МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Державний біотехнологічний університет
ФАКУЛЬТЕТ МЕХАТРОНІКИ ТА ІНЖИНІРИНГУ

Кафедра фізики і математики

МАТЕМАТИЧНА СТАТИСТИКА

Основи теорії та методика розв'язування задач
з варіантами індивідуальних завдань

для студентів першого (бакалаврського) рівня вищої освіти
денної та заочної форм навчання інженерних спеціальностей

Затверджено
Рішенням Науково-методичної
ради ФМІ ДБТУ
Протокол № 8 від 14.04.2022 р.

Харків
2022

УДК 512.64(076.2)
Л-59

Схвалено
на засіданні кафедри фізики і математики
Протокол № 10 від 19.05.2022 р.

Математична статистика: основи теорії та методика розв'язування задач з варіантами індивідуальних завдань для студентів першого (бакалаврського) рівня вищої освіти денної та заочної форм навч. інж. спец./ О.І.Завгородній [та ін.]; Держ. біотехнол. ун-т.– Харків: ДБТУ, 2022. – 57 с.

Навчально-методичний посібник призначений для вивчення статистичних методів досліджень у різних галузях народногосподарського виробництва. Для самостійної роботи і придбання практичних навиків розв’язку задач наведена необхідна теоретична інформація та варіанти індивідуальних завдань. Розібрана достатня кількість прикладів і приведено зразок виконання індивідуального завдання. Для більш глибокого розуміння матеріалу і поліпшення ефективності роботи над індивідуальним завданням показана комп’ютерна технологія розв’язку задач засобами «Mathcad».

Розрахована на студентів першого (бакалаврського) рівня вищої освіти денної та заочної форм навчання інженерних спеціальностей.

Рецензенти:

А. О. Пак, доктор техн. наук, доцент кафедри фізики і математики Державного біотехнологічного університету

О. А. Макаров, кандидат фіз.-мат. наук, доцент кафедри прикладної математики Харківського національного університету ім. В. Н. Каразіна.

Відповідальний за випуск О. І. Завгородній, д-р техн. наук., проф.

З М І С Т

ПЕРЕДМОВА

Аналітичне дослідження закономірностей в природі, технологічних процесів у виробництві тощо часто відрізняється високим рівним складності, що гальмує оволодіння відповідними знаннями про процеси, які розглядаються. В цих випадках досить успішно застосовуються статистичні методи пізнання. Тому уміння використовувати статистичні методи досліджень є досить важливим, а відповідна тематика міститься в тих розділах вищої математики, які є обов'язковими для вивчення студентами закладів вищої освіти технічних спеціальностей.

Метою посібника є прививання студентам навичок у розв'язуванні задач математичної статистики у різних галузях народногосподарського виробництва. Для спрощення сприйняття матеріалу значна увага надавалась його графічній інтерпретації.

Перший розділ присвячено суттєвості вибіркового методу, як основного в статистичних дослідженнях. Другий – статистичному розподілу вибірок та їх графічному зображенню. Окремо, в третьому розділі, розглянута статистична функція розподілу і її особливості для дискретних та неперервних випадкових величин. В четвертому розділі розглянуті точечні та інтервальні оцінки числових характеристик випадкових величин, в тому числі – вимоги до вказаних оцінок. П'ятий розділ присвячений перевірці статистичних гіпотез про закони розподілу за допомогою критеріїв узгодженості. Розглянуто критерії Пірсона та Колмогорова. В шостому розділі даються варіанти індивідуальних завдань та приклади їх виконання. З метою поліпшення ефективності роботи над індивідуалками та більш глибокого оволодіння студентами сучасних комп'ютерних технологій, наведено приклад виконання завдання в середовищі «Mathcad». В останній частині посібника приводяться додатки з таблицями математичної статистики, необхідними для розв'язку задач.

Посібник відповідає навчальній програмі з вищої математики і буде корисним для студентів денної і заочної форм навчання.

1. Вибірковий метод

Задачею математичної статистики є розробка методів відбору та обробки статистичних даних для виявлення наукових і практичних висновків.

Нехай треба дослідити сукупність однорідних об′єктів відносно деякої якісної або кількісної ознаки. Наприклад, якщо маємо партію деталей, то якісною ознакою може бути стандартність деталі, а кількісною – її розмір.

Інколи проводять повне дослідження, тобто обстежують кожний із об′єктів сукупності відносно ознаки, якою цікавляться, але на практиці можливість провести таке дослідження виникає достатньо рідко. Наприклад, якщо сукупність включає дуже велику кількість об′єктів, то провести повне обстеження фізично неможливо. Якщо ж дослідження об′єкта зв′язано з його знищенням, або потребує великих матеріальних затрат, то проводити повне дослідження немає сенсу. В цих умовах з усієї сукупності випадково відбирають обмежене число об′єктів і проводять дослідження над ними.

