Я.М.Мамчич

ТЕОРІЯ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНА СТАТИСТИКА

Лекції

Луцьк 2018

Лекція 1

Елементи комбінаторики

Правило додавання. Якщо деякий об'єкт A можна вибрати із сукупності об'єктів m способами, а інший об'єкт B можна вибрати n способами, то або A, або B можна вибрати $m+n$ способами.

Приклад 1. У класі 12 хлопчиків і 15 дівчаток.

Скількома способами можна вибрати одного учня?

Розв'язок. 12+15=27 способами.

Правило множення (основне правило комбінаторики). Якщо деякий об'єкт A можна вибрати із сукупності об'єктів m способами і після кожного такого вибору об'єкт B можна вибрати n способами, то пару об'єктів A, а потім B можна вибрати $m \cdot n$ способами.

Приклад 2. У класі 12 хлопчиків і 15 дівчаток.

Скількома способами можна вибрати одного хлопчика і одну дівчинку?

Розв'язок. $12 \cdot 15 = 180$ способами.

Приклад 3. Скільки є двоцифрових чисел?

Розв'язок. Першу цифру можна задати 9 способами (цифри від 1 до 9). Другу цифру можна задати 10 способами (цифри від 0 до 9). Значить двоцифрове число можна створити $9 \cdot 10 = 90$ способами, тобто існує 90 двоцифрових чисел.

Основне правило комбінаторики у загальному вигляді. Нехай необхідно виконати одну за одною k дій. Якщо першу дію можна виконати n_1 способами, другу – n_2 способами, третю – n_3 способами і так до k– ї дії, яку можна виконати n_k способами, то усі k дій разом можуть бути виконані

$$n_1 \cdot n_2 \cdot n_3 \cdot \ldots \cdot n_k$$

способами.

Перестановки. Перестановками називаються набори елементів, які складаються з одних і тих самих n різних елементів і відрізняються лише порядком їхнього розташування. Число перестановок з n елементів

$$P_n = n!,$$

де $n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n.$

Приклад 4. Скільки перестановок можна скласти із трьох прапорців різного кольору?

Розв'язок. На перше місце можна вибрати один із трьох прапорців, на друге місце – один із двох, що залишилися, на третє – один останній. Таким чином

$$P_n = 1 \cdot 2 \cdot 3 = 3!.$$

Розміщення. Розміщеннями із n елементів по k елементів називають набори елементів, які містять k елементів, вибраних із n різних елементів. При цьому ці набори повинні відрізнятися або складом елементів, або порядком їх розташування. Число усіх можливих розміщень із n елементів по k елементів

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

Приклад 5. Є 6 прапорців різного кольору.

Скільки можна скласти різних сигналів із 3 прапорців, якщо порядок їх розташування грає роль?

Розв'язок. На перше місце можна вибрати один із 6 прапорців, на друге – один із 5, що залишились, на третє – один із 4, які залишаться після вибору двох попередніх. Таким чином, усього можна скласти

$$A_6^3 = 6 \cdot 5 \cdot 4 = 120 \text{ сигналів.}$$

Комбінації (сполучення). Комбінаціями із n елементів по k елементів називають набори елементів, які містять k елементів, вибраних із n різних елементів. При цьому ці набори повинні відрізнятися хоча б одним елементом. Порядок розташування елементів ролі не грає. Число комбінацій із n елементів по k елементів

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

Приклад 6. У групі 20 студентів. Скількома способами можна скласти команду із 5 чоловік?

Розв'язок. Так як різні команди обов'язково повинні відрізнятися хоча б одним учасником, то використовуємо формулу для числа комбінацій

$$C_{20}^5 = \frac{20!}{5!\cdot(20-5)!} = \frac{20!}{5!\cdot 15!} = \frac{20\cdot 19\cdot 18\cdot 17\cdot 16}{5\cdot 4\cdot 3\cdot 2\cdot 1} = 15504$$

(чисельник і знаменник скоротили на 15!).

Розбиття на групи. Число способів, якими можна розбити множину із n елементів на m груп, кожна з яких містить відповідно $k_1, k_2, k_3, \ldots, k_m$ елементів, дорівнює

$$C_n(k_1, k_2, \ldots, k_m) = \frac{n!}{k_1!k_2!\ldots k_m!}.$$

Приклад 7. Скількома способами можна розселити 8 студентів у трьох кімнатах: одномісній, тримісній та чотиримісній?

Розв'язок. Вісім студентів необхідно розбити на три групи: з 1– го, 3– х та 4– х чоловік, тому використовуємо формулу для розбиття на групи. Число можливих способів дорівнює

$$C_8(1,3,4)=\frac{8!}{1!\cdot 3!\cdot 4!}=280.$$

Перестановки з повтореннями. Число різних перестановок, які можна скласти з n елементів, серед яких є k_1 однакових елементів першого типу, k_2 однакових елементів другого типу, … , k_m однакових елементів $m-$ го типу, дорівнює

$$C_n(k_1,k_2,...,k_m)=\frac{n!}{k_1!k_2!..k_m!}.$$

Приклад 8. Скільки різних слів можна отримати, переставляючи букви слова “математика”?

Розв’язок. Слово “математика” містить 10 букв, серед них 3 букви ”а”, 2 – “м”, 2 – “т”, 1 – “е”, 1 – “и”, 1 – “к”. Шукане число слів дорівнює

$$C_{10}(3,2,2,1,1,1)=\frac{10!}{3!\cdot 2!\cdot 2!}=151200$$

Лекція 2

Класичне означення ймовірності. Статистичне означення ймовірності. Геометрична ймовірність

Класичне означення ймовірності. Ймовірністю події A називають відношення числа елементарних результатів досліду, які сприяють настанню події A, до загального числа можливих елементарних результатів досліду

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

Приклад 1. Кидають гральний кубик.

Знайти ймовірність того, що а) випаде 5 очок, б) випаде більше 5 очок

Розв'язок. а) Загальне число можливих результатів дорівнює 6 (кількість випавших очок може бути від 1 до 6). Сприятливим є лише один результат – випало 5 очок. Таким чином, ймовірність випадання 5 очок дорівнює P=1/6.

б) Загальне число можливих результатів дорівнює 6. Сприятливими є два результати – випало 5 очок або 6 очок. Ймовірність того, що випаде більше 5 очок дорівнює P=2/6=1/3.

Приклад 2. Кидають два гральних кубики.

Знайти ймовірність того, що сума очок на випавших гранях а) дорівнює 12, б) більше 10.

Розв'язок. а) Кожен із результатів кидання "одного" кубика може сполучатися із кожним результатом кидання "іншого" кубика. Таким чином, загальне число можливих елементарних результатів випробування дорівнює $6 \cdot 6 = 36$. Сприятливим є тільки один результат – на обох кубиках випало число 6. Ймовірність того, що сума очок на випавших гранях дорівнює 12: P=1/36.

б) Загальне число можливих елементарних результатів випробування дорівнює $6 \cdot 6 = 36.$ Сприятливими є такі три результати: 1) 5,6; 2) 6,5 та 3) 6,6. Таким чином, ймовірність того, що сума очок на випавших гранях більша 10 буде дорівнювати Р=3/36=1/12.

Приклад 3. У групі 15 студентів. Серед них 3 відмінники. Навмання відбирають 6 студентів.

Знайти ймовірність того, що серед відібраних 2 відмінники.

Розв'язок. Загальна кількість способів, якими можна відібрати 6 студентів із 15 дорівнює C_{15}^{6}. Підрахуємо кількість сприятливих варіантів, коли серед відібраних є 2 відмінники і 4 не відмінники. Двох відмінників із трьох можна вибрати C_3^2 способами. Залишається вибрати ще чотири студенти із тих, що є не відмінниками. Це можна зробити C_{13}^{4} способами, тому кількість сприятливих варіантів дорівнює $C_3^2 \cdot C_{13}^4$.

Ймовірність того, що серед шести відібраних студентів буде двоє відмінників дорівнює

$$P = C_3^2 \cdot C_{13}^4 / C_{15}^6 .$$

Статистичне означення ймовірності. При статистичному означенні в якості ймовірності події приймають її відносну частоту, яка визначається формулою

$$W(A) = \frac{m}{n},$$

де m – число дослідів, у яких подія А відбулась, n – загальне число проведених дослідів.

Геометрична ймовірність. Нехай відрізок l складає частину відрізка L. На відрізок L навмання ставлять точку. Якщо вважати, що ймовірність попадання точки на відрізок l пропорційна довжині цього відрізку і не залежить від його розміщення відносно відрізка L, то ймовірність попадання точки на відрізок l визначається рівнянням

$$P = \frac{l}{L}.$$

Нехай плоска фігура s складає частину плоскої фігури S. На фігуру S навмання ставлять точку. Якщо вважати, що ймовірність попадання точки на фігуру s пропорційна площі цієї фігури і не залежить ні від її розміщення відносно S, ні від форми s, то ймовірність попадання точки у фігуру s визначається рівнянням

$$P = \frac{s}{S}.$$

Аналогічно визначається ймовірність попадання точки у просторову фігуру v, яка складає частину фігури V.

Приклад 1. На відрізку L довжиною 25 см знаходиться менший відрізок l довжиною 5 см. Знайти ймовірність того, що точка, навмання поставлена на великий відрізок, потрапить і на малий відрізок. Вважаємо, що ймовірність попадання точки на відрізок l пропорційна довжині цього відрізку і не залежить від його розміщення відносно відрізка L,

Розв'язок. Шукана ймовірність дорівнює

$$P = \frac{l}{L} = \frac{5}{25} = 0{,}2.$$

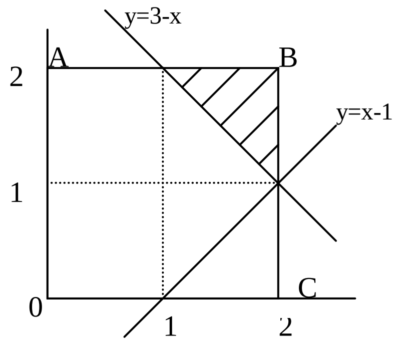
Приклад 2. У квадрат із стороною a=20 см вписано коло. У квадрат навмання ставиться точка. Знайти ймовірність того, що ця точка потрапить у середину кола. Вважаємо, що ймовірність попадання точки у круг пропорційна площі круга і не залежить від його розташування.

Розв'язок. Радіус вписаного кола $r = a/2 = 10$ ńě. Шукана ймовірність дорівнює

$$P = \frac{\text{Площа круга}}{\text{Площа квадрата}} = \frac{\pi r^2}{a^2} = \frac{314}{400} = 0{,}785.$$

Приклад 3. Навмання беруть два додатних числа x та y, кожне з яких не перевищує двійки. Знайти ймовірність того, що сума цих чисел $x+y$ більша трьох, а різниця $x–y$ менша за одиницю.

Розв'язок. За умовою задачі беруть числа, які задовільняють нерівностям $0<x\leq 2$ та $0<y\leq 2$. Розглянемо прямокутну систему координат $x0y$. У цій системі координат наведеним вище подвійним нерівностям задовільняють координати будь–якої точки, що належить квадрату $OABC$. Таким чином, цей квадрат можна роглядати як фігуру S, координати точок якої представляють усі можливі пари значень чисел x та y.



Пара чисел x та y повинна також задовольняти нерівностям $x+y>3$ та $x–y<1$. Перепишемо нерівності у такому вигляді $y>3–x$, $y>x–1$. Ці нерівності виконуються для координат тих точок фігури S, які лежать вище прямих $y=3–x$ та $y=x–1$ і утворюють фігуру s (заштрихований трикутник).

