієнта кореляції для великого обсягу сукупності

Розгляньте приклад і запишіть його у зошит.

Приклад. Дослідника-ентомолога цікавить питання про взаємозв'язок (кореляцію) між довжиною крила і довжиною хоботка у бджіл. Для цього проведено відповідне обстеження 45 комах і результати в міліметрах (ряд x – довжина крила, ряд y – довжина хоботка) записано в табл. 19 в порядку проведення вимірювань.

Спочатку визначимо межі варіації як по ряду x , так і по ряду y . Найменше значення варіанти в ряді x дорівнює 9,02, а найбільше – 9,87.

Тепер поділимо ряд x на класи. Для $n = 45$ беремо 5 класів з класовим інтервалом $\lambda_x = 0,2$ мм. Дістаємо класи: 9,00–9,19; 9,20–9,39; 9,40–9,59; 9,60–9,79; 9,80–9,99. Ряд y зручно поділити на 4 класи з класовим інтервалом $\lambda_y = 0,2$ мм. Маємо класи: 6,10–6,29; 6,30–6,49; 6,50–6,69; 6,70–6,89.

Таблиця 19

Довжина крила (x)	Довжина хоботка (y)	Довжина крила (x)	Довжина хоботка (y)	Довжина крила (x)	Довжина хоботка (y)
9,74	6,44	9,68	6,69	9,74	6,63
9,59	6,54	9,77	6,70	9,67	6,68
9,71	6,64	9,60	6,51	9,49	6,71
9,56	6,55	9,59	6,62	9,45	6,50
9,61	6,61	9,54	6,68	9,55	6,55
9,74	6,67	9,74	6,74	9,25	6,35
9,20	6,22	9,02	6,35	9,73	6,75
9,48	6,62	9,68	6,53	9,59	6,70
9,78	6,64	9,59	6,79	9,70	6,61
9,56	6,62	9,87	6,63	9,84	6,70
9,64	6,70	9,21	6,17	9,72	6,65
9,72	6,75	9,61	6,57	9,50	6,55
9,81	6,71	9,55	6,64	9,61	6,59
9,64	6,45	9,53	6,43	9,65	6,77
9,08	6,25	9,71	6,55	9,67	6,58

Встановивши класи, внесемо дані вимірювань обох рядів у кореляційну решітку. Зручніше спочатку накреслити решітку без граф обробки, тільки з клітинками класів, і кожену пару вимірів (для кожної комахи) проставляти у відповідній клітинці окремою крапкою або паличкою (за методом конвертів). Потім можна рознести дані олівцем безпосередньо в кореляційну решітку.

Далі приступають до складання кореляційної решітки в повному вигляді і проводять відповідну обробку по графах. У кореляційній решітці (табл. 20) модальні класи обведено товщими лініями.

Таблиця 20

ряд x		Класи ряду x					f_{y_i}	a_{y_i}	$f_{y_i} a_{y_i}$	$f_{y_i} a_{y_i}^2$
		9,00– 9,19	9,20– 9,39	9,40– 9,59	9,60– 9,79	9,80– 9,99				
Класи ряду y	6,10– 6,29	1	2				3	–2	–6	12
	6,30– 6,49	1	1	1	2		5	–1	–5	5
	6,50– 6,69			10	15	1	26	0	0	0
	6,70– 6,89			3	6	2	11	1	11	11
f_{x_i}		2	3	14	23	3	45		0	28
a_{x_i}		–3	–2	–1	0	1				
$f_{x_i} a_{x_i}$		–6	–6	–14	0	3		–23		
$f_{x_i} a_{x_i}^2$		18	12	14	0	3			47	

Випишемо з кореляційної решітки потрібні значення тих сум, які входять у формулу; вони містяться в правих нижніх клітинках по краях виїмки решітки:

$$\sum f_{x_i} \cdot a_{x_i} = -23; \quad \sum f_{y_i} \cdot a_{y_i} = 0; \quad \sum f_{x_i} \cdot a_{x_i}^2 = 47; \quad \sum f_{y_i} \cdot a_{y_i}^2 = 28; \quad n = 45.$$

Залишилося обчислити значення суми потрійного добутку $\sum f_{xy_i} a_{x_i} a_{y_i}$. Для цього частоту варіант у кожній заповненій клітинці решітки f_{xy_i} треба помножити на відповідні значення a_{x_i} і a_{y_i} , а потім підрахувати загальну їх суму $\sum f_{xy_i} a_{x_i} a_{y_i}$.

