

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ АКАДЕМІЯ НАУК
ВИЩОЇ ШКОЛИ УКРАЇНИ**

**М.А. Мартиненко, О.М. Нецадим, О.І.
Радзівська, В.М. Сафонов**

МАТЕМАТИЧНА СТАТИСТИКА

*Навчальний посібник для студентів вищих
навчальних закладів*

2-ге видання, доповнене

Рекомендовано Міністерством освіти і науки України

Київ
ЦП “КОМПРИНТ”
2011

УДК519.22/.25 ББК 22.18я73 М33

Гриф надано Міністерством освіти і науки України (Лист від 26.08.2005 р. №14/18.2-1187)

Рецензенти:

О.М. Станжицький, д-р фіз.-мат. наук, проф. (Національний університет ім. Т.Г. Шевченка);

Т.І. Олешко, проф. (Національний авіаційний університет)

Мартиненко М.А., Нецадим О.М., Радзівська О.І., Сафонов В.М.-

М 33 Математична статистика: Навчальний посібник. - 2-ге вид., переробл. та доповн. - К.: ЦП "КОМПРИНТ", 2011. - 216 с.

КВИ 978-966-97188-7-7

Розглядаються основи математичної статистики згідно програми дисципліни "Теорія ймовірностей і математична статистика". Теоретичний матеріал подається у формі лекцій і супроводжується прикладами типових задач із економічної практики.

Наприкінці кожної лекції наведено перелік основних питань для самоконтролю знань і достатню кількість завдань для проведення практичних занять. Багато задач подано із детальними розв'язками.

В посібнику вміщено завдання для індивідуальної роботи студентів та модульного контролю.

Розрахований на студентів економічних спеціальностей вищих навчальних закладів усіх форм навчання.

УДК 519.22/.25

ББК22.18я73

КІШ 978-966-97188-7-7

© Мартиненко М.А., Нецадим О.М.,
© Радзівська О.І., Сафонов В.М., 2011

ПЕРЕДМОВА до другого видання

В другому виданні перероблено і доповнено матеріал лекції 8, в якій викладаються основи кореляційного та регресійного аналізу. Суттєво перероблено пункт 8.6 (нелінійна регресія) і добавлено новий пункт, присвячений кореляційному відношенню і його властивостям. Деякі виправлення внесені в тексті інших лекцій посібника.

Автори

ПЕРЕДМОВА

В основу навчального посібника покладено багаторічний досвід авторів, якого вони набули при викладанні дисципліни “Теорія ймовірностей і математична статистика” в Національному університеті харчових технологій та інших навчальних закладах України. Цей посібник відповідає сучасним вимогам навчальної програми дисципліни для підготовки фахівців-економістів. За задумом авторів посібник має сприяти підвищенню рівня фундаментальної математичної підготовки студентів, посиленню прикладного економічного спрямування набутих знань.

Матеріал посібника подається у вигляді лекцій та практичних занять і охоплює основні теми розділу “Математична статистика”, а саме: вибірки та їх представлення, числові характеристики статистичного розподілу вибірки, статистичне оцінювання параметрів генеральної сукупності, перевірка достовірності статистичних гіпотез, елементи дисперсійного аналізу, основи кореляційного і регресійного аналізу. При написанні посібника автори намагались досягти максимальної доступності матеріалу для студентів, зберігаючи необхідний рівень математичної строгості. Успішному засвоєнню теоретичного матеріалу має сприяти велика кількість розглянутих прикладів. Значну кількість задач запропоновано в кінці кожної лекції для проведення практичних занять та самостійного розв’язування. Необхідні для розв’язання задач основні таблиці математичної статистики наведені в додатках.

В кінці посібника представлені індивідуальні контрольні завдання та розгорнутий предметний показник основних термінів.

Автори

Лекція 5. Перевірка статистичних гіпотез.

5.1. Перевірка статистичної гіпотези про рівність двох дисперсій.

Гіпотези про дисперсії виникають досить часто, оскільки дисперсія характеризує такі важливі показники, як точність машин, приладів, технологічних процесів, ризик інвестицій тощо.

Нехай є дві нормально розподілені сукупності з ознаками X і Y , дисперсії яких дорівнюють σ_x^2 і σ_y^2 . Необхідно перевірити нульову гіпотезу про рівність дисперсій $H_0 : \sigma_x^2 = \sigma_y^2$ за альтернативної гіпотези $H_1 : \sigma_x^2 \neq \sigma_y^2$ або $H_1 : \sigma_x^2 > \sigma_y^2$.

Для перевірки гіпотези H_0 з цих сукупностей зроблені дві незалежні вибірки обсягів n_1 і n_2 , для яких знайдені виправлені вибіркові дисперсії S_x^2 і S_y^2 . Більшу з вибіркових дисперсій позначимо через S_1^2 , а меншу – S_2^2 .

За статистичний критерій береться випадкова величина

$$F = \frac{S_1^2}{S_2^2}, \quad (5.1)$$

яка має розподіл Фішера-Снедекора з $k_1 = n_1 - 1$ і $k_2 = n_2 - 1$ ступенями волі, де n_1 і n_2 – обсяги вибірок, за якими обчислені відповідно S_1^2 і S_2^2 .

В залежності від вигляду гіпотези H_1 будується критична область. Критична точка $f_{кр}$ правобічної критичної області визначається із таблиці розподілу Фішера-Снедекора (додаток 7) за відомими значеннями рівня значущості α та k_1, k_2 :

$f_{кр} = F(\alpha, k_1, k_2)$. У випадку двобічної критичної області $f_{кр}$ знаходиться

аналогічно, тільки замість α беруть $\frac{\alpha}{2}$: $f_{кр} = F\left(\frac{\alpha}{2}, k_1, k_2\right)$.

Приклад 5.1. Одна і та ж продукція виробляється на двох технологічних лініях. На другій лінії зроблено вдосконалення, яке привело до скорочення варіації

часу виготовлення продукції. Вибіркова перевірка варіації часу виготовлення на першій і другій лініях дала значення $S_x^2 = 1,05$ при 13 спостереженнях і $S_y^2 = 0,6$ при 12 спостереженнях. Чи можна вважати суттєвою рубіжність між варіаціями тривалості процесу виготовлення продукції на першій і другій лініях.

Розв'язання. Сформулюємо нульову гіпотезу $H_0 : D(X) = D(Y)$, де X, Y – тривалість технологічного процесу відповідно на першій і другій лініях.

За альтернативну візьмемо гіпотезу $H_1 : D(X) > D(Y)$ – після вдосконалення на другій лінії зменшились коливання тривалості процесу виготовлення продукції.

За статистичний критерій візьмемо випадкову величину (5.1). Обчислюємо спостережене значення критерію:

$$f^* = \frac{1,05}{0,6} = 1,75.$$

Число ступенів волі для більшої виправленої дисперсії $S_1^2 = S_x^2$ становить $k_1 = 13 - 1 = 12$ для меншої $S_2^2 = S_y^2 - k_2 = 12 - 1 = 11$.

За вибраним рівнем значущості $\alpha = 0,01$, ступенями волі $k_1 = 12; k_2 = 11$ з таблиці Фішера-Снедекора визначаємо критичну точку $f_{кр} = F(0,01; 12; 11) = 4,4$ і будуємо правобічну критичну область (рис. 5.1)

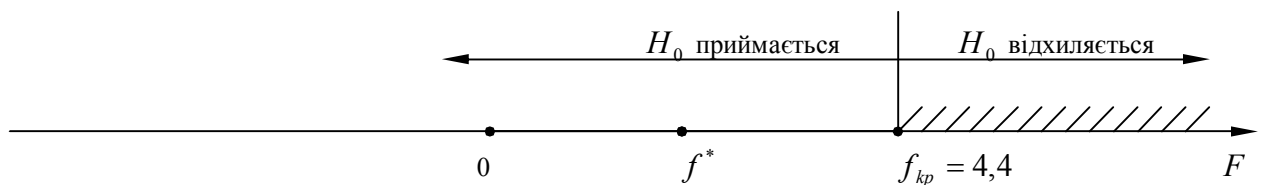


Рис. 5.1

Оскільки $f^* < f_{кр}$, то f^* не належить критичній області і гіпотеза H_0 приймається.

