МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний університет «Чернігівська політехніка»

МАТЕМАТИЧНА СТАТИСТИКА

МЕТОДИЧНІ ВКАЗІВКИ ТА ЗАВДАННЯ
ДО САМОСТІЙНОЇ РОБОТИ
З ДИСЦИПЛІНИ „ТЕОРІЯ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНА СТАТИСТИКА”
для студентів інженерних спеціальностей

Обговорено і рекомендовано
на засіданні кафедри АТ та ГМ,
протокол № 13 від 14.06. 2021р.

Чернігів - 2021

Математична статистика. Методичні вказівки та завдання до самостійної роботи з дисципліни “Теорія ймовірностей та математична статистика ”, для студентів інженерних спеціальностей./ Укл.: Казнадій С.П., Корнієнко С.П., Мурашковська В.П.— Чернігів: НУ «Чернігівська політехніка», 2021. — 45с.

Укладачі: Казнадій Світлана Петрівна, старший викладач
Корнієнко Світлана Петрівна, кандидат технічних наук, доцент
Мурашковська Віра Петрівна, старший викладач

Відповідальний за випуск: Кальченко Віталій Іванович, завідувач кафедри автомобільного транспорту та галузевого машинобудування, професор, доктор технічних наук

Рецензент: Венжега Володимир Іванович – доцент, кандидат технічних наук кафедри автомобільного транспорту та галузевого машинобудування

Зміст

Вступ

Математична статистика широко застосовується, як в прикладній математиці, так і в різних природничих дисциплінах. Методи математичної статистики складають математичні методи систематизації, обробки і аналізу будь-яких статистичних даних. Математичні методи безпосередньо зв’язані з імовірнісною оцінкою результатів спостереження і визначення величини математичної імовірності. В зв’язку з цим висновки математичної статистики відносно масових явищ і процесів носять імовірнісний характер і опираються на апарат теорії ймовірностей.

Теоретичною основою курса “Математична статистика ” є курс теорії імовірностей – складова частина курсу вищої математики, який вивчає закономірності випадкових явищ.

Мета цих методичних вказівок – надати студентам теоретичну та практичну допомогу в самостійній роботі по вивченню розділу “Математична статистика ” з дисципліни “Теорія ймовірностей та математична статистика ”. Методичні вказівки містять теоретичний матеріал, необхідний для самостійного виконання студентами індивідуальних домашніх завдань. Наведені приклади розв’язання основних стандартних задач.

В розділі “Завдання до самостійної роботи ” наведено 30 варіантів індивідуальних завдань. Завдання охоплюють в цілому матеріал розділу, який в тому чи іншому обсязі вивчається студентами.

Методичні вказівки можуть бути використані під час аудиторних занять, як довідковий матеріал та в якості задачника при проведенні самостійних та модульно-тестових робіт.

Частина 1
КОРОТКІ ТЕОРЕТИЧНІ ВІДОМОСТІ

1.1 Статистичні розподіли вибірок та їх числові характеристики
1.1.1 Генеральна та вибіркова сукупність

Нехай потрібно дослідити сукупність однорідних об'єктів (статистичну сукупність) відносно деякої якісної або кількісної ознаки, що характеризує ці об'єкти. Повне дослідження (коли досліджується кожний з об'єктів сукупності) вимагає великих матеріальних витрат. Тому вибирають випадковим чином з усієї сукупності обмежену кількість об'єктів і їх досліджують.

Вибірковою сукупністю (вибіркою) називають сукупність випадково відібраних однорідних об'єктів.

Генеральною сукупністю називають сукупність всіх однорідних об'єктів, з яких створюється вибірка.

Обсягом сукупності (вибіркової або генеральної) називають кількість об'єктів цієї сукупності.

Вибірки бувають двох типів:

1) *повторна* - відібраний об'єкт повертається в генеральну сукупність;

2) *без повторна* - відібраний об'єкт не повертається в генеральну сукупність.

Вибірка повинна правильно відображати пропорції генеральної сукупності, тобто бути *репрезентативною,* або *представницькою.* Згідно з законом великих чисел вибірка буде репрезентативною, якщо її здійснюють випадково.

1.1.2 Дискретний статистичний розподіл вибірки та її числові характеристики

Кількісна ознака елементів генеральної сукупності можуть бути одновимірними і багатовимірними, дискретними і неперервними.

Коли реалізується вибірка, кількісна ознака набуває конкретних числових значень x_i, яку називають *варіантою.*

Зростаючий числовий ряд варіант називають *варіаційним рядом (точковим).*

Кожна варіанта 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, що різняться числовим значенням; n - обсяг вибірки.

Різниця $x_k - x_1$, має *назву розмаху* варіаційного ряду. Наприклад, для варіаційного ряду 1, 4, 8, 9, 12 розмах дорівнює 12 - 1 = 11.Числа спостережень називають *частотами,* а їх відношення до обсягу вибірки $\frac{n_i}{n} = w_i$ називають *відносними частотами***.** Для кожної вибірки виконується рівність $\sum_{i=1}^{k} w_i = 1.$

Дискретним статистичним розподілом вибірки називають перелік варіант x_i та відповідних їм частот n_i або відносних частот $\frac{n_i}{n} = w_i$.

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k
w_i	w_1	w_2	…	w_k

Емпіричною функцією розподілу (функцією розподілу вибірки) називають функцію $F^*(x)$, що визначає для кожного x відносну частоту події $X < x$, тобто

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

де n_x — кількість варіант статистичного розподілу вибірки, значення яких менше за фіксовану варіанту x; n - обсяг вибірки.

$F^*(x)$ - називають ще *функцією нагромадження відносних частот*.

Властивості функції $F^*(x)$:

1. $0 \le F^*(x) \le 1$;
2. $F^*(x_{min}) = 0$, де x_{min} є найменшою варіантою варіаційного ряду;
3. $F^*(x_{max}) = 1$, де x_{max} є найбільшою варіантою варіаційного ряду;
4. $F^*(x)$ є неспадною функцією аргументу x, а саме: $F^*(x_2) \ge F^*(x_1)$ при $x_2 \ge x_1$.

Зображення варіаційних рядів у табличній та графічній формах дозволяє одержати лише перше уявлення про найбільш загальні характерні властивості досліджуваного розподілу.

Всебічна характеристика рядів розподілу передбачає з'ясування умов, під впливом яких сформувався досліджуваний розподіл, вираження основних особливостей числовими характеристиками.

Медіаною Me^* називають варіанту, яка поділяє варіаційний ряд на дві частини, рівні за кількістю варіант. Якщо кількість варіант непарна, тобто $n = 2к + 1$, то $Me = x_k + 1$; при парному $n=2k$ медіана $Me = (x_k + x_{k+1})/2$.
Наприклад, для ряду 1, 4, 8, 9, 12 , $Me^* = 8$, для ряду 1, 4, 8, 9 $Me^* = (4 + 8) / 2 = 6$.

Модою Mo^* називають варіанту, яка має найбільшу частоту появи.

Вибірковою середньою $\overline{x_B}$ називають середнє арифметичне значення варіант вибірки.

Якщо всі значення $x_1, x_2, \ldots, x_n$ ознаки вибірки обсягу n різні, то

$$\overline{x_B} = \frac{(x_1 + x_2 + \ldots + x_n)}{n}.$$

Якщо значення ознаки $x_1, x_2, \ldots, x_к$ мають відповідно частоти $n_1, n_2, \ldots, n_к$, при чому $n_1 + n_2 + \ldots + n_к = n$, то

$$\overline{x_B} = \frac{n_1x_1 + n_2x_2 + \ldots + n_kx_k}{n} = \frac{\sum_{i=1}^{k} n_i x_i}{n}.$$

Зауваження. Якщо варіанти x_i — великі числа, то для спрощення розрахунків доцільно відняти від кожної варіанти одне і те саме число c, тобто перейти до умовних варіант $u_i = x_i - c$ (за c зручно взяти число, близьке до вибіркової середньої; оскільки вибіркова середня невідома, за С вибирають число з найбільшою частотою).

$$\text{Тоді}\quad \overline{x_B} = C + \frac{\sum_{i=1}^{k} n_i u_i}{n}.$$

Відхиленням варіант називають різницю (x_i - $\overline{x_B}$).
Сума відхилень усіх варіант варіаційного ряду вибірки завжди дорівнює нулеві:

$$\sum_{i=1}^{n}\left(x_i - \overline{x_B}\right)n_i = \sum_{i=1}^{n} x_i n_i - \sum_{i=1}^{n} \overline{x_B} n_i = n x_B - n x_B = 0.$$

Дисперсія вибирається для вимірювання розсіювання варіант вибірки відносно $\overline{x_B}$.

Дисперсія вибірки – це середнє арифметичне квадратів відхилень відносно $\overline{x_B}$, яке обчислюється за формулою

$$D_B = \frac{\sum_{i=1}^{n}(x_i - \overline{x_B})^2 n_i}{n} \quad \text{або} \quad D_B = \frac{\sum_{i=1}^{n} x_i^2 n_i}{n} - \left(\overline{x_B}\right)^2.$$

Середнє квадратичне відхилення вибірки σ_B. При обчисленні D_B відхилення підноситься до квадрата, а отже, змінюється одиниця виміру ознаки *X*, тому на основі дисперсії вводиться середнє квадратичне відхилення $\sigma_B = \sqrt{D_B}$, яке вимірює розсіювання варіант вибірки відносно $\overline{x_B}$, але в тих самих одиницях, в яких вимірюється ознака *X*.

Виправлену вибіркову дисперсію позначають

$$S^2 = \frac{n}{n-1} D_B = \frac{1}{n-1}\sum_{i=1}^{m} n_i (x_i - \overline{x_B})^2$$

Виправленим середнім квадратичним відхиленням вибірки буде s=$\sqrt{S^2}$.

При досить великих *n* (обсяг вибірки) вибіркова дисперсія D_B, та виправлена вибіркова дисперсія S^2 різняться дуже мало. Тому в практичних задачах виправлену дисперсію S^2 та виправлене середнє квадратичне відхилення вибірки *s* використовують лише за обсягу вибірки *n* <30.

Для грубого оцінювання розсіювання варіант відносно $\overline{x_B}$ застосовується величина, яка дорівнює різниці між найбільшою x_{max} і найменшою x_{min} варіантами варіаційного ряду. Ця величина називається *розмахом.* $R = x_{max} - x_{min}$.

Коефіцієнт варіації V. Для порівняння оцінок варіацій статистичних рядів із різними значеннями $\overline{x_B}$, які не дорівнюють нулеві, вводиться коефіцієнт варіації, який обчислюється за формулою $V = \frac{\sigma_B}{\overline{x_B}} \cdot 100\%$.

Приклад 1. У результаті вибіркового аналізу добової виручки підприємства отримана вибірка обсягом *n* = 40 (млн. грн.):

0,87	0,94	0,99	0,90	0,90	0,87	0,85	0,87	0,90	0,94
0,87	0,87	0,82	0,90	0,94	0,90	0,85	0,85	0,87	0,94
0,81	0,82	0,87	0,97	0,90	0,94	0,85	0,81	0,87	0,85
0,90	0,82	0,99	0,90	0,94	0,82	0,97	0,81	0,85	0,87

Скласти: а) варіаційний ряд; б) дискретний статистичний розподіл вибірки; в) побудувати емпіричну функцію дискретного розподілу $F^*(x)$. Обчислити: г) Me^*, д) Mo^*, е) $\overline{x_B}$, ж) D_B, з) σ_B, і) R, к) V.

Розв'язання.

а) випишемо різні значення варіант, що потрапили до вибірки: 0,87; 0,94; 0,99; 0,90; 0,85; 0,82; 0,81; 0;97. Запишемо їх в порядку зростання, дістанемо дискретний варіаційний ряд:

0,81; 0,82; 0,85; 0,87; 0,90; 0,94; 0,97; 0,99.

б) підрахуємо частоту кожної варіанти з варіаційного ряду і занесемо дані до таблиці:

x_i	0,81	0,82	0,85	0,87	0,90	0,94	0,97	0,99
n_i	3	4	6	9	8	6	2	2

Таким чином, отримуємо дискретний статистичний розподіл вибірки.

в) емпіричну функцію шукаємо за формулою

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

Найменша варіанта дискретного статистичного розподілу вибірки дорівнює $x_{min} = 0,81$, отже, $F^*(x) = 0$ при $x \le 0,81$.

Значення $X < 0,82$, а саме $x_1 = 0,81$ спостерігалося 3 рази, отже,

$F^*(x) = \frac{3}{40} = 0,075$, при $0,81 < x \le 0,82$.

Значення $X < 0,85$, а саме $x_1 = 0,81$, $x_2 = 0,82$ спостерігалися 3 + 4 = 7 раз, отже,

$F^*(x) = \frac{7}{40} = 0,175$, при $0,82 < x \le 0,85$.

Значення $X<0,87$, а саме $x_1 = 0,81$, $x_2 = 0,82$, $x_3 = 0,85$ спостерігалися 3+4+6=13 разів, отже,

$F^*(x) = \frac{13}{40} = 0,325$, при $0,85 < x < 0,87$.

Значення $X<0,90$, а саме $x_1 = 0,81$, $x_2 = 0,82$, $x_3 = 0,85$, $x_4 = 0,87$ спостерігалися 3+4+6+9=22 рази, отже,

$F^*(x) = \frac{22}{40} = 0,55$, при $0,87<x<0,90$.