Виборочною сукупністю – вибіркою називають сукупність випадково відібраних об′єктів.

Генеральною сукупністю називають сукупність об′єктів із яких проводиться вибірка.

Об′ємом або обсягом сукупності (виборочної чи генеральної) називають число об′єктів цієї сукупності.

Наприклад, якщо з 1000 деталей відібрано для обстеження 100, то об′єм генеральної сукупності $N = 1000$, а об′єм вибірки $n = 100$.

Для того, щоб по даним вибірки можна було впевнено стверджувати про ознаку генеральної сукупності, яка нас цікавить, треба щоб об′єкти вибірки правильно її представляли. Ця вимога коротко формулюється так: **вибірка повинна бути репрезентативною**. Такою ця вибірка буде лише тоді, коли всі об′єкти генеральної сукупності вибираються випадково, тобто тоді, коли вони мають однакову ймовірність попасти в вибірку.

На практиці застосовують різні способи здійснення вибірок: з застосуванням пронумерованих карточок, таблиць випадкових чисел, механічних прийомів та ін. Деякі вибірки проводять у відповідності з введеними державними стандартами. З ними можна ознайомитись при вивченні відповідних дисциплін чи із збірників галузевих видань.

2. Статистичний розподіл вибірки та його зображення. Формула Стерджеса

Нехай кількісна ознака X генеральної сукупності є випадковою дискретною величиною. Припустимо, що над цією ознакою проведено n незалежних спостережень і при цьому величина X прийняла n_1 разів значення x_1, n_2 разів – значення $x_2, \ldots$ і, нарешті, n_k разів – значення x_k. В результаті одержана вибірка об′єму $n = n_1 + n_2 + \cdots + n_k$.

Числа $n_i, i = 1, 2, \ldots, k$ називають частотами, а відношення $W_i = n_i / n$ – відносними частотами.

Очевидно, що:

$$W_1 + W_2 + \cdots + W_k = \frac{n_1}{n} + \frac{n_2}{n} + \cdots + \frac{n_k}{n} = \frac{n}{n} = 1$$

Упорядковану за величиною послідовність випадкових значень **називають варіаційним рядом**, а самі ці значення – **варіантами**. В, наведному вище, прикладі варіантами є числа $x_1, x_2, \ldots, x_k$.

Емпіричним або статистичним розподілом дискретної випадкової величини називають сукупність спостережених значень ($x_i, i = 1, 2, \ldots, k$), розміщених в зростаючому порядку, з указанням відповідних відносних частот ($W_i, i = 1, 2, \ldots, k$).

Розглянемо, наприклад, роботу однотипних машин на деякому проміжку часу *t*. Число відмов кожної машини на цьому проміжку часу напевне передбачити не можна, тобто воно є випадковим явищем, яке можна дослідити саме шляхом спостережень. Статистичний розподіл вказаної дискретної випадкової величини – числа відмов, виниклих на проміжку часу *t*, можна представити у вигляді таблиці.

Таблиця 1

Число відмов – x_i	1	2	3	4	5	6	7	8	Примітки
Число машин – n_i	1	8	20	20	37	10	3	1	$\sum n_i = 100$
Відносна частота – W_i	0,01	0,08	0,20	0,20	0,37	0,10	0,03	0,01	$\sum W_i = 1$

Крім табличного способу задання емпіричного розподілу дискретної випадкової величини існують графічні. Один з них – полігон розподілу. Він є аналогом многокутника розподілу [1, 4, 7], але по осі ординат замість ймовірностей p_i відкладають відносні частоти W_i. На рис.1 показано полігон розподілу для випадкової величини, заданої таблицею 1.

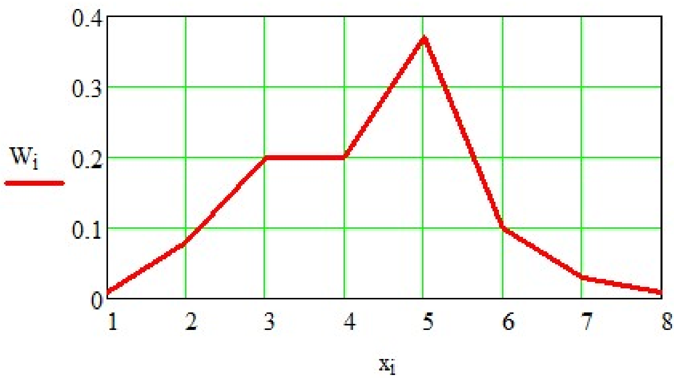


Рис.1

При табличному зображенні емпіричного розподілу неперервної випадкової величини в таблиці вказують відносні частоти не кожного окремого значення, а значень, що належать прийнятим інтервалам.