Отже, при 1%-му рівні значущості немає підстав для відхилення гіпотези H_0 : вдосконалення не суттєво вплинуло на розбіжність між дисперсіями.

5.2. Перевірка статистичної гіпотези про рівність генеральних середніх двох сукупностей.

Нехай задано дві генеральні сукупності, ознаки яких X і Y мають нормальний закон розподілу і при цьому незалежні одна від одної. З кожної сукупності зроблено незалежні вибірки: з першої – обсягу n_1 , а з другої – обсягу n_2 і обчислено \bar{x}_e та \bar{y}_e .

Потрібно за вибірковими середніми при обраному рівні значущості перевірити достовірність гіпотези $H_0 : M(X) = M(Y)$. За статистичний критерій візьмемо випадкову величину

$$U = \frac{\bar{X}_b - \bar{Y}_b}{\sigma(\bar{X}_b - \bar{Y}_b)} \quad (5.2)$$

При цьому можуть спостерігатися наступні випадки.

Випадок 1. Відомі значення дисперсій ознак генеральних сукупностей: $D(X) = D_x$ і $D(Y) = D_y$.

Оскільки для незалежних вибірок

$$D(\bar{X}_b - \bar{Y}_b) = D(\bar{X}_b) + D(\bar{Y}_b) = \frac{D(X)}{n_1} + \frac{D(Y)}{n_2}, \quad \text{то}$$

критерій (5.2) набуває вигляду:

$$Z = \frac{\bar{X}_b - \bar{Y}_b}{\sqrt{\frac{D_x}{n_1} + \frac{D_y}{n_2}}}, \quad (5.3)$$

причому $M(Z) = 0$, $\sigma(Z) = 1$.

Залежно від формулювання альтернативної гіпотези H_1 будуються відповідно правобічна, лівобічна та двобічна критичні області. Критичні точки знаходять за таблицею значень функції Лапласа при заданому рівні значущості α .

Зауваження. Статистичним критерієм (5.3) можна користуватись при перевірці гіпотези H_0 і в тому випадку, коли немає підстав вважати, що випадкові величини X та Y мають нормальний закон розподілу. На відміну від попереднього, в цьому разі обсяги вибірок повинні бути великими ($n_1, n_2 > 30$).

Випадок 2. Дисперсії невідомі, але рівні між собою: $D(X) = D(Y) = \sigma^2$. В цьому випадку

$$\sigma(\bar{X}_e - \bar{Y}_e) = \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (5.4)$$

Замінімо невідому дисперсію σ^2 генеральної сукупності її виправленою вибірковою оцінкою S^2 . Краща оцінка дисперсії досягається шляхом додавання двох вибірових дисперсій σ_{ex}^2 і σ_{ey}^2 :

$$S^2 \approx \frac{n_1 \sigma_{ex}^2 + n_2 \sigma_{ey}^2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)S_x^2 + (n_2 - 1)S_y^2}{n_1 + n_2 - 2}. \quad (5.5)$$

Враховуючи (5.2), (5.4), (5.5) приходимо до статистичного критерію

$$t = \frac{\bar{x}_b - \bar{y}_b}{\sqrt{n_1 \sigma_{ex}^2 + n_2 \sigma_{ey}^2}} \cdot \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}. \quad (5.6)$$

Випадкова величина (5.6) при малих обсягах вибірок має розподіл Стюдента з $k = n_1 + n_2 - 2$ ступенями волі.

Далі перевірка достовірності нульової гіпотези проводиться за відомою схемою з використанням таблиць розподілу Стюдента.

Зауваження. При великих обсягах вибірок ($n_1, n_2 > 30$) розподіл випадкової величини t асимптотично наближається до нормального. В цьому разі критичні точки визначаються за таблицею значень функції Лапласа.

Випадок 3. Дисперсії $D(X)$, $D(Y)$ невідомі і різні, обсяги вибірок великі.

Якщо незалежні вибірки мають великі обсяги, а значення генеральних дисперсій невідомі, то у формулі (5.3) генеральні дисперсії можна замінити

відповідними виправленими вибірковими дисперсіями S_x^2 , S_y^2 . Одержана внаслідок цього випадкова величина

$$Z = \frac{\bar{x}_e - \bar{y}_e}{\sqrt{\frac{S_x^2}{n_1} + \frac{S_y^2}{n_2}}} \quad (5.7)$$

має нормований нормальний закон розподілу і обирається за статистичний критерій.

Зауваження. При виборі належного статистичного критерію перевірки нульової гіпотези за умови, що генеральні дисперсії невідомі, попередньо потрібно встановити, чи рівні між собою генеральні дисперсії. Така проблема розглядалася в п.5.1.

Приклад 5.2. Виробляється продукція харчування. Для дослідження її якості зроблено дві вибірки по 10 одиниць із кожної послідовної серії і визначено процентний вміст контрольованої сировини в кожній вибірці. В першій вибірці середній відсоток становив 68,2% із стандартним відхиленням 0,7%. Для другої серії – 67% із стандартним відхиленням 0,74%. Чи є підстави вважати, що процентний вміст сировини у двох серіях різний?

Розв'язання. Позначимо процентний вміст контрольованої сировини в першій серії продукції через X , а в другій – через Y . Сформуємо нульову гіпотезу $H_0 : M(X) = M(Y)$ – процентний вміст сировини у двох серіях однаковий; альтернативну $H_1 : M(X) \neq M(Y)$ – процентний вміст сировини різний.

За умовою задачі відомо: $\bar{x}_b = 68,2\%$; $\sigma_{ex} = 0,7\%$; $\bar{y}_b = 67\%$; $\sigma_{ey} = 0,74\%$.

Оскільки генеральні дисперсії невідомі, то спочатку перевіримо гіпотезу про рівність двох генеральних дисперсій: $H_0 : D(X) = D(Y)$ при альтернативній гіпотезі $H_1 : D(X) \neq D(Y)$.

Вибираємо статистичний критерій (5.1):

$$F = \frac{S_1^2}{S_2^2}.$$

Знайдемо спостережуване значення критерію. Для цього обчислимо виправлені вибіркві дисперсії:

$$S_x^2 = \frac{n_1}{n_1 - 1} \sigma_{bx}^2 = \frac{10}{9} \cdot 0,7^2 \approx 0,544,$$

$$S_y^2 = \frac{n_2}{n_2 - 1} \sigma_{by}^2 = \frac{10}{9} \cdot (0,74)^2 \approx 0,608.$$

$$\text{Отже, маємо } f^* = \frac{S_1^2}{S_2^2} = \frac{S_y^2}{S_x^2} = \frac{0,608}{0,544} = 1,118.$$

За рівнем значущості $\alpha = 0,02$, числами ступенів волі $k_1 = n_1 - 1 = 9$; $k_2 = n_2 - 1 = 9$ та із таблиці Фішера-Снедекора визначаємо критичну точку двобічної критичної області

$$f_{kp} = F\left(\frac{\alpha}{2}; k_1; k_2\right) = F(0,01; 9; 9) = 5,35.$$

Одержуємо $f^* < f_{kp}$, тому нульова гіпотеза приймається: дві генеральні дисперсії можна вважати рівними.

Перевіримо, далі, гіпотезу про рівність генеральних середніх. При рівні значущості $\alpha = 0,05$ скористаємося критерієм (5.6) з числом $k = n_1 + n_2 - 2 = 10 + 10 - 2 = 18$ ступенів волі. За таблицю розподілу Стьюдента знаходимо критичну точку двобічної області: $t_{kp} = t(\alpha; k) = t(0,05; 18) = 2,1$.