У модальних класах відхилення a_{x_i} , або a_{y_i} дорівнює нулю.

Отже, добуток $a_{x_i} a_{y_i}$, коли в нього входить хоч би одне значення модального класу, також дорівнюватиме нулю. Усі інші клітинки решітки, якщо вони заповнені певною кількістю варіант, дають відповідні значення добутку $a_{x_i} a_{y_i}$, які є цілими числами.

Отже, кореляційну решітку модальні (нульові) класи поділяють на 4 квадранти. Нульові класи обведено товщими лініями. Кожен квадрант характеризується своїм знаком:

I квадрант	+	II квадрант	-
III квадрант	-	IV квадрант	+

Величину суми $\sum f_{xy_i} a_{x_i} a_{y_i}$ визначають за квадрантами. Знаки в квадрантах відомі, отже, величини a_{x_i} і a_{y_i} можна перемножити, не зважаючи на знак, який встановлюють тільки в кінці операції підсумовування.

$$\text{У нашому випадку для I квадранта } \sum f_{xy_i} a_{x_i} a_{y_i} \text{ (I) дорівнює: } \left\{ \begin{array}{l} 1 \cdot 3 \cdot 2 = 6 \\ 2 \cdot 2 \cdot 2 = 8 \\ 0 \cdot 2 \cdot 2 = 0 \\ 1 \cdot 3 \cdot 1 = 3 \\ 1 \cdot 2 \cdot 1 = 2 \\ 1 \cdot 1 \cdot 1 = 1 \end{array} \right.$$

$$\sum = 6 + 8 + 3 + 2 + 1 = 20 \text{ (знак в квадранті плюс)}$$

Для II квадранта $\sum f_{xy_i} a_{x_i} a_{y_i}$ (II) = 0, бо в ньому немає жодної заповненої клітинки (знак в квадранті мінус).

Для III квадранта $\sum f_{xy_i} a_{x_i} a_{y_i}$ (III) дорівнює $3 \cdot 1 \cdot 1 = 3$ (знак в квадранті мінус).

Для IV квадранта $\sum f_{xy_i} a_{x_i} a_{y_i}$ (IV) дорівнює $2 \cdot 1 \cdot 1 = 2$ (знак в квадранті плюс).

Загальна сума значень по всіх квадрантах, враховуючи їх знаки, становить $\sum f_{xy_i} a_{x_i} a_{y_i} = +20 - 0 - 3 + 2 = +19$.

Тепер залишається тільки підставити всі визначені величини в загальну формулу коефіцієнта кореляції:

$$r = \frac{1}{45} \left(\frac{19 - \left(\frac{-23 \cdot 0}{45} \right)}{\sqrt{\frac{47}{45} - \left(\frac{-23}{45} \right)^2} \cdot \sqrt{\frac{28}{45} - \left(\frac{0}{45} \right)^2}} \right) = +0,605.$$

Кореляція статистично істотна, оскільки коефіцієнту кореляції 0,605 відповідає значення $z = 0,70$ (у табл. 11 [4]), для якого коефіцієнт Ст'юдента $t = z\sqrt{n-3} = 0,70\sqrt{42} = 4,54$, а йому, в свою чергу, відповідає висока ймовірність $p > 0,999$ (у табл. 7 [4]).

Практична робота №11

Тема: Ранговий коефіцієнт кореляції та кореляція за альтернативними і якісними показниками

Мета: Навчитись підраховувати ранговий коефіцієнт кореляції та коефіцієнт кореляції за альтернативними і якісними показниками.

Матеріал і обладнання: калькулятори, комп'ютери, статистичні таблиці.

Контрольні питання:

1. Ранговий коефіцієнт кореляції Спірмена.
2. Процентна характеристика сукупності.
3. Оцінка вірогідності різниці між процентними характеристиками двох альтернативних сукупностей.
4. Обчислення коефіцієнта кореляції r при альтернативній мінливості.

Інформаційний матеріал

Ранговий коефіцієнт кореляції Спірмена.

Цей коефіцієнт служить показником сили та напрямку для порядкових даних, а також для попереднього оцінювання кореляцій між кількісними (інтервальними) показниками. Для його підрахунку кожна з двох ознак ранжирується і підраховується d – різниця в кожній парі (n – число пар):

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n}.$$

Для оцінювання отриманого результату використовують таблицю 13 [4].