Значення $X<0,94$, а саме $x_1 = 0,81$, $x_2 = 0,82$, $x_3 = 0,85$, $x_4 = 0,87$, $x_5 = 0,90$ спостерігалися 3+4+6+9+8=30 разів, отже,

$F^*(x) = \frac{30}{40} = 0,75$, при $0,90<x<0,94$.

Значення $X <0,97$, а саме $x_1 = 0,81$, $x_2 = 0,82$, $x_3 = 0,85$, $x_4 = 0,87$, $x_5 = 0,90$, $x_6 = 0,94$ спостерігалися 3+4+6+9+8+6 = 36 раз, отже,

$F^*(x) = \frac{36}{40} = 0,9$ при $0,94 < x < 0,97$.

Значення X < 0,99, а саме x_1 =0,81, x_2 =0,82, x_3 =0,85, x_4 =0,87, x_5 =0,90, x_6= 0,94, x_7 =0,97 спостерігалися 3+4+6+9+8+6+2=38 раз, отже,

$F^*(x) = \frac{38}{40} = 0{,}95$, при 0,97<x<0,99.

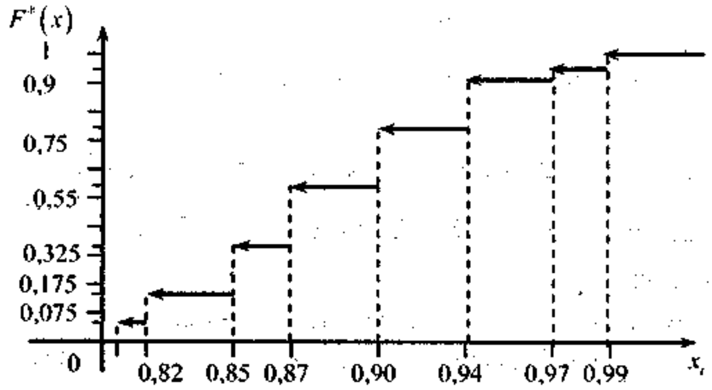
Значення X>0,99, а саме x_1 =0,81, x_2 =0,82, x_3 =0,85, x_4 =0,87, x_5 =0,90, x_6 =0,94, x_7=0,97, x_8 =0,99 спостерігалися 3+4+6+9+8+6+2+2 = 40 разів, отже,

$F^*(x) = \frac{40}{40} = 1$, при x>0,99.

Згідно з означенням $F^*(x)$має такий вигляд:

$$F^*(x) = \begin{cases} 0, & x \le 0{,}81; \\ 0{,}075, & 0{,}81 < x \le 0{,}82; \\ 0{,}175, & 0{,}82 < x \le 0{,}85; \\ 0{,}325, & 085 < x \le 0{,}87; \\ 0{,}55, & 0{,}87 < x \le 0{,}90; \\ 0{,}75, & 0{,}90 < x \le 0{,}94; \\ 0{,}9, & 0{,}94 < x \le 0{,}97; \\ 0{,}95, & 0{,}97 < x \le 0{,}99 \\ 1, & x > 0{,}99. \end{cases}$$

Побудуємо графік цієї функції:



г) кількість варіант парна n=2k, тому медіана $Me^*=(x_k + x_{k+1})/2 =$ (0,87+0,90)/2 =0,89, Me^*=0,89.

д) найбільша частота появи n_4=9 у варіанти 0,87, тому мода Mo^*= 0,87.

е) $\overline{x_B} = \frac{\sum_{i=1}^{k} n_i x_i}{n} =$

$$= \frac{0{,}81 \cdot 3 + 0{,}82 \cdot 4 + 0{,}85 \cdot 6 + 0{,}87 \cdot 9 + 0{,}90 \cdot 8 + 0{,}94 \cdot 6 + 0{,}97 \cdot 2 + 0{,}99 \cdot 2}{40} = 0{,}885.$$

$\overline{x_B} = 0{,}885$;

ж)

$$D_B = \frac{\sum_{i=1}^{n} x_i^2 n_i}{n} - (\overline{x}_B)^2 =$$

$$= \frac{(0,81)^2 \cdot 3 + (0,82)^2 \cdot 4 + (0,85)^2 \cdot 6 + (0,87)^2 \cdot 9 + (0,90)^2 \cdot 8 + (0,94)^2 \cdot 6 + (0,97)^2 \cdot 2 + (0,99)^2 \cdot 2}{40} - (0,885)^2 =$$

$$0,785715 - 0,783225 = 0,0025$$

D_B=0,0025;

з) $\sigma_B = \sqrt{D_B} = \sqrt{0,0025} = 0,05;$

і) $R = x_{max} - x_{min} = 0,99 - 0,81 = 0,18;$

к) $V = \frac{\sigma_B}{\overline{x}_B} \cdot 100\% = \frac{0,05}{0,885} \cdot 100\% = 5,65.$

1.1.3 Інтервальний статистичний розподіл. Полігон і гістограма частот

В тому випадку, коли дискретний варіаційний ряд має велику кількість варіант, чи побудова його взагалі не рентабельна, оскільки у вибірці майже всі елементи різні, будують *інтервальний варіаційний ряд.* Орієнтовану кількість *m* часткових інтервалів можна визначити за формулою

$m = [\sqrt{n}]$ (ціла частина $\sqrt{n}$), або за формулою Стерджерса $m = 1+1,4\cdot ln(n)$.

Як правило інтервал береться однаковим. Тут $h = x_i - x_{i-1}$ є довжиною часткового i-го інтервалу. Інтервальний статистичний розподіл можна задати також у вигляді послідовності інтервалів та відповідних їм частот або відносних частот.

Інтервальний статистичний розподіл має вигляд:

Інтервал	$x_0 \div x_1$	$x_1 \div x_2$	$x_2 \div x_3$	…	$x_{k-1} \div x_k$
n_i	n_1	n_2	…	n_k	n_i
w_i	w_1	w_2	…	w_k	w_i

За інтервальним варіаційним рядом можна побудувати точковий варіаційний ряд, замінивши інтервал одним з його представників. Якщо випадкова величина набуде *m* значень, рівних x_k, то у випадку парного значення *m* половину цих значень можна віднести до інтервалу $x_{i-1} \div x_i$, а другу половину – до інтервалу $x_i \div x_{i+1}$, $1 \le i \le l-1$. При непарному *m* до одного з цих двох інтервалів можна віднести $\frac{m+1}{2}$ значень, а до другого - $\frac{m-1}{2}$ значень. При великому обсязі *n* вибірки немає істотного значення, до якого з інтервалів віднесено більшу кількість значень.

Емпірична функція $F^*(x)$ (комулята). При побудові комуляти $F^*(x)$ для інтервального статистичного розподілу вибірки за основу береться припущення, що ознака на кожному частинному інтервалі має рівномірну щільність імовірностей. Тому комулята матиме вигляд ламаної лінії, яка зростає на кожному частковому інтервалі і наближається до одиниці.

Аналогом емпіричної функції $F^*(x)$ у теорії ймовірностей є інтегральна функція $F(x) = P(X<x)$.

Для визначення *медіани* інтервального статистичного розподілу вибірки необхідно визначити медіанний частковий інтервал. Якщо, наприклад на i-му інтервалі $[x_{i-1}\text{-}x_i]$ $F^*(x_{i-1})<0{,}5$ і $F^*(x_i)>0{,}5$, то беручи до уваги, що досліджувана ознака X є неперервною і при цьому $F^*(x)$ є неспадною функцією, всередині інтервалу $[x_{i-1} \div x_i]$ неодмінно існує таке значення $X = Me$, де $F^*(Me)=0{,}5$.

Для визначення *моди* інтервального статистичного розподілу необхідно знайти модальний інтервал, тобто такий частинний інтервал, що має найбільшу частоту появи.

Використовуючи лінійну інтерполяцію, *моду* обчислимо за формулою

$$Mo^* = x_{i-1} + \frac{n_{Mo} - n_{Mo-1}}{2n_{Mo} - n_{Vo-1} - n_{Vo+1}} h, \text{ де}$$

x_{i-1} - початок модального інтервалу;
h – довжина, або крок, часткового інтервалу;
n_{Mo} – частота модального інтервалу;
n_{Mo-1} - частота до модального інтервалу;
n_{Mo+1} - частота після модального інтервалу.

Для визначення $\overline{x_B}$, D_B, σ_B перейдемо від інтервального розподілу до дискетного, варіантами якого є середина часткових інтервалів $x_i^* = x_{i-1} + \frac{h}{2} = x_i - \frac{h}{2}$ і який має такий вигляд:

$x_i^* = x_{i-1} + \frac{h}{2} = x_i - \frac{h}{2}$	x_1^*	x_2^*	x_3^*	…	x_k^*
h_i	h_1	h_2	h_3	…	h_k

Тоді $\overline{x_B}$, D_B, σ_B обчислюються за формулами:

$$\overline{x_B} = \frac{\sum_{i=1}^{k} n_i x_1^*}{h}; \; D_B = \frac{\sum_{i=1}^{n} (x_1^*)^2 \, n_i}{h} - \left(\overline{x_B}\right)^2; \; \sigma_B = \sqrt{D_B}.$$

Дискретний статистичний розподіл вибірки можна зобразити графічно у вигляді ламаної лінії, відрізки якої сполучають координати точок (x_i, n_i), або (x_i, w_i). У першому випадку ламану лінію називають *полігоном частот*, у другому - *полігоном відносних частот.*

Для побудови полігону частот відкладемо на осі абсцис варіанти x_i а на осі ординат - відповідні їм частоти n_i і, з'єднавши точки (x_i, n_i) відрізками прямих, отримаємо шуканий полігон частот.

Для ілюстрації розподілу неперервної випадкової величини має сенс будувати діаграми, які називаються гістограмами. Інтервал, де знаходяться всі значення ознаки, які спостерігаються, розбивається на кілька часткових інтервалів завдовжки h. Для кожного часткового інтервалу знаходять n_i - суму частот варіант, що потрапили в i-й інтервал.

Гістограмою частот називають східчасту фігуру, що складається з прямокутників, основами яких є інтервали завдовжки $h_i = x_i - x_{i-1}$, а висоти дорівнюють $\frac{n_i}{h_i}$ (щільність частоти).Площа часткового i-го прямокутника дорівнює $h_i \cdot \left(\frac{n_i}{h_i}\right) = n_i$ - сумі частот варіант, що потрапили в i-й інтервал. Площа гістограми частот дорівнює сумі всіх частот, тобто обсягу вибірки.

Гістограмою відносних частот називають ступінчату фігуру, що складається з прямокутників, основами яких є часткові інтервали завдовжки h_i, а висоти дорівнюють відношенню $\frac{w_i}{h_i}$ (щільність відносної частоти).

Площа часткового i-го прямокутника дорівнює $h_i \cdot \left(\frac{w_i}{h_i}\right) = w_i$ - відносній частоті варіант, що потрапили в i-й інтервал. Площа гістограми відносних частот дорівнює сумі відносних частот, тобто одиниці.

Приклад. На кондитерській фабриці автомат розливає рідкий шоколад у форми для одержання шоколадних плиток. Для контролю навмання відібрано 20 плиток із партії готової продукції. Результати обстеження наведені в таблиці:

Маса, *г*	96-98	98-100	100-102	102-104
К-ть плиток	4	8	6	2

Побудувати: 1) полігон і гістограму відносних частот; 2) кумулятивну криву.

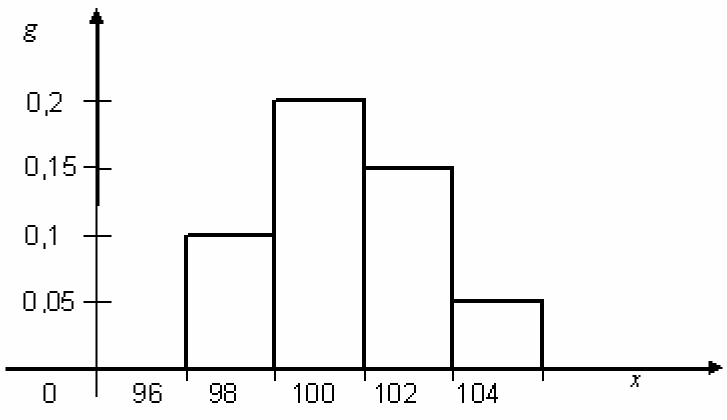
Розв'язання. Для кількісної ознаки X – маса шоколаду у формах, маємо неперервний статистичний розподіл. Обчислимо відносні частоти $w_i = \frac{n_i}{n}$ появи ознаки X на відповідних інтервалах; маємо ряд розподілу:

Інтервал	96-98	98-100	100-102	102-104
w_i	0,2	0,4	0,3	0,1

Довжина кожного інтервалу $h_i = x_i - x_{i-1} = 2$, $i = 1,\dots, 4$. Визначаємо щільності g_i:

$g_i = \frac{0,2}{2} = 0,1; g_2 = \frac{0,4}{2} = 0,2$; $g_3 = \frac{0,3}{2} = 0,15; g_4 = \frac{0,1}{2} = 0,05.$