Шукана ймовірність

$$P = \frac{s}{S} = \frac{1}{8} = 0{,}125.$$

Лекція 3

Протилежні події. Теореми про додавання та множення ймовірностей

Протилежні події. Подія $\bar{A}$ називається протилежною до події A, якщо вона полягає у тому, що подія A не відбулась.

Приклад 1. Кидають гральний кубик. Подія A – випаде парне число очок, значить протилежна подія $\bar{A}$ – парне число очок не випаде.

Якщо у результаті експерименту відбулась подія A, то, відповідно, подія $\bar{A}$ не відбулась. І навпаки, якщо відбулась подія $\bar{A}$, то подія A не відбулась. Звідси випливає, що

$$P(A)+P(\bar{A})=1.$$

Прийнято ймовірність події A позначати через p, а ймовірність протилежної події $\bar{A}$ – через q. Тоді отримуємо

$$p+q=1.$$

Приклад 2. Ймовірність того, що солдат влучить у мішень p=0,7. Знайти ймовірність того, солдат у мішень не влучить.

Розв'язок. Події "влучив у мішень" та "не влучив" протилежні, тому шукана ймовірність дорівнює

$$q=1-p=1-0{,}7=0{,}3.$$

Сумісні та несумісні події. Події A та B називаються сумісними, якщо вони можуть відбутись одночасно в одному і тому ж експерименті. Події A та B називаються несумісними, якщо вони не можуть відбутись одночасно в одному і тому ж експерименті.

Приклад 3. Кидають гральний кубик. Подія A – випаде число 6, подія B – випаде число очок більше 4. Події A та B сумісні, так як якщо випаде число очок 6, то відбудуться обидві події A та B.

Приклад 4. Кидають гральний кубик. Подія A – випаде парне число очок, подія B – випаде непарне число очок. Події A та B несумісні, так як не може одночасно випасти парна та непарна кількість очок.

Залежні та незалежні події. Події A та B називаються незалежними, якщо ймовірність однієї із них не залежить від настання чи ненастання іншої. Відповідно, події A та B називаються залежними, якщо ймовірність однієї із них залежить від настання чи ненастання іншої.

Умовна ймовірність. Умовною ймовірністю $P_A(B)$ називають ймовірність події B, обчислену при умові, що подія A уже відбулась.

Приклад 5. Гральний кубик кидають два рази. Подія A – випало 5 очок при першому киданні. Подія B – випало 6 очок при другому киданні. Події A та B незалежні, так як ймовірність кожної з цих подій не залежить від настання іншої, завжди залишається однаковою і дорівнює 1/6.

Приклад 6. Студент знає 20 із 25 питань програми. Він отримує два питання. Позначимо: подія A – студент знає перше питання, подія B – студент знає друге питання. Події A та B залежні, так як ймовірність події B залежить від того, відбулась подія A чи ні. Якщо студент знає перше питання, то серед 24 питань, що залишились, буде 19, які він знає і ймовірність події B буде дорівнювати

$$P_A(B) = \frac{19}{24} = 0{,}79.$$

Якщо студент не знає перше питання, то серед 24 питань, що залишились, буде 20, які він знає і ймовірність події B буде дорівнювати

$$P_A(B) = \frac{20}{24} = 0{,}83.$$

Добуток подій. Добутком двох подій A та B називається така подія C, яка полягає у тому, що відбулись обидві події A та B.

Теорема 1. Ймовірність того, що відбулись дві незалежні події *A* та *B* дорівнює добутку ймовірностей цих подій

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

Приклад 7. Кидають два гральні кубики. Знайти ймовірність того, що на першому кубику випаде число 6, а на другому – число 5.

Розв'язок. Позначимо: подія *A* – на першому кубику випаде число 6, подія *B* – на другому випаде число 5.

Повинні відбутись і подія *A* і подія *B*. Отже маємо добуток двох подій *AB*. Події *A* та *B* незалежні. За теоремою 1 обчислюємо шукану ймовірність

$$P(AB) = P(A)P\ (B) = \frac{1}{6} \cdot \frac{1}{6} = 0{,}028.$$

Теорема 2. Ймовірність того, що відбулись дві залежні події *A* та *B* дорівнює добутку ймовірності однієї з цих подій на умовну ймовірність іншої при умові, що перша подія вже відбулась,

$$P(AB)=P(A)P_A(B).$$

Приклад 8. Студент знає 20 із 25 питань програми. Він отримує два питання. Знайти ймовірність того, що студент буде знати обидва питання.

Розв'язок. Позначимо: подія *A* – студент знає перше питання, подія *B* – студент знає друге питання. Події *A* та *B* залежні, так як ймовірність події *B* залежить від того, відбулась подія *A* чи ні.

Застосовуючи теорему 2 отримаємо шукану ймовірність

$$P(AB) = P(A)P_A(B) = \frac{20}{25} \cdot \frac{19}{24} = 0{,}63.$$

Сума подій. Сумою двох подій A та B називається така подія C, яка полягає у тому, що відбулась або подія A або подія B або обидві події A та B разом.

Теорема 3. Ймовірність суми двох несумісних подій A та B дорівнює сумі ймовірностей цих подій

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

Приклад 9. Кидають гральний кубик. Знайти ймовірність того, що випаде або число 6 або непарна кількість очок.

Розв'язок. Позначимо: подія A – випало число 6, подія B – випала непарна кількість очок. Події A та B несумісні. Застосовуючи теорему 3 отримуємо шукану ймовірність

$$P(A+B)=P(A)+P(B)=\frac{1}{6}+\frac{3}{6}=0{,}67.$$

Теорема 4. Ймовірність суми двох сумісних подій A та B дорівнює сумі ймовірностей подій A та B без ймовірності добутку цих подій

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

Приклад 10. Ймовірність, що перший студент здасть екзамен дорівнює 0,7. Ймовірність, що другий студент здасть екзамен дорівнює 0,8. Знайти ймовірність того, що екзамен здасть хоча б один студент.

Розв'язок. Позначимо: подія A – перший студент здав екзамен, подія B – другий студент здав екзамен. Події A та B сумісні, тому

$$P(A+B)=P(A)+P(B)-P(AB)=0{,}7+0{,}8-0{,}7\cdot 0{,}8=0{,}94.$$

Лекція 4

Формула повної ймовірності. Формула Бейеса. Повторні незалежні випробування.

Формула повної ймовірності. Ймовірність події A, яка може відбутись лише після настання однієї із несумісних подій (гіпотез) $B_1, B_2, ..., B_n$, які утворюють повну групу подій $(P(B_1)+P(B_2)+...+P(B_n)=1)$, дорівнює сумі добутків ймовірностей кожної із гіпотез на відповідну умовну ймовірність події A

$$P(A)=P(B_1)\cdot P_{B_1}(A)+P(B_2)\cdot P_{B_2}(A)+...+P(B_n)\cdot P_{B_n}(A).$$

Приклад 1. На курсі чотири групи, у кожній з яких 25 студентів. У першій групі 5 відмінників, у другій – 6, у третій – 4, у четвертій – 3. Навмання вибирають групу а потім з цієї групи навмання вибирають студента.

Знайти ймовірність того, що вибраний студент відмінник.

Розв'язок. Використаємо формулу повної ймовірності. Події (гіпотези) B_1, B_2, B_3, B_4 – це вибір першої, другої, третьої або четвертої групи

$$P(B_1)=P(B_2)=P(B_3)=P(B_4)=\frac{1}{4}=0,25.$$

Якщо вибрали першу групу, то ймовірність вибрати відмінника з цієї групи дорівнює

$$P_{B_1}(A)=\frac{5}{25}=0,2.$$

Якщо вибрали другу групу, то ймовірність вибрати відмінника з цієї групи дорівнює

$$P_{B_2}(A)=\frac{6}{25}=0{,}24.$$

Відповідно для третьої і четвертої груп отримуємо

$$P_{B_3}(A)=\frac{4}{25}=0{,}16,\;\; P_{B_4}(A)=\frac{3}{25}=0{,}12.$$

Підставимо отримані дані у формулу повної ймовірності

$$P(A)=P(B_1)\cdot P_{B_1}(A)+P(B_2)\cdot P_{B_2}(A)+P(B_3)\cdot P_{B_3}(A)+P(B_4)\cdot\cdot P_{B_4}(A)=$$

$$=0{,}25\cdot 0{,}2+0{,}25\cdot 0{,}24+0{,}25\cdot 0{,}16+0{,}25\cdot 0{,}12=0{,}18.$$

Формула Бейеса. Нехай подія *А* може відбутись лише після настання однієї із несумісних подій (гіпотез) B_1, B_2, ..., B_n, які утворюють повну групу подій. Якщо подія *А* вже відбулась, то ймовірності гіпотез можуть бути переоцінені за формулою Бейеса

$$P_A(B_i)=\frac{P(B_i)\cdot P_{B_i}(A)}{P(A)} \quad (i=1,2,\ldots,n),$$

де

$$P(A)=P(B_1)\cdot P_{B_1}(A)+P(B_2)\cdot P_{B_2}(A)+\ldots+P(B_n)\cdot P_{B_n}(A).$$

Приклад 2. На курсі чотири групи, у кожній з яких 25 студентів. У першій групі 5 відмінників, у другій – 6, у третій – 4, у четвертій – 3. Навмання вибирають групу а потім з цієї групи навмання вибирають студента. Вибраний студент виявився відмінником.

Знайти ймовірність того, що він з третьої групи.

Розв'язок. Використаємо дані із попереднього прикладу і підставимо їх у формулу Бейеса

$$P_A(B_3)=\frac{P(B_3)\cdot P_{B_3}(A)}{P(A)}=\frac{0{,}25\cdot 0{,}16}{0{,}18}=0{,}22..$$

Формула Бернуллі. Ймовірність того, що в n незалежних дослідах, у кожному з яких ймовірність настання події дорівнює p ($0<p<1$), подія відбудеться рівно k разів (незалежно у якій послідовності), дорівнює

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

де $q=1-p$.

Ймовірність того, що в n незалежних дослідах, у кожному з яких ймовірність настання події дорівнює p ($0<p<1$), подія відбудеться не менше k_1 і не більше k_2 разів дорівнює

$$P_n(k_1)+P_n(k_1+1)+P_n(k_1+2)+...+P_n(k_2).$$

Приклад 1. Монету кидають п'ять разів. Знайти ймовірність того, що герб випаде: а)рівно три рази; б)не менше трьох разів.

Розв'язок. Ймовірність випадання герба у кожному киданні постійна $p=1/2$ і не залежить від результатів інших кидань монети, тому можемо використати формулу Бернуллі. Таким чином,

а) ймовірність того, що герб випаде рівно три рази дорівнює

$$P_5(3)=C_5^3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^{5-3}=\frac{5!}{3!\cdot(5-3)!}\cdot\left(\frac{1}{2}\right)^5=0{,}31;$$

б) ймовірність того, що герб випаде не менше трьох разів дорівнює

$$P_5(3)+P_5(4)+P_5(5)=\frac{10}{32}+\frac{5}{32}+\frac{1}{32}=0{,}5.$$

У випадку, коли число n досить велике, то замість формули Бернуллі використовують локальну та інтегральну формули Лапласа.