Обчислюємо спостережуване значення статистичного критерію:

$$t^* = \frac{\bar{x}_b - \bar{y}_b}{\sqrt{n_1 \sigma_{ax}^2 + n_2 \sigma_{ay}^2}} \cdot \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} = \frac{68,2 - 67}{\sqrt{10 \cdot 0,7^2 + 10 \cdot 0,74^2}} \sqrt{\frac{10^2 \cdot 18}{10 + 10}} = 3,53.$$

Оскільки $t^* > t_{kp}$, то на 5%-вому рівні значущості відхиляємо нульову гіпотезу і приймаємо альтернативну (рис. 5.2): дві серії мають різний вміст даної сировини.

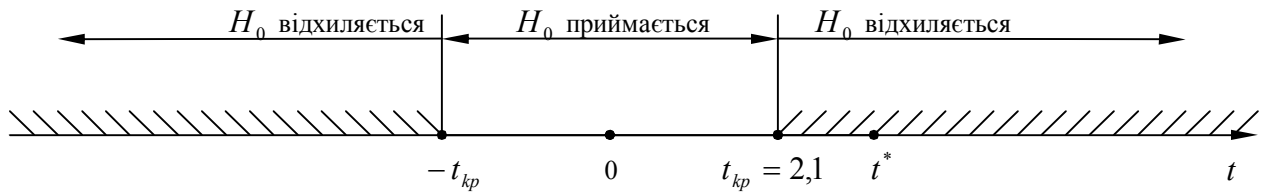


Рис. 5.2

5.3. Перевірка статистичної гіпотези про рівність часток ознаки у двох сукупностях.

Задача про порівняння часток ознаки у двох сукупностях досить часто зустрічається на практиці. Наприклад, вибіркова частка ознаки однієї сукупності відмінна від частки цієї ознаки в іншій сукупності. Чи свідчить це про те, що наявність ознаки в одній сукупності дійсно є більш ймовірною, чи отримана розбіжність часток є випадковою?

Нехай є дві генеральні сукупності, ймовірності появи ознаки (частки ознаки) X в яких дорівнюють відповідно p_1 і p_2 . Потрібно перевірити нульову гіпотезу про рівність часток ознаки X для цих сукупностей, тобто $H_0 : p_1 = p_2$. Для перевірки гіпотези з цих сукупностей зроблено дві незалежні вибірки обсягів n_1 , n_2 . Вибіркові частки появи ознаки дорівнюють відповідно $w_1 = \frac{m_1}{n_1}$ і $w_2 = \frac{m_2}{n_2}$, де m_1 і m_2 – число варіант з даною ознакою в цих вибірках.

При великих обсягах вибірок розподіл відносних частот наближається до нормального закону розподілу з параметрами:

$$M(W_1) = p_1, M(W_2) = p_2; D(W_1) = \frac{p_1 q_1}{n_1}, D(W_2) = \frac{p_2 q_2}{n_2},$$

де $q_1 = 1 - p_1$ і $q_2 = 1 - p_2$.

За статистичний критерій перевірки гіпотези H_0 візьмемо випадкову величину

$$Z = \frac{w_1 - w_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}. \quad (5.8)$$

Якщо справджується гіпотеза $H_0 : p_1 = p_2 = p$, тоді випадкова величина Z має нормальний розподіл з параметрами $M(Z) = 0$ і $D(Z) = 1$. Дійсно, оскільки $M(W_1 - W_2) = p - p = 0$, то дістанемо:

$$M(Z) = M\left(\frac{W_1 - W_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}\right) = \frac{M(W_1) - M(W_2)}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = 0;$$

$$D(Z) = D\left(\frac{W_1 - W_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}\right) = \frac{D(W_1) + D(W_2)}{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \frac{\frac{pq}{n_1} + \frac{pq}{n_2}}{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 1,$$

За невідоме значення ймовірності p у формулі (5.8) беруть її найкращу оцінку, що дорівнює вибірковій частці ознаки, якщо дві вибірки об'єднати в одну, тобто

$$p = \frac{m_1 + m_2}{n_1 + n_2}.$$

Знаходження критичних тичок і вибір критичної області проводиться згідно методики, розглянутої раніше.

Критерії перевірки статистичних гіпотез про порівняння параметрів двох сукупностей наведено в табл. 5.1.

Приклад 5.3. Податкова інспекція перевіряє фінансову документацію підприємства. Детально вивчена випадкова вибірка $n_1 = 50$ закінчених рахунків. В чотирьох із них виявили неправильне нарахування розміру податку. Інспектори запропонували вдосконалити процедуру оподаткування. Після закінчення відведеного терміну знову перевірили випадкову вибірку $n_2 = 60$ завершених рахунків; в трьох із них виявили порушення при нарахуванні податку. Чи є підстави вважати, що нова процедура привела до зменшення порушень?

Розв'язання. Сформулюємо нульову гіпотезу $H_0 : p_1 = p_2 = p$ – обидві вибірки взяті із генеральних сукупностей з рівними частками порушень. За альтернативу візьмемо гіпотезу $H_1 : p_1 > p_2$ – вдосконалення процедури нарахування податку зменшили частку помилок.

За рівнем значущості, наприклад, $\alpha = 0,05$ із таблиць значень функції Лапласа знаходимо критичну точку для правобічної критичної області:

$$\Phi(z_{kp}) = \frac{1 - 2\alpha}{2} = 0,45; \text{ звідки } z_{kp} = 1,645.$$

$$\text{Оскільки за умовою } m_1 = 4, m_2 = 3, \text{ то } p = \frac{m_1 + m_2}{n_1 + n_2} = \frac{7}{110} \approx 0,0636, \quad q = 0,9364.$$

Обчислюємо спостережуване значення критерію (5.8):

$$z^* = \frac{\frac{4}{50} - \frac{3}{60}}{\sqrt{0,0636 \cdot 0,9364 \left(\frac{1}{50} + \frac{1}{60} \right)}} \approx 0,64.$$

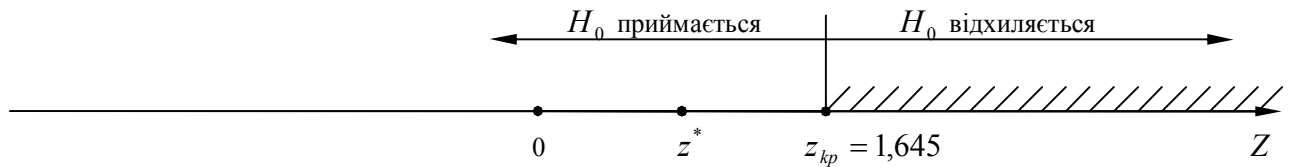


Рис. 5.3

Видно, що $z^* < z_{kp}$, тобто z^* належить області прийняття нульової гіпотези.

Отже, на 5%-вому рівні значущості немає підстав вважати, що вдосконалення в системі нарахування податку привели до зменшення кількості помилок.

5.4. Перевірка непараметричних статистичних гіпотез. Критерій погодженості.

Важливою задачею математичної статистики є встановлення теоретичного закону розподілу ознаки генеральної сукупності. Зокрема, багато із розглянутих вище критеріїв перевірки параметричних статистичних гіпотез ґрунтувалися на припущенні, що ознака генеральної сукупності має нормальний закон розподілу. Тому використовувати ці критерії можливо у разі достатньої упевненості, що спостережувана ознака генеральної сукупності має нормальний закон розподілу або близький до нього.

Якщо закон розподілу ознаки невідомий, то на основі дослідження вибіркової сукупності одержують емпіричний розподіл, за яким робиться припущення про *вигляд і параметри* невідомого розподілу ознаки генеральної сукупності. Припущення про закон розподілу ознаки може бути зроблено за наявними теоретичними передумовами про характер зміни ознаки. Зокрема, виконання умов центральної граничної теореми свідчить про нормальний закон розподілу. В деяких випадках підставою для висування гіпотези про теоретичний закон розподілу можуть бути деякі формальні властивості статистичного розподілу, а

саме: рівність нулю коефіцієнтів A та E свідчить про нормальний розподіл; рівність \bar{x}_b і σ_b є ознакою показникового розподілу. Інколи припущення про вигляд теоретичного розподілу може ґрунтуватись на графічному зображенні статистичного розподілу вибірки.