Нехай, наприклад, на токарному станку виготовляється деталь, розмір якої задається по кресленню. Дійсний розмір виготовлених деталей буде мати випадкові відхилення в межах технічного допуску і є неперервною випадковою величиною. Емпіричний розподіл розміру деталі показано в таблиці 2.

Таблиця 2

Інтервал		Середина інтервалу x_i	Частоти n_i	Відносні частоти W_i	Накоплені відносні частоти
від	до				
20,00	20,05	20,025	2	0,02	0,02
20,05	20,10	20,075	10	0,10	0,12
20,10	20,15	20,125	24	0,24	0,36
20,15	20,20	20,175	30	0,30	0,66
20,20	20,25	20,225	22	0,22	0,88
20,25	20,30	20,275	10	0,10	0,98
20,30	20,35	20,325	2	0,02	1
			$\sum n_i = 100$	$\sum W_i = 1$	

Графічний розподіл неперервної ознаки будують переважно у вигляді гістограми та полігона частот n_i або відносних частот W_i. Інколи доцільно використовувати також полігон щільності відносних частот W_i/h (h – довжина інтервалу). На рис.2 графічний розподіл зображено за даними таблиці 2.

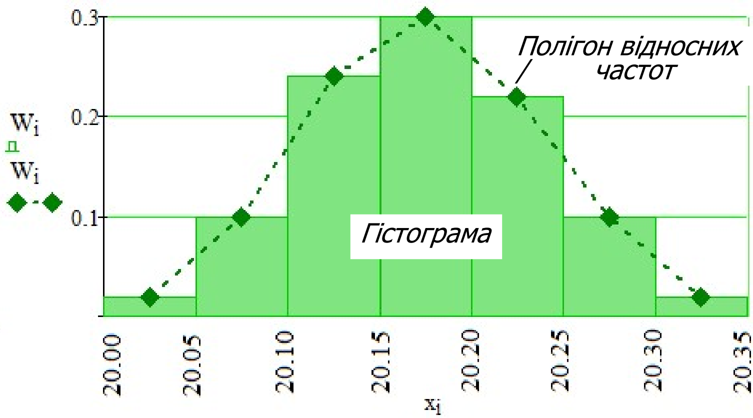


Рис.2

Як бачимо, неперервна випадкова ознака потребує побудови варіаційного ряду, як це було і для дискретної ознаки. Відмінність лише в тому, що варіаційний ряд для неперервної ознаки є інтервальним. Для визначення раціонального числа інтервалів – такого, при якому побудована гістограма і полігон розподілу не були б дуже громіздкими і в той же час дозволяли виявити характерні риси явища, яке досліджується, слід застосовувати емпіричну формулу Стерджеса:

$$\boxed{k = 1 + 3,322 \lg n}\ . \qquad (1)$$

Тут n – обсяг вибірки, а k – розрахункове число інтервалів.

Формула (1) дає дробове значення, тому його необхідно округляти до цілого. На одержане, таким чином, значення орієнтуються при виборі числа інтервалів, тобто, як правило, приймають його за одержаним підрахунком. Але в окремих випадках буває доцільним прийняти його з деяким незначним відхиленням від розрахункового значення. Така необхідність може трапитися при бажанні зменшити кількість розрахунків, або підлаштувати їх до вимог певних статистичних критеріїв (з ними будемо знайомитись нижче). В нашому випадку $1 + 3,322 \lg 100 = 7,64$ (див. табл.2) тому для побудови розподілу на рис.2 прийнято число інтервалів $k = 7$.

За формулою Стерджеса маємо таку відповідність (табл.3):

Таблиця 3

n	6-11	12-22	23-44	45-90	91-180	181-359
k	4	5	6	7	8	9

Отже, знаючи об′єм вибірки n, число інтервалів k легко знаходиться з другого рядка таблиці 3. Оскільки об′єм вибірки $n = 22$ є досить малим, а об′єм $n = 180$ в більшості випадків є достатнім, то найбільш розповсюджені значення чисел інтервалів знаходяться у межах від 6 до 8.

3. Емпірична функція розподілу

Нехай статистичний розподіл кількісної ознаки X – відомий. Найбільш повно характеризує ознаку X емпірична функція розподілу.