Локальна формула Лапласа. Ймовірність того, що в n незалежних дослідах, у кожному з яких ймовірність настання події дорівнює p ($0<p<1$), подія відбудеться рівно k разів (незалежно у якій послідовності), наближено дорівнює (тим точніше, чим більше n)

$$P_n(k)=\frac{1}{\sqrt{npq}}\varphi(x),$$

де

$$q=1-p,\quad \varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},\quad x=\frac{k-np}{\sqrt{npq}}.$$

Значення функції Гауса $\varphi(x)$ наведені у додатку 1. Функція $\varphi(x)$ парна, тобто $\varphi(-x)=\varphi(x)$.

Приклад 2. Гральний кубик кидають 100 разів. Знайти ймовірність того, що число 6 випаде 30 разів.

Розв'язок. Число дослідів $n=100$. Кількість випадань числа 6 $k=30$. Ймовірність випадання числа 6 при кожному киданні $p=1/6=0{,}17$, відповідно $q=1-0{,}17=0{,}83$. Обчислимо значення числа x

$$x=\frac{k-np}{\sqrt{npq}}=\frac{30-100\cdot 0{,}17}{\sqrt{100\cdot 0{,}17\cdot 0{,}83}}=\frac{13}{3{,}76}=3{,}46.$$

Із додатку 1 знаходимо $\varphi(x)=\varphi(3{,}46)=0{,}0011$. Підставимо отримані значення у локальну формулу Лапласа

$$P_n(k)=\frac{1}{\sqrt{npq}}\varphi(x)=P_{100}(30)=\frac{1}{3{,}76}\cdot 0{,}0011=0{,}0003.$$

Інтегральна формула Лапласа. Ймовірність того, що в n незалежних дослідах, у кожному з яких ймовірність настання події дорівнює p $(0<p<1)$, подія відбудеться не менше k_1 і не більше k_2 разів наближено дорівнює

$$P_n(k_1;k_2)=\Phi(x^{''})-\Phi(x^{'}),$$

де

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{z^2}{2}}d(z)-$$ функція Лапласа,

$$x^{'}=\frac{k_1-np}{\sqrt{npq}},\quad x^{''}=\frac{k_2-np}{\sqrt{npq}}.$$

Таблиця значень функції Лапласа для додатних значень x $(0\le x\le 5)$ наведена у додатку 2. Для значень $x>5$ приймають $\Phi(x)=0{,}5$. Для від’ємних значень x використовують цю ж таблицю, враховуючи, що функція Лапласа непарна $[\Phi(-x)=-\Phi(x)]$.

Приклад 3. Гральний кубик кидають 100 разів. Знайти ймовірність того, що число 6 випаде в межах від 30 до 50 разів.

Розв’язок. Число дослідів $n=100$. Кількість випадань числа 6 знаходиться в межах від $k_1=30$ до $k_2=50$ разів. Ймовірність випадання числа 6 при кожному киданні $p=1/6=0{,}17$, відповідно $q=1-0{,}17=0{,}83$. Обчислимо значення параметрів $x^{'}$ та $x^{''}$

$$x^{'}=\frac{k_1-np}{\sqrt{npq}}=\frac{30-100\cdot 0{,}17}{\sqrt{100\cdot 0{,}17\cdot 0{,}83}}=\frac{13}{3{,}76}=3{,}46;$$

$$x'' = \frac{k_2 - np}{\sqrt{npq}} = \frac{50 - 100 \cdot 0{,}17}{\sqrt{100 \cdot 0{,}17 \cdot 0{,}83}} = \frac{37}{3{,}76} = 9{,}8.$$

Із додатку 2 знаходимо $\Phi(x')$=Φ(3,46)=0,499, $\Phi(x'')$= Φ(9,8)=0,5. Підставимо отримані значення в інтегральну формулу Лапласа

$$P_{100}(30;50) = \Phi(9{,}8) - \Phi(3{,}46) = 0{,}5 - 0{,}499 = 0{,}001.$$

Лекція 5

Числові характеристики дискретних випадкових величин.

Закон розподілу дискретної випадкової величини. Законом розподілу дискретної випадкової величини називають перелік її можливих значень і відповідних їм ймовірностей. Найчастіше закон розподілу задають у вигляді таблиці

x_i	x_1	x_2	...	x_n
p_i	p_1	p_2	...	p_n

$x_1, x_2, \dots x_n$ – це значення, яких набирає випадкова величина X,
$p_1, p_2, \dots p_n$ – відповідні цим значенням ймовірності, причому $p_1+p_2+\dots+p_n=1$.

Математичне сподівання. Нехай дискретна випадкова величина X може приймати лише значення x_1, x_2, ..., x_n, ймовірності яких відповідно рівні p_1, p_2, ..., p_n. Математичним сподіванням дискретної випадкової величини X називають суму добутків усіх її можливих значень на відповідні ймовірності

$$M(X)=x_1p_1+x_2p_2+\dots+x_np_n.$$

Властивості математичного сподівання.

Властивість 1. Математичне сподівання постійної величини дорівнює самій постійній величині

$$M(C)=C.$$

Властивість 2. Постійний множник можна виносити за знак математичного сподівання

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

Властивість 3. Математичне сподівання суми двох випадкових величин дорівнює сумі математичних сподівань цих величин

$$M(X+У)=M(X)+M(У).$$

Властивість 4. Математичне сподівання добутку двох незалежних випадкових величин дорівнює добутку математичних сподівань цих величин

$$M(XУ)=M(X)M(У).$$

Властивість 5. Математичне сподівання числа настання події A в n незалежних дослідах, у кожному з яких ймовірність настання події p постійна, дорівнює добутку числа випробувань на ймовірність настання події у кожному випробуванні

$$M(X)=np.$$

Відхилення. Відхиленням називають різницю між випадковою величиною і її математичним сподіванням $X-M(X)$.

Дисперсія. Дисперсією дискретної випадкової величини називають математичне сподівання квадрату відхилення випадкової величини від її математичного сподівання

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

Для обчислення дисперсії зручно використовувати формулу

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

$M(X^2)$ знаходять за формулою

$$M(X^2)=x_1^2p_1+x_2^2p_2+...+x_n^2p_n.$$

Властивості дисперсії.

Властивість 1. Дисперсія постійної величини дорівнює нулю

$$D(C)=0.$$

Властивість 2. Постійний множник можна виносити за знак дисперсії, піднісши його до квадрату

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

Властивість 3. Дисперсія суми двох незалежних випадкових величин дорівнює сумі дисперсій цих величин

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

Властивість 4. Дисперсія різниці двох незалежних випадкових величин дорівнює сумі дисперсій цих величин

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

Властивість 5. Дисперсія числа настання події A в n незалежних дослідах, у кожному з яких ймовірність настання події p постійна, дорівнює добутку числа випробувань на ймовірності настання і ненастання події у кожному випробуванні

$$D(X)=npq.$$

Середнє квадратичне відхилення. Середнім квадратичним відхиленням називають квадратний корінь із дисперсії

$$\sigma=\sqrt{D}.$$

Приклад 1. Дискретна випадкова величина X задана законом розподілу

x_i	1	2	3
p_i	0,2	0,5	0,3

Знайти математичне сподівання, дисперсію та середнє квадратичне відхилення.

Розв'язок. Математичне сподівання знаходимо за означенням

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

Дисперсію знаходимо за формулою

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

Спочатку знайдемо $M(X^2)$

$$M(X^2) = 1^2 \cdot 0{,}2 + 2^2 \cdot 0{,}5 + 3^2 \cdot 0{,}3 = 4{,}9.$$

Тоді

$$D(X) = 4{,}9 - 2{,}1^2 = 0{,}49.$$

Середнє квадратичне відхилення – це квадратний корінь із дисперсії

$$\sigma = \sqrt{0{,}49} = 0{,}7.$$

Лекція 6

Функції розподілу і щільності розподілу ймовірностей випадкової величини. Нормальний розподіл

Функція розподілу. Функцією розподілу ймовірностей випадкової величини називають функцію $F(x)$, яка визначає ймовірність того, що випадкова величина X у результаті випробування набере значення менше за x

$$F(x)=P(X<x).$$

Властивості функції розподілу.

Властивість 1. Функція розподілу набирає значень з відрізку [0,1]

$$0 \le F(x) \le 1.$$

Властивість 2. $F(x)$ – неспадна функція, тобто якщо $x_2>x_1$ то $F(x_2) \ge F(x_1)$.

Наслідок. Ймовірність того, що випадкова величина набере значення з інтервалу (a,b) дорівнює приросту функції розподілу на цьому інтервалі

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

Властивість 3. Якщо можливі значення випадкової величини належать інтервалу (a,b), то: 1)$F(x)=0$ при $x \le a$; 2)$F(x)=1$ при $x \ge b$.

Приклад 1. Задано закон розподілу випадкової величини X

x_i	– 4	– 1	2	6	9	13
p_i	0,1	0,2	0,1	0,3	0,1	0,2

Записати у вигляді формули функцію $F(x)$ та побудувати її графік.

Розв'язок. Згідно з властивостями $F(x)$, дістаємо наведені далі співвідношення.

1) $F(-4) = P(X < -4) = 0$;

2) $F(-1) = P(X < -1) = P(X = -4) = 0{,}1$;

3) $F(2) = P(X < 2) = P(X = -4) + P(X = -1) = 0{,}1 + 0{,}2 = 0{,}3$;

4) $F(6) = P(X < 6) = P(X = -4) + P(X = -1) + P(X = 2) = 0{,}1 + 0{,}2 + 0{,}1 = 0{,}4$;

5) $F(9) = P(X < 9) = P(X = -4) + P(X = -1) + P(X = 2) + P(X = 6) = 0{,}1 + 0{,}2 + 0{,}1 + 0{,}3 = 0{,}7$;

6) $F(12) = P(X < 13) = P(X = -4) + P(X = -1) + P(X = 2) + P(X = 9) = 0{,}1 + 0{,}2 + 0{,}1 + 0{,}3 + 0{,}1 = 0{,}8$;

7) $F(x)|_{x>13} = P(X > 13) = P(X = -4) + P(X = -1) + P(X = 2) + P(X = 9) + P(X = 13) = 0{,}1 + 0{,}2 + 0{,}1 + 0{,}1 + 0{,}3 + 0{,}1 + 0{,}2 = 1$.

Графік функції $F(x)$ зображено на рис.1

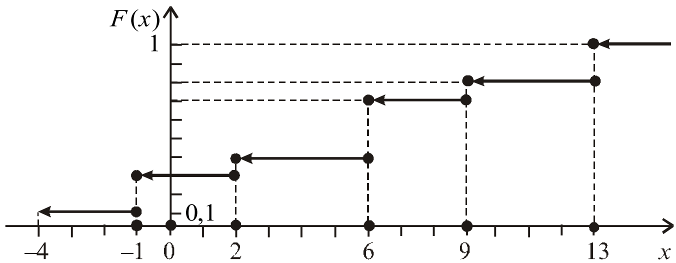


Рис. 1

Щільність розподілу. Щільністю розподілу ймовірностей неперервної випадкової величини X називають функцію $f(x)$ – першу похідну від функції розподілу $F(x)$

$$f(x)=F'(x).$$

Властивості щільності розподілу.