Якщо вигляд теоретичного розподілу ознаки встановлений, то за статистичним розподілом вибірки знаходять найкращі оцінки невідомих параметрів цього розподілу.

Після цього виникає задача перевірки правильності вибору виду розподілу, погодженості дійсного теоретичного розподілу з емпіричним. Перевірка гіпотези про погодженість вибіркового розподілу з теоретичним проводиться на основі спеціально підібраної випадкової величини – *критерію погодженості*. Одним з найбільш поширених таких критеріїв є критерій χ^2 Пірсона, який визначається за формулою

$$\chi^2 = \sum_{i=1}^l \frac{(n_i - m_i)^2}{m_i}, \quad (5.9)$$

де l – число часткових проміжків інтервального статистичного розподілу вибірки; n_i – спостережувана частота для відповідного проміжку; m_i – теоретична частота, знайдена за вибраним розподілом. Для кожного інтервалу $(\alpha_{i-1}; \alpha_i)$ теоретичні частоти знаходять за формулою $m_i = np_i$, де n – обсяг вибірки; $p_i = P(\alpha_{i-1} < X < \alpha_i)$, $i = \overline{1, l}$. Зокрема для нормального розподілу:

$$m_i = n \left[\Phi \left(\frac{\alpha_i - \bar{x}_b}{\sigma_b} \right) - \Phi \left(\frac{\alpha_{i-1} - \bar{x}_b}{\sigma_b} \right) \right]. \quad (5.10)$$

Випадкова величина (5.9) має розподіл χ^2 з числом ступенів волі $k = l - r - 1$, де r – число параметрів, якими визначається закон розподілу генеральної сукупності. Так, наприклад, для розподілу Пуассона, який

характеризується одним параметром λ , $r = 1$; для нормального розподілу $r = 2$, оскільки цей закон визначається двома параметрами μ і σ .

Якщо усі емпіричні частоти збігаються з теоретичними, то з формули (5.9) маємо $\chi^2 = 0$, у супротивному разі $\chi^2 > 0$. Тому критична область завжди буде правобічною. При заданих рівні значущості α і числі ступенів волі за таблицею (додаток 8) визначається критична точка: $\chi_{kp}^2 = \chi^2(\alpha; k)$. Гіпотеза H_0 про закон розподілу генеральної сукупності відхиляється, якщо спостережуване значення χ_*^2 критерію (5.9) виявиться більшим або рівним χ_{kp}^2 , і приймається H_0 , якщо $\chi_*^2 < \chi_{kp}^2$.

Застосування критерію погодженості χ^2 вимагає дотримання таких умов:

- 1) обсяг вибірки повинен бути достатньо великий ($n \geq 50$);
- 2) інтервали з малочисельними частотами ($n_i < 5$) попередньо потрібно об'єднати.

Перевірку гіпотези H_0 , що генеральна сукупність має нормальний розподіл ознаки X , виконують за такою схемою.

1. Обчислюються $\bar{x}_b, \sigma_b, u_i = \frac{\alpha_i - \bar{x}_b}{\sigma_b}, i = \overline{0; l}$.
2. Обчислюються теоретичні частоти $m_i = n[\Phi(u_i) - \Phi(u_{i-1})], i = \overline{1; l}$, причому $u_0 = -\infty, u_l = +\infty$.
3. Обчислюється спостережуване значення критерію χ_*^2 .
4. Визначається критична точка χ_{kp}^2 правобічної критичної області.
5. Робиться висновок щодо гіпотези H_0 .

Приклад 5.4. За поданим інтервальним статистичним розподілом вибірки при рівні значущості $\alpha = 0,05$ перевірити достовірність гіпотези H_0 про нормальний закон розподілу ознаки X .

X	80-90	90-100	100-110	110-120	120-130
n_i	6	10	50	25	9

Розв'язання. Значення \bar{x}_b, σ_b обчислюємо за дискретним статистичним розподілом

x_i	85	95	105	115	125
n_i	6	10	50	25	9

$$\bar{x}_b = \frac{1}{n} \sum_{i=1}^l x_i n_i = (85 \cdot 6 + 95 \cdot 10 + 105 \cdot 50 + 115 \cdot 25 + 125 \cdot 9) / 100 = 107,1;$$

$$D_b = \frac{1}{n} \sum_{i=1}^l x_i^2 n_i - (\bar{x}_b)^2 = \frac{1}{100} (85^2 \cdot 6 + 95^2 \cdot 10 + 105^2 \cdot 50 + 115^2 \cdot 25 + 125^2 \cdot 9) - (107,1)^2 = 90,59;$$

$$\sigma_b = \sqrt{90,59} = 9,52.$$

Обчислюємо теоретичні частоти за формулою (5.10) і результати обчислення запишемо у табл. 5.2:

Таблиця 5.2

α_{i-1}	α_i	n_i	u_{i-1}	u_i	$\Phi(u_{i-1})$	$\Phi(u_i)$	$\Phi(u_i) - \Phi(u_{i-1})$	m_i
30	90	6	-2,85	-1,796	-0,4977	-0,4641	0,04	4
90	100	10	-1,796	-0,746	-0,4641	-0,2734	0,191	19
100	110	50	-0,746	0,305	-0,2734	0,1217	0,395	40
110	120	25	0,305	1,402	0,1217	0,4192	0,298	30

120	130	9	1,402	2,405	0,4192	0,4918	0,073	7
-----	-----	---	-------	-------	--------	--------	-------	---

За рівнем значущості $\alpha = 0,05$ і числом ступенів волі $k = 5 - 2 - 1 = 3$ знаходимо з таблиці критичну точку $\chi_{kp}^2 = \chi^2(0,05; 2) = 6$.

Обчислюємо спостережуване значення критерію

$$\chi_*^2 = \sum_{i=1}^5 \frac{(n_i - m_i)^2}{m_i} = \frac{(6-4)^2}{4} + \frac{(10-19)^2}{17} + \frac{(50-40)^2}{40} + \frac{(25-30)^2}{30} + \frac{(9-7)^2}{7} \approx 9,163.$$

Оскільки $\chi_*^2 > \chi_{kp}^2$ (рис. 5.4), то гіпотезу H_0 відхиляємо; немає підстав вважати що ознака X має нормальний закон розподілу в генеральній сукупності.

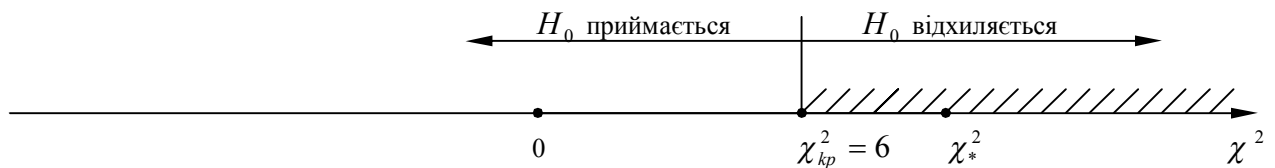


Рис. 5.4

Запитання до самоконтролю

1. Сформулюйте задачі перевірки статистичних гіпотез про рівність двох дисперсій.
2. Який статистичний критерій береться при перевірці гіпотези $H_0 : \sigma_x^2 = \sigma_y^2$?
3. Який статистичний критерій вибирається для перевірки гіпотези $H_0 : M(X) = M(Y)$, якщо обсяг вибірки великий і відомі дисперсії?
4. Як здійснюється перевірка гіпотези $H_0 : M(X) = M(Y)$, якщо генеральні дисперсії невідомі і рівні між собою, а обсяги вибірок малі?