Процентна характеристика сукупності

Статистичну обробку даних при альтернативній класифікації зводять до аналізу кількісного співвідношення між чисельностями даних в обох групах і встановлення процентів наявності p і відсутності q досліджуваного ефекту або ознаки. Якщо кількість випадків з наявним ефектом позначити n^+ , а кількість випадків з відсутнім ефектом n^- , то процент наявності ефекту

буде $p = \frac{n^+}{n} \cdot 100\%$, а процент відсутності ефекту q дорівнюватиме $q = \frac{n^-}{n} \cdot 100\%$, де n – повний обсяг сукупності.

У біологічній практиці часто трапляється така ситуація, коли експериментальні дані поділяються тільки на два класи так, що в одному класі концентруються об'єкти з наявністю певної ознаки, а в другому – з відсутністю цієї ознаки. У багатьох випадках така класифікація відбиває саму природу явищ. Наприклад, ми спостерігаємо розподіл біологічних об'єктів на жіночі і чоловічі особини, проводимо класифікацію людей на хворих і здорових, поділяємо тварин на корисних і шкідливих тощо. Така класифікація, коли увага зосереджується тільки на двох альтернативних можливостях і сукупність варіант поділяється на дві категорії за наявністю або відсутністю певної ознаки, називається альтернативною; мінливість даних при цій класифікації також називається альтернативною. Очевидно, що $q = 100\% - p$ процент наявного ефекту p є оцінкою вибірки в цілому. Мінливість процентної характеристики альтернативної сукупності оцінюють обчисленням стандартної похибки процента наявного або відсутнього ефекту за такою формулою: $m_p = m_q = \sqrt{\frac{p \cdot q}{n}}$.

Вона аналогічна стандартній похибці середнього арифметичного при кількісній мінливості. (Кількісною мінливістю ми називатимемо всі раніше розглянуті випадки, коли статистичний аналіз ведеться безпосередньо з числовими значеннями варіант) і є оцінкою розсіяння процента досліджуваного ефекту завдяки статистичній природі подій в сукупностях.

Статистично оброблені результати в альтернативному експерименті записують так: $p \pm m_p$ (із зазначенням n)

Середнє квадратичне відхилення S_x при альтернативній мінливості не має практичного застосування. Це зрозуміло, тому що S_x є мірою індивідуальної мінливості, тоді як процент наявності ефекту p є характеристикою сукупності в цілому, для якої мірою мінливості є стандартна похибка процента m_p .

Оцінка вірогідності різниці між процентними характеристиками двох альтернативних сукупностей.

Щоб встановити ступінь вірогідності різниці між процентними характеристиками досліджуваних двох груп, достатньо провести це порівняння для одного показника альтернативної мінливості – p або q . Наприклад, якщо обстежують процент певного захворювання у представників двох різних видів тварин і дослідника цікавить, чи справді вид А більше піддається даному захворюванню, ніж вид В, то достатньо проаналізувати вірогідність різниці в проценті *хворих* тварин між обома видовими групами (p), тому що процент здорових тварин $q = 100\% - p$.

Як і в загальному випадку, ймовірність доведення різниці між процентами наявності ефекту здійснюють обчисленням коефіцієнта Ст'юдента t , який визначають як відношення величини цієї різниці до її стандартної похибки $t = \frac{d}{m_d}$,

де $d = |p_1 - p_2|$. Стандартну похибку різниці між процентами двох

досліджуваних груп m_d обчислюють за формулою $m_d = \sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$, де

\bar{p} – середній процент наявності ефекту для обох груп. Його обчислюють, в загальному випадку, беручи до уваги можливу різну кількість дослідів в обох

групах, за формулою $\bar{p} = \left(\frac{n_1^+ + n_2^+}{n_1 + n_2}\right) \cdot 100\%$.

Якщо ж $n_1 \approx n_2$, то середній процент p можна обчислити за спрощеною формулою: $\bar{p} = \frac{p_1 + p_2}{2}$. В обох випадках середній процент відсутності ефекту \bar{q} обчислюють як $100\% - p$.

Обчислення коефіцієнта кореляції r при альтернативній мінливості.