Властивість 1. Щільність розподілу – функція невід'ємна

$$f(x) \geq 0.$$

Властивість 2. (Умова нормування). Невласний інтеграл від щільності розподілу у межах від – ∞ до +∞ дорівнює одиниці

$$\int_{-\infty}^{\infty} f(x)dx = 1\,.$$

Якщо неперервна випадкова величина X визначена лише на проміжку $[a;\, b]$, то умова нормування має такий вигляд

$$\int_{\acute{r}}^{b} f(x)dx = 1.$$

Ймовірність того, що неперервна випадкова величина X набере значень з інтервалу (a,b), дорівнює

$$P(a < X < b) = \int_{a}^{b} f(x)dx.$$

Якщо щільність розподілу $f(x)$ відома, то функція розподілу $F(x)$ знаходиться за формулою

$$F(x) = \int_{-\infty}^{x} f(x)dx.$$

Математичне сподівання. Математичне сподівання неперервної випадкової величини Х обчислюють за формулою

$$M(X) = \int_{-\infty}^{\infty} x\, f(x)dx\,.$$

У випадку, коли усі можливі значення величини X належать інтервалу (a,b), то

$$M(X)=\int_{ŕ}^{b} x\, f(x)dx.$$

Дисперсія. Дисперсія неперервної випадкової величини X обчислюється за формулою

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

або рівносильною формулою

$$D(X)=\int_{-\infty}^{\infty}x^2 f(x)dx-M^2(X)\,.$$

У випадку, коли усі можливі значення величини X належать інтервалу (a,b), то використовують формулу

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

або рівносильну формулу

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

Середнє квадратичне відхилення. Середнє квадратичне відхилення неперервної випадкової величини X обчислюється за формулою

$$\sigma\left(X\right)=\sqrt{D(X)}\ .$$

Нормальний розподіл. Нормальним називають такий розподіл ймовірностей неперервної випадкової величини, який описується щільністю

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

де параметр a – математичне сподівання нормального розподілу,
σ – середнє квадратичне відхилення.

Ймовірність попадання у заданий інтервал нормальної випадкової величини. Ймовірність того, що розподілена за нормальним законом випадкова величина X набере значення з інтервалу (α, β) дорівнює

$$P(\alpha < X < \beta)=\Phi\left(\frac{\beta-a}{\sigma}\right)-\Phi\left(\frac{\alpha-a}{\sigma}\right),$$

де $\Phi(x)$ – функція Лапласа.

Обчислення ймовірності заданого відхилення. Ймовірність того, що відхилення нормально розподіленої випадкової величини X за абсолютною величиною менше заданого додатного числа δ, дорівнює

$$P\left(\left|X-a\right|<\delta\right)=2\Phi\left(\frac{\delta}{\sigma}\right).$$

Правило трьох сігм. Ймовірність того, що відхилення нормально розподіленої випадкової величини X за абсолютною величиною менше потроєного середнього квадратичного відхилення дорівнює 0,9973:

$$P(|X-a|<3\sigma)=0,9973.$$

Лекція 7

Статистичний розподіл вибірки

Генеральна сукупність. Генеральна сукупність – множина однотипних об'єктів, кількісна чи якісна ознака яких підлягає вивченню.

Вибіркова сукупність – підмножина об'єктів, відібраних у відповідний спосіб із генеральної сукупності.

Варіанти. Коли реалізується вибірка, кількісна ознака X набуває конкретних числових значень $x_1, x_2, ..., x_k$, які називаються варіантами.

Варіаційний ряд. Числовий ряд варіант, розташованих у зростаючому порядку, називається варіаційним.

Частоти. Кожна варіанта x_i вибірки може бути спостереженою n_i раз ($n_i \geq 1$). Число n_i називають частотою варіанти x_i.

При цьому

$$n = \sum_{i=1}^{k} n_i ,$$

де k – кількість варіант, які відрізняються числовим значенням,
n – об'єм вибірки.

Відносні частоти. Відношення частоти n_i варіанти x_i до об'єму вибірки n називають її відносною частотою і позначають через w_i

$$w_i = \frac{n_i}{n} .$$

Для кожної вибірки виконується рівність

$$\sum_{i=1}^{k} w_i = 1 .$$

Накопичені частоти. Накопичена частота N_i варіанти x_i – це сума частот варіант, які не перевищують x_i.

Накопичені відносні частоти. Накопичена відносна частота W_i варіанти x_i – це сума відносних частот варіант, які не перевищують x_i.

Дискретний статистичний розподіл вибірки. Дискретним статистичним розподілом вибірки називають перелік варіант варіаційного ряду і відповідних їм частот або відносних частот.

У табличній формі він має вигляд

x_i	x_1	x_2	x_3	…	x_k
n_i	n_1	n_2	n_3	…	n_k

або такий вигляд

x_i	x_1	x_2	x_3	…	x_k
w_i	w_1	w_2	w_3	…	w_k

Інтервальний статистичний розподіл вибірки. У разі, коли досліджувана ознака X є неперервною величиною і кількість варіант досить велика, результати вибірки подають інтервальним рядом. Для цього область значень розбивають на k інтервалів і для кожного інтервалу визначають частоти. Кількість інтервалів визначають за формулою Стерджеса

$$k=1+3{,}3\lg n,$$

де n – об’єм вибірки.

Довжину інтервалів h найчастіше беруть однаковою і визначають за формулою

$$h=(x_{max}-x_{min})/k,$$

де x_{max} та x_{min} – це максимальне та мінімальне значення у вибірці.

Інтервальним статистичним розподілом вибірки називають перелік часткових інтервалів і відповідних їм частот або відносних частот.

У табличній формі цей розподіл має вигляд

інтервал	$x_0 - x_1$	$x_1 - x_2$	$x_2 - x_3$	…	$x_{k-1} - x_k$
n_i	n_1	n_2	n_3	…	n_k

Полігон частот і відносних частот. Дискретний статистичний розподіл вибірки можна зобразити графічно у вигляді ламаної лінії.

Полігон частот – це ламана, відрізки якої сполучають точки з координатами (x_i; n_i). Полігон відносних частот – це ламана, відрізки якої сполучають точки з координатами (x_i; W_i).

Гістограма частот і відносних частот. Для графічного зображення інтервального статистичного розподілу вибірки використовують гістограму частот і гістограму відносних частот.

Гістограмою частот називають ступінчасту фігуру, що складається з прямокутників, основами яких є часткові інтервали довжини $h=x_i-x_{i-1}$, а висотами відношення n_i/h (щільність частоти). Площа гістограми частот дорівнює об'єму вибірки.

Гістограмою відносних частот називають ступінчасту фігуру, що складається з прямокутників, основами яких є часткові інтервали довжини $h=x_i-x_{i-1}$, а висотами відношення w_i/h (щільність відносної частоти). Площа гістограми відносних частот дорівнює одиниці.

Кумулята. Кумулята частот – це ламана, відрізки якої сполучають точки з координатами (x_i; N_i). Кумулята відносних частот – це ламана, відрізки якої сполучають точки з координатами (x_i; W_i).

У випадку, коли статистичний розподіл вибірки є інтервальним, то у якості варіант приймають верхні межі інтервалів.

При збільшенні об'єму вибірки гістограма відносних частот наближається до функції щільності розподілу, а кумулята відносних частот – до функції розподілу.

Емпірична функція розподілу $F^*(x)$ та її властивості. Емпірична функція розподілу $F^*(x)$ – це функція аргументу x, що визначає відносну частоту події $X < x$, тобто

$$F^*(x) = W(X < x) = \frac{n_x}{n},$$

де n — об'єм вибірки, n_x — кількість варіант статистичного розподілу вибірки, значення яких менші за фіксовану варіанту x.

Властивості $F^*(x)$:

1) $0 \le F^*(x) \le 1$;

2) $F^*(x)=0$ для $x \le x_{min}$, де $x_{\min}$ - найменша варіанта варіаційного ряду;

3) $F^*(x)=1$ для $x > x_{max}$, де $x_{\max}$ - найбільша варіанта варіаційного ряду;

4) $F^*(x)$ є неспадна функція аргументу x, а саме: $F(x_2) \ge F(x_1)$ при $x_2 \ge x_1$.

Приклад 1. У цеху встановлено 5 верстатів. Протягом 25 днів реєструвалась кількість верстатів, які не працювали. Отримали такі значення: 0, 1, 2, 1, 1, 2, 3, 2, 1, 4, 2, 0, 0, 2, 2, 3, 3, 1, 0, 1, 2, 1, 3, 5, 0.

Записати статистичний розподіл вибірки, побудувати полігон і кумуляту частот, записати емпіричну функцію розподілу і побудувати її графік.

Розв'язок. Досліджувана величина X (кількість верстатів, які не працювали) може набирати значення 0,1,2,3,4,5. Загальний об'єм вибірки n=25.

На підставі отриманих даних складемо статистичний розподіл вибірки і додамо до отриманої таблиці рядок з накопиченими частотами

x_i	0	1	2	3	4	5
n_i	5	7	7	4	1	1
N_i	5	12	19	23	24	25

На основі цієї таблиці будуємо полігон частот (рис.1) і кумуляту накопичених частот (рис.2).

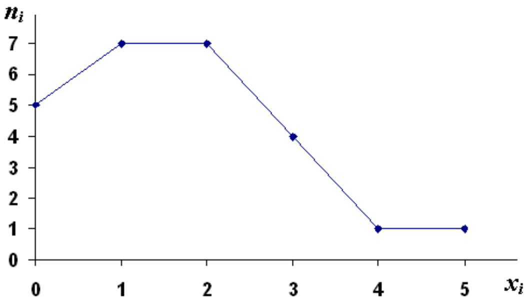


Рис.1

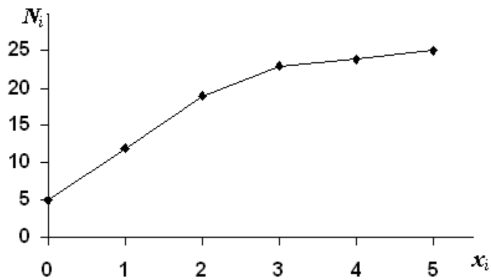


Рис.2.

Запишемо емпіричну функцію розподілу, скориставшись формулою $F^*(x) = W(X < x) = \dfrac{n_x}{n}$

$$F^*(x) = \begin{cases} 0, & \text{якщо } x \le 0; \\ 0{,}2, & \text{якщо } 0 < x \le 1; \\ 0{,}48, & \text{якщо } 1 < x \le 2; \\ 0{,}76, & \text{якщо } 2 < x \le 3; \\ 0{,}92, & \text{якщо } 3 < x \le 4; \\ 0{,}96, & \text{якщо } 4 < x \le 5; \\ 1, & \text{якщо } x > 5. \end{cases}$$

Графік емпіричної функції розподілу наведено на рис.3.

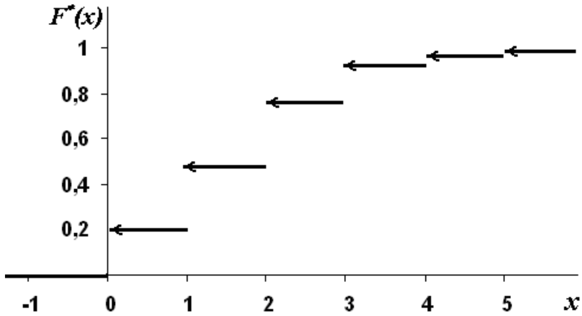


Рис.3.

П р и к л а д 2. Магазин за тиждень реалізував 30 пар жіночого взуття наступних розмірів: 37, 35, 33, 36, 41, 37, 36, 36, 39, 38, 34, 40, 37, 35, 39, 36, 38, 37, 38, 35, 34, 36, 39, 38, 35, 37, 38, 39, 38, 37.

Записати інтервальний статистичний розподіл вибірки. Побудувати гістограму частот.