5. Який статистичний критерій використовується при перевірці гіпотези $H_0 : M(X) = M(Y)$, якщо дисперсії невідомі і рівні між собою, а обсяги вибірок великі?
6. Яким буде критерій перевірки статистичної гіпотези $H_0 : M(X) = M(Y)$, якщо $D(X), D(Y)$ невідомі і різні, а обсяги вибірок великі?
7. Як перевіряється гіпотеза про рівність часток ознаки у двох сукупностях?
8. Чим керуються при висуненні гіпотези про закон розподілу ознаки генеральної сукупності?
9. Що беруть за значення невідомих параметрів теоретичного розподілу?
10. Що називають теоретичними частотами?
11. Як обчислюються теоретичні частоти, якщо теоретичний розподіл є нормальним?
12. Що таке критерій погодженості?
13. Як визначається критерій погодженості Пірсона?
14. Наведіть загальну схему перевірку гіпотези про нормальний закон розподілу генеральної сукупності.
15. Яким критеріям повинні задовольняти статистичні розподіли при застосуванні критерію Пірсона?

Практичне заняття № 5

Статистичні гіпотези (продовження)

№5.1. Певну продукцію виробляють на двох технологічних лініях. Зроблені на другій лінії вдосконалення привели до скорочення варіації часу випуску. Вибіркове вимірювання варіації часу випуску продукції на першій і другій лініях відповідно дало $S_x^2 = 2,52$ хв² при 10 спостереженнях і $S_y^2 = 0,6$ хв² при 9 спостереженнях. Чи можна за рівня значущості $\alpha = 0,01$ вважати суттєвою розбіжністю між варіаціями тривалості процесу виготовлення продукції на цих лініях?

№5.2. Під час дослідження стабільності температури в термостаті дістали результати: 21,2; 21,8; 21,3; 21; 21,4; 21,3. З метою стабілізації температури було використано удосконалений пристрій і після цього заміри температури показали результати 37,7; 37,6; 37,6; 37,4. Чи можна за рівня значущості $\alpha = 0,01$ вважати використання удосконаленого пристрою для стабілізації температури ефективним?

№5.3. Досліджувалось споживання масла за добу одним мешканцем у двох мікрорайонах міста. Результати вимірювання подано такими статистичними розподілами:

x_i , МГ	15	18	21	24
n_{xi}	1	5	4	2

y_i , МГ	14	20	26	30
n_{yi}	2	3	5	1

Ознаки X та Y (добове споживання масла) є незалежними нормально розподіленими випадковими величинами. При рівні значущості $\alpha = 0,1$ перевірити правильність основної гіпотези $H_0 : D(X) = D(Y)$, якщо альтернативна гіпотеза $H_1 : D(X) \neq D(Y)$.

Розв'язання. Обчислюємо: $n_x = 1 + 5 + 4 + 2 = 12$, $n_y = 2 + 3 + 5 + 1 = 11$;

$$S_x^2 = \frac{n_x}{n_x - 1} \left[\frac{\sum x_i^2 \cdot n_{xi}}{n_x} - \left(\frac{\sum x_i \cdot n_{xi}}{n_x} \right)^2 \right] =$$

$$= \frac{12}{11} \cdot \left(\frac{225 + 5 \cdot 324 + 4 \cdot 441 + 2 \cdot 576}{12} - \left(\frac{15 + 18 \cdot 5 + 21 \cdot 4 + 24 \cdot 2}{12} \right)^2 \right) =$$

$$= \frac{12}{11} \cdot (396,75 - 390,0625) \approx 7,3;$$

$$S_y^2 = \frac{n_y}{n_y - 1} \left[\frac{\sum y_i^2 \cdot n_{yi}}{n_y} - \left(\frac{\sum y_i \cdot n_{yi}}{n_y} \right)^2 \right] =$$

$$= \frac{11}{10} \cdot \left(\frac{2 \cdot 196 + 3 \cdot 400 + 5 \cdot 676 + 900}{11} - \left(\frac{14 \cdot 2 + 20 \cdot 3 + 26 \cdot 5 + 30}{11} \right)^2 \right) \approx 23,02.$$

Видно, що $S_y^2 > S_x^2$, тому за статистичний критерій візьмемо випадкову величину $F = \frac{S_y^2}{S_x^2}$.

Обчислюємо спостережуване значення критерію $f^* = \frac{S_y^2}{S_x^2} = \frac{23,02}{7,3} \approx 3,15$.

Згідно альтернативної гіпотези критична область – двобічна. Отже, при відшуванні критичної точки беруть рівень значущості вдвічі менший заданого:

$\frac{\alpha}{2} = \frac{0,1}{2} = 0,05$. Враховуючи ступені волі $k_1 = n_y - 1 = 10$, $k_2 = n_x - 1 = 11$, за

таблицею знаходимо $f_{kp} = F(0,05;10;11) = 2,86$ і будуємо критичну область (рис. 5.5).

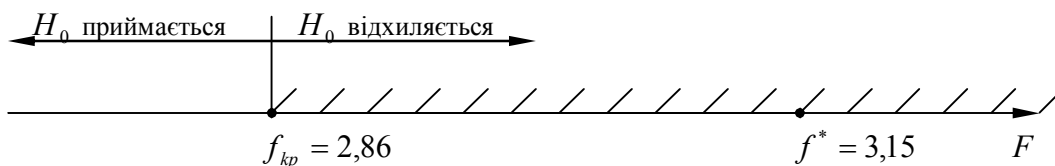


Рис. 5.5

Оскільки $f^* > f_{kp}$, то основна гіпотеза про рівність генеральних дисперсій відхиляється.

№5.4. Заміри висоти окремих паростків розсади, що вирощувались у двох парниках, подано двома статистичними розподілами:

X_i , мм	68	72	76	80
n'_i	1	2	4	5

y_j , мм	66	68	70	72
n''_j	2	3	4	6

Ознаки X і Y (довжини паростків) є незалежними нормально розподіленими випадковими величинами. При рівні значущості $\alpha = 0,1$ перевірити гіпотезу $H_0 : D(X) = D(Y)$ за альтернативної гіпотези $H_1 : D(X) \neq D(Y)$.

№5.5. Визначення вмісту (в грамах) харчових добавок у двох партіях продукції подано двома розподілами:

x_i , г	4,36	4,96	5,46	5,96
n'_i	1	2	3	6

y_j, Γ	4,44	4,84	5,24	6,64
n_j''	6	5	2	1

Ознаки X і Y (вміст добавок) є незалежними нормально розподіленими випадковими величинами. При рівні значущості $\alpha = 0,01$ перевірити гіпотезу $H_0 : D(X) = D(Y)$ за альтернативної $H_1 : D(X) > D(Y)$. Розв'язати цю задачу при рівні значущості $\alpha = 0,1$ і альтернативній гіпотезі $H_1 : D(X) \neq D(Y)$.

№5.6. Для перевірки ефективності нової технології виробництва відібрано дві групи робітників: перша працювала за новою технологією, а друга – за старою. Результати виробітку продукції на одного робітника наведено у вигляді розподілів:

x_i	53	56	59	62	65
n_{x_i}	4	6	10	12	4

y_j	48	50	52	54	56
n_{y_j}	2	5	14	6	3

Ознаки X і Y (виробіток продукції на одного працівника) є незалежними нормально розподіленими випадковими величинами зі значеннями $\sigma_x = 72$ та $\sigma_y = 50$. При рівні значущості $\alpha = 0,01$ перевірити гіпотезу $H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза $H_1 : M(X) > M(Y)$.

Розв'язання. Оскільки ознаки X та Y мають нормальний закон розподілу і

відомі їхні дисперсії, то скористаємося критерієм (5.3):
$$Z = \frac{\bar{x}_b - \bar{y}_b}{\sqrt{\frac{D_x}{n_1} + \frac{D_y}{n_2}}}$$

За рівнянням $\Phi(z_{kp}) = \frac{1 - 2 \cdot 0,01}{2} = 0,49$ із таблиці знаходимо критичну точку

$z_{kp} = 2,33$ і будуємо правобічну критичну область.