Емпіричною функцією розподілу називають функцію $\boldsymbol{F^*(x)}$, що визначає для кожного значення $\boldsymbol{x}$ відносну частоту події $\boldsymbol{X < x}$. Тобто:

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

де $\boldsymbol{n_x}$ – число спостережень при яких ознака $\boldsymbol{X}$ приймала значення, менші за $\boldsymbol{x}$.

Аналогом емпіричної функції $F^*(x)$ в теорії ймовірностей є інтегральна функція розподілу $F(x) = P(x < X)$. В статистиці її називають також теоретичною функцією або функцією розподілу генеральної сукупності. Відмінність між ними полягає в тому, що $F^*(x)$ визначає відносну частоту події $x < X$, а $F(x)$ – ймовірність цієї ж події. Але із збільшенням об'єму вибірки згадані частоти та ймовірності зближаються (по ймовірності), а точність апроксимації теоретичної функції емпіричною – зростає. Властивості емпіричної функції $F^*(x)$ такі ж як і теоретичної:

1) $0 \le F^*(x) \le 1$;
2) $F^*(x_2) \ge F^*(x_1)$ при $x_2 > x_1$ – неспадна функція;
3) $F^*(x)\Big|_{x \le x_{\min}} = 0$, де $x_{\min}$ – найменша варіанта ряду;
4) $F^*(x)\Big|_{x > x_{\max}} = 1$, де $x_{\max}$ – найбільша варіанта.

При цьому для дискретних випадкових величин $F^*(x)$ має ступінчатий вид, а для неперервних вона є неперервною неспадаючою кривою.

Для побудови емпіричної функції розподілу дискретної випадкової ознаки використаємо дані таблиці 1.

$F^*(x) = 0$ при $x \le 1$ ($x_{\min} = 1$ – властивість 3).

Значення $X < 2$, а саме $x = 1$ спостерігалось 1 раз з частотою 0,01, тому $F^*(x) = 0{,}01$ при $1 < x \le 2$.

Значення $X<3$, а саме $x=1$ та $x=2$ спостерігалось $1+8=9$ разів з частотою $0,01+0,08=0,09$, тому $F^*(x)=0,09$ при $2<x\le 3$.

Аналогічно одержимо:

$F^*(x)=0,29$ при $3<x\le 4$; $F^*(x)=0,49$ при $4<x\le 5$;
$F^*(x)=0,86$ при $5<x\le 6$; $F^*(x)=0,96$ при $6<x\le 7$;
$F^*(x)=0,99$ при $7<x\le 8$;

Нарешті $F^*(x)=1$ при $x>8$ ($x_{\max}=8$ – властивість 4).

Отже, шукана емпірична функція розподілу має вид:

$$F^*(x)=\begin{cases}0 & \textit{при} & x\le 1;\\ 0,01 & \textit{при} & 1<x\le 2;\\ 0,09 & \textit{при} & 2<x\le 3;\\ 0,29 & \textit{при} & 3<x\le 4;\\ 0,49 & \textit{при} & 4<x\le 5;\\ 0,86 & \textit{при} & 5<x\le 6;\\ 0,96 & \textit{при} & 6<x\le 7;\\ 0,99 & \textit{при} & 7<x\le 8;\\ 1 & \textit{при} & x>8.\end{cases}$$

Виходячи з суті утворення емпіричної функції розподілу, її ще називають функцією накоплення відносних частот. Графік знайденої функції зображено нижче на рис. 3. Затушований кружечок означає, що відповідна точка належить області визначення функції, а стрілка – відсутність точки в указаній області.