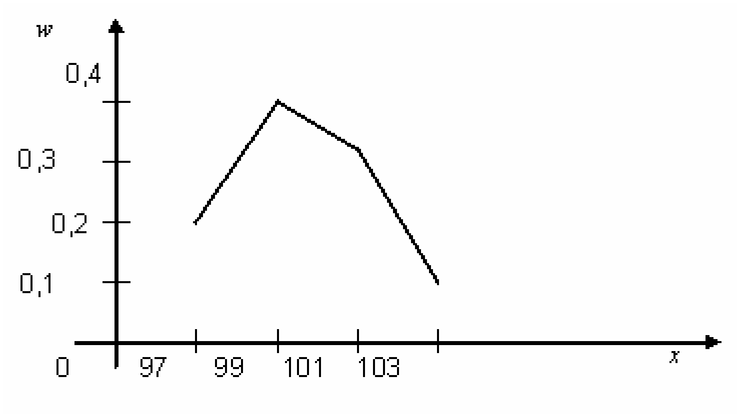
Гістограма відносних частот має вигляд



Перейдемо від інтервального до дискретного статистичного розподілу відносних частот

x_i	97	99	101	103
wi	0,2	0,4	0,3	0,1

Побудуємо полігон відносних частот :



а) якщо $0\leq x\leq 96$, то $F(x)=\frac{n_x}{n}=\frac{0}{20}=0$;

б) якщо $x = 98$, то $F(x)=\frac{n_1}{n}=\frac{4}{20}=0{,}2$;

в) якщо $x = 100$, то $F(x)=\frac{n_1+n_2}{n}=\frac{4+8}{20}=0{,}6$;

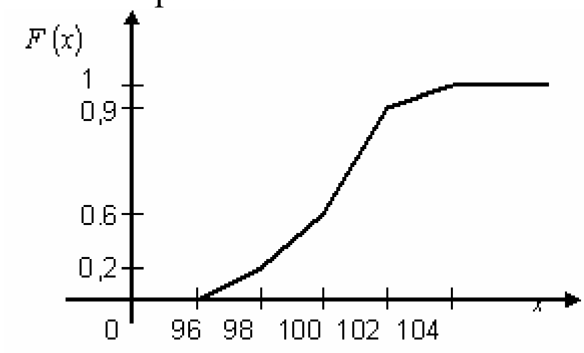
г) якщо $x = 102$, то $F(x)=\frac{n_1+n_2+n_3}{n}=\frac{4+8+6}{20}=0{,}9$;

д) якщо $x\geq 104$, то $F(x) = 1$.

Отже, маємо емпіричну функцію

$$F(x)=\begin{cases} 0, & при\ x\in [0;96]; \\ 0{,}2, & при\ x=98; \\ 0{,}6, & при\ x=100; \\ 0{,}9, & при\ x=102; \\ 1, & при\ x\geq 104. \end{cases}$$

Її графіком буде кумулятивна крива:



1.1.4 Емпіричні моменти

Початковим емпіричним моментом k-го порядку $v^*{}_k$ називають середнє зважене значення варіант у степені k(k=1,2,…). Обчислюється за формулою

$$\nu_k^* = \frac{\sum_{i=1}^{n} x_i^k n_i}{n}.$$

При k=1дістанемо початковий момент першого порядку: $\nu_1^* = \frac{\sum_{i=1}^{n} x_i n_i}{n} = \overline{x_b}$.

При k=2обчислимо початковий момент другого порядку: $\nu_2^* = \frac{\sum_{i=1}^{n} x_i^2 n_i}{n}$.

Отже, дисперсію вибірки можна подати через початкові моменти першого та другого порядки, а саме: $D_B = v^*{}_2 - (v^*{}_1)^2$.

Центральним емпіричним моментом k-го порядку називають середнє зважене відхилення варіант у степені k (k=1,2,…). Обчислюється за формулою

$$\mu_k^* = \frac{\sum_{i=1}^{n} (x_i - \overline{x}_B)^k \cdot n_i}{n}.$$

При k=1дістанемо центральний момент першого порядку:

$$\mu_1^* = \frac{\sum_{i=1}^{n} (x_i - \overline{x}_B) \cdot n_i}{n} = \frac{\sum_{i=1}^{n} x_i \cdot n_i}{n} - \overline{x}_B \cdot \frac{\sum_{i=1}^{n} n_i}{n} = \overline{x}_B - \overline{x}_B = 0.$$

При k=2 маємо:

$$\mu_2^* = \frac{\sum_{i=1}^{n} (x_i - \overline{x}_B)^2 \cdot n_i}{n} = D_B.$$

Центральний емпіричний момент третього порядку застосовується для обчислення коефіцієнта асиметрії.

Коефіцієнтом асиметрії статистичного розподілу називається число, яке обчислюється за формулою:

$$A_s^* = \frac{\mu_3^*}{\sigma_B^3} \quad ,$$

де $\mu_3^* = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x}_B)^3 \cdot n_i = \nu_3^* - 3\nu_2^* \cdot \nu_1^* + 2(\nu_1^*)^2$.

При $A_s = 0$ варіанти статистичного розподілу вибірки симетрично розміщені відносно $\bar{x}_B$.

При $A_s < 0$ варіанти статистичного розподілу $x_i < \overline{x_B}$ переважають варіанти $x_i > \overline{x_B}$, таку асиметрію називають від'ємною.

При $A_s > 0$ варіанти статистичного розподілу $x_i > \overline{x_B}$ переважають варіанти $x_i < \overline{x_B}$, таку асиметрію називають додатною.

Центральний емпіричний момент четвертого порядку застосовується для обчислення ексцесу.

Ексцесом статистичного розподілу називається число, яке обчислюється за формулою

$$E_s^* = \frac{\mu_4^*}{\sigma_B^4} - 3 \quad ,$$

де $\mu_4^* = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_B)^4 \cdot n_i = \nu_4^* - 4\nu_3^* \cdot \nu_1^* + 6\nu_2^* \cdot (\nu_1^*)^2 - 3(\nu_1^*)^4$.

E_s^*, як правило, використовується при дослідженні неперервних ознак генеральних сукупностей, оскільки він оцінює крутизну закону розподілу неперервної випадкової величини порівняно з нормальним. Для нормального закону розподілу, як відомо, $E_s^* = 0$.Якщо $E_s^* > 0$ ($E_s^* < 0$), то статистичний розподіл має більш загострену (пологу) вершину порівняно із нормальною кривою.

Приклад. Для вибірки з прикладу (п.1.1.2.) обчислити $A^*_{s,}$ E_s^*.

Розв’язання. Знайдемо початкови момент першого, другого, третього та четвертого порядку:

$\nu_1^* = \frac{1}{n}\sum_{i=1}^{8} x_i n_i = \overline{x_b}$ =0,885;

$\nu_2^* = \frac{1}{n}\sum_{i=1}^{8} x_i^2 n_i = \frac{1}{40}(0{,}81^2 \cdot 3 + 0{,}84^2 \cdot 4 + 0{,}85^2 \cdot 6 + 0{,}87^2 \cdot 9 + 0{,}90^2 \cdot 8 + 0{,}94^2 \cdot 6 + 0{,}97^2 \cdot 2 +$
$+ 0{,}99^2 \cdot 2) = 0{,}7857$;

$\nu_3^* = \frac{1}{n}\sum_{i=1}^{8} x_i^3 n_i = \frac{1}{40}(0{,}81^3 \cdot 3 + 0{,}84^3 \cdot 4 + 0{,}85^3 \cdot 6 + 0{,}87^3 \cdot 9 + 0{,}90^3 \cdot 8 + 0{,}94^3 \cdot 6 + 0{,}97^3 \cdot 2 +$
$+ 0{,}99^3 \cdot 2) = 0{,}6998$;

$\nu_4^* = \frac{1}{n}\sum_{i=1}^{8} x_i^4 n_i = \frac{1}{40}(0{,}81^4 \cdot 3 + 0{,}84^4 \cdot 4 + 0{,}85^4 \cdot 6 + 0{,}87^4 \cdot 9 + 0{,}90^4 \cdot 8 + 0{,}94^4 \cdot 6 + 0{,}97^4 \cdot 2 +$
$+ 0{,}99^4 \cdot 2) = 0{,}6253$.

Знайдемо центральні емпенричні моменті третього та четвертого порядку:

$\mu_3^* = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_B)^3 \cdot n_i = \nu_3^* - 3\nu_2^* \cdot \nu_1^* + 2(\nu_1^*)^2 = 0{,}6998 - 3 \cdot 0{,}7857 \cdot 0{,}885 + 2 \cdot (0{,}885)^2 = 0{,}1803$.

$\mu_4^* = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_B)^4 \cdot n_i = \nu_4^* - 4\nu_3^* \cdot \nu_1^* + 6\nu_2^* \cdot (\nu_1^*)^2 - 3(\nu_1^*)^4$=0,6253-4·0,6998·0,885+6·0,7857·(0,885)^2 - 3·(0,885)^4=1,839.

Тоді $A_s^* = \frac{\mu_3^*}{\sigma_B^3} = \frac{0,1803}{(0,05)^3}$= 1442,4;

$E_s^* = \frac{\mu_4^*}{\sigma_B^4} - 3 = \frac{0,6253}{(0,05)^4} - 3$=100045.

1.2 Точкові оцінки параметрів розподілу

1.2.1 Класифікація оцінок

Нехай *X*— деяка генеральна сукупність (випадкова величина), а $x_1, x_2, ..., x_n$ -деяка вибіркова сукупність. Нехай, крім того, відомий характер закону розподілу генеральної сукупності *X*, але параметри цього розподілу $a_1, a_2, ..., a_n$ — невідомі. Позначимо $\theta = (a_1, a_2, ..., a_n)$— невідомий вектор параметрів.

Статистичною оцінкою θ^* невідомого параметра θ теоретичного розподілу називають функцію $f(x_1, x_2, ..., x_n)$ від випадкових величин $x_1, x_2, ..., x_n$ які спостерігаються.

Статистичні оцінки параметрів поділяються на *точкові* та *інтервальні.*

Точковою називають статистичну оцінку, яка визначається одним числом $\theta^* = f(x_1, x_2, ..., x_n)$, де $x_1, x_2, ..., x_n$ — результати *n* спостережень над кількісною ознакою *X*.

Інтервальна оцінка визначається двома числами — θ_1^* і θ_2^*, які є кінцями проміжку, що покриває параметр, який ми оцінюємо.

Для того, щоб статистичні оцінки давали кращі наближення невідомих параметрів, вони повинні задовольняти певні вимоги.

Точкова оцінка θ^* параметра θ є наближеним значенням параметраθ, тобто $\theta \approx \theta^*$. Як розуміти «наближене значення»? Потрібно ввести якусь міру відхилення точкової оцінки θ^* від параметра θ. Розглянемо похибку θ^*- θ, яка теж є випадковою величиною. Цю міру можна ввести, наприклад, за потреби, щоб математичне сподівання випадкової похибки дорівнювало 0.

Незміщеною називають точкову оцінку $\theta^* = \theta^*\left(\bar{X}\right)$параметра θ, якщо $M(\theta^*) = \theta$ або $M(\theta^* - \theta) = 0$ за будь-якого обсягу вибірки. У противному разі оцінку називають *зміщеною*, тобто коли $M(\theta^*) \neq \theta$. Різниця $\theta^* - \theta = \delta$ називається *зміщенням статистичної оцінки* θ^* .

Оцінювальний параметр може мати кілька точкових незміщених статистичних оцінок.

Точкова статистично оцінка називається *ефективною*, якщо при заданому обсязі вибірки вона має мінімальну дисперсію серед усіх інших оцінок. Ефективні оцінки мають особливо великий практичний інтерес, оскільки вони дають в середньому найменшу похибку.

Точкова статистична оцінка θ^* називається *консистенційною (ґрунтовною)*, якщо у разі необмеженого збільшення обсягу вибірки θ^* наближається до оцінювального параметра θ, а саме:

$$\lim_{n\to\infty} P(|\theta^* - \theta| < \varepsilon) = 1.$$

1.2.2 Властивості статистичних оцінок

Нехай розподіл дискретної генеральної сукупності X, з якої одержано вибірку $x_1, x_2, ..., x_n$ залежить від параметра θ, значення якого нам невідоме, і нехай $\theta^* = f(x_1, x_2, ..., x_n)$, де $f(x_1, x_2, ..., x_n)$ — деяка відома нам функція елементів даної вибірки (деяка статистика). Розглянемо, як за результатами дослідів дістати оцінки для математичного сподівання $M(X) = m$ та дисперсії $D(X) = D$.

Нехай з генеральної сукупності вибрана повторна вибірка обсягу n зі значеннями ознаки $x_1, x_2, ..., x_n$, які усі різні. Потрібно оцінити за даними вибірки невідому генеральну середню $\overline{x_Г}$. Оцінкою генеральної середньої $\overline{x_Г}$ (математичне сподівання) є вибіркова середня $\overline{x_в}$.

Ця оцінка є *незміщеною,* тобто $M(\overline{x_B}) = \overline{x_Г}$. Розглядатимемо $\overline{x_B}$ як випадкову величину і $x_1, x_2, ..., x_n$ незалежні, однаково розподілені випадкові величини, математичне сподівання яких $M(x_i) = m$ (отже, $\overline{x_Г} = m$). Оскільки

$$M(\overline{x_B}) = \frac{1}{n} M(\sum_{i=1}^{n} x_i) = \frac{1}{n}\sum_{i=1}^{n} M(x_i) = \frac{1}{n}\sum_{i=1}^{n} m = m = \overline{x_Г},$$

то оцінка $\overline{x_B}$ для $\overline{x_Г}$, є *незміщеною.*

Вибіркова середня є *консистенційною* оцінкою генеральної середньої, оскільки згідно з законом великих чисел при $n \to \infty$ вона збігається за ймовірністю до математичного сподівання $M(x_i)=m$ або, що те саме, до генеральної середньої $\overline{x_Г}$, (або $\overline{x_Г} = m$).