При альтернативній мінливості, коли експериментальні дані групуються за типом дворозрядної класифікації, їх зручно подавати за табличною формою. У табл. 21 наведено чотириклітинну кореляційну решітку, у клітини якої записують дані про кількість випадків з позитивним і негативним ефектами n^+ і n^- для одного і другого варіаційних рядів, один з яких (x) розміщений по горизонталі, а другий (y) – по вертикалі. Для того щоб легко було користуватися формулою коефіцієнта кореляції, клітини кореляційної решітки умовно позначають, наприклад, a, b, v, z .

З таблиці 21 видно, що кореляційна решітка складається, по суті, усього з чотирьох клітин, а в крайніх клітинах подано значення відповідних сум. У нижній правій клітині зазначено суму всіх даних, тобто n .

Таблиця 21

Ряд x Ряд y	Наявність ефекту (+)	Відсутність ефекту (-)	$\sum y$
Наявність ефекту (+)	a	b	$a+b$
Відсутність ефекту (-)	v	z	$v+z$
$\sum x$	$a+v$	$b+z$	$a+b+v+z = n$

Використовуючи умовні позначення кореляційної решітки, коефіцієнт кореляції при альтернативній мінливості обчислюють за формулою

$$r = \frac{az - bv}{\sqrt{(a+v)(b+z)(a+b)(v+z)}}.$$

Для доведення вірогідності кореляції при альтернативній мінливості переводимо r в z (табл. 11 [4]) і обчислюємо коефіцієнт Ст'юдента t за формулою: $t = z\sqrt{n-3}$.

Якщо обчислене t відповідає ймовірності $p > 0,95$ за таблицею Ст'юдента, то кореляція статистично істотна.

Для цієї мети можна також використати критерій Пірсона, який обчислюють за формулою $\chi^2 = n \cdot r^2$.

Хід роботи

Робота №1. Обчислення рангового коефіцієнту кореляції Спірмена

Розгляньте приклад і запишіть його у зошит.

Приклад. Розмір черепа (x) і вага головного мозку (y) (табл. 22).

Таблиця 22

x	155	190	170	165	160	195	180	185	175
y	1300	1650	1440	1386	1393	1725	1502	1550	1455
R_x	1	8	4	3	2	9	6	7	5
R_y	1	8	4	2	3	9	6	7	5
d	0	0	0	+1	-1	0	0	0	0
d^2	0	0	0	1	1	0	0	0	0

Підраховуємо $\sum d^2 = 2$; $r_s = 1 - 0,00833 \times 2 = +0,9834$. Для отриманого результату при $n=9$ та $P_o=0,01$ (за табл. 13 [4]) $r_s = 0,83$, оскільки $0,98 > 0,83$; таким чином $P_o < 0,01$.

Робота №2. Статистична обробка даних при альтернативній класифікації

Розгляньте приклад і запишіть його у зошит.

Приклад. Дослідних лабораторних мишей (56 шт.) опромінювали рентгенівськими променями у відповідній дозі. Через певний строго визначений час загинуло 19 мишей. Треба зробити узагальнюючий висновок про процент смертності мишей від опромінення заданою дозою рентгенівських променів. Експериментальні дані наведено в табл. 23.

Таблиця 23

Класи	Кількість	Процент
Наявний ефект (+)	19	34
Відсутній ефект (-)	37	66
Всього	56	100

Середня смертність мишей при даній дозі $p = \frac{19}{56} \cdot 100\% = 34\%$,

отже, $p \pm m_p = 34 \pm \sqrt{\frac{34 \cdot 66}{56}} = 34 \pm 6,3\%$.

Можна зробити висновок з вірогідністю, наприклад, 0,95 ($t = 2,0$), що смертність мишей характерна для заданої дози рентгенівських променів, лежить у межах $p \pm 2 \cdot m_p = 21,4 \div 12,6\%$.

Робота №3. Оцінка вірогідності різниці між процентними характеристиками двох альтернативних сукупностей

Розгляньте приклад і запишіть його у зошит.

Приклад. Вивчалася мутагенна дія ультрафіолету з довжиною хвилі λ_1 і λ_2 на культуру мікроорганізмів. У першому досліді (λ_1) в чашці Петрі спостерігалися 18 мутантних і 75 нормальних колоній. У другому досліді (λ_2) з 147 колоній мутантних було 41. Треба встановити, чи можна вважати вірогідно доведеним, що досліджувані довжини хвилі ультрафіолету мають різну мутагенну дію. Отже, для першої групи $n_1 = 93$, $n_1^+ = 18$. Процент мутацій у цій групі $p_1 = \frac{n_1^+}{n_1} \cdot 100\% = \frac{18}{93} \cdot 100\% = 19,4\%$. Процент відсутності $q_1 = 100\% - p_1 = 80,6\%$.