Р о з в' я з о к. Мінімальне значення вибірки дорівнює 33, максимальне значення дорівнює 41. Хоча обчислена за формулою Стерджеса кількість інтервалів для вибірки об'ємом 30 дорівнює 6, однак, виходячи з практичних міркувань розіб'ємо наш діапазон значень на 4 інтервали однакової довжини: 33-35, 35-37, 37-39, 39-41. Підраховуючи кількість значень з кожного інтервалу будемо дотримуватись наступного правила: якщо значення співпадає з лівим краєм інтервалу, то це значення відносять до даного інтервалу, якщо ж значення співпадає з правим краєм інтервалу, то це значення відносять до наступного інтервалу (крім останнього).

Враховуючи усе вищесказане запишемо інтервальний статистичний розподіл вибірки

Розмір (інтервал)	33-35	35-37	37-39	39-41
Кількість пар, n_i	3	9	12	6

Гістограму частот для отриманого розподілу наведено на рис.4.

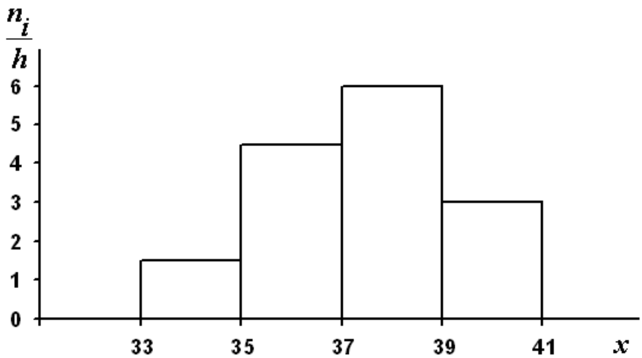


Рис.4.

Лекція 8

Вибіркові характеристики

Середнє вибіркове дискретного ряду. Середнє вибіркове визначають як відношення суми окремих значень вибірки до кількості цих значень. Розрізняють середнє вибіркове просте і зважене. Середнє вибіркове просте визначають за формулою

$$\overline{x_в} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n},$$

де x_i – окремі значення (варіанти), n – об'єм вибірки.

Середнє вибіркове зважене застосовують тоді, коли значення досліджуваної ознаки $x_1, x_2, ..., x_k$ повторюються відповідно з частотами $n_1, n_2, ..., n_k$ (тобто задано дискретний статистичний розподіл вибірки).

Середнє вибіркове зважене визначають за формулою

$$\overline{x_в} = \frac{x_1 n_1 + x_2 n_2 + ... + x_k n_k}{n} = \frac{\sum_{i=1}^{k} x_i n_i}{n},$$

де x_i – варіанти, n_i, - частоти, n – об'єм вибірки.

Приклад 1. Студент здав 4 екзамени і отримав оцінки: 4,3,4,4. Обчислити середній бал.

Розв'язок. Використаємо формулу для середнього вибіркового простого

$$\overline{x_в} = \frac{4+3+4+4}{4} = 3{,}75.$$

Приклад 2. Результати здачі екзамену 10 студентами: 3,4,3,4,4,5,3,5,4,4. Обчислити середній бал.

Розв'язок. Запишемо статистичний розподіл вибірки:

Оцінка	3	4	5
Кількість студентів	3	5	2

Використаємо формулу для середнього вибіркового зваженого

$$\overline{x_в} = \frac{\sum_{i=1}^{k} x_i n_i}{n} = \frac{3 \cdot 3 + 4 \cdot 5 + 5 \cdot 2}{10} = 3{,}9 .$$

Середнє вибіркове інтервального ряду. Середнє вибіркове інтервального ряду визначають наступним способом. Спочатку записують дискретний статистичний розподіл, значення якого – це середини часткових інтервалів вихідного інтервального ряду, а частоти це – відповідно частоти для часткових інтервалів вихідного ряду. Після цього середнє вибіркове визначають за формулою для середнього вибіркового зваженого.

У випадку, коли перший та останній інтервали ряду є відкритими, тобто не мають чіткої нижньої та верхньої межі, їхні середні значення визначають наступним чином. Щоб знайти середнє значення для першого відкритого інтервалу, від його верхньої межі віднімають половину величини наступного інтервалу, а щоб знайти середнє значення останнього відкритого інтервалу, до його нижньої межі додають половину величини попереднього інтервалу.

Приклад 3. На виробництві 70 робітників. Розподіл кількості робітників залежно від стажу роботи наведено в таблиці

Стаж роботи	менше 5	5-10	10-15	15-20	20-25	Більше 25
Кількість робітників	8	12	20	16	9	5

Обчислити середній стаж роботи робітника.

Розв'язок. Перейдемо від інтервального до дискретного статистичного розподілу вибірки, записавши його у таблицю

Стаж роботи	2,5	7,5	12,5	17,5	22,5	27,5
Кількість робітників	8	12	20	16	9	5

Використаємо формулу для середнього вибіркового зваженого

$$\overline{x_в} = \frac{\sum_{i=1}^{k} x_i n_i}{n} = \frac{2{,}5\cdot 8 + 7{,}5\cdot 12 + 12{,}5\cdot 20 + 17{,}5\cdot 16 + 22{,}5\cdot 9 + 27{,}5\cdot 5}{70} = 14.$$

Середнє геометричне. Середнє геометричне використовується для визначення середніх темпів зростання (спадання), тобто коли загальний об'єм явищ становить не суму, а добуток значень досліджуваної ознаки.

Середнє геометричне обчислюють за формулою

$$\overline{x}_{геом} = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n}\ .$$

Мода. Модою називають значення, яке найчастіше повторюється у досліджуваній сукупності. У дискретному варіаційному ряду це варіанта, яка має найбільшу частоту.

Приклад 9. У місті 10 фірм реалізують деякий товар за наступними цінами: 100, 110, 90, 110, 110, 120, 90, 100, 110, 100 грн.

Знайти модальну ціну на даний товар.

Розв'язок. Запишемо статистичний розподіл вибірки

Ціна, грн	90	100	110	120
Кількість фірм	2	3	4	1

З розподілу видно, що модальною буде ціна 110 грн, так як вона зустрічається найчастіше

Медіана. Медіаною називають таке значення досліджуваної ознаки, яке ділить ранжируваний (записаний у порядку зростання або спадання) ряд значень вибірки на дві рівні за кількістю частини. При непарній кількості значень у вибірці за медіану приймають центральне значення ранжируваного ряду $M_e=x_{(n+1)/2}$, а при парній кількості – середню арифметичну двох центральних значень $M_e=(x_{n/2}+x_{(n+2)/2})/2$.

Приклад 11. У місті 9 фірм реалізують деякий товар за наступними цінами: 10, 11, 9, 11, 11, 12, 9, 10, 11 грн.

Знайти медіанне значення ціни на даний товар.

Розв'язок. Запишемо значення цін у вигляді варіаційного ряду 9, 9, 10, 10, <u>11</u>, 11, 11, 11,12. Медіаною буде значення 11 (виділене в ряду), так як воно займає центральне місце у варіаційному ряду.

Приклад 12. У місті 10 фірм реалізують деякий товар за наступними цінами: 100, 110, 90, 110, 110, 120, 90, 100, 110, 100 грн.

Знайти медіану даного цінового ряду.

Розв'язок. Запишемо значення цін у вигляді варіаційного ряду 90, 90, 100, 100, <u>100</u>, <u>110</u>, 110, 110, 110,120. Два центральних значення – це 100 і 110 (виділені в ряду). Медіаною буде середня арифметична цих значень, тобто число 105.

Варіація. Варіація – це зміна величини ознаки у досліджуваній сукупності.

Розмах варіації. Розмах варіації – це різниця між найбільшим та найменшим значенням варіюючої ознаки

$$R=x_{max}-x_{min}.$$

Середнє лінійне відхилення. Середнє лінійне відхилення – це середнє арифметичне з абсолютних відхилень усіх варіант від середнього значення варіюючої ознаки. Його визначають за формулою

$$\overline{d}=\frac{|x_1-\overline{x_в}|n_1+|x_2-\overline{x_в}|n_2+...+|x_k-\overline{x_в}|n_k}{n}=\frac{\sum_{i=1}^{k}|x_i-\overline{x_в}|n_i}{n},$$

де $\overline{x_в}$ - вибіркова середня, x_i – варіанти, n_i – частоти, n – об'єм вибірки.

Лінійний коефіцієнт варіації. Лінійний коефіцієнт варіації – це відношення середнього лінійного відхилення до середнього значення ознаки

$$\lambda=\frac{\overline{d}}{\overline{x_в}}100\%.$$

За допомогою лінійного коефіцієнта варіації можна порівнювати варіацію різних сукупностей, так як на відміну від середнього лінійного відхилення його значення не залежить від одиниць вимірювання.

Вибіркова дисперсія. Дисперсія – це середнє арифметичне квадратів відхилень усіх значень ознаки від її середньої величини. Її обчислюють за такою формулою

$$D_в = \frac{\left(x_1 - \bar{x}_в\right)^2 n_1 + \left(x_2 - \bar{x}_в\right)^2 n_2 + \ldots + \left(x_k - \bar{x}_в\right)^2 n_k}{n} = \frac{\sum_{i=1}^{k}\left(x_i - \bar{x}_в\right)^2 n_i}{n}$$

,

де $\bar{x}_в$ – вибіркова середня, x_i – варіанти, n_i – частоти, n – об'єм сукупності.

Для спрощення обчислень використовують формулу

$$D_в = \frac{\sum_{i=1}^{m} x_i^2\ n_{x_i}}{n} - (\bar{x}_в)^2 = \overline{x^2} - \left(\bar{x}_в\right)^2,$$

де $\overline{x^2}$ – середнє арифметичне квадратів значень досліджуваної ознаки.

"Виправлена" вибіркова дисперсія Вибіркова дисперсія є зміщеною оцінкою дисперсії генеральної сукупності (дає занижені значення для генеральної дисперсії). Тому вибіркову дисперсію "виправляють" таким чином, щоб вона стала незміщеною оцінкою. Для цього вибіркову дисперсію множать на коефіцієнт Бесселя $\frac{n}{n-1}$ і позначають

$$s^2 = \frac{n}{n-1} D_в = \frac{1}{n-1} \sum_{i=1}^{k}\left(x_i - \bar{x}_в\right)^2 n_i.$$

Середнє квадратичне відхилення. Вибіркове середнє квадратичне відхилення – це корінь квадратний з дисперсії

$$\sigma_в = \sqrt{D_в}\ .$$

"Виправлене" вибіркове середнє квадратичне відхилення – це корінь квадратний з "виправленої" дисперсії

$$s = \sqrt{s^2} \ .$$

Коефіцієнт варіації. Коефіцієнт варіації (квадратичний коефіцієнт варіації) – це відношення середнього квадратичного відхилення до середнього значення ознаки

$$\nu = \frac{\sigma_в}{\overline{x}_в} 100\% \, .$$

Якщо коефіцієнт варіації менший або дорівнює 33%, то варіація вважається слабкою і можна говорити про однорідність даних, якщо більший за 33% , то варіація вважається сильною. У випадку сильної варіації досліджувана сукупність вважається неоднорідною (наприклад, розбита на групи), а середнє значення - нетиповим і його не можна використовувати як узагальнюючий показник цієї сукупності.

Приклад 2. Задано статистичний розподіл вибірки

x_i	0	1	2	4
n_i	2	3	3	2

Обчислити дисперсію, виправлену дисперсію, середнє квадратичне відхилення, виправлене середнє квадратичне відхилення, коефіцієнт варіації.