Вибіркова дисперсія дає занижені значення для дисперсії *D(X)* генеральної сукупності. Вона буде *зміщеною оцінкою D(X).* Але математичне сподівання D_B буде $M(D_B) = \frac{n-1}{n} D(X)$. Тому вибіркову дисперсію доцільно виправити таким чином, аби вона стала незміщеною оцінкою. Для цього достатньо D_B, помножити на дріб $\frac{n}{n-1}$.

Виправлену вибіркову дисперсію позначають

$$S^2 = \frac{n}{n-1} D_B = \frac{1}{n-1}\sum_{i=1}^{m} n_i (x_i - \overline{x_B})^2$$

Виправлена вибіркова дисперсія є *незсуненою оцінкою* генеральної дисперсії. *Виправленим середнім квадратичним відхиленням вибірки* буде $s = \sqrt{S^2}$.

Приклад. При проведенніли обстеження урожайності зернових культур отримали наступний розподіл:

Урожайність, x_i, ц/га	Кількість участків n_i, га
15-17	100
17-19	250
19-21	450
21-23	200
Всього	1000

Знайти ймовірність того, що середній урожай на усій території в 10000га відхиляється від середнього урожаю по вибірці не більш ніж на 0,1 ц, якщо вибірка повторна.

Розв´язання. Знайдемо середню вибіркову і середню дисперсію. Перейдемо від інтервального ряду до дискретного и складемо розрахункову таблицю.

с- значення x_i з найбільшою частотою, тоб то c=20. $u_i = \dfrac{x_i - c}{h}$, де h=2.

x_i	n_i	u_i	$u_i \cdot n_i$	$u_i^2 \cdot n_i$
16	100	-2	-200	400
18	250	-1	-250	250
20	450	0	0	0
22	200	1	200	200
	1000		-250	850

Обчислимо моменти першого і другого порядку:

$$\nu_1^* = \frac{\sum_{i=1}^{n} u_i \cdot n_i}{n} = -0{,}25; \quad \nu_2^* = \frac{\sum_{i=1}^{n} (u_i)^2 \cdot n_i}{n} = 0{,}85.$$

$$\overline{x_B} = \nu_1^* h + c = 19{,}5; \quad \sigma_B^2 = (\nu_2^* - (\nu_1^*)^2)h^2 = 3{,}15.$$

Тоді $\sigma^2_{\overline{x_B}} = \sqrt{\dfrac{3{,}15}{1000}} = \sqrt{0{,}00315} \approx 0{,}056$ ц,

$$P(\left|\overline{x_0} - 19{,}5\right| \le 0{,}1) \approx 2\varPhi\left(\frac{0{,}1}{0{,}056}\right) = 2\varPhi(1{,}79) = 0{,}9265.$$

1.2.3 Методи визначення точкових статистичних оцінок

Серед методів визначення точкових статистичних оцінок найчастіше використовують методи аналогій, максимальної правдоподібності і найменших квадратів.

Метод аналогій.К.Пірсон в 1894 році для точкової оцінки параметрів розподілу генеральної сукупності запропонував метод аналогій , суть якого полягає в тому,що ці оцінки шукаємо як розв’язок системи рівнянь, яку отримуємо в результаті прирівнювання такої кількості перших теоретичних та емпіричних моментів відповідальних порядків, скільки параметрів хочемо оцінити.

Оцінка одного параметра. Нехай задано щільність розподілу $f(x, \theta,)$, що визначається одним невідомим параметром θ. Потрібно знайти точкову оцінку параметра θ.

Відповідно до методу аналогій для оцінки невідомого параметра θ прирівнюють один теоретичний момент одному емпіричному моменту того самого порядку. Наприклад, можна прирівняти початковий теоретичний момент першого порядку початковому емпіричному моментові першого порядку: $\nu_1 = \nu_1^*$ Враховуючи, що $\nu_1 = M(X)$ і $\nu_1^* = \overline{x_B}$, дістанемо

$$M(X) = \overline{x_B}.$$

Математичне сподівання $M(X)$, як видно зі співвідношення

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

є функцією від невідомого параметра θ, тому, розв'язавши рівняння відносно параметра θ, отримаємо його точкову оцінку θ^*.

Оцінка двох параметрів. Нехай задано щільність розподілу $f(x, \theta_1, \theta_2)$, що визначається невідомими параметрами θ_1, θ_2.

Відповідно до методу аналогій прирівнюємо, наприклад, початковий теоретичний момент першого порядку початковому емпіричному моментові першого порядку і центральний теоретичний момент другого порядку центральному емпіричному моментові другого порядку:

$\nu_1 = \nu_1^*, \quad \mu_2 = \mu_2^*.$

Враховуючи, що $\nu_1 = M(X), \quad \mu_2 = D(X), \quad \nu_1^* = \overline{x_B}, \quad \mu_2^* = D_{B,}$, дістанемо

$$\left.\begin{aligned} M(X) &= \overline{x_B} \\ D(X) &= D_B \end{aligned}\right\}$$

Математичне сподівання та дисперсія є функціями параметрів θ_1 і θ_2. Тому, розв'язавши систему двох рівнянь з двома невідомими θ_1 і θ_2, отримаємо їх точкові оцінки θ_1^* і θ_2^*.

Приклад. Випадкова величина X розподілена по закону Пуассона з невідомим параметром λ. По емпіричному розподілу

x_i	0	1	2	3	4	5	6	7	8
n_i	186	176	85	32	13	3	2	2	1

методом аналогій знайти точкову оцінку параметра λ.

Розв´язання. Знайдемо початковий момент першого порядку:

$$\nu_1^* = \frac{\sum_{i=1}^{n} x_i^1 n_i}{n} = \frac{0\cdot186+1\cdot176+2\cdot85+3\cdot32+4\cdot13+5\cdot3+6\cdot2+7\cdot2+8\cdot1}{500} = 1{,}086$$

$\nu_1 = \lambda$.

Порівнюючи теоретичний і емпіричний моменти, одержимо: λ=1,086.

<u>Метод максимальної правдоподібності.</u> Цей метод розроблено Р.Фішером в 1912 році і посідає центральне місце в теорії статистичного оцінювання параметрів θ.

Розглянемо випадкову вибірку з щільністю розподілу ймовірностей $f(x,\theta)$. У разі реалізації вибірки з варіантами $x_1, x_2, \ldots, x_n$ щільність імовірностей вибірки буде такою: $f(x_1, x_2, \ldots, x_n, \theta^*)=f(x_1,\theta^*)\, f(x_2, \theta^*) \ldots \cdot f(x_n, \theta^*)$.

При цьому варіанти розглядаються як незалежні випадкові величини, котрі мають один і той самий закон розподілу, що й ознака генеральної сукупності X.

Суть методу полягає в тому, що фіксуючи значення варіант $x_1, x_2, \ldots, x_n$, визначають таке значення параметра θ^*, при якому функція максимізується. Вона називається *функцією максимальної правдоподібності* і позначається так: $L=L(\theta^*)$.

Для визначення статистичних оцінок θ_1^*, θ_2^*, …,θ_n^* складається система рівнянь (згідно з необхідною умовою екстремуму)
$$\begin{cases} \dfrac{\partial L}{\partial \theta_1^*} = 0, \\ \dfrac{\partial L}{\partial \theta_2^*} - 0, \\ \ldots, \\ \dfrac{\partial L}{\partial \theta_n^*} = 0. \end{cases}$$

Ті розв’язки системи, при яких функція $L(\theta^*)$ досягає максимуму, називають *оцінками максимальної правдоподібності*.

На практиці зручно від функції правдоподібності $L(\theta^*)$ перейти до її логарифма $\ln L(\theta^*)$, тоді система рівнянь буде така:
$$\begin{cases} \dfrac{\partial \ln L}{\partial \theta_1^*} = 0, \\ \dfrac{\partial \ln L}{\partial \theta_2^*} = 0, \\ \ldots, \\ \dfrac{\partial \ln L}{\partial \theta_n^*} = 0. \end{cases}$$

Оцінки θ_1^*, θ_2^* , …, θ_n^* є асимптотична ефективними, тобто при $n \to \infty$ прямують до ефективних оцінок.

Приклад. Випадкова величина X розподілена по показниковому закону з щільністю ймовірностей $f(x, \theta)=\theta\cdot e^{-\theta x}$. За результатом вибірки $x_1, x_2, \ldots, x_n$ і методом максимальної правдоподібності оценіти параметр θ.

Розв'язання. . Складемо функцію $L(\theta)$:

$$L(\theta)=\theta \cdot e^{-\theta\cdot x_1} \cdot \theta \cdot e^{-\theta\cdot x_2} \cdot \ldots \cdot \theta \cdot e^{-\theta\cdot x_n} = \theta^n \cdot e^{-\theta\cdot\sum_{i=1}^{n} x_i}.$$

Знайдемо $\ln L=n\cdot\ln\theta-\sum_{i=1}^{n} x_i$; $\dfrac{d\ln L(\theta)}{d\theta} = \dfrac{n}{\theta} - \sum_{i=1}^{n} x_i = 0$; $\dfrac{n}{\theta} = \sum_{i=1}^{n} x_i$;

$$\theta^* = \frac{1}{\overline{x}_B}.$$

Метод найменших квадратів. Суть методу полягає в оцінки параметрів θ_1^*, θ_2^*, …, θ_n^* таким чином, щоб сума квадратів відхилень варіант вибірки від емпіричних значень била найменшою. Це означає, що функція $R(\theta^*) = \sum_{i=1}^{n} (x_i - \theta^*)^2 \cdot n_i$ повинна бути мінімальною. Використовуючи умову екстремуму, дістанемо: $\dfrac{\partial R}{\partial \theta^*} = -2\sum_{i=1}^{n} (x_i - \theta^*) n_i = 0,\quad \sum_{i=1}^{n} x_i \cdot n_i - \sum_{i=1}^{n} n_i \cdot \theta^* = 0,$

$$\theta^* = \frac{1}{n}\sum_{i=1}^{n} x_i n_i = \overline{x_B}\ .$$

Звідси точковою статистичною оцінкою буде $\theta^* = \overline{x_B}$ - вибіркова середня.

1.3 Інтервальні оцінки параметрів розподілу

1.3.1 Основні статистичні розподіли

Найбільш важливим у статистичних дослідженнях є нормальний закон розподілу. Розглянемо декілько випадкових величин, розподілених за нормальним законом, яки частіше зустрічаються у математичної статистиці.

1. *Розподіл* χ^2_k (хі-квадрат з k ступенями свободи)

Якщо кожна із X_i (i=1, 2, …, n) незалежних випадкових величин характеризується нормальним законом розподілу ймовірностей $N(0,1)$, то випадкова величина матиме розподіл χ^2.

$$\chi^2 = \sum_{i=1}^{n} x_i^2$$

Числові характеристики будуть таки: $M(\chi^2) = n$, $D(\chi^2) = 2n$.

2. *Розподіл Стьюдента* t_n.

Нехай $x_0, x_1, x_2, \ldots, x_n$ - незалежні випадкові величини, тоді випадкова величина

$$t_n = \frac{x_0}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}}$$

має розподіл Стьюдента з n ступенями свободи: $M(t_n) = 0$, $D(t_n) = \dfrac{n}{n-2}$.

При $n \to \infty, P_n(x) \to \varphi(x)$ при $n>30$ розподіл Стьюдента можна замінити на стандартний нормальний розподіл.

1.3.2 Інтервальні оцінки параметрів розподілу

Раніше розглядали оцінки параметрів генеральної сукупності, яки визначались одним числом θ^* (так звані точкові оцінки). Якщо обсяг вибірки малий, то точкова оцінка параметра може значно відрізнятися від самого параметра. Тому у випадку невеликих вибірок застосовують так звану інтервальну оцінку. Статистична оцінка, яка визначається двома числами – θ_1^* і θ_2^* - кінцями інтервалу, який покриває оцінювальний параметр, називається *інтервальною*.

Різниця між статистичною оцінкою θ^* та її оцінювальним параметром θ, взята за абсолютним значенням, називається *точністю оцінки*, а саме:
$|\theta^* - \theta| < \delta$,
де δ є точністю оцінки.

Оскільки θ^* є випадковою величиною, то і δ буде випадковою, тому остання нерівність справджуватимется з певною ймовірністю:
$P(|\theta^* - \theta| < \delta) = \gamma$,
де γ називається *довірчою ймовірністю* (*надійністю*).

На практиці надійність оцінки задається наперед. Значення γ беруть близьким до одиниці. Як правило, вважають, що γ дорівнює 0,95, 0,99, 0,999.

Довірчим (*надійним*) називають інтервал $(\theta^* - \delta,\ \theta^* + \delta)$, що покриває невідомий параметр θ з довірчою ймовірністю (надійністю) γ.

Очевидно, що довірчий інтервал має випадкові кінці, оскільки за різними вибірками будуть визначатися різні значення θ^*. На практиці довжина довірчого інтервалу відіграє важливу роль. Чим менша довжина довірчого інтервалу, тим точніша оцінка параметра.

Довжина довірчого інтервалу 2δ визначається двома величинами: довірчою ймовірністю γ та обсягом вибірки n.

Щоб знайти довірчу ймовірність оцінки, потрібно знайти її закон розподілу. Але в багатьох випадках він виявляється близьким до нормального.