Для другої групи $n_2 = 147$, $n_2^+ = 41$, $p_2 = \frac{41}{147} \cdot 100\% = 27,9\%$,
 $q_2 = 100\% - p_2 = 72,1\%$.

В обох дослідях різниця між процентами мутацій $d = 27,9\% - 19,4\% = 8,5\%$.

Яка ймовірність того, що визначена різниця в мутагенній дії не випадкова, а справді показує різний вплив ультрафіолету досліджених довжин хвиль на генетичні структури мікроорганізмів?

Обчислимо стандартну похибку встановленої різниці

$$m_d = \sqrt{\bar{p} \cdot \bar{q} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

Спочатку визначимо середні проценти \bar{p} і \bar{q} :

$$\bar{p} = \left(\frac{n_1^+ + n_2^+}{n_1 + n_2} \right) \cdot 100\% = \frac{18 + 41}{93 + 147} \cdot 100\% = 24,5\% ;$$

$$\bar{q} = 100\% - 24,5\% = 75,5\%.$$

$$\text{Тоді } m_d = \sqrt{24,5 \cdot 75,5 \cdot \left(\frac{1}{93} + \frac{1}{147}\right)} = 5,7\% ;$$

$$t = \frac{d}{m_d} = \frac{8,5}{5,7} = 1,5.$$

В таблиці нормального інтеграла ймовірностей (табл. 3 [4]) знаходимо, що значенню $t = 1,5$ відповідає ймовірність 0,8664.

Оскільки вона менша за 0,95 (граничне значення, при якому різницю вважають статистично істотною), проведені обстеження не дають підстав стверджувати, що досліджувані довжини хвиль ультрафіолету λ_1 і λ_2 відрізняються за мутагенною дією.

Зрозуміло, що величини різниці між процентами наявності і відсутності ефекту ідентичні:

$$d_p = 27,9 - 19,4 = 8,5; \quad d_q = 80,6 - 72,1 = 8,5$$

У формулу m_d входять середні проценти наявності і відсутності ефекту \bar{p} і \bar{q} , отже, встановити ймовірність різниці між обома групами можна як за наявністю ефекту, так і за його відсутністю.

Робота №4. Обчислення коефіцієнта кореляції при альтернативній мінливості

Розгляньте приклад і запишіть його у зошит.

Приклад. Дослідника-радіобіолога цікавить, чи пов'язані між собою поява злоякісних новоутворень у віддалений після опромінення період та різке зниження лейкоцитів безпосередньо після опромінення (лейкопенія) і наскільки тісний цей зв'язок. Щоб з'ясувати це питання, було проведено 96 дослідів на білих мишах. Тварин опромінювали відповідною дозою рентгенівських променів і підраховували у них лейкоцити після опромінення. Встановлювали загальну кількість тварин, в яких розвинулася лейкопенія. Через відповідний проміжок часу після опромінення підраховували загальну кількість тварин, в яких з'явилися злоякісні пухлини, причому як у групі, в якій спостерігалася лейкопенія. Так і в групі, в якій кількість лейкоцитів залишалася в межах норми.

Коефіцієнт кореляції між названими радіобіологічними показниками визначали за приведеною вище схемою. Цифрові дані, внесені безпосередньо в кореляційну решітку, наведено в табл. 24.

Таблиця 24

Ряд у	Ряд x	Поява злоякісних пухлин (+)	Відсутність злоякісних пухлин (—)	$\sum y$
Наявність лейкопенії	(+)	27	13	40
Відсутність лейкопенії	(—)	1	55	56
$\sum x$		28	68	$n = 96$

З таблиці видно, що з 28 тварин, у яких з'явилися злоякісні пухлини після опромінення, тільки в одному випадку кількість лейкоцитів не зменшилася. З іншого боку, лейкопенія спостерігалася також і в тих випадках, які не супроводжувалися злоякісним ростом, таких випадків було 13 із 40. Обчислимо коефіцієнт кореляції між досліджуваними явищами

$$r = \frac{27 \cdot 55 - 1 \cdot 13}{\sqrt{28 \cdot 68 \cdot 40 \cdot 56}} = 0,71.$$

Він свідчить, що між смертністю мишей від дії рентгенівських променів і появою в них лейкопенії після опромінення існує позитивний кореляційний зв'язок середньої тісноти, вірогідність якого доведено з імовірністю $p > 0,99$, тому що для $r = 0,71$ $z = 0,89$, а $t = 0,89 \cdot \sqrt{93} = 8,58$.