Розв'язок. Знаходимо дисперсію за спрощеною формулою

$$D_в = \overline{x^2} - \left(\overline{x}_в\right)^2 .$$

Спочатку обчислюємо $\overline{x}_в$

$$\overline{x}_в = \frac{0 \cdot 2 + 1 \cdot 3 + 2 \cdot 3 + 4 \cdot 2}{10} = 1{,}7 \, .$$

Для полегшення обчислення $\overline{x^2}$ додамо рядок із квадратами значень варіант

x_i	0	1	2	4
n_i	2	3	3	2
x_i^2	0	1	4	16

Величину $\overline{x^2}$ знайдемо за формулою для середнього арифметичного зваженого

$$\overline{x^2} = \frac{0\cdot 2+1\cdot 3+4\cdot 3+16\cdot 2}{10} = 4{,}7\ .$$

Тоді шукане значення дисперсії дорівнюватиме

$$D_в = 4{,}7\text{-}1{,}7^2 = 1{,}81.$$

"Виправлена" дисперсія

$$s^2 = \frac{n}{n-1} D_в = \frac{10}{9}\cdot 1{,}81 = 2{,}01.$$

Середнє квадратичне відхилення дорівнює

$$\sigma_в = \sqrt{D_в} = \sqrt{1{,}81} = 1{,}35\,.$$

"Виправлене" середнє квадратичне відхилення

$$s = \sqrt{s^2} = \sqrt{2{,}01} = 1{,}42\,.$$

Коефіцієнт варіації

$$\nu = \frac{\sigma_в}{\overline{x}_в}100\% = \frac{1{,}35}{1{,}7}\cdot 100\% = 79{,}4\%.$$

Асиметрія. Асиметрія (скошеність) емпіричного розподілу – це відношення центрального емпіричного моменту третього порядку до куба середнього квадратичного відхилення

$$A_s = \frac{\mu_3}{\sigma_в^{\ 3}} = \frac{\sum_{i=1}^{k}\left(x_i - \overline{x}_в\right)^3 n_i}{n\sigma_в^{\ 3}}.$$

Якщо асиметрія дорівнює нулю, то розподіл симетричний (для нормального розподілу A_s=0), якщо не дорівнює нулю – асиметричний. У випадках, коли A_s>0, розподіл має правосторонню асиметрію (права вітка кривої розподілу довша за ліву), коли A_s<0 – лівосторонню асиметрію (ліва вітка кривої розподілу довша за праву).

Прийнято вважати, що при значенні асиметрії $|A_s|<0{,}25$ асиметрія є незначною, при значенні $|A_s|>0{,}5$ – емпіричний розподіл відрізняється від нормального значним зміщенням.

При симетричному розподілі $\overline{x}_в = M_o = M_e$. Для правосторонньої асиметрії виконується таке співвідношення $\overline{x}_в > M_e > M_o$, для лівосторонньої таке $\overline{x}_в < M_e < M_o$.

Ексцес. Ексцес емпіричного розподілу – це відношення центрального емпіричного моменту четвертого порядку до середнього квадратичного відхилення у четвертому степені мінус три

$$E_s = \frac{\mu_4}{\sigma_в^{\ 4}} - 3 = \frac{\sum_{i=1}^{k}\left(x_i - \overline{x}_в\right)^4 n_i}{n\sigma_в^{\ 4}} - 3.$$

При нормальному розподілі E_s=0, при гостровершинному (вершина емпіричного розподілу виступає над вершиною нормального) E_s>0, при плосковершинному (вершина емпіричного розподілу знаходиться нижче вершини нормального розподілу) E_s<0.

У тих випадках, коли $|E_s|<0,4$, крива емпіричного розподілу вважається слабоексцесивною.

Якщо у вибірці A_s та E_s близькі до нуля, то можна припустити, що генеральна сукупність розподілена за нормальним законом.

Інтервальні оцінки. Інтервальною називають оцінку, яка визначається двома числами – кінцями інтервалу, який покриває досліджуваний параметр θ.

Точність оцінки. Точність оцінки – це таке додатне число δ, яке характеризує величину відхилення досліджуваного параметра θ від параметра θ^*, яким оцінюють θ, тобто виконується нерівність

$$|\theta - \theta^*| < \delta.$$

Чим менше δ, тим вища точність оцінки. Із збільшенням об'єму вибірки точність оцінки підвищується.

Надійність оцінки. Надійність оцінки – це таке число γ, яке дорівнює ймовірності того, що виконується нерівність $|\theta - \theta^*| < \delta$. Записують це таким чином

$$P(|\theta - \theta^*| < \delta) = \gamma \text{ або } P(\theta^* - \delta < \theta < \theta^* + \delta) = \gamma.$$

Як правило, надійність оцінки вибирають рівною 0,95, 0,99 або 0,999.

Число $\alpha = 1 - \gamma$ - це ймовірність похибки при оцінюванні.

Довірчий інтервал. Довірчим називають інтервал (θ^*-δ;θ^*+δ), який із заданою надійністю γ покриває досліджуваний параметр θ. Довірчий інтервал часто записують у такому вигляді $\theta^*-\delta<\theta<\theta^*+\delta$.

Довірчий інтервал для оцінки математичного сподівання** а **нормально розподіленої ознаки X при відомому середньому квадратичному відхиленні σ.

Для оцінки математичного сподівання a нормально розподіленої ознаки X при відомому середньому квадратичному відхиленні σ служить довірчий інтервал

$$\overline{x}_в - t\frac{\sigma}{\sqrt{n}} < a < \overline{x}_в + t\frac{\sigma}{\sqrt{n}},$$

де $t\frac{\sigma}{\sqrt{n}} = \delta$ – точність оцінки, n – об'єм вибірки, t – аргумент функції Лапласа, для якого $\Phi(t) = \frac{\gamma}{2}$, γ - надійність оцінки.

Параметр t знаходять із таблиці для значень функції Лапласа $\Phi(x)$ (додаток 2).

Таким чином можна записати

$$P\left(\overline{x}_в - t\frac{\sigma}{\sqrt{n}} < a < \overline{x}_в + t\frac{\sigma}{\sqrt{n}}\right) = \gamma.$$

Зміст цього виразу полягає у тому, що довірчий інтервал $\left(\overline{x}_в - t\frac{\sigma}{\sqrt{n}}, \overline{x}_в + t\frac{\sigma}{\sqrt{n}}\right)$ покриває оцінюваний параметр a з надійністю γ. Точність оцінки $\delta = t\frac{\sigma}{\sqrt{n}}$.

Лекція 9

Елементи теорії кореляції.

***Двовимірний статистичний розподіл вибірки*.** Двовимірним статистичним розподілом вибірки, елементам якої притаманні кількісні ознаки X і Y, називають перелік варіант x_i, y_i та відповідних цим парам варіант частот n_i.

У табличній формі розподіл має такий вигляд

x_i	x_1	x_2	...	x_k
y_i	y_1	y_2	...	y_k
n_i	n_1	n_2	...	n_k

де $n_1+n_2+...+n_k=n$ – об’єм вибірки.

***Кореляційна залежність*.** Кореляційною називають таку залежність між ознаками X та Y, коли при зміні однієї з ознак змінюється середнє значення іншої.

***Кореляційне поле*.** Кореляційне поле ознак X та Y – це графічне представлення результатів досліджень на координатній площині xOy у вигляді точок з координатами $(x_1 y_1)$, $(x_2 y_2)$,..., $(x_n y_n)$. На основі аналізу кореляційного поля можна вирішити питання про наявність чи відсутність залежності між ознаками, прослідкувати характер залежності (лінійна, нелінійна, функціональна чи статистична) та її тенденцію (додатну чи від’ємну).

Приклад 1. Задано двовимірну вибірку

x_i	1	2	3	4	5	6	7	8	9	10
y_i	2	3,7	6,2	7,9	9,9	12	14,1	16,3	17,8	19,9

Побудувати кореляційне поле.

Розв'язок. Відкладемо на площині xOy точки з координатами (1;2), (2;3,7), (3;6,2) та ін. Отримаємо кореляційне поле для значень ознак X та Y, на якому чітко видно лінійну залежність Y від X (рис.5).

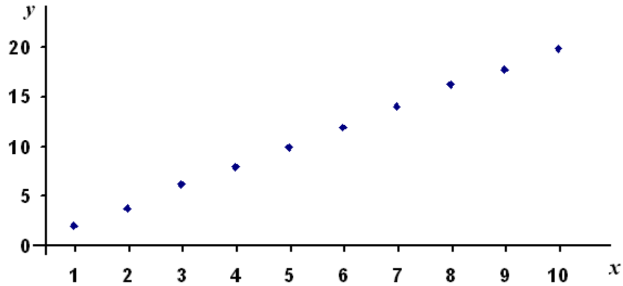


Рис.5.

Вибірковий коефіцієнт кореляції. Вибірковий коефіцієнт кореляції (коефіцієнт кореляції Пірсона) використовують для кількісної оцінки прямолінійного зв'язку між ознаками X та Y, які мають двовимірний нормальний розподіл. Вибірковий коефіцієнт кореляції обчислюють за формулою

$$r_в = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}_в\right)\left(y_i - \overline{y}_в\right)}{n\sigma_x\sigma_y},$$

де σ_x та σ_y – це вибіркові середні квадратичні відхилення ознак X та Y.

На практиці для обчислення коефіцієнта кореляції незгрупованих даних використовують формулу

$$r_в = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}_в\overline{y}_в}{n\sigma_x\sigma_y},$$

для згрупованих даних

$$r_в = \frac{\sum_{i=1}^{k} x_i y_i n_i - n\bar{x}_в \bar{y}_в}{n\sigma_x \sigma_y}.$$

Вибірковий коефіцієнт кореляції знаходиться в межах від 0 до +1 при прямій залежності та від 0 до –1 при зворотній залежності. Чим ближчий коефіцієнт кореляції до ±1, тим тісніший зв'язок між ознаками *X* та *Y* і навпаки, чим ближчий коефіцієнт кореляції до 0, тим слабший зв'язок між ознаками.

Якщо вибірковий коефіцієнт кореляції дорівнює ±1, то між досліджуваними ознаками існує функціональний лінійний зв'язок.

Для оцінювання сили зв'язку між корелюючими ознаками використовують шкалу Чеддока: якщо $|r_в| = 0{,}1 - 0{,}3$, то лінійний зв'язок дуже слабкий, якщо $|r_в| = 0{,}3 - 0{,}5$ – зв'язок слабкий, якщо $|r_в| = 0{,}5 - 0{,}7$ – зв'язок середній, якщо $|r_в| = 0{,}7 - 0{,}9$ – зв'язок сильний, якщо $|r_в| > 0{,}9$ – зв'язок дуже сильний.

Приклад 2. Задано двовимірну вибірку об'ємом n=11

x_i	-4	-3	-2	-1	0	1	2	3	4	5	6
y_i	25	16	9	4	1	0	1	4	9	16	25

Обчислити вибірковий коефіцієнт кореляції. Перевірити, чи існує залежність між ознаками *X* та *Y*.

Розв'язок. Знайдемо значення $\bar{x}_в$ та $\bar{y}_в$

$$\bar{x}_в = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{-4+(-3)+(-2)+(-1)+0+1+2+3+4+5+6}{11} = 1.$$

$$\bar{y}_в = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{25+16+9+4+1+0+1+4+9+16+25}{11} = 10.$$

Обчислимо коваріацію за формулою для незгрупованих даних

$$\operatorname{cov}(X,Y) = \frac{\sum_{i=1}^{n} x_i y_i}{n} - \bar{x}_в \bar{y}_в = \frac{-4\cdot 25+(-3)\cdot 16+(-2)\cdot 9+(-1)\cdot 4+0\cdot 1}{11} +$$

$$+\frac{1\cdot 9+2\cdot 1+3\cdot 4+4\cdot 9+5\cdot 16+6\cdot 25}{11} - 1\cdot 10 = 10-10 = 0.$$

Вибірковий коефіцієнт кореляції

$$r_в = \frac{\operatorname{cov}(X,Y)}{\sigma_x \sigma_y} = \frac{0}{\sigma_x \sigma_y} = 0.$$

Коваріація і вибірковий коефіцієнт кореляції дорівнюють нулю, отже ознаки X та Y є лінійно незалежними.