Розглянемо три випадки:

1. Нехай ознака X генеральної сукупності має нормальній закон розподілу і відоме середнє квадратичне відхилення генеральної сукупності $\sigma_Г$, із заданою надійністю γ. Побудуємо довірчий інтервал для $\overline{X}_Г$. Оскільки $\overline{x}_В$ як точкова незміщена статистична оцінка для $\overline{X}_Г = M(X)$ має нормальний закон розподілу з числовими характеристиками $M(\overline{x}_в) = \overline{X}_Г = a$, $\sigma(\overline{x}_в) = \dfrac{\sigma_Г}{\sqrt{n}}$, то скориставшись формулою $P(\theta^* - \delta < \theta < \theta^* + \delta) = \gamma$, дістанемо
$P(|\overline{x}_В - a| < \delta) = \gamma$.

Випадкова величина $\overline{x}_в - a$ має нормальний закон розподілу з числовими характеристиками $M(\overline{x}_В - a) = M(\overline{x}_В) - a = a - a = 0$; $D(\overline{x}_В - a) = D(\overline{x}_В) = \dfrac{D_г}{n}$; $\sigma(\overline{x}_в) = \dfrac{\sigma_Г}{\sqrt{n}}$.

Тому $\dfrac{\overline{x_В} - a}{\frac{\sigma_г}{\sqrt{n}}}$ матиме нормований нормальний закон розподілу $N(0;1)$.

Назначивши $\dfrac{\delta}{\frac{\sigma_г}{\sqrt{n}}} = x$, тоді $P\left(\left|\dfrac{\overline{x_В} - a}{\frac{\sigma_г}{\sqrt{n}}}\right| < x\right) = \gamma$ або $P\left(\overline{x_В} - \dfrac{x \cdot \sigma_г}{\sqrt{n}} < a < \overline{x_В} + \dfrac{x \cdot \sigma_г}{\sqrt{n}}\right) = \gamma$.

Використовує формулу $P(|\overline{x}_В - a| < \delta) = 2\Phi\left(\dfrac{\delta}{\sigma}\right)$, отримуємо

$$P\left(\left|\frac{\overline{x_B}-a}{\frac{\sigma_r}{\sqrt{n}}}\right|<x\right)=2\Phi(x)=\gamma .$$

Знаходимо аргумент x, а саме: $2\Phi(x)=\gamma$, $\Phi(x)=0{,}5\gamma$. Аргумент x знаходимо за значенням функції Лапласа, яка дорівнює $0{,}5\gamma$ за таблицею (доданок 2).

Отже, довірчий інтервал дорівнюватиме:

$$\overline{x_B}-\frac{x\cdot\sigma_r}{\sqrt{n}}<a<\overline{x_B}+\frac{x\cdot\sigma_r}{\sqrt{n}} .$$

Величина $\frac{x\cdot\sigma_r}{\sqrt{n}}$ називається *точністю оцінки,або похибкою вибірки*.

Приклад. Машина, яка фасує цукор у пакети, тривалий час забезпечувала нормальний розподіл маси із стандартним відхиленням σ = 2,25 *г*. Після переходу на новий вид пакетів для контролю навмання відібрали 25 нових пакетів, в яких середня маса виявилась $\overline{x}_B$=1002 г. З надійністю γ= 0,95 знайти надійний інтервал для середньої маси пакета цукру в генеральної сукупності.

Розвязання**.** Із рівняння $\Phi(x)=\frac{0{,}95}{2}=0{,}475$ за таблицею значень функції Лапласа визначаємо x=1,96; звідки $\delta=\frac{x\sigma}{\sqrt{n}}=\frac{1.96\cdot 2.25}{\sqrt{25}}=0{,}882$.

Маємо надійний інтервал для генеральної середньої μ:

$1002 - 0{,}882 < \mu < 1002 + 0{,}882$ або $1001{,}118 < \mu < 1002{,}882$.

При відомих γ і δ мінімальний обсяг вибірки, який забезпечує задану точність δ, визначається формулою $n_{\min}=\left(\frac{x\sigma}{\delta}\right)^2$

2. Побудуємо довірчий інтервал для $\overline{X}_r=a$, при невідомому значенні середнього квадратичного відхилення генеральної сукупності $\sigma_г$. Для оцінки обираємо випадкову величину $t=\frac{\overline{x_B}-a}{\frac{S}{\sqrt{n}}}$, що має розподіл Стьюдента з $k=n-1$ступенями свободи, тоді $P\left(\overline{x_B}-\frac{t_\gamma\cdot S}{\sqrt{n}}<a<\overline{x_B}+\frac{t_\gamma\cdot S}{\sqrt{n}}\right)=\gamma$.

Довірчий інтервал для оцінки а буде таким:

$$\overline{x_B}-\frac{t_\gamma\cdot S}{\sqrt{n}}<a<\overline{x_B}+\frac{t_\gamma\cdot S}{\sqrt{n}} .$$

Параметри $\overline{x_B}, S, \sqrt{n}$ знаходяться з даної вибірці, t_γ - за розподілом Стьюдента (доданок 3). При великих вибірках (n>30) розподіл Стьюдента наближається до нормального, і тоді можна користуватися теоремами о нормальному розподілу.

Приклад. Ознака X генеральної сукупності розподілена нормально. За

вибіркою обсягу n =16 знайдено вибіркову середню $x_В$ = 20,2 і виправлене середнє квадратичне відхилення S = 0,8. Оцінити математичне сподівання випадкової величини X за допомогою надійного інтервалу з надійністю γ=0,95.
Розв'язання. З таблиці розподілу Стьюдента визначаємо $t_\gamma = t((0{,}95;16-1) = 2{,}13$.

Обчислюємо $\frac{t_\gamma \cdot S}{\sqrt{n}} = \frac{2.13 \cdot 0{,}3}{\sqrt{16}} = 0{,}426$ і маємо надійний інтервал

$20{,}2 - 0{,}426 < \mu < 20{,}2 + 0{,}426$ або $19{,}774 < \mu < 20{,}626$.

3. Нехай ознака X генеральної сукупності має нормальній закон розподілу с невідомими параметрами a і $\sigma_Г$. Для знаходження довірчого інтервалу застосовуємо випадкову величину

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

що має розподіл χ^2 із $k=n-1$ступенями свободи. Тоді

$$P(|\sigma_Г - s| < \delta) = \gamma, \quad s - \delta < \sigma_Г < s + \delta, \quad \frac{\delta}{s} = q,$$

Отримуємо довірчий інтервал:

$$\left(1 - \frac{\delta}{s}\right)s < \sigma_г < \left(1 + \frac{\delta}{s}\right)s \text{ або } s(1-q) < \sigma_Г < s(1+q).$$

Величина *знаходиться* за таблицею (додаток 5) по відомим n і γ.
Приклад. Кількісна ознака X генеральної сукупності розподілена нормально. За вибіркою об'єму n=25 знайдено виправлене середнє квадратичне відхилення S=0,8. Знайти надійний інтервал, який покриває генеральне середнє квадратичне відхилення $\sigma_Г$ з надійністю γ = 0,95.
Розв'язання. Із таблиці (додаток 4) знаходимо
$q(\gamma;k) = q(0{,}95;25-1) = q(0{,}95;24) = 0{,}32$.

За формулою $\left(1 - \frac{\delta}{s}\right)s < \sigma_г < \left(1 + \frac{\delta}{s}\right)s$ записуємо надійний інтервал:

$0{,}8(1-0{,}32) < \sigma_Г < 0{,}8(1+0{,}32)$ або $0{,}544 < \sigma_Г < 1{,}056$.

1.4 Перевірка статистичних гіпотез

1.4.1 Статистична гіпотеза. Основні поняття

Інформація, яку дістають на підставі вибірки, реалізованої із генеральної сукупності, може бути використована для формулювання певних суджень про всю генеральну сукупність. Таки судження про вигляд невідомого розподілу генеральної сукупності або про параметри відомих розподілів називають *статистичними гіпотезами*.

Наприклад, статистичними будуть гіпотези:

- генеральна сукупність розподілена за законом Пуассона;
- дисперсії двох нормальних сукупностей рівні між собою.

У першій гіпотези зроблено припущення про вигляд невідомого розподілу, у другій – про параметри двох відомих розподілів. Гіпотеза «на міжнародній олімпіаді переможе команда з України» не є статистичною, оскільки в неї не мовиться ні про вигляд, ні про параметри розподілу.

Разом з висунутою гіпотезою завжди можна розглядати протилежну її гіпотезу. Якщо висунута гіпотеза буде відхилена, то має місце протилежна гіпотеза. Тому ці гіпотез доцільно розрізняти.

Статистичні гіпотези про значення параметрів ознак генеральної сукупності називаються *параметричними*.

Статистичні гіпотези, що висуваються на підставі обробки вибірки про закон розподілу ознаки генеральної сукупності, називаються *непараметричними*.

Нульовою (*основною*) називають висунуту гіпотезу H_0.

Альтернативною (*конкурентною*) називають гіпотезу H_a, що суперечить нульовій гіпотезі. Кожній нульовій гіпотезі можна протиставити кілька альтернативних гіпотез.

Простою називають гіпотезу, якщо вона містить лише одне припущення. Наприклад, якщо λ-параметр показникового розподілу, то гіпотеза H_0: λ=2 - проста.

Складною називають гіпотезу, якщо вона складається зі скінченної або нескінченної кількості простих гіпотез. Наприклад, складна гіпотеза H: $\lambda>2$ складається з нескінченної множини простих гіпотез H_n: $\lambda=3$, $\lambda=4$, $\lambda=5$,… .

Висунута нульова гіпотеза може бути правильною або неправильною, тому виникає необхідність її перевірки. Оскільки перевірку гіпотези здійснюють статистичними методами, її називають *статистичною*. В результаті статистичної перевірки нульової гіпотези у двох випадках можна зробити хибний висновок, тобто можуть бути допущені помилки першого та другого роду.

Помилка першого роду полягає в тому, що нульова гіпотеза відхиляється, хоча насправді вона має місце.

Помилка другого роду полягає в тому, що нульова гіпотеза приймається, хоча вона не є істинною.

Ймовірність α припустити *помилку першого роду*, тобто $\alpha=P(H_1/H_0)$, називається *рівнем значущості*.

Величина α задається наперед і за звичай приймається рівною 0,001; 0,005; 0,01; 0,05.

Ймовірність припустити помилку другого роду позначають через β : $\beta=P(H_1/H_0)$.

Ймовірність $1-\beta=P(H_1/H_1)$ не припустити *помилку другого роду*, тобто відхилити гіпотезу H_0 , коли вона невірна, називається *потужністю критерію*.

Очевидно, що при заданому рівні значущості α , потужність бажано мати найбільшою, що гарантує найменшу ймовірність помилки другого роду.

1.4.2 Алгоритм перевірки статистичної гіпотези

Для перевірки нульової гіпотези використовують спеціально підібрану випадкову величину, точний або наближений розподіл якої невідомий.

Статистичним критерієм називають випадкову величину *K*, яка застосовується для перевірки нульової гіпотези. Якщо статистична характеристика розподілена нормально, то критерій позначають *U* або *Z*; якщо вона має розподіл Фішера-Снедокора, то її позначають *F*, за законом Стьюдента – *T*, за законом «хі-квадрат» - χ^2.

Наприклад, якщо перевіряють гіпотезу про рівність дисперсій двох нормальних генеральних сукупностей (H_0: D(x)=D(y)), то за критерій вибирають відношення виправлених вибіркових дисперсій: $F=\frac{S_1^2}{S_2^2}$.

Ця величина випадкова, бо в різних дослідах дисперсії будуть набувати різних, наперед невідомих значень; вона розподілена за законом Фішера-Снедекора. Для перевірки нульової гіпотези за даними вибірок обчислюють часткові значення величин у складі критерію, і таким чином отримують спостережуване значення критерію.

Спостережуваним значення $K_{сп}$ називають те значення критерію, яке обчислене за даними вибірок.

Наприклад, якщо за даними вибірок із двох нормальних генеральних сукупностей знайдено виправлені вибіркові дисперсії $S_1^2=20$ та $S_2^2=5$, то спостережуване значення критерію дорівнює:

$$F_{сп}=\frac{S_1^2}{S_2^2}=\frac{20}{5}=4.$$

Після обрання певного критерію *K* множину всіх можливих його значень поділяють на дві неперервні підмножини: одна з них містить значення критерію, при яких нульова гіпотеза відхиляється, а друга - при яких вона приймається.

Критичною областю називають сукупність значень критерію, при яких нульову гіпотезу відхиляють.

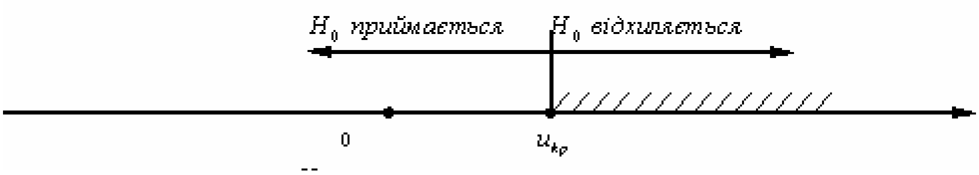
Областю прийняття гіпотези (областю допустимих значень) називають сукупність значень критерію, при яких нульову гіпотезу приймають.

Основний принцип перевірки статистичних гіпотез такий: якщо спостережуване значення критерію належить критичній області, нульову гіпотезу відхиляють; якщо спостережуване значення критерію належить області прийняття гіпотези, нульову гіпотезу приймають. Оскільки критерій *K* – одномірна випадкова величина, всі її можливі значення належать деякому інтервалу. Тому критична область та область прийняття гіпотези також є інтервалами і, отже, існують точки, які їх відокремлюють.