Обчислюємо: $n_1 = 4 + 6 + 10 + 12 + 4 = 36$, $n_2 = 2 + 5 + 14 + 6 + 3 = 30$;

$$\bar{x}_b = \frac{53 \cdot 4 + 56 \cdot 6 + 59 \cdot 10 + 62 \cdot 12 + 65 \cdot 4}{36} = \frac{212 + 336 + 590 + 744 + 260}{36} = 59,5;$$

$$\bar{y}_b = \frac{48 \cdot 2 + 50 \cdot 5 + 52 \cdot 14 + 54 \cdot 6 + 56 \cdot 3}{30} = \frac{96 + 250 + 728 + 324 + 168}{30} = 52,2;$$

спостережуване значення критерію $z^* = \frac{59,5 - 52,2}{\sqrt{\frac{5184}{36} + \frac{2500}{30}}} \approx 0,48$.

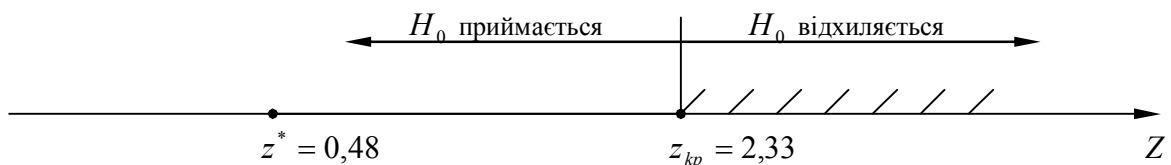


Рис. 5.6

Оскільки $z^* < z_{kp}$ (рис. 5.6), то основна гіпотеза приймається; нова технологія не привела до підвищення середнього виробітку робітників.

№5.7. Результати вивчення втрат виробництва (y %) на підприємстві після і до вдосконалення технологічного процесу наведені у вигляді розподілів:

$x_i, \%$	6,5	6,6	6,8	7	7,2
n_{x_i}	6	8	10	4	2

$y_j, \%$	6,6	6,7	6,8	7	7,1
n_{y_j}	4	7	10	8	6

При рівні значущості $\alpha = 0,01$ перевірити гіпотезу $H_0 : M(X) = M(Y)$ за альтернативної гіпотези $H_1 : M(X) \neq M(Y)$. Розв'язати задачу при $\alpha = 0,05$ і $H_1 : M(X) > M(Y)$. Вважаються відомими $D_x = 50$, $D_y = 60$.

№5.8. У двох регіонах країни досліджувались витрати праці (в год.) при виробництві продукції на підприємствах галузі. Результати представлено двома статистичними розподілами:

x_i	8,9	9,2	9,5	9,8	10,1
n_{x_i}	6	7	10	8	4

y_j	9,7	9,8	9,9	10	10,1	10,2
n_{y_j}	3	5	7	8	4	3

Вважаючи, що ознаки X , Y мають нормальний закон розподілу із значеннями дисперсій $D_x = 2,2 \text{ год}^2$, $D_y = 2,8 \text{ год}^2$. при рівні значущості $\alpha = 0,001$ перевірити гіпотезу $H_0 : M(X) = M(Y)$ за альтернативної $H_1 : M(X) < M(Y)$. Розв'язати задачу при $\alpha = 0,01$ і $H_1 : M(X) \neq M(Y)$.

№5.9. Витрати сировини на одиницю продукції (в умовних одиницях) за новою та старою технологіями відповідно подано розподілами:

x_i	115	118	121	124	127	130
n_{x_i}	1	3	6	4	3	1

y_j	114	116	118	120	122	124
n_{y_j}	2	4	6	5	2	1

Вважаючи X і Y незалежними і нормально розподіленими випадковими величинами, при рівні значущості $\alpha = 0,001$ перевірити гіпотезу $H_0 : M(X) = M(Y)$ за альтернативної $H_1 : M(X) \neq M(Y)$.

№5.10. Вимірювалась пружність зразків, виготовлених з однієї марки сталі і вибраних із двох партій. Результати подано статистичними розподілами:

x_i	34,2	38,2	42,2	46,2	50,2
n_{x_i}	2	5	10	4	4

y_j	36,8	38,8	40,8	42,8	44,8
n_{y_j}	2	4	6	5	3

Зважаючи, що ознаки X та Y – незалежні і мають нормальний розподіл при рівні значущості $\alpha = 0,01$ перевірити гіпотезу $H_0 : M(X) = M(Y)$ за альтернативної $H_1 : M(X) < M(Y)$. Розв'язати задачу при $\alpha = 0,001$ і $H_1 : M(X) \neq M(Y)$.

№5.11. Два малих підприємства виробляють однотипні батарейки. Для перевірки тривалості роботи взяли випадкову вибірку $n_x = 35$ батарейок першого підприємства. Середній термін роботи виявився 193,8 год. із вибірковою стандартним відхиленням $\sigma_{ax} = 5,8$ год. Подібна перевірка при $n_y = 40$ батарейок другого підприємства виявила: середній термін роботи 198 год., вибіркоче стандартне відхилення $\sigma_{ay} = 8,7$ год.

Вважаючи, що обидві вибірки взяті із нормальних генеральних сукупностей, при рівні значущості $\alpha = 0,01$ перевірити твердження, що термін роботи батарейок першого підприємства менший, ніж термін другого.

Розв'язання. Сформулюємо статистичну гіпотезу $H_0 : M(X) = M(Y)$ – середній

термін роботи батарейок обох підприємств однаковий і альтернативну $H_1 : M(X) < M(Y)$.

Генеральні дисперсії невідомі, тому попередньо перевіримо гіпотезу про рівність дисперсій:

$$H_0 : \sigma_x^2 = \sigma_y^2 \text{ за альтернативної } H_1 : \sigma_x^2 \neq \sigma_y^2.$$

Скористаємось F-критерієм (5.1) при рівні значущості $\alpha = 0,1$:

$$S_x^2 = \frac{n_x}{n_x - 1} \sigma_{bx}^2 = \frac{35}{34} \cdot 5,8^2 \approx 34,63;$$

$$S_y^2 = \frac{n_y}{n_y - 1} \sigma_{by}^2 = \frac{40}{39} \cdot 8,7^2 \approx 77,63.$$

Оскільки $S_y^2 > S_x^2$, то $k_1 = n_y - 1 = 39$, $k_2 = n_x - 1 = 34$.

За таблицю визначаємо критичну точку

$$f_{kp} = F\left(\frac{\alpha}{2}; k_1; k_2\right) = F(0,05; 39; 34) \approx \frac{1}{2}(F(0,05; 30; 35) + F(0,05; 50; 35)) = \frac{1,786 + 1,703}{2} = 1,74.$$

Обчислюємо спостережуване значення критерію: $f^* = \frac{S_y^2}{S_x^2} = \frac{77,63}{34,63} \approx 2,24$.

Оскільки $f^* > f_{kp}$, то гіпотеза H_0 відхиляється.

Враховуючи, що $D(X) \neq D(Y)$ і обсяги вибірок великі, для перевірки гіпотези H_0 скористаємось статистичним критерієм (5.7):

$$Z = \frac{\bar{x}_e - \bar{y}_e}{\sqrt{\frac{S_x^2}{n_1} + \frac{S_y^2}{n_2}}}$$

При альтернативній гіпотезі $H_1 : M(X) < M(Y)$ будуємо лівобічну критичну область. Критичну точку визначаємо із таблиці за рівністю

$$\Phi(z_{kp}) = -\frac{1-2\alpha}{2} = -\frac{1-2 \cdot 0,01}{2} = -0,49; \text{ тобто } z_{kp} = -2,32.$$

За даними вибірок обчислюємо значення критерію

$$z^* = \frac{193,8 - 198}{\sqrt{\frac{34,63}{35} + \frac{77,63}{40}}} = -\frac{4,2}{1,7} \approx -2,47.$$

Оскільки $z^* < z_{kp}$ (рис. 5.7), то гіпотеза H_0 відхиляється. Отже, приймається гіпотеза H_1 : середній термін роботи батарейок першого підприємства менший.