Практична робота №12

Тема: Лінійна регресія. Побудова теоретичної і емпіричної ліній регресії

Мета: Навчитись будувати теоретичні і емпіричні лінії регресії. Вміти розв'язувати задачі на побудову рівняння лінійної регресії для двох взаємозалежних ознак.

Матеріал і обладнання: калькулятори, комп'ютери, розрахункові таблиці.

Контрольні питання:

1. Рівняння лінійної регресії.

2. Теоретична лінія регресії.
3. Емпірична лінія регресії.
4. Коефіцієнт регресії.

Хід роботи

Робота №1. Складання рівняння лінійної регресії між двома ознаками

Розгляньте приклад і запишіть його у зошит.

Приклад. Розглянемо дані (табл. 19), що були отримані при вивченні залежності між величиною амплітуди пікового біоелектричного потенціалу (x) і вмістом заліза в м'язі (y). Для цього скористаємося вже обчисленими значеннями коефіцієнта кореляції r та відповідних середніх квадратичних відхилень S_x і S_y : $S_x = 8,8$; $S_y = 4,1$; $r = 0,896$.

$$\text{Тоді за формулою } b_{y/x} = r \frac{S_y}{S_x} = 0,896 \cdot \frac{4,1}{8,8} = 0,42.$$

Рівняння лінійної регресії між величиною амплітуди пікового біоелектричного потенціалу (x) і вмістом заліза в м'язі (y) має такий вигляд:

$$y = \bar{y} + b_{y/x}(x - \bar{x}) = 33 + 0,42(x - 80) = 0,42x - 0,6.$$

Тепер можна для будь-якого значення x вказати, яке йому відповідає, в середньому, значення y .

Робота №2. Побудова теоретичної лінії регресії

Розгляньте приклад і запишіть його у зошит.

Приклад. Повернемося до попередніх даних (із табл. 19), про взаємну залежність між амплітудою біоелектричного потенціалу та вмістом заліза в м'язі і відкладемо добуті числові дані на графіку. Рівняння лінійної регресії, розраховане попередньо, має такий вигляд: $y = 0,42x - 0,6$.

Обчислимо за цим рівнянням два теоретичні значення y , наприклад для $x_1 = 60$ і $x_2 = 100$. Вони будуть відповідно $y_1 = 24,6$ і $y_2 = 41,4$. Відкладемо їх на графіку у вигляді хрестиків і проведемо між ними пряму лінію. Це і є теоретична лінія регресії, яка показує встановлену залежність вмісту заліза в м'язі від зміни амплітуди біоелектричного потенціалу.

Робота №3. Визначення емпіричної і теоретичної ліній регресії на основі кореляційної решітки

Розгляньте приклад і запишіть його у зошит.

Приклад. У прикладі про довжину крила (x) і довжину хоботка (y) у бджіл (табл. 19) при обробці кореляційної решітки (табл. 20) були одержані такі величини: $n = 45$; $\lambda_x = 0,2$; $\lambda_y = 0,2$; $\sum f \cdot a_x = -23$; $\sum f \cdot a_y = 0$;

$$\sum f \cdot a_x^2 = 47; \sum f \cdot a_y^2 = 28.$$

Визначаємо за відповідними формулами середні арифметичні та середні квадратичні відхилення, коефіцієнт кореляції: $\bar{x} = 9,6$; $S_x = 0,18$; $\bar{y} = 6,6$; $S_y = 0,16$; $r = 0,605$.

Тоді коефіцієнт регресії $b_{y/x} = r \cdot \frac{S_y}{S_x} = 0,54$, а відповідне рівняння лінійної регресії для залежності y від x має вигляд $y = \bar{y} + b(x - \bar{x}) = 6,6 + 0,54(x - 9,6) = 0,54x + 1,4$.