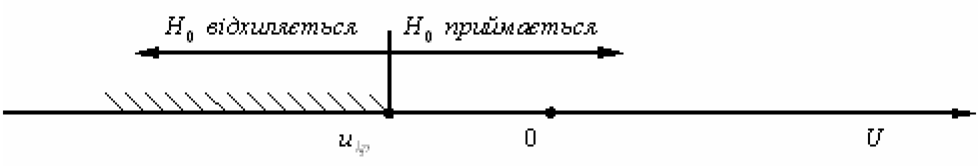
Критичними точками (границями) $u_{кр}$ називають точки, які відокремлюють критичну область від області прийняття гіпотези.

Розрізняють однобічну (правобічну або лівобічну) та двобічну критичні області.

Правобічною називають критичну область, що визначається нерівністю $K>u_{кр}$, де $u_{кр}$ - додатне число.

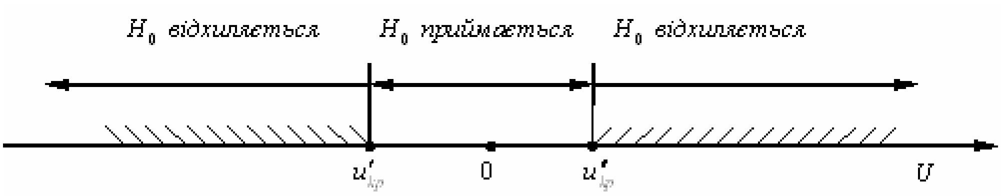


Лівобічною називають критичну область, що визначається нерівністю $K<u_{кр}$, де $u_{кр}$ - від'ємне число.



Однобічною називають критичну область, яка є правобічною або лівобічною.

Двобічною називають критичну область, яка визначається нерівностями $K> u'_{кр}$, $K< u''_{кр}$, де $u''_{кр} > u'_{кр}$. Якщо критичні точки симетричні відносно нуля, то двобічна критична область визначається нерівностями ($u_{кр}$ > 0): $K< -u_{кр}$, $K>u_{кр}$ або рівносильною нерівністю $|K| > u_{кр}$.



Для знаходження критичної області задають достатньо малу ймовірність помилки першого роду – рівень значущості α і шукають критичні точки:

а) для правобічної критичної області $P(K> u_{кр})=\alpha$ ($u_{кр} >0$);

б) для лівобічної критичної області $P(K< u_{кр})=\alpha$ ($u_{кр} < 0$);

в) для двобічної симетричної області $P(K> u_{кр})=\frac{\alpha}{2}$ ($u_{кр} >0$), $P(K< -u_{кр})=\frac{\alpha}{2}$.

Для кожного критерію є відповідні таблиці, що дають змогу знайти критичну точку $u_{кр}$, яка задовольняє потрібну умову (додаток 6).

Коли критична точка вже знайдена, обчислюють за даними вибірок спостережуване значення критерію і, якщо, $K_{сп} > K$, нульову гіпотезу відхиляють; якщо ж $K_{сп} < K$ – нульову гіпотезу приймають.

Ймовірність здійснити помилку першого роду обчислюють так:

$$P(\overline{x_B} > K_{кр}) = \int_{K_{кр}}^{\infty} f(x;a)dx = \alpha .$$

Ймовірність здійснити помилку другого роду позначають β.

$$\int_{-\infty}^{K_{кр}} f(x;в)dx = \beta .$$

Різницю π =1-β називають імовірністю обгрунтованного відхилення H_0, або *потужністю критерію*. Іншими словами, потужність критерію – це ймовірність того, що нульову гіпотезу буде відхилено, у той час як буде правильною альтернативна гіпотеза. Якщо рівень значущості α вже обрано, то критичну область доцільно будувати так, щоб потужність критерію була максимальною. Виконання цієї вимоги забезпечує мінімальну ймовірність помилки другого роду. Єдиним способом одночасного зменшення ймовірностей помилок першого та другого роду є збільшення обсягу вибірки.

Алгоритм перевірки правильності нульової статистичної гіпотези H_0:

1. Сформулювати H_0 й одночасно альтернативну гіпотезу H_a.
2. Вибрати статистичний критерій, який відповідав би сформульованій нульовій гіпотезі.
3. Залежно від змісту нульової та альтернативної гіпотез будується правобічна, лівобічна або двобічна критична область, а саме: нехай H_0: $\overline{x_r} = a$, тоді, якщо
H_0: $\overline{x_r} > a$, то вибирається правобічна критична область, якщо
H_0: $\overline{x_r} < a$, то вибирається лівобічна критична область, якщо
H_0: $\overline{x_r} \neq a$, то вибирається двобічна критична область.
4. Для побудови критичної області (лівобічної, правобічної чи двобічної) необхідно знайти критичні точки. За вибраним статистичним критерієм та рівнем значущості α знаходяться критичні точки.
5. За результатами вибірки обчислюється спостережуване значення критерію $K_{сп}^*$.
6. Відхиляють чи приймають нульову гіпотезу на підставі таких міркувань:

у разі, коли $K^* \in \overline{A}$, це є малоймовірною випадковою подією, Р($K^* \in \overline{A}$)=α і, незважаючи на це, вона відбулася, то в цьому разі H_0 відхиляється:
для лівобічної критичної області $P(K_{сп}^* < K) = \alpha$;
для правобічної критичної області $P(K_{сп}^* > K) = \alpha$;
для двобічної критичної області $P(K_{сп}^* < K') + P(K_{сп}^* > K'') = \alpha$
або $P(K_{сп}^* < K') = P(K_{сп}^* > K'') = \frac{\alpha}{2}$, ураховуючи ту обставину, що критичні точки $u'_{кр}$ і $u''_{кр}$ симетрично розташовані відносно нуля.

1.4.3 Перевірка гіпотези про значення ймовірності (частку ознаки) в генеральній сукупності

Нехай задано генеральну сукупність з ознакою X . Зроблено вибірку обсягом n. При заданому рівні значущості α необхідно перевірити достовірність гіпотези $H_0 : p = p_0$ – ймовірність p появи ознаки X в генеральній сукупності дорівнює заданій величині p_0 . Розглянемо випадкову величину $W = \frac{m}{n}$, яка має біномний розподіл. При великих $n (n > 30)$ можна наближено вважати нормальним закон розподілу величини W з математичним сподіванням $M(W)=p_0$, дисперсією $D(W)=\frac{p_0 \cdot q_0}{n}$ і середнім квадратичним відхиленням

$\sigma(W)=\sqrt{\dfrac{p_0\cdot q_0}{n}}$, де q = 1- p .За статистичний критерій приймемо нормовану нормально розподілену випадкову величину $Z \sim N(0;1)$:

$$Z=\frac{W-p_0}{\sigma(W)}=\frac{W-p_0}{\sqrt{p_0q_0}}\sqrt{n}\,.$$

Критична область будується в залежності від змісту альтернативної гіпотези H_1 аналогічно тому, як при перевірці гіпотези про значення генеральної середньої.

Критерії перевірки гіпотез про числові значення параметрів нормального закону розподілу наведено в таблиці

H_0	Припущення	Статистичний критерій	H_1	Критична точка	Область прийняття H_0
M(X)=a	σ - відоме	$Z=\dfrac{\overline{x_B}-\mu}{\sigma/\sqrt{n}}$	M(X)>μ M(X)<μ M(X)≠μ	$\Phi(z_{кр})=\frac{1-2\alpha}{2}$ $\Phi(z_{кр})=-\frac{1-2\alpha}{2}$ $\Phi(z_{кр})=\frac{1-\alpha}{2}$	$z< z_{кр}$ $z> - z_{кр}$ $\lvert z\rvert < z_{кр}$
	σ - невідоме	$t=\dfrac{\overline{x_B}-\mu}{s/\sqrt{n}}$	M(X)>μ M(X)<μ M(X)≠μ	$t_{кр}=t(\alpha;n-1)$ $t_{кр}= - t(\alpha;n-1)$ $t_{кр}=t(\frac{\alpha}{2};n-1)$	$t < t_{кр}$ $t > - t_{кр}$ $\lvert t\rvert < t$
$p=p_0$	n>30 $np_0>5$ $nq_0>5$	$Z=\dfrac{w-p_0}{\sqrt{p_0q_0/n}}$ $w=\frac{m}{n},q_0=1-p$	$p>p_0$ $p<p_0$ $p\neq p_0$	$\Phi(z_{кр})=\frac{1-2\alpha}{2}$ $\Phi(z_{кр})=-\frac{1-2\alpha}{2}$ $\Phi(z_{кр})=\frac{1-\alpha}{2}$	$z< z_{кр}$ $z> - z_{кр}$ $\lvert z\rvert < z_{кр}$

Приклад. При виробництві продукції харчування можлива поява неякісної. Ведеться контроль якості, щоб частка неякісної продукції булла менша, ніж *3%*. Серед відібраних *200* зразків продукції *4* виявились неякісними. Чи є підстави вважати, що виробничий процес вийшов з-під контролю і виробляється багато неякісної продукції?

Розв'язання. Перевіримо гіпотезу H_0: p = 0,03– частка неякісної продукції становить *3%* (p_0 =0,03) , при альтернативній гіпотезі H_1: p > 0,03 – зросла частка неякісної продукції.

За змістом альтернативної гіпотези будуємо правобічну критичну область. Для цього знаходимо z_{kp} з рівняння $\Phi(z_{кр})=\frac{1-2\alpha}{2}$ при вибраному рівні значущості α= 0,01: $\Phi(z_{кр})=\frac{1-2\cdot 0{,}01}{2}$=0,49; $z_{кр}$= 2,33.

Обчислюємо спостережуване значення критерію при m=4; n=200; $W=\frac{m}{n}$=0,02; p_0= 0,03; q_0= 0,97.

$$z^*=\frac{0{,}02-0{,}03}{\sqrt{0{,}03\cdot 0{,}97}}\sqrt{200}=-\frac{0{,}01}{\sqrt{0{,}03\cdot 0{,}97}}\cdot 10\sqrt{2}\approx -0{,}83$$

Отже, $z^* < z_{кр}$ і гіпотеза H_0 приймається.

Тому немає підстав вважати, що виробничий процес вийшов з-під контролю і дає більше, ніж *3%* неякісної продукції.

1.4.4 Порівняння двох дисперсій нормальних генеральних сукупностей

Порівняння вибіркових дисперсій дає можливість визначити, чи можна вважати вибіркові дисперсії статистичними оцінками однієї і тієї самої дисперсії генеральної сукупності.

Нехай генеральні сукупності *X* і *У* розподілені нормально і перша вибірка здійснена з генеральної сукупності з ознакою *X*, дисперсія якої дорівнює *D(X)*, а друга – з генеральної сукупності з ознакою *У,* дисперсія якої дорівнює *D(У)*. За незалежними вибірками обсягів n_1 і n_2 знайдено виправлені дисперсії S_x^2 і S_y^2, які потрібно порівняти:

а) для того, щоб при заданому рівні значущості α перевірити нульову гіпотезу H_0: D(*X*)=D(*У*) про рівність генеральних дисперсій нормальних сукупностей при альтернативній гіпотезі H_1: D(*X*)>D(*У*), потрібно обчислити спостережуване значення критерію (відношення більшої виправленої дисперсії до меншої) $F_{сп}=\frac{S_Б^2}{S_М^2}$ і за таблицею критичних точок розподілу Фішера-Снедекора, за заданим рівнем значущості α і ступенями вільності $k_1=n_1-1$, $k_2=n_2-1$ ($к_1$- ступінь вільності більшої виправленої дисперсії) знайти критичну точку $F_{кр}(\alpha, к_1, к_2)$.

Якщо $K_{сп}<K$, то гіпотеза H_0 приймається.

Якщо $K_{сп}>K$, то гіпотезу H_0 відхиляють;

б) для того, щоб при заданому рівні значущості α перевірити нульову гіпотезу про рівність генеральних дисперсій нормальних сукупностей при альтернативній гіпотезі H_1: $D(X)\neq D(У)$, потрібно обчислити відношення більшої виправленої дисперсії до меншої, $F_{сп}=\frac{S_Б^2}{S_М^2}$, і за таблицею критичних точок розподілу Фішера-Снедекора, за заданим рівнем значущості $\frac{\alpha}{2}$ (вдвічі менше заданого) і ступенями вільності k_1, k_2 (k_1- ступінь вільності більшої дисперсії) знайти критичну точку $F_{кр}(\frac{\alpha}{2}, к_1, к_2)$.

Якщо $K_{сп}<K$, то гіпотеза H_0 приймається.

Якщо $K_{сп}>K$, то гіпотезу H_0 відхиляють.

Приклад. Досліджуються два прилади, яки вимірюють будь-яку величину. За допомогою першого приладу виконано n_1 =13 вимірювань, отримано

S_1^2=0,72; другого – n_2=18, S_2^2=0,20. При α=0,01 перевірити гіпотезу H_0: *D(X)=D(У)* при H_1: *D(X)>D(У)* (другий прилад удосконалено).
Розв'язання.

Розглянемо такі гіпотези: H_0: *D(X)=D(У)* і H_1: *D(X)>D(У).* З рівняння $F_{сп} = \frac{S_Б^2}{S_M^2} = \frac{0{,}72}{0{,}20} = 3{,}6$. Знайдемо за таблицею (додаток) критичну точку $F_{кр}$(α=0,01, k_1=13-1=12, k_2=18-1=17)=3,45.
Оскільки $F_{сп}>F_{кр}$, то гіпотеза H_0 відхиляється, гіпотеза H_1 приймається, отже другий прилад удосконалено.