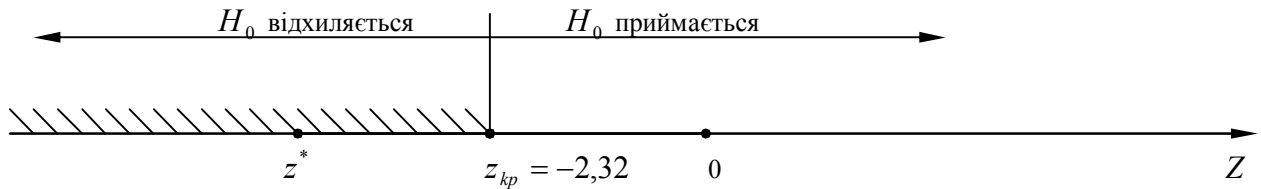


Рис. 5.7

№5.12. Супермаркет реалізує гречану крупу, яка фасується двома товариствами з обмеженою відповідальністю. Виникли підозри, що друге товариство в середньому наповнює пакунки меншою кількістю крупи, ніж перше. Із партії пакунків двох постачальників були зроблені випадкові вибірки і отримані наступні результати: $n_x = 30$, $\bar{x}_b = 1003,6$ г, $\sigma_{bx} = 5,03$ г – для першого товариства; $n_y = 40$, $\bar{y}_b = 998,2$ г, $\sigma_{by} = 4,39$ г – для другого. Чи є підстави для подібних підозр?

№5.13. В партії з 500 деталей, які виготовлені першим автоматом, виявилось 60 нестандартних, а із 600 деталей, виготовлених другим автоматом, виявилось 40 нестандартних. При рівні значущості 0,01 перевірити гіпотезу $H_0 : p_1 = p_2 = p$ (про рівність ймовірностей виготовлення нестандартної деталі кожним автоматом), якщо $H_1 : p_1 \neq p_2$. Розв'язати задачу при $H_1 : p_1 > p_2$.

№5.14. Зовнішній аудитор перевіряє систему обліку малого підприємства. Перша перевірка 100 обладунк показала, що 56 із них містять помилки.

Підприємству надали місяць для виправлення системи. Повторна перевірка виявила 28 помилкових серед 78 перевірених обладок. Чи є підстави вважати, що частка помилок зменшилась?

№5.15. За статистичним розподілом вибірки

$[\alpha_{i-1}; \alpha_i)$	0-6	6-12	12-18	18-24	24-30	30-36
n_i	8	12	30	36	10	4

визначити теоретичний закон розподілу ознаки генеральної сукупності X та при рівні значущості $\alpha = 0,01$ перевірити правильність висунутої нульової гіпотези.

Розв'язання. За заданим інтервальним розподілом будуюмо гістограму частот (рис. 5.8).

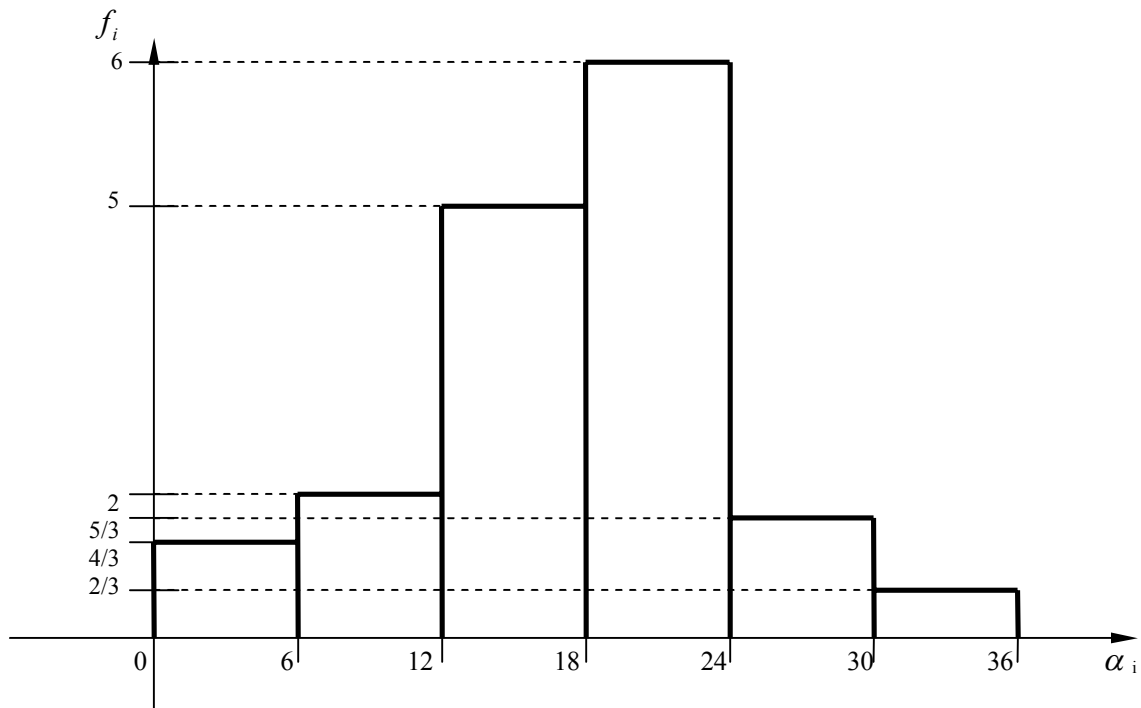


Рис. 5.8

Тут $h = 6$, $f_1 = \frac{4}{3}$, $f_2 = 2$, $f_3 = 5$, $f_4 = 6$, $f_5 = \frac{5}{3}$, $f_6 = \frac{2}{3}$.

За формою гістограми частот можна припустити, що ознака X розподілена за

нормальним законом. Отже, висуваємо гіпотезу H_0 : ознака X має нормальний закон розподілу. Для перевірки висунутої нульової гіпотези використаємо критерій погодженості Пірсона.

Перейдемо від інтервального до дискретного розподілу і знайдемо \bar{x}_b, σ_b .

x_i	3	9	15	21	27	33
n_i	8	12	30	36	10	4

$$\bar{x}_b = \frac{3 \cdot 8 + 9 \cdot 12 + 15 \cdot 30 + 21 \cdot 36 + 27 \cdot 10 + 33 \cdot 4}{100} = 17,4,$$

$$\sigma_b = \sqrt{\frac{9 \cdot 8 + 81 \cdot 12 + 225 \cdot 30 + 441 \cdot 36 + 729 \cdot 10 + 1089 \cdot 4}{100} - 17,4^2} \approx 7,1.$$

Обчислюємо теоретичні частоти (табл. 5.3):

Табл. 5.3

α_{i-1}	α_i	n_i	$u_{i-1} = \frac{\alpha_{i-1} - \bar{x}_b}{\sigma_b}$	$u_i = \frac{\alpha_i - \bar{x}_b}{\sigma_b}$	$\Phi(u_{i-1})$	$\Phi(u_i)$	$m_i = n(\Phi(u_i) - \Phi(u_{i-1}))$
0	6	8	-2,45	-1,61	-0,4929	-0,4463	5
6	12	12	-1,61	-0,76	-0,4463	-0,2764	17
12	18	30	-0,76	0,08	-0,2764	0,0319	31
18	24	36	0,08	0,93	0,0319	0,3238	29
24	30	10	0,93	1,77	0,3238	0,4616	14
30	36	4	1,77	2,62	0,4616	0,4956	3

Знаходимо спостережуване значення критерію (табл. 5.4):

Таблиця 5.4

$n_i - m_i$	$(n_i - m_i)^2$	$\frac{(n_i - m_i)^2}{m_i}$
3	9	1,8
-5	25	1,47
-1	1	0,03
7	49	1,69
-4	16	1,14
1	1	0,33

$$\chi^2_* = \sum_{i=1}^6 \frac{(n_i - m_i)^2}{m_i} = 1,8 + 1,47 + 0,03 + 1,69 + 1,14 + 0,33 = 6,46.$$

За таблицею (додаток 8) при відомих $\alpha = 0,01$, $k = 6 - 2 - 1 = 3$ для розподілу χ^2 відшукаємо $\chi^2_{кр.} = \chi^2(\alpha; k) = \chi^2(0,01; 3) = 11,3$.