Використаємо кореляційне поле для перевірки наявності іншого типу зв’язку між цими ознаками. Кореляційне поле (рис.6) показує, що між X та Y існує параболічна залежність.

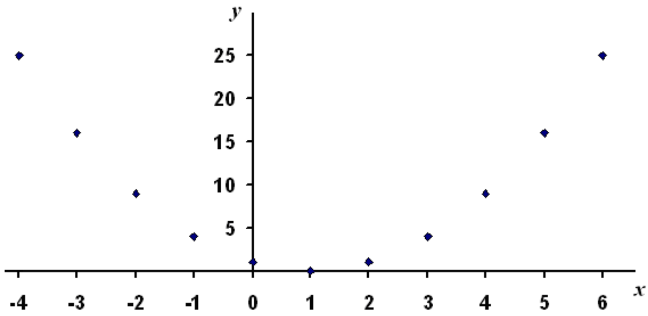


Рис.6.

Аналізуючи додатково дані з таблиці неважко встановити, що це функціональний зв'язок $Y = (X-1)^2$.

Лекція 10

Перевірка статистичних гіпотез

Основні поняття

Статистична гіпотеза. Статистичною гіпотезою H називається припущення відносно виду або параметрів розподілу досліджуваної ознаки X, яке може бути перевірене за результатами вибірки.

Статистичні гіпотези поділяються на параметричні і непараметричні. Параметричні гіпотези передбачають, що вигляд закону розподілу відомий і перевірка зводиться до перевірки значень невідомих параметрів. Непараметричні гіпотези передбачають встановлення вигляду закону розподілу

Прості і складні статистичні гіпотези. Проста статистична гіпотеза – це гіпотеза, що містить лише одне припущення. Складною статистичною гіпотезою називають гіпотезу, що складається з скінченного або нескінченного числа простих гіпотез.

Нульова гіпотеза. Нульова гіпотеза – це припущення, яке вибирається за основне і підлягає перевірці. Нульова гіпотеза позначається через H_0.

Конкуруюча гіпотеза. Конкуруюча (альтернативна) гіпотеза – це припущення, що суперечать нульовій гіпотезі. Конкуруюча гіпотеза позначається через H_1.

Статистичний критерій. Статистичним критерієм називають випадкову величину $K(X)$, яка служить для перевірки нульової гіпотези. Емпіричним (спостереженим) значенням $K_{\text{спост}}$ називають значення критерію, що обчислене за даними вибірки.

Критична область. Критична область – це множина значень критерію, при яких нульову гіпотезу H_0 відкидають. Якщо спостережне значення критерію не належить критичній області, то H_0 приймається.

Критичні точки. Критичними точками (границями) $K_{\text{кр}}$ називають точки, які відділяють критичну область від області прийняття нульової гіпотези.

Рівень значущості. Рівень значущості – це ймовірність того, що відкидається вірна гіпотеза H_0, а приймається невірна конкуруюча гіпотеза H_1. Рівень значущості позначають через α. Найбільш вживані такі його значення: 0,05; 0,01; 0,001.

Перевірка статистичних гіпотез про дисперсії та середні

Перевірка гіпотези про рівність двох дисперсій. Критерій Фішера. Порівняння генеральних дисперсій D_X та D_Y двох нормально розподілених ознак X та Y здійснюється зіставленням виправлених вибіркових дисперсій s_x^2 та s_y^2.

За нульову приймають гіпотезу H_0: $D_X = D_Y$.

За статистичний критерій береться випадкова величина F, яка має розподіл Фішера–Снедекора із k_1 і k_2 ступенями вільності

$$F = \frac{s_б^2}{s_м^2},$$

де $s_б^2$ є більшою з виправлених дисперсій, $s_м^2$ є меншою з виправлених дисперсій. Числа ступенів вільності знаходять із співвідношень: $k_1 = n_б - 1$ і $k_2 = n_м - 1$, де $n_б$ – об’єм вибірки із більшою виправленою дисперсією, n_i – об’єм вибірки із меншою виправленою дисперсією.

При конкуруючій гіпотезі H_1: $D_X > D_Y$ критичне значення критерію $F_{кр}(\alpha;\ k_1;\ k_2)$ знаходять за таблицею критичних точок розподілу Фішера–Снедекора (додаток 7) відповідно до заданого рівня значущості α і чисел ступенів вільності k_1 та k_2.

При конкуруючій гіпотезі H_1: $D_X \neq D_Y$ критичну точку $F_{кр}(\alpha/2;\ k_1;\ k_2)$ шукають при рівні значущості $\alpha/2$ (удвічі меншому за заданий).

Якщо $F_{спост} < F_{кр}$ – нема підстав відкидати нульову гіпотезу про рівність генеральних дисперсій. Різниця між емпіричними значеннями дисперсій несуттєва і є результатом дії випадкових причин.

Якщо $F_{спост} > F_{кр}$ – нульову гіпотезу відкидають і приймають альтернативну гіпотезу H_1.

Приклад 1. Дисперсія такого показника, як стресостійкість для бухгалтерів дорівнює $D_X = 6{,}17$**,** а для менеджерів – $D_Y = 4{,}41$**.** Об'єми вибірок однакові і дорівнюють $n_x = n_y = 12$.

Визначити, чи можна вважати дисперсії показника стресостійкості для бухгалтерів і менеджерів приблизно однаковими при рівні значущості 0,05?

Розв'язок. Перевіримо за допомогою критерію Фішера нульову гіпотезу H_0: $D_X = D_Y$ при конкуруючій гіпотезі H_1: $D_X > D_Y$.

Знайдемо спостережене значення критерію

$$F_{спост} = \frac{s_б^2}{s_м^2} = \frac{6{,}17}{4{,}41} = 1{,}4\,.$$

Знайдемо за таблицею критичних точок розподілу Фішера–Снедекора (додаток 7) відповідно до заданого рівня значущості $\alpha = 0{,}05$ і чисел ступенів вільності $k_1 = k_2 = 11$ критичне значення критерію $F_{кр}(\alpha;\ k_1;\ k_2) = F_{кр}(0{,}05;11;11) = 2{,}82$.

Так як $F_{спост} < F_{кр}$ – нема підстав відкидати нульову гіпотезу про рівність генеральних дисперсій. Різниця між емпіричними значеннями дисперсій показника стресостійкості несуттєва.

Перевірка гіпотези про рівність генеральної дисперсії $D_г$ заданому числу D_0. Для того, щоб при заданому рівні значущості α перевірити нульову гіпотезу $H_0 : D_г = D_0$, якщо конкуруюча гіпотеза має вигляд $H_1 : D_г > D_0$, необхідно обчислити спостережене значення критерію

$$\chi^2_{спост} = \frac{(n-1)s^2}{D_0},$$

де n – об'єм вибірки, s^2 – "виправлена" вибіркова дисперсія.

За таблицею критичних точок розподілу χ^2 (див. додаток 5) за заданим рівнем значущості α і числу ступенів вільності $k = n - 1$, де n – об'єм вибірки, знайти критичну точку $\chi^2_{кр}(\alpha; k)$. Порівняти знайдені значення критерію $\chi^2_{спост}$ та $\chi^2_{кр}(\alpha; k)$.

Якщо $\chi^2_{спост} < \chi^2_{кр}$ – нема підстав відкидати нульову гіпотезу. Якщо $\chi^2_{спост} > \chi^2_{кр}$ – нульову гіпотезу відкидають.

При конкуруючій гіпотезі $H_1 : D_г < D_0$ знаходять критичну точку $\chi^2_{кр}(1-\alpha; k)$. Якщо $\chi^2_{спост} > \chi^2_{кр}$ – нема підстав відкидати нульову гіпотезу.

При конкуруючій гіпотезі $H_1 : D_{ă} \neq D_0$ знаходять лівосторонню критичну точку $\chi^2_{лів.кр}(1-\alpha; k)$ і правосторонню критичну точку $\chi^2_{правкр}(\alpha; k)$. Якщо $\chi^2_{лів.кр}(1-\alpha; k) < \chi^2_{спост} < \chi^2_{правкр}$ – нема підстав відкидати нульову гіпотезу. В інших випадках нульову гіпотезу відкидають.

Приклад 2. Точність роботи ваг перевіряється за величиною дисперсії похибок при зважуванні еталонного зразка. Провели 20 контрольних зважувань еталонного зразка. Значення похибок зважування X (мг) наведені у таблиці

x_i	-3	-2	-1	0	1	2	3
n_i	1	2	5	7	3	1	1

Згідно з технічними характеристиками дисперсія не повинна перевищувати 1 мг^2.

Перевірити на рівні значущості 0,05 нульову гіпотезу $H_0 : D_г = 1$ при конкуруючій гіпотезі $H_0 : D_г > 1$.

Розв'язок. Знайдемо середнє вибіркове та "виправлену" вибіркову дисперсію:

$$\bar{x}_в = \frac{\sum_{i=1}^{k} x_i n_i}{n} = \frac{(-3\cdot 1) + (-2\cdot 2) + (-1\cdot 5) + 1\cdot 3 + 2\cdot 1 + 3\cdot 1}{20} = -0{,}2,$$

$$D_в = \frac{\sum_{i=1}^{k} x_i^2 \, n_i}{n} - (\bar{x}_в)^2 = \frac{9\cdot 1 + 4\cdot 2 + 1\cdot 5 + 1\cdot 3 + 4\cdot 1 + 9\cdot 1}{20} - (-0{,}2)^2 = 1{,}86.$$

$$s^2 = \frac{n}{n-1} D_в = \frac{20}{19}\cdot 1{,}86 = 1{,}96.$$

Знайдемо спостережене значення критерію

$$\chi^2_{спост} = \frac{(n-1)s^2}{D_0} = \frac{(20-1)\cdot 1{,}96}{1} = 37{,}24.$$

За таблицею критичних точок розподілу χ^2 (див. додаток 5) при заданому рівні значущості α=0,05 і числу ступенів вільності k=20–1=19 знаходимо критичну точку $\chi^2_{кр}(0{,}05;19) = 30{,}1.$

Так як $\chi^2_{спост} > \chi^2_{кр}$ – нульову гіпотезу відкидаємо і приймаємо конкуруючу гіпотезу $H_0 : D_г > 1$. Отже ваги потребують додаткового настроювання.

***Перевірка гіпотези про рівність генеральної середньої заданому числу a*₀.** Для перевірки при заданому рівні значущості α правильності нульової гіпотези $H_0 : \overline{x}_г = a_0$, де a_0 є певним числом, обчислюють спостережене значення критерію *T*, який має розподіл Стьюдента з $k = n - 1$ ступенями вільності

$$T_{спост} = \frac{(\overline{x}_в - a_0)\sqrt{n}}{s},$$

де $\overline{x}_в$ – середнє вибіркове, *s* – "виправлене" середнє квадратичне відхилення, *n* – об'єм вибірки.