1.4.5 Статистична перевірка непараметричних гіпотез. Критерій узгодженості Пірсона

Одної з главных задач математической статистики является установление істинного закона распределения случайной величины на основании экспериментальных данных. Однако, как бы мы не выбирали вид закона распределения и его параметры, полной уверенности в том, что в результате получится истинный закон распределения , к которому принадлежит имеющаяся выборка, не существует. Поэтому, может ставиться вопрос лишь о том, что на определенном уровне доверия закон согласуется с данными выборки. Тому, якщо закон розподілу генеральної сукупності невідомий, але є підстави припустити, що він має певний вигляд (назвемо його *A*), то перевіряють нульову гіпотезу H_0: генеральна сукупність розподілена за законом *A*.

Перевірка гіпотези про передбачуваний закон невідомого розподілу здійснюється так само, як і перевірка гіпотези про параметри розподілу, тобто за допомогою спеціально підібраної випадкової величини – критерію.

Критерієм узгодженості називають критерій, за допомогою якого перевіряють справедливость гіпотези H_0. про передбачуваний закон невідомого розподілу генеральної сукупності. Критерій узгодженості це особливо підібрана випадкова величина з відомим законом розподілу.

Найбільш використовуемим є критерії узгодженості: χ^2 («хі-квадрат») Пірсона, Колмогорова, Романовського, Смірнова та інших.

Критерій узгодженості Пірсона. Він застосовується для перевірки гіпотези про нормальний розподіл генеральної сукупності. Нехай вибірка x_1, x_2, ..., x_m обсягу *n* має теоретичні частоти: $n_1, n_2, ..., n_m$.

Потрібно з рівнем значущості α перевірити основну гіпотезу H_0 : генеральна сукупність розподілена нормально.
Критерієм перевірки цієї гіпотези беруть випадкову величину

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n_i^{'})^2}{n_i^{'}},$$

де $n_i^{'}$ - емпіричні частоти, обчислені за формулою $n_i^{'} = np_i$.

Кількість ступенів вільності знаходять з рівності $k=m-1-r$, де m – кількість варіант вибірки або часткових інтервалів варіант; r – кількість параметрів передбачуваного розподілу. Наприклад, нормальний розподіл має два параметри (математичне сподівання та середньоквадратичне відхилення), тому $r=2$, $к=s-1-2=s-3$.Величина χ^2 – випадкова, оскільки в різних дослідах вона набуватиме різних наперед невідомих значень. Чим більше узгоджуються емпіричні та теоретичні розподіли, тим менше будуть розрізнятись емпіричні та теоретичні частоти, тим менше буде величина χ^2. Тому можна зробити висновок, що критерій χ^2 певною мірою характеризує близькість емпіричного та теоретичного розподілів.

Критичне значення випадкової величини χ^2 залежить від рівня значущості α та ступенів вільності її розподілу $к$: $\chi^2_{кр} = \chi^2(\alpha, к)$.

Правило Пірсона. Щоб при заданому рівні значущості α перевірити нульову гіпотезу H_0 про нормальний розподіл генеральної сукупності, потрібно:
1) обчислити теоретичні частоти $n_к'$ для варіант вибірки;
2) обчислити спостережуване значення критерію χ^2 за формулою

$$\chi^2 = \sum_{к=1}^{m} \frac{(n_к - n_к')^2}{n_к'};$$

3) знайти ступінь вільності χ^2 ;
4) знайти з таблиці (доданок Д) критичну точку $\chi^2_{кр}$, яка відповідає заданому рівню значущості α та ступеню вільності $к$;
5) порівняти $\chi^2_{св}$ та $\chi^2_{кр}$ і зробити висновок: якщо $\chi^2_{св} < \chi^2_{кр}$, то гіпотезу H_0 треба прийняти; якщо $\chi^2_{св} > \chi^2_{кр}$, то гіпотезу H_0 треба відхилити.

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

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

Розв'язання. Значення $\overline{x_B}$, $\sigma_в$ обчислюємо за дискретним статистичним розподілом.

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

$\overline{x_B} = \frac{1}{n}\sum_{i=1}^{5} x_i \cdot n_i$ = (85·6+95·10+105·50+115·25+125·9)/100=107,1;

$D_B = \frac{1}{n}\sum_{i=1}^{5} x_i^2 \cdot n_i - (\overline{x_B})^2 = \frac{1}{100}(85^2 \cdot 6+95^2 \cdot 10+105^2 \cdot 50+115^2 \cdot 20+125^2 \cdot 9)-(107,1)^2 =$ 90,59;

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

Обчислюємо теоретичні частоти: $u_i = \dfrac{x_i - \overline{x_B}}{\sigma_B}$; Φ($u_i$)-функцію Лапласа, яку визначаємо з таблиці (додаток Б); m_i=n·(Ф(u_{i-1})- Ф(u_i)) і результати запишемо у таблицю:

x_{i-1}	x_i	n_i	u_{i-1}	u_i	$\Phi(u_{i-1})$	$\Phi(u_i)$	$\Phi(u_{i-1})$ - $\Phi(u_i)$	m_i
80	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 α = 0,05 і числом степенів свободи k = 5 - 2 -1 = 3 знаходимо з таблиці (додаток Д) критичну точку $\chi^2_{кр} = \chi^2$ (0,05;2) = 6.

Обчислюємо спостережуване значення критерію

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

Оскільки $\chi^2 > \chi^2_{кр}$, то гіпотезу H_0 відхиляємо; немає підстав вважати що ознака X має нормальний закон розподілу в генеральній сукупності.

Частина 2
МЕТОДИЧНІ ВКАЗІВКИ ДО ПРАКТИЧНИХ ЗАНЯТЬ

2.1 Статистичні розподіли вибірок та їх числові характеристики

Теоретичні запитання до теми

1. Дати визначення генеральної сукупності..
2. Яка сукупність називається вибірковою?
3. Яка вибірка називається репрезентативною?
4. Що називається варіантою? Який ряд називається варіаційним?
5. Дискретним статистичним розподілом вибірки називається…
6. Який статистичний розподіл вибірки називається інтервальним?
7. Як можна визначити кількість інтервалів для неперервного статистичного розподілу вибірки?
8. Що називається емпіричною функцією розподілу? Властивості *F*(x)*.
9. Що таке полігон частот, відносних частот?
10. Що називають гістограмою частот, відносних частот?
11. Як визначається вибіркова середня величина $\overline{x_B}$?
12. Як визначається *Me* для інтервального статистичного розподілу?
13. Як визначається *Mo* для інтервального статистичного розподілу?
14. Що називається гістограмою частот і відносних частот?

15. Якому інтервалу належать значення емпіричної функції?
16. Емпіричний кореляційний момент та його властивості
17. Як визначається і що характеризує коефіцієнт асиметрії статистичного розподілу?
18. Як обчислюється коефіцієнт ексцесу статистичного розподілу і що він характеризує?

Завдання для аудиторної роботи

1. Записати вибірку 10, 12, 3, 4, 10, 5, 8, 9, 11, 8, 10, 7, 8, 5, 9, 11, 5, 3, 1, 8, 2, 9, 8, 7, 10, 9, 2, 8, 6, 2, 9, 3, 9, 6, 8, 7, 4, 6, 5, 7 у вигляді: а) варіаційного; б) статистичного ряду.

2. Через кожну годину вимірювалась напруга в електромережі. Результати вимірювання напруги у вольтах наведені у вигляді статистичного ряду:
222, 220, 219, 216, 224, 221, 220, 216, 218, 211, 217, 219,221, 220, 220, 221, 215, 222, 218, 218, 223, 221, 225, 219, 220, 216, 222, 217, 220, 215.

Потрібно:

1.) побудувати дискретний статистичний розподіл для цієї вибірки, полігон частот, гістограму відносних частот і $F^*(x)$;.

2.) обчислити $\overline{x_B}$, D_B, σ_B, R, V;

3.) знайти Me^*, Mo^*.

3. Відсоток виконання плану підприємства за рік та кількість підприємств, що виконують цей план, наведено у вигляді інтервального статистичного розподілу:

x_i, % h=10%	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100	100-110	110-120
n_i	2	6	13	16	25	12	10	8	5	3	1

Потрібно:

1.) побудувати гістограму відносних частот і комуляту $F^*(x)$;.

2.) обчислити $\overline{x_B}$, σ_B, E_S^*, A_S^*, Mo^*, Me^*.

2.2 Точкові оцінки параметрів розподілу

Теоретичні запитання до теми

1. Що називається точковою статистичною оцінкою?
2. Яка статистичну оцінку називають незміщеною?
3. Яке рівняння достатньо мати для оцінки одного параметра?
4. Що називають ефективною точковою статистичною оцінкою?
5. Що називають ґрунтовною точковою статистичною оцінкою?
6. У чому сутність методу моментів?
7. У чому сутність методу найменших квадратів?
8. У чому сутність методу максимальної правдоподібності?

9. Які рівняння треба мати для оцінки двох параметрів?
10. Чому дорівнює виправлена дисперсія?

Завдання для аудиторної роботи

1.Знайти зміщену і незміщену точкові оцінки дисперсії варіаційного ряду, яка складена за даними вибірки:

x	1	2	5	8	9
n_i	3	4	6	4	3

2. Граничне навантаження на сталевий болт x_i, що вимірювалося в лабораторних умовах, задано як інтервальний статистичний розподіл

Інтервал, $км/мм^2$	4,5-5,5	5,5-6,5	6,5-7,5	7,5-8,5	8,5-9,5
n_i	40	32	28	24	20

Інтервал, $км/мм^2$	9,5-10,5	10,5-11,5	11,5-12,5	12,5-13,5	13,5-14,5
n_i	18	16	12	8	4

Визначити точкові оцінки для генерального середнього та генеральної дисперсії.

3. Випадкова величина X розподілена по біноміальному закону з невідомим параметром p. Проведено 100 дослідів, отриман емпірічниї розподіл появи події А.

x_i	0	1	2	3	4	5	6	7	8	9	10
n_i	2	5	8	12	25	20	11	8	5	3	1

В першому рядку x_i – кількість появи події А в одному з 10 дослідів, в другому рядку n_i – кількість дослідів в яких з’явилась подія А.

Використовуючи метод моментів, знайти точкову статистичну оцінку параметра p.

4. Знайти методом моментів за заданою вибіркою точкові оцінки невідомих параметрів a і σ нормального розподілу

x_i	6,5	8,5	10,5	12,5	14,5	16,5
n_i	15	70	100	50	10	5

5. Випадкова величина X розподілена по показниковому закону з щільністю ймовірностей $f(x, \theta)=\theta e^{-\theta x}$.

Вибірка має вигляд:

x_i	52	48	49	49	52	50	47	48

Використовуючи метод максимальної правдоподібності, знайти точкову статистичну оцінку параметра θ.

2.3 Інтервальні оцінки параметрів розподілу

Теоретичні запитання до теми

1. Сформулюйте основні статистичні розподіли.
2. Що таке початкові статистичні оцінки і які вимоги до них?
3. Що називають точністю і надійністю оцінки?
4. Дайте тлумачення ґрунтовності статистичної оцінки.
5. Що називають довірчим інтервалом?
6. Яка статистична оцінка називається ефективною?
7. В чому полягає метод аналогій визначення точкових статистичних оцінок?
8. В чому сутність методу найменших квадратів визначення статистичних оцінок?
9. В чому полягає метод максимальної правдоподібності статистичного оцінювання параметрів?
10. Який довірчий інтервал з надійністю γ покриває невідомий параметр a?
11. Як можна знайти мінімальний обсяг вибірки при заданій граничній помилці δ і довірчій імовірності γ?
12. За якою формулою знаходять довірчий інтервал для оцінки $\sigma_г$?

Завдання для аудиторної роботи

1. Знайти довірчий інтервал для оцінки з надійністю 0,99 невідомого математичного сподівання a нормально розподіленої ознаки Х генеральної сукупності, якщо дано генеральне квадратичне відхилення σ, вибіркове середнє $\overline{x_в}$ та обсяг вибірки n: а) $\sigma = 4; \overline{x_в} = 10{,}2; n = 16$; б) $\sigma = 5; \overline{x_в} = 16{,}8; n = 25$.

2. Верстат-автомат штампує валики. За вибіркою обсягом n=100 обчислене вибіркове середнє діаметрів виготовлених валиків. Знайти з надійністю 0,95 точність δ, з якою вибіркова середня оцінює математичне сподівання діаметрів виготовлених валиків, знаючи, що їхнє середнє квадратичне відхилення σ=2мм.

3. Знайти мінімальний обсяг вибірки, за якого з надійністю 0,925 точність оцінки математичного сподівання нормально розподіленої генеральної сукупності за вибірковою середньою буде рівне 0,2, якщо відомо середнє квадратичне відхилення генеральної сукупності σ=1,5.

4. У 30 телевізорів була виміряна чутливість x_i . Результати вимірювання подано як дискретний статистичний розподіл:

x_i, мкВ	200	250	300	350	400	450	500
n_i	2	7	6	8	4	2	1

Із надійністю γ=0,99 побудувати довірчий інтервал для $\overline{X_Г}$, якщо $\sigma_Г$=4.