Обчисливши значення y для двох точок x (наприклад, $x_1 = 9$ і $x_2 = 10$), можна побудувати теоретичну лінію регресії. Аналогічні обчислення і побудову можна зробити і при вивченні залежності ознаки x від ознаки y .

На основі кореляційної решітки можна також побудувати емпіричну лінію регресії. Для цього треба обчислити для кожного з класів ряду y часткові середні значення x , які позначимо через \bar{x}' . Для цього користуємося відповідними центральними значеннями класів.

У табл. 25 наведено кореляційну решітку для відомих нам даних про залежність між довжиною крила (x) і довжиною хоботка (y) у бджіл. У табл. 25 для обох рядів наведено центральні значення класів y_{ci} та x_{ci} . Оперуючи ними і частотами варіант u відповідних клітинках решітки, обчислюємо часткові середні значення (\bar{x}') для кожного з класів ряду y (табл. 25, праворуч) і часткові середні значення \bar{y}' для кожного з класів ряду x (табл. 25, унизу).

Таблиця 25

ряд x		Центральні значення класів ряду x (x_{ci})					Часткові середні (\bar{x}')
		9,1	9,3	9,5	9,7	9,9	
Центральні значення класів ряду y (y_{ci})	6,2	1	2				9,13
	6,4	1	1	1	2		9,46
	6,6			10	15	1	9,63
	6,8			3	6	2	9,68
Часткові середні (\bar{y}')		6,3	6,27	6,63	6,64	6,67	

Наприклад, часткове середнє значення \bar{x}' для класу ряду y з центральним значенням $y_u = 6,4$ обчислюють так:

$$\bar{x}'(y_u = 6,4) = \frac{1}{n} \sum f_i \cdot x_{u_i} = \frac{1 \cdot 9,1 + 1 \cdot 9,3 + 1 \cdot 9,5 + 2 \cdot 9,7}{5} = 9,46.$$

$$\text{Аналогічно } \bar{y}'(x_u = 9,7) = \frac{1}{n} \sum f_i \cdot y_{u_i} = \frac{2 \cdot 6,4 + 15 \cdot 6,6 + 6 \cdot 6,8}{23} = 6,64.$$

За частковими середніми значеннями будують емпіричні лінії регресії. Так, емпірична лінія регресії y/x будується для кожного фіксованого кроку ряду x (центральні значення x_u відкладено обчислені часткові середні значення \bar{y}').

Слід звернути увагу на те, що емпірична лінія регресії є ламаною лінією, бо на неї впливають випадкові фактори статистичної природи. Теоретична лінія регресії згладжує цю ламану лінію до прямої, що проходить на найменшій відстані між експериментальними точками (точніше кажучи, сума квадратів відхилень експериментальних точок від лінії регресії є меншою, ніж від будь-якої іншої лінії, проведеної через дану сукупність точок).

Література

Основна:

1. Атраментова Л. О. Біометрія: підруч. для студ. вищ. навч. закладів / Л. О. Атраментова, О. М. Утєвська. – Харків : Ранок, 2007. – 176 с.
2. Гумецький, Р. Я. Математичні методи в біології : Теоретичні відомості, програмований практикум, комп'ютерні тести: Навч. посібник / Р. Я. Гумецький, Б. М. Паляниця, М. Є. Чабан.– Львів: ЛНУ, 2004. – 112 с.
3. Горошко, М.П. Біометрія: Навчальний посібник / М. П. Горошко, С. І. Миклуш, П. Г. Хомюк. – Львів: Камула. 2004. – 236 с.
4. Зінченко О. П. Біометрія. Ч. III. Статистичні таблиці: Метод. матеріали до викон. лабораторних робіт / О. П. Зінченко, Я. В. Степанюк. – Луцьк : РВВ «Вежа» Волин. нац. ун-ту ім. Лесі Українки, 2010.– 28 с.
5. Зінченко О. П. Математичні методи в біології. Основні статистичні формули та статистичні таблиці : Метод. матеріали до викон. лабораторних робіт / Зінченко О. П., Степанюк Я. В. – Луцьк : Медіа, 2016. – 24 с.
6. Калінін М. І. Біометрія: Підручник для студентів вузів біологічних і екологічних напрямків / М. І. Калінін, В. В. Єлісеєв. – Миколаїв: Вид-во МФ НаУКМА, 2000. – 204 с. – Режим доступу: <https://lib.chmnu.edu.ua/index.php?m=1&b=3>