Оскільки $\chi^2_* < \chi^2_{кр.}$, то немає підстав для відхилення нульової гіпотези про нормальний закон розподілу ознаки X .

№5.16. Вимірювання зросту юнаків випускних класів школи дав такі результати:

$[\alpha_{i-1}; \alpha_i)$, см	154-158	158-162	162-166	166-170	170-174	174-178	178-182	182-186
n_i	8	14	20	32	12	8	4	2

Підібрати теоретичний закон розподілу ознаки X – зріст юнака. При рівні значущості $\alpha = 0,01$ перевірити правильність висунутої гіпотези.

№5.17. Контролювався вміст нітратів (в умовних одиницях) у сировині, що надходила на переробку. Результати наведені в таблиці:

$[\alpha_{i-1}; \alpha_i)$	10,4-10,6	10,6-10,8	10,8-11,0	11,0-11,2	11,2-11,4
n_i	40	100	200	40	20

При рівні значущості $\alpha = 0,05$ перевірити гіпотезу H_0 про нормальний закон розподілу ознаки X .

№5.18. За інтервальним статистичним розподілом вибірки

$[\alpha_{i-1}; \alpha_i)$	80-90	90-100	100-110	110-120	120-130
n_i	2	14	60	20	4

перевірити гіпотезу H_0 про нормальний закон розподілу ознаки X . Прийняти рівень значущості $\alpha = 0,01$.

№5.19. За статистичним розподілом

$[\alpha_{i-1}; \alpha_i)$	0-10	10-20	20-30	30-40	40-50
n_i	40	30	20	6	4

з'ясувати теоретичний закон розподілу випадкової величини X . При рівні значущості $\alpha = 0,05$ перевірити правильність такого припущення.

Розв'язання. Будуємо гістограму частот (рис. 5.9).

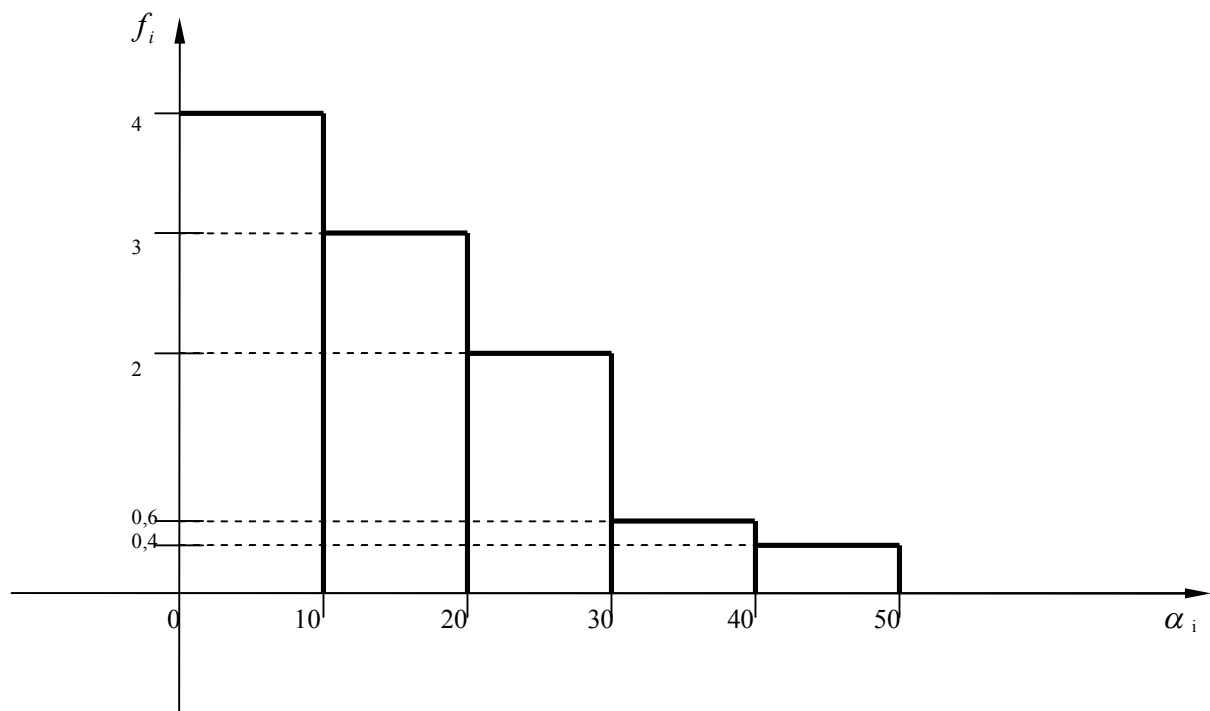


Рис. 5.9

За формою отриманої гістограми можна стверджувати, що ознака X має показниковий закон розподілу. Для перевірки цього твердження застосуємо критерій Пірсона. Теоретичні частоти в цьому випадку обчислюються за формулою

$$m_i = n(e^{-\lambda\alpha_{i-1}} - e^{-\lambda\alpha_i}), \text{ де } \lambda = \frac{1}{\bar{x}_b}.$$

Перейдемо до дискретного розподілу і знайдемо \bar{x}_b :

x_i	5	15	25	35	45
n_i	40	30	20	6	4

$$\bar{x}_b = \frac{5 \cdot 40 + 15 \cdot 30 + 25 \cdot 20 + 35 \cdot 6 + 45 \cdot 4}{100} = 15,4;$$

$$\text{тоді } \lambda = \frac{1}{15,4} \approx 0,065.$$

Далі обчислюємо теоретичні частоти (табл. 5.5).

Таблиця 5.5

α_{i-1}	α_i	n_i	$e^{-\lambda\alpha_{i-1}}$	$e^{-\lambda\alpha_i}$	$m_i = n(e^{-\lambda\alpha_{i-1}} - e^{-\lambda\alpha_i})$
0	10	40	1	0,522	48
10	20	30	0,522	0,273	25
20	30	20	0,273	0,142	13
30	40	6	0,142	0,074	7
40	50	4	0,074	0,0039	7

Обчислюємо спостережуване значення критерію (табл. 5.6).

Таблиця 5.6

$n_i - m_i$	$(n_i - m_i)^2$	$\frac{(n_i - m_i)^2}{m_i}$
-8	64	1,33
5	25	1
7	49	3,77
-1	1	0,14
-3	9	1,29

$$\chi^2_* = \sum_{i=1}^5 \frac{(n_i - m_i)^2}{m_i} = 1,33 + 1 + 3,77 + 0,14 + 1,29 = 7,53.$$

Відшукуємо критичну точку для розподілу χ^2 (при $\alpha = 0,05$; $k = 5 - 1 - 1 = 3$):

$$\chi^2_{кр.} = \chi^2(\alpha; k) = \chi^2(0,05; 3) = 7,8.$$

Оскільки $\chi^2_* < \chi^2_{кр.}$, то нульова гіпотеза про показниковий закон розподілу ознаки X приймається.

№5.20. За статистичним розподілом

$[\alpha_{i-1}; \alpha_i)$	0-8	8-16	16-24	24-32	32-40
n_i	40	30	20	8	2

з'ясувати теоретичний закон розподілу ознаки X і при рівні значущості $\alpha = 0,01$ перевірити правильність такого припущення.