5. У 30 випадково вибраних валиках виміряні відхилення їх діаметрів від номіналу x_i . Результати вимірювань наведено як інтервальний статистичний розподіл:

x_i, мм	5-10	10-15	15-20	20-25	25-30
n_i	2	6	10	8	4

З надійністю γ=0,999 побудувати довірчий інтервал для $\overline{X_Г}$, якщо $\sigma_Г$=0,8 мм.

2.4 Перевірка статистичних гіпотез

Теоретичні запитання до теми

1. Дати визначення нульової та альтернативної гіпотез.
2. Які гіпотези називають параметричними?
3. Які статистичні гіпотези називають простою та складною?
4. Що називають статистичним критерієм?
6. Які зустрічаються закони розподілу статистичних критеріїв?
7. Дайте поняття критичної області, області допустимих значень та критичної точки.
8. Які ви знаєте критичні області?
9. Алгоритм перевірки правильності нульової гіпотези.
10. Що розуміють під помилками першого та другого роду?
11. Що таке рівень значущості α?
12. Що таке потужність критерію і якою вона має бути?
13. З якої рівності знаходять число ступенів вільності?
14. В якому випадку нульову гіпотезу відхиляють?
15. Що таке критерії погодженості?
16. Як визначається критерій погодженості Пірсона?

Завдання для аудиторної роботи

1. При підкиданні монети 10 разів герб випав Х раз. Класифікувати такі гіпотези:

$H^{(1)}$: Х має біноміальний розподіл В(10, ½);

$H^{(2)}$: Х має біноміальний розподіл В(10,р), де 1/3≤р≤⅔;

$H^{(3)}$: Р(Х≤3)> ½;

$H^{(4)}$: монета несиметрична.

2. Велика партія виробів може м ати деяку частку дефектних. Постачальник стверджує, що ця частка становить 5%; покупець припускає, що частка дефектних виробів становить 10%. Умови поставки: з партії випадковим чином вибирається і перевіряється 10 виробів; партія приймається на умовах постачальника, якщо при перевірці виявлено не більше одного дефектного виробу; в противному разі партія приймається на умовах покупця. Сформулювати цю задачу в термінах теорії перевірки статистичних гіпотез і відповісти на такі запитання:

а) Які статистика критерію, область її значень, критична область?

б) Який розподіл має статистика критерію?

в) В чому полягають нульова та альтернативна гіпотези?

г) В чому полягають помилки першого та другого роду і які їх імовірності?

3. Проведено 25 незалежних вимірювань випадкової величини X, що має нормальний закон розподілу зі значенням $\sigma_Г$=2:

x_i,	2,4	5,4	8,4	11,4	14,4	17,4
n_i	2	3	10	6	3	1

При рівні значущості α=0,001 перевірити правильність нульової гіпотези

$H^{(0)}$:M(X)=10,5, якщо альтернативна гіпотеза

$H^{(0)}$:M(X)<10,5.

4. Маємо дані про розподіл підприємств певної області за зростанням виробітку на одного працівника у відсотках до наступного року:

x_i, %	75	85	95	105	115	125
n_i	5	8	10	5	2	1

Ураховуючи, що ознака має нормальний закон розподілу зі значенням $\sigma_Г$=6, перевірити правильність нульової гіпотези при рівні значущості α=0,01.

$H^{(0)}$:M(X)=90, якщо альтернативна гіпотеза

$H^{(0)}$:M(X)$\neq$90.

5. Результати вимірювання зросту дівчаток віком 16 років дали такі показники:

h, см	160-164	164-168	168-172	172-176	176-180
n_i	4	6	20	4	2

Вважаючи, що випадкова величина X- зріст дівчаток – має нормальний закон розподілу, при рівні значущості α=0,001 перевірити правильність нульової гіпотези

$H^{(0)}$:M(X)=180, якщо альтернативна гіпотеза

$H^{(0)}$:M(X)$\neq$180.

6. Вимірювання зросту юнаків віком 17 років дали такі результати:

h, 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

Перевірити, використовуючи критерій Пірсона, за рівнем значущості α=0,01, чи узгоджується гіпотеза про нормальній розподіл генеральної сукупності X з заданим емпіричним розподілом.

7. Використовуючи критерій Пірсона, за рівнем значущості α=0,05, перевірити гіпотезу про нормальний розподіл сукупності, якщо відомі емпіричні(перший рядок) та теоретичні (другий рядок) частоти:

n_i	6	12	16	40	13	8	5
n_i'	4	11	15	43	15	6	6

8. Користуючись критерієм Пірсона, визначити, за рівнем значущості α=0,01, чи не суперечить вибірковим даним гіпотеза про те, що випадкова величина X в генеральної сукупності розподілена за нормальним законом.

x_i	65-67,5	67,5-70	70-72,5	72,5-75	75-77,5	77,5-80	80-82,5	5
n_i	*3*	3	5	5	9	12	16	6

x_i	82,5-85	85-87,5	87,5-90	90-92,5	92,5-95	95-97,5	97,5-100	5
n_i	*15*	10	8	5	4	3	2	6

Рекомендована література

1. Вишневский Л.Д., Гусак Д.В., Погребецкая Т.А., Тер-Саакянц Г.Л. Математическая статистика и случайные процессы:Практикум:Учеб.пособие.- К.: Высш. шк., 1992.- 143с.
2. Вища математика. Основні означення, приклади і задачі (у 2- х ч.) К., “ Либідь”, 1992, ч.2.
3. Гмурман В.Е. Теория вероятностей и математическая статистика: Учеб. Пособие для вузов. - М.: Высш. шк., 2002. - 479с.
4. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. - М.: Высш. шк., 1975. – 333 с.
5. Жлуктенко В.І., Наконечний С.І., Савіна С.С. Теорія ймовірностей і мат-математична статистика:Навч.-метод. Посібник: У 2-х ч. – Ч.ІІ. Математична статистика. –К.:КНЕУ,2001. – 336 с.
6. Лапач С.Н., Чубенко А.В., Бабич П.Н. Статистика в науке и би знесе. –К.: Морион, 2002. - 640 с.
7. Крамер Г. Математические методы статистики. –М.: Висш. шк.,1975. -436с.
8. Ластівка І.О., Мартиненко В.П., Паламарчук Ю.А., Шевченко І.В. Вища математика. Модуль 8. Теорія ймовірностей. Випадкові події: Навч. Посібник. – К.: Книжкове вид-во НАУ,2006. – 108 с.
9. Ластівка І.О., Мартиненко В.П., Паламарчук Ю.А., Шевченко І.В. Вища математика. Модуль 9. Теорія ймовірностей. Випадкови величини.: Навч. Посібник. – К.: Книжкове вид-во НАУ,2007. – 110 с.
10. Ластівка І.О., Ковалюк В.С., Паламарчук Ю.А., Трофименко В.І. Вища математика. Модуль 10. Математична статистика.: Навч. Посібник. – К.: Книжкове вид-во НАУ,2007. – 100 с.
11. Турчин В.М. Математична статистика: Навч. Посібник. - К.: Видавничий центр «Академія», 1999. - 240 с.

Додаток А

Таблиця значень функції Гаусса $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

x	0	1	2	3	4	5	6	7	8	9
0,0	3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3835
0,3	3814	3802	3790	3778	3765	3752	3739	3725	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	i 276	1257	1238	1EI9	1200	1182	1163	1145	11 27
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0 154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
4,0	0001	0001	0001	0000	0000	0000	0000	0000	0000	0000

Додаток Б

Таблиця значень функції Лапласа $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{x^2}{2}} dx$

x	0	1	2	3	4	5	6	7	8	9
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0,2	07926	08318	08706	09095	09483	09871	10257	10642	11026	11409
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240
0,6	22507	22907	23237	23565	23891	24215	24537	24857	25175	25490
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1,1	36433	36650	36864	37076	37286	37499	37698	37900	38100	38298
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1,7	45543	45637	45728	45818	45907	45994	4608-	46164	46246	46327
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
20	47725	47778	47831	47822	47932	47982	48030	48077	48124	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2.4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861

Додаток В

Таблиця значень $t_\gamma = t(\gamma, n)$

n	γ			n	γ		
	0,95	0,99	0,999		0,95	0,99	0,999
5	2,78	4,60	8,61	20	2,093	2,861	3,883
6	2,57	4,03	6,86	25	2,064	2,797	3,745
7	2,45	3,71	5,96	30	2,045	2,756	3,659
8	2,37	3,50	5,41	35	2,032	2,720	3,600
9	2,31	3,36	5,04	40	2,023	2,708	3,558
10	2,26	3,25	4,78	45	2,016	2,692	3,527
11	2,23	3,17	4,59	50	2,009	2,679	3,502
12	2,20	3,11	4,44	60	2,001	2,662	3,464
13	2,18	3,06	4,32	70	1,996	20,649	3,439
14	2,16	3,01	4,22	80	1,991	2,640	3,418
15	2,15	2,98	4,14	90	1,987	2,633	3,403
16	2,13	2,95	4,07	100	1,984	2,627	3,392
17	2,12	2,92	4,02	120	1,980	2,617	3,374
18	2,11	2,90	3,97	∞	1,960	2,576	3,291
19	2,10	2,88	3,92				

Додаток Г

Таблиця значень $q = q(\gamma, n)$

n	γ			n	γ		
	0,95	0,99	0,999		0,95	0,99	0,999
5	1,37	2,67	5,64	20	0,37	0,58	0,88
6	1,09	2,01	3,88	25	0,32	0,49	0,73
7	0,92	1,62	2,98	30	0,28	0,43	0,63
8	0,80	1,38	2,42	35	0,26	0,38	0,56
9	0,71	1,20	2,06	40	0,24	0,35	0,50
10	0,65	1,08	1,80	45	0,22	0,32	0,46
11	0,59	0,98	1,60	50	0,21	0,30	0,43
12	0,55	0,90	1,45	60	0,188	0,269	0,38
13	0,52	0,83	1,33	70	0,174	0,245	0,34
14	0,48	0,78	1,23	80	0,161	0,224	0,31
15	0,46	0,73	1,15	90	0,151	0,211	0,29
16	0,44	0,70	1,07	100	0,143	0,198	0,27
17	0,42	0,66	1,01	150	0,115	0,160	0,211
18	0,40	0,63	0,96	200	0,099	0,136	0,185
19	0,39	0,60	0,92	250	0,098	0,120	0,162

Додаток Д

Таблиця значень $\chi^2_{кр} = \chi^2(\alpha, k)$

Число ступенів свободи, k	Рівень значущості, α					
	0,01	0,025	0,05	0,95	0,975	0,999
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	902	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

Додаток Е

Таблиця критичних точок розподілу Стьюдента (t-розподілу)

Число ступенів свободи, k	Рівень значущості, α				
	0,10	0,05	0,02	0,01	0,001
1	6,3138	12,7062	31,8205	63,6567	636,6192
2	2,9200	4,3027	6,9646	9,9248	31,5991
3	2,3534	3,1824	4,5407	5,8409	12,924
4	2,1318	2,7764	3,7469	4,6041	8,6103
5	2,0150	2,5706	3,3649	4,0321	6,8688
6	1,9432	2,4469	3,1427	3,7074	5,9588
7	1,8946	2,3646	2,9980	3,4995	5,4079
8	1,8595	2,3060	2,8965	3,3554	5,0413
9	1,8331	2,2622	2,8214	3,2498	4,7809
10	1,8125	2,2281	2,7638	3,1693	4,5869
11	1,7959	2,2010	2,7181	3,1058	4,4370
12	1,7823	2,1788	2,6810	3,0545	4,3178
13	1,7709	2,1604	2,6503	3,0123	4,2208
14	1,7613	2,1448	2,6245	2,9768	4,1405
15	1,7531	2,1314	2,6025	2,9467	4,0728
16	1,7459	2,1199	2,5835	2,9208	4,0150
17	1,7396	2,1098	2,5669	2,8982	3,9651
18	1,7341	2,1009	2,5524	2,8784	3,9216
19	1,7291	2,0930	2,5395	2,8609	3,8834
20	1,7247	2,0860	2,5280	2,8453	3,8495
21	1,7207	2,0796	2,5176	2,8314	3,8193
22	1,7171	2,0739	2,5083	2,8188	3,7921
23	1,7139	2,0687	2,4999	2,8073	3,7676
24	1,7109	2,0639	2,4922	2,7969	3,7454
25	1,7081	2,0595	2,4851	2,7874	3,7251
26	1,7056	2,0555	2,4786	2,7787	3,7066
27	1,7033	2,0518	2,4727	2,7707	3,6896
28	1,7011	2,0484	2,4671	2,7633	3,6739
29	1,6991	2,0452	2,4620	2,7564	3,6594
30	1,6973	2,0423	2,4573	2,7500	3,6460
35	1,6896	2,0301	2,4377	2,7238	3,5911
40	1,6839	2,0211	2,4233	2,7045	3,5510
45	1,6794	2,0141	2,4121	2,6896	3,5203
50	1,6759	2,0086	2,4033	2,6778	3,4960
55	1,6730	2,004	2,3961	2,6682	3,4764
60	1,6706	2,0003	2,3901	2,6603	3,4602
70	1,6669	1,9944	2,3808	2,6479	3,4350
80	1,6641	1,9901	2,3739	2,6387	3,4163
90	1,6620	1,9867	2,3685	2,6316	3,4019
100	1,6602	1,9840	2,3642	2,6259	3,3905
110	1,6588	1,9818	2,3607	2,6213	3,3812
120	1,6577	1,9799	2,3578	2,6174	3,3735
∞	1,6448	1,9600	2,3263	2,5758	3,2905