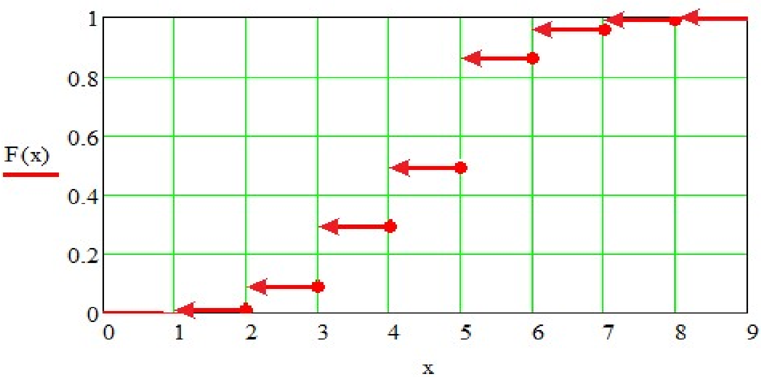


Рис.3

Розглянемо тепер емпіричну функцію розподілу неперервної випадкової ознаки, заданої інтервальним рядом (табл.2). Для підрахунку значень функції використаємо, приведені в цій таблиці величини накоплених частот (останній стовпчик). Значення функції $F^*(x)$ лівіше точки $x = 20,00$ та правіше точки $x = 20,35$ відомі: $F^*(x) = 0$ при $x \le 20,00$; $F^*(x) = 1$ при $x > 20,35$. Оскільки, як уже розглядалося, $F^*(x)$ є функцією накоплених частот, то значення цієї функції у вузлових точках знаходимо з табл.2: $F^*(20,05) = 0,02$; $F^*(20,10) = 0,12$; $F^*(20,15) = 0,36$; $F^*(20,20) = 0,66$; $F^*(20,25) = 0,88$; $F^*(20,30) = 0,98$. Закон зміни значень функції між вузловими точками достеменно невідомий, але при побудові графіка за основу береться припущення, що випадкова ознака на кожному частинному інтервалі має рівномірну щільність ймовірностей. Тоді графіком функції буде ламана лінія, яка проходить через вузлові точки (рис.4).

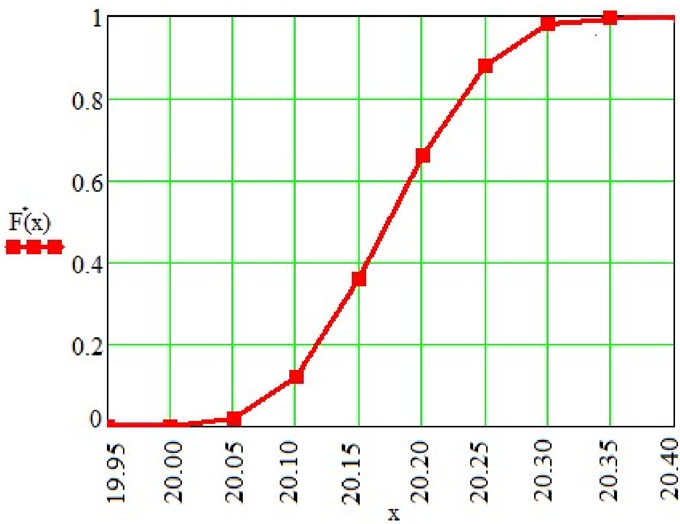


Рис.4

4. Статистичні оцінки параметрів розподілу

Нехай X – випадкова величина з законом розподілу $F(x,\theta)$, де θ – параметр розподілу, числове значення якого невідоме. Про величину θ судять за результатами обстеження вибірки.

Будь-яку однозначно визначену функцію результатів спостережень, за допомогою якої судять про значення параметра θ, називають **оцінкою або статистикою параметра** θ. Вибіркову оцінку параметра θ будемо позначати $\overline{\theta}$.

Вибір оцінки, яка дозволяє одержати хороше наближення невідомого параметра розподілу – основна задача теорії оцінювання.

4.1. Точечні оцінки параметрів розподілу

Якщо оцінка параметра θ задається одним числом, вона називається точечною оцінкою цього параметра.

Щоб точечні вибіркові оцінки були якомога ближчими до відповідних параметрів генеральної сукупності, до них висуваються такі вимоги:

– ***спроможність;***
– ***ефективність;***
– ***незміщеність.***

Оцінка називається ***спроможною***, якщо із збільшенням об′єму вибірки n вона наближається (збігається по ймовірності) до параметра, який оцінюється, тобто, коли за формулою Бернуллі $P(\left|\overline{\theta}-\theta\right|<\varepsilon)\to 1$ при $n\to\infty$

Виконання цієї вимоги дає можливість забезпечення необхідної точності вибіркової оцінки за рахунок вірного підбору об′єму вибірки.

Оцінка називається ***ефективною***, якщо вона має найменшу дисперсію серед дисперсій всіх можливих оцінок параметра θ.

Оцінка $\overline{\theta}$ будь-кого параметра θ, одержана за результатом обстеження вибірки, є випадковою, адже кожна повторна вибірка дає іншу (може навіть дуже близьку) оцінку того ж параметра. Це значить, що оцінка $\overline{\theta}$ володіє дисперсією. Отже, ефективна оцінка має найменше розсіювання значень, що об′єктивно покращує оцінювання.

Оцінка називається ***незміщеною***, якщо її математичне сподівання дорівнює параметру, який оцінюється: $\boldsymbol{M}[\overline{\theta}] = \theta$.

Нехтування цією вимогою заздалегідь спричиняє наявність похибки у вигляді систематичних відхилень при визначенні $\overline{\theta}$.