Додаткова:

7. Барковський В. В. Теорія ймовірностей та математична статистика / В. В. Барковський. – Київ: Центр учбової літератури, 2010. – 424 с.
8. Біометрія: теоретичні відомості та лабораторний практикум: Навч. посібник / М. В. Дика, М. М. Тарновська, М. М. Яремчук, А. Б. Генега, Д. І. Санагурський. – Львів: ЛНУ, 2016. – 100 с.
9. Боровиков В.П. Популярное введение в современный анализ данных в системе STATISTICA / В.П. Боровиков. – М.: Горячая линия-Телеком, 2013. – 288 с. – Режим доступу: <https://www.twirpx.com/file/2289362/>
10. Вуколов З. А. Основи статистического анализа. Практикум по

- статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL: учебное пособие / З. А. Вуколов. – М.: ФОРУМ. 2008. – 464 с. – Режим доступа: <https://www.twirpx.com/file/43323/>
11. Горкавий, В. К. Статистика : підручник / В. К. Горкавий. – К. : Аграрна освіта, 2009. – 511 с
 12. Гроссман С. Математика для биологов / С. Гроссман, Дж. Тернер. – М. : Высш. шк., 1983.– 383 с.
 13. Донченко В. С. Теорія ймовірностей та математична статистика / В. С. Донченко, М. В.-С. Сидоров, М. М. Шарапов. – Альма-матер. – К.: «Академія», 2009. – 288 с.
 14. Кармелюк Г. І. Теорія ймовірностей та математична статистика. Посібник з розв'язання задач / Г. І. Кармелюк. – К.: Центр учбової літератури, 2007. – 576 с.
 15. Лакин Г. Ф. Биометрия / Г. Ф. Лакин. – М. : Высш. школа, 1990. – 352 с.
 16. Сеньо П. С. Теорія ймовірностей та математична статистика / П. С. Сеньо. – К.: Знання, 2007. – 556 с.
 17. Теорія ймовірностей, математична статистика та імовірнісні процеси: навч. посіб. / Ю. М. Слюсарчук, Й. Я. Хром'як, Л. Л. Джавала, В. М. Цимбал ; М-во освіти і науки України, Нац. ун-т «Львів. політехніка». – Львів: Вид-во Львів. політехніки, 2015. – 364 с.
 18. Чепур С.С. Біометрія: Методичний посібник / С.С. Чепур. – Ужгород: Видавництво УжНУ «Говерла», 2015. – 40 с.

Зміст

Передмова.....	3
<i>Модуль 1. Основи теорії ймовірностей.....</i>	<i>4</i>
<i>Практична робота №1. Вибіркові простори і простори рівних ймовірностей.....</i>	<i>4</i>
<i>Практична робота №2. Скінченні простори ймовірностей.....</i>	<i>8</i>
<i>Практична робота №3. Умовна ймовірність.....</i>	<i>11</i>
<i>Практична робота №4. Теорема Байєса. Повторні випробування.....</i>	<i>13</i>
<i>Модуль 2. Основи варіаційної статистики.....</i>	<i>16</i>
<i>Практична робота №5. Фіксація і початкове впорядкування даних. Побудова графіків варіаційних рядів</i>	<i>16</i>
<i>Практична робота №6. Показники положення і мінливості.....</i>	<i>20</i>
<i>Практична робота №7. t–критерій Ст’юдента. F–критерій Фішера.....</i>	<i>25</i>
<i>Практична робота №8. Оцінка характеру розподілу. Розв’язання прикладів на належність експериментального розподілу частот у класах до теоретично відомого стандарту та нормальний закон розподілу.....</i>	<i>30</i>
<i>Практична робота №9. Оцінка характеру розподілу. Розв’язання прикладів на біноміальний закон розподілу, закон розподілу Пуассона. Порівняння за характером розподілу.....</i>	<i>33</i>
<i>Практична робота №10. Знаходження коефіцієнту кореляції (на конкретних біологічних прикладах).....</i>	<i>38</i>
<i>Практична робота №11. Ранговий коефіцієнт кореляції та кореляція за альтернативними і якісними показниками</i>	<i>43</i>
<i>Практична робота №12. Лінійна регресія. Побудова теоретичної і емпіричної ліній регресії.....</i>	<i>50</i>
Література.....	53