При конкуруючій гіпотезі $H_1 : \overline{x}_г \neq a_0$ із таблиці критичних точок розподілу Стьюдента (додаток 6), за заданим рівнем значущості α і числу ступенів вільності $k = n_x + n_y - 2$ спочатку знаходять правосторонню критичну точку $t_{пр.кр}(\alpha/2; k)$, а потім лівосторонню критичну точку $t_{лів.кр}(\alpha/2; k) = -t_{пр.кр}(\alpha/2; k)$.

Якщо $t_{лів.кр}(\alpha/2; k) < T_{спост} < t_{пр.кр}(\alpha/2; k)$ – нульову гіпотезу приймають, в іншому випадку нульову гіпотезу відкидають.

При конкуруючій гіпотезі $H_1 : \overline{x}_г > a_0$ із таблиці критичних точок розподілу Стьюдента (додаток 6), за заданим рівнем значущості α і числу ступенів вільності $k = n - 1$ знаходять правосторонню критичну точку $t_{пр.кр}(\alpha; k)$. Якщо $T_{спост} < t_{пр.кр}$ – нульову гіпотезу приймають, якщо $T_{спост} > t_{пр.кр}$ – нульову гіпотезу відкидають.

При конкуруючій гіпотезі $H_1 : \bar{x}_г < a_0$ спочатку знаходять правосторонню критичну точку $t_{пр.кр}(\alpha; k),$ а потім – лівосторонню $t_{лів.кр}(\alpha; k) = -t_{пр.кр}(\alpha; k).$ Якщо $T_{спост} > t_{лів.кр}$, то нульову гіпотезу приймають, якщо $T_{спост} < t_{лів.кр}$ – нульову гіпотезу відкидають.

Приклад 3. Провели 20 контрольних зважувань еталонного зразка. Значення похибок зважування X (мг) наведені у таблиці

x_i	-3	-2	-1	0	1	2	3
n_i	1	2	5	7	3	1	1

Згідно з технічними характеристиками середнє значення похибки зважування повинно дорівнювати 0.

Перевірити на рівні значущості 0,05 нульову гіпотезу $H_0 : \bar{x}_г = 0,$ при конкуруючій гіпотезі $H_1 : \bar{x}_г \neq 0.$

Розв'язок. Знайдемо середнє вибіркове, вибіркову дисперсію "виправлену" вибіркову дисперсію та "виправлене" середнє квадратичне відхилення:

$$\bar{x}_в = \frac{\sum_{i=1}^{k} x_i n_i}{n} = \frac{(-3 \cdot 1) + (-2 \cdot 2) + (-1 \cdot 5) + 1 \cdot 3 + 2 \cdot 1 + 3 \cdot 1}{20} = -0{,}2,$$

$$D_в = \frac{\sum_{i=1}^{k} x_i^2 \, n_i}{n} - (\bar{x}_в)^2 = \frac{9 \cdot 1 + 4 \cdot 2 + 1 \cdot 5 + 1 \cdot 3 + 4 \cdot 1 + 9 \cdot 1}{20} - (-0{,}2)^2 = 1{,}86,$$

$$s^2 = \frac{n}{n-1} D_в = \frac{20}{19} \cdot 1{,}86 = 1{,}96,$$

$$s=\sqrt{s^2}=\sqrt{1{,}96}=1{,}4\,.$$

Знайдемо спостережене значення критерію

$$T_{спост}=\frac{\left(\overline{x}_в-a\right)\sqrt{n}}{s}=\frac{\left(-0{,}2-0\right)\sqrt{20}}{1{,}4}=-0{,}64\,.$$

Конкуруюча гіпотеза $H_1:\overline{x}_г\neq 0$. Із таблиці критичних точок розподілу Стьюдента (додаток 6), за заданим рівнем значущості $\alpha=0{,}05$ і числу ступенів вільності $k=n-1=19$ знайдемо критичні точки

$$t_{пр.кр}(\alpha/2;k)=t_{пр.кр}(0{,}025;19)=2{,}09$$

та

$$t_{лів.кр}(0{,}025;19)=-2{,}09.$$

Так як $t_{лів.кр}(\alpha/2;k)<T_{спост}<t_{пр.кр}(\alpha/2;k)$, то нульову гіпотезу приймаємо.

Можна вважати, що середнє значення похибки зважування дорівнює 0 і відповідає технічним характеристикам ваг.

Перевірка гіпотези про рівність середніх двох генеральних сукупностей. Для того, щоб перевірити нульову гіпотезу $H_0:\overline{x}_г=\overline{y}_г$ про рівність середніх значень $\overline{x}_г$ та $\overline{y}_г$ двох нормально розподілених генеральних сукупностей ознак *X* та *Y*, генеральні дисперсії яких невідомі, але вважаються однаковими, необхідно знайти спостережене значення критерію *T*, який має розподіл Стьюдента з $k=n_x+n_y-2$ ступенями вільності

$$T_{спост} = \frac{\overline{x}_в - \overline{y}_в}{(n_x - 1)s_x^2 + (n_y - 1)s_y^2} \sqrt{\frac{n_x n_y (n_x + n_y - 2)}{n_x + n_y}} ,$$

де n_x та n_y – об'єми вибірок, $\overline{x}_в$ та $\overline{y}_в$ – середні вибіркові, s_x^2 та s_y^2 – "виправлені" вибіркові дисперсії відповідно для ознак X та Y.

При конкуруючій гіпотезі $H_1 : \overline{x}_г \neq \overline{y}_г$ із таблиці критичних точок розподілу Стьюдента (додаток 6), за заданим рівнем значущості α і числу ступенів вільності $k = n_x+n_y - 2$ спочатку знаходять правосторонню критичну точку $t_{пр.кр}(\alpha/2;k)$, а потім лівосторонню критичну точку $t_{лів.кр}(\alpha/2;k) = -t_{пр.кр}(\alpha/2;k)$.

Якщо $t_{лів.кр}(\alpha/2;k) < T_{спост} < t_{пр.кр}(\alpha/2;k)$ – нульову гіпотезу приймають, в іншому випадку нульову гіпотезу відкидають.

При конкуруючій гіпотезі $H_1 : \overline{x}_г > \overline{y}_г$ із таблиці критичних точок розподілу Стьюдента (додаток 6), за заданим рівнем значущості α і числу ступенів вільності $k = n_x+n_y - 2$ знаходять правосторонню критичну точку $t_{пр.кр}(\alpha;k)$. Якщо $T_{спост} < t_{пр.кр}$ – нульову гіпотезу приймають, якщо $T_{спост} > t_{пр.кр}$ – нульову гіпотезу відкидають.

При конкуруючій гіпотезі $H_1 : \overline{x}_г < \overline{y}_г$ спочатку знаходять правосторонню критичну точку $t_{пр.кр}(\alpha;k)$, а потім – лівосторонню $t_{лів.кр}(\alpha;k) = -t_{пр.кр}(\alpha;k)$. Якщо $T_{спост} > t_{лів.кр}$, то нульову гіпотезу приймають, якщо $T_{спост} < t_{лів.кр}$ – нульову гіпотезу відкидають.

Приклад 4. Провели контроль роботи двох відділень банку з метою встановлення середнього часу обслуговування клієнтів. Результати наведено у таблицях перше відділення

час ослуговування, x_i	5	7	10	12	15	20
кількість клієнтів, n_i	2	4	9	4	2	4

друге відділення

час ослуговування, y_i	5	7	9	15	20	25
кількість клієнтів, n_i	3	5	10	3	3	1

При рівні значущості $\alpha = 0{,}1$ перевірити нульову гіпотезу $H_0 : \overline{x}_г = \overline{y}_г$ про рівність середнього часу обслуговування клієнтів в обох відділеннях банку. За альтернативну прийняти гіпотезу $H_1 : \overline{x}_г > \overline{y}_г$.

Розв'язок. Знайдемо середнє вибіркове, вибіркову дисперсію та "виправлену" вибіркову дисперсію для першого відділення

$$\overline{x}_в = \frac{\sum_{i=1}^{k} x_i n_i}{n} = \frac{5\cdot 2 + 7\cdot 4 + 10\cdot 9 + 12\cdot 4 + 15\cdot 2 + 20\cdot 4}{25} = 11{,}44,$$

$$D_{в_x} = \frac{\sum_{i=1}^{k} x_i^2\ n_i}{n} - (\overline{x}_в)^2 = \frac{25\cdot 2 + 49\cdot 4 + 100\cdot 9 + 144\cdot 4 + 225\cdot 2 + 400\cdot 4}{25} -$$

$$-(11{,}44)^2 = 20,$$

$$s_x^2 = \frac{n}{n-1} D_в = \frac{25}{24}\cdot 20 = 20{,}83.$$

Знайдемо середнє вибіркове, вибіркову дисперсію та "виправлену" вибіркову дисперсію для другого відділення

$$\overline{y}_в = \frac{\sum_{i=1}^{k} y_i n_i}{n} = \frac{5\cdot 3 + 7\cdot 5 + 9\cdot 10 + 15\cdot 3 + 20\cdot 3 + 25\cdot 1}{25} = 10{,}8,$$

$$D_{в_y} = \frac{\sum_{i=1}^{k} y_i^2\; n_i}{n} - (\overline{y}_в)^2 = \frac{25\cdot 3 + 49\cdot 5 + 81\cdot 10 + 225\cdot 3 + 400\cdot 3 + 625\cdot 1}{25} -$$

$$-(10{,}8)^2 = 28{,}56,$$

$$s_y^2 = \frac{n}{n-1} D_в = \frac{25}{24}\cdot 228{,}56 = 29{,}75.$$

Знайдені виправлені дисперсії різні, тому перевіримо попередньо гіпотезу про рівність генеральних дисперсій за допомогою критерію Фішера–Снедекора.

Знайдемо відношення більшої дисперсії до меншої

$$F_{спост} = \frac{s_б^2}{s_м^2} = \frac{29{,}75}{20{,}83} = 1{,}43.$$

Знайдемо за таблицею критичних точок розподілу Фішера–Снедекора (додаток 7) відповідно до заданого рівня значущості $\alpha = 0{,}1$ і чисел ступенів вільності $k_1 = k_2 = 24$ критичне значення критерію $F_{кр}(\alpha;\ k_1;\ k_2) = F_{кр}(0{,}1;24;24) = 1{,}7.$

Так як $F_{спост} < F_{кр}$ – нема підстав відкидати нульову гіпотезу про рівність генеральних дисперсій.

Припущення про рівність генеральних дисперсій виконується, тому можна порівнювати середні.

Знайдемо спостережене значення критерію

$$T_{спост} = \frac{\overline{x}_в - \overline{y}_в}{(n_x - 1)s_x^2 + (n_y - 1)s_y^2} \sqrt{\frac{n_x n_y (n_x + n_y - 2)}{n_x + n_y}} =$$

$$= \frac{11{,}44 - 10{,}8}{(25-1)\cdot 20{,}83 + (25-1)\cdot 29{,}75} \cdot \sqrt{\frac{25\cdot 25\cdot (25+25-2)}{25+25}} = 0{,}013.$$

Конкуруюча гіпотеза має вигляд $H_1 : \overline{x}_г > \overline{y}_г$, Із таблиці критичних точок розподілу Стьюдента (додаток 6), за заданим рівнем значущості α =0,1 і числу ступенів вільності $k = n_x + n_y - 2 = 48$ знаходимо правосторонню критичну точку $t_{пр.кр}(\alpha; k) = t_{пр.кр}(0{,}1;48) = 1{,}3$. Так як $T_{спост} < t_{пр.кр}$ – нульову гіпотезу приймаємо. Таким чином, можна вважати, що середній час обслуговування клієнтів в обох відділеннях однаковий.