Варто підкреслити, що ідеальним є виконання одночасно всіх вказаних вище вимог, але на практиці це вдається не завжди.

Найбільш важливими числовими характеристиками випадкової величини X, без яких важко одержати об′єктивне уявлення про її властивості, є математичне сподівання $-M[X]$ та дисперсія – $D[X]$. Оскільки зміщені точечні оцінки можуть давати значні похибки, особливо при малих об′ємах вибірок, проведемо, перш за все, дослідження оцінок для $M[X]$ та $D[X]$ на “незміщеність”.

Для оцінки математичного сподівання $M[X]$ використовують середнє арифметичне значення:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\left(x_1 + x_2 + \cdots + x_n\right). \tag{2}$$

Якщо кожне значення x_i спостерігалось неодноразово – n_i разів $(\sum n_i = n,\ i = 1, 2, \ldots, k)$, то формула для середнього арифметичного буде мати вигляд:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \frac{1}{n}\left(x_1 n_1 + x_2 n_2 + \cdots + x_k n_k\right). \tag{3}$$

З метою спрощення перетворень але без втрати загальності міркувань далі при дослідженнях будемо використовувати формулу (2).

Нехай випадкова величина X має математичне сподівання a і дисперсію σ^2. Розглянемо тепер математичне сподівання середнього арифметичного:

$$M[\overline{x}] = M\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n} M[x_i] = \frac{1}{n}\sum_{i=1}^{n} a = \frac{na}{n} = a\,. \qquad (4)$$

Таким чином, оцінка $\overline{x}$ є незміщеною оцінкою математичного сподівання. Можна показати, що вона є також спроможною і ефективною.

В співвідношенні (4) використано такі властивості:

$$M[CX] = CM[X]; \quad M\left[\sum_i X\right] = \sum_i M[X], \quad \left(C = \text{const}\right). \quad (5)$$

Розглянемо далі вибіркову дисперсію $\hat{S}_X^2$, як оцінку генеральної дисперсії σ^2:

$$\hat{S}_X^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = \frac{1}{n}\left[\left(x_1 - \overline{x}\right)^2 + \left(x_2 - \overline{x}\right)^2 + \cdots + \left(x_n - \overline{x}\right)^2\right]. \ (6)$$

Перетворимо формулу вибіркової дисперсії.

$$\hat{S}_X^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\left(x_i - a\right) - \left(\overline{x} - a\right)\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - a\right)^2 - \frac{2}{n}\left(\overline{x} - a\right)\sum_{i=1}^{n}\left(x_i - a\right) +$$

$$+\frac{1}{n}\sum_{i=1}^{n}\left(\overline{x} - a\right)^2 \qquad (7)$$

Оскільки $\sum_i x_i = n\overline{x}$, то:

$$\frac{2}{n}\left(\overline{x} - a\right)\sum_{i=1}^{n}\left(x_i - a\right) = \frac{2}{n}\left(\overline{x} - a\right)\left(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} a\right) = \frac{2}{n}\left(\overline{x} - a\right)\left(n\overline{x} - na\right) =$$

$$= 2\left(\overline{x} - a\right)^2 \qquad (8)$$

Отже:

$$\hat{S}_X^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - a\right)^2 - 2\left(\overline{x} - a\right)^2 + \frac{1}{n}n\left(\overline{x} - a\right)^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(x_i - a\right)^2 - \left(\overline{x} - a\right)^2. \qquad (9)$$

Обчислимо спочатку математичне сподівання першого доданку виразу (9):

$$M\left[\frac{1}{n}\sum_{i=1}^{n}\left(x_i-a\right)^2\right]=\frac{1}{n}\sum_{i=1}^{n}M\left[\left(x_i-a\right)^2\right]=\frac{1}{n}\sum_{i=1}^{n}\sigma^2=\frac{n\sigma^2}{n}=\sigma^2. \quad (10)$$

Математичне сподівання другого доданку:

$$M\left[\left(\overline{x}-a\right)^2\right]=D\left[\overline{x}\right]=D\left[\frac{1}{n}\sum_{i=1}^{n}x_i\right]=\frac{1}{n^2}\sum_{i=1}^{n}D\left[x_i\right]=\frac{n\sigma^2}{n^2}=\frac{\sigma^2}{n}. \quad (11)$$

Враховуючи співвідношення (9)-(11), остаточно одержимо:

$$M\left[\hat{S}_X^2\right]=M\left[\frac{1}{n}\sum_{i=1}^{n}\left(x_i-a\right)^2\right]-M\left[\left(\overline{x}-a\right)^2\right]=\sigma^2-\frac{\sigma^2}{n}=\frac{n-1}{n}\sigma^2. \quad (12)$$

В співвідношенні (11) використано властивості дисперсії:

$$D[CX]=C^2D[X]; \quad D\left[\sum_i X\right]=\sum_i D[X]. \quad (13)$$

Одержаний результат показує, що математичне сподівання вибіркової дисперсії не дорівнює дисперсії генеральної сукупності (множник $(n-1)/n$ не дорівнює одиниці). Тому **вибіркова дисперсія є зміщеною оцінкою**. Щоб уникнути цього недоліку в розрахунках використовують виправлену дисперсію:

$$S_X^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2=\frac{1}{n-1}\left[\left(x_1-\overline{x}\right)^2+\left(x_2-\overline{x}\right)^2+\cdots+\left(x_n-\overline{x}\right)^2\right] \quad (14)$$

або

$$S_X^2=\frac{1}{n-1}\sum_{i=1}^{k}\left(x_i-\overline{x}\right)^2 n_i=$$

$$=\frac{1}{n-1}\left[\left(x_1-\overline{x}\right)^2 n_1+\left(x_2-\overline{x}\right)^2 n_2+\cdots+\left(x_k-\overline{x}\right)^2 n_k\right]. \quad (15)$$

В співвідношенні (15) мається на увазі, що значення x_i у випробуваннях спостерігається n_i разів ($\sum n_i=n,\ i=1,2,\ldots,k$).

За формулою (14) маємо:

$$S_X^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2=\frac{n}{n-1}\cdot\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2=\frac{n}{n-1}\hat{S}_X^2. \quad (16)$$

Тепер підрахуємо математичне сподівання виправленої дисперсії:

$$M\left[S_X^2\right]=M\left[\frac{n}{n-1}\hat{S}_X^2\right]=\frac{n}{n-1}M\left[\hat{S}_X^2\right]=\frac{n}{n-1}\cdot\frac{n-1}{n}\sigma^2=\sigma^2\,. \quad (17)$$

Як бачимо, виправлена дисперсія є незміщеною оцінкою. Можна довести що вона є також спроможною, але не задовольняє вимогам ефективності.

Якщо об′єм вибірки достатньо великий $n \geq 30$, то виправлена S_X^2 і вибіркова $\hat{S}_X^2$ дисперсії практично співпадають. В цих умовах можна застосовувати в розрахунках формулу (6) для вибіркової дисперсії без суттєвих втрат точності.

Для оцінки генерального середнього квадратичного відхилення використовують відповідну вибіркову оцінку, яку ще називають стандартним відхиленням:

$$S_X=\sqrt{S_X^2}\,. \quad (18)$$

Для інших числових характеристик генеральної сукупності застосовують такі вибіркові оцінки:

– для коефіцієнта варіації

$$V_X=\frac{S_X}{\overline{x}}\,, \quad (19)$$

– для асиметрії

$$A_X=\frac{\frac{1}{n}\sum_{i=1}^{k}\left(x_i-\overline{x}\right)^3 n_i}{S_X^3}=$$

$$=\frac{1}{nS_X^3}\left[\left(x_1-\overline{x}\right)^3 n_1+\left(x_2-\overline{x}\right)^3 n_2+\cdots+\left(x_k-\overline{x}\right)^3 n_k\right], \quad (20)$$

– для ексцесу

$$E_X=\frac{\frac{1}{n}\sum_{i=1}^{k}\left(x_i-\overline{x}\right)^4 n_i}{S_X^4}-3=$$

$$\frac{1}{nS_X^4}\left[\left(x_1-\overline{x}\right)^4 n_1+\left(x_2-\overline{x}\right)^4 n_2+\cdots+\left(x_k-\overline{x}\right)^4 n_k\right]-3\,. \quad (21)$$

Вони є аналогами відповідних числових характеристик в теорії ймовірностей.

Відмітимо, що формули (3), (15), (18)-(21) правомірно застосовувати як для дискретних, так і для неперервних випадкових величин, які, як відомо, задаються інтервальним рядом. Треба тільки врахувати, що у випадку неперервних випадкових величин значення x_i трактуються, як середини відповідних інтервалів, а k, як число цих інтервалів. Наведемо далі розрахунки точечних оцінок числових характеристик випадкової величини за даними таблиці 2 де, очевидно, $k = 7$:

– середнє значення

$$\overline{x} = \frac{1}{100}\sum_{i=1}^{7} x_i n_i = \frac{1}{100}\big(20,025\cdot 2 + 20,075\cdot 10 + 20,125\cdot 24 + 20,175\cdot 30 +$$

$$+20,225\cdot 22 + 20,275\cdot 10 + 20,325\cdot 2 = 20,174;$$

– виправлена дисперсія

$$S_X^2 = \frac{1}{99}\sum_{i=1}^{7}\left(x_i - \overline{x}\right)^2 n_i = \frac{1}{99}\Big[\left(20,025 - 20,174\right)^2\cdot 2 + \left(20,075 - 20,174\right)^2\cdot 10 +$$

$$+\left(20,125 - 20,174\right)^2\cdot 24 + \left(20,175 - 20,174\right)^2\cdot 30 + \left(20,225 - 20,174\right)^2\cdot 22 +$$

$$+\left(20,275 - 20,174\right)^2\cdot 10 + \left(20,325 - 20,174\right)^2\cdot 2\Big] = 0,00409;$$

– стандартне відхилення

$$S_X = \sqrt{S_X^2} = \sqrt{0,00409} = 0,064;$$

– вибірковий коефіцієнт варіації

$$V_X = \frac{S_X}{\overline{x}} = \frac{0,064}{20,174} = 0,00317;$$

– вибіркова асиметрія

$$A_X = \frac{\frac{1}{100}\sum_{i=1}^{7}\left(x_i - \overline{x}\right)^3 n_i}{S_X^3} = \frac{1}{100\cdot 0,064^3}\Big[\left(20,025 - 20,174\right)^3\cdot 2 +$$

$$+\left(20,075 - 20,174\right)^3\cdot 10 + \left(20,125 - 20,174\right)^3\cdot 24 + \left(20,175 - 20,174\right)^3\cdot 30 +$$

$$+\left(20,225 - 20,174\right)^3\cdot 22 + \left(20,275 - 20,174\right)^3\cdot 10 + \left(20,325 - 20,174\right)^3\cdot 2\Big] =$$

$$= 0,00369;$$

– вибірковий ексцес

$$E_X = \frac{\frac{1}{100}\sum_{i=1}^{7}\left(x_i - \overline{x}\right)^4 n_i}{S_X^4} - 3 = \frac{1}{100\cdot 0,064^4}\Big[\left(20,025-20,174\right)^4\cdot 2 +$$

$$+\left(20,075-20,174\right)^4\cdot 10 + \left(20,125-20,174\right)^4\cdot 24 + \left(20,175-20,174\right)^4\cdot 30 +$$

$$+\left(20,225-20,174\right)^4\cdot 22 + \left(20,275-20,174\right)^4\cdot 10 + \left(20,325-20,174\right)^4\cdot 2\Big] -$$

$$-3 = -0,421.$$

Одержані результати статистичного дослідження вибірки показують, що середнє значення ($\overline{x} = 20,174$) практично співпадає з серединою інтервалу (20,175), а величина асиметрії близька до нуля ($A_X = 0,00369$). Це говорить про те, що ознака X (розмір деталі) генеральної сукупності з великою ймовірністю може мати симетричний розподіл. При цьому розсіювання значень цієї ознаки незначна ($V_X = 0,00317$), а гостровершинність близька до нормального закону розподілу, але дещо поступається йому у цьому показнику ($E_X = -0,421 < 0$). Ці висновки не суперечать вигляду гістограми і полігону розподілу, побудованих на рис.2 за тими ж даними.

4.2. Інтервальні оцінки параметрів розподілу

При малих об′ємах вибірок точечні оцінки можуть привести до грубих помилок, тому в цих обставинах слід використовувати інтервальні оцінки.

Інтервальною називають оцінку параметра θ, яка визначається двома числами – кінцями інтервалу.

Нехай за результатами обстеження вибірки знайдена статистична характеристика θ^*, яка служить оцінкою невідомого параметра θ. Якщо $\left|\theta - \theta^*\right| < \delta$, то чим менше δ, тим точніше оцінка. Але статистичні оцінки не дозволяють категорично визнавати справедливість цієї нерівності. Можна говорити лише про ймовірність α, з якою виконується відмічена нерівність.

Надійністю оцінки параметра θ по θ^* називають ймовірність α, з якою виконується нерівність $\left|\theta-\theta^*\right|<\delta$, а ймовірність $\gamma=\mathbf{1}-\alpha$ називають **рівнем значущості**.

На практиці найчастіше використовуються такі значення надійності: $\alpha=0,95$; $\alpha=0,99$.

Отже, можна записати $P(\left|\theta-\theta^*\right|<\delta)=\alpha$ або, що те саме, $P(\theta^*-\delta<\theta<\theta^*+\delta)=\alpha$, тобто ймовірність того, що інтервал $(\theta^*-\delta,\ \theta^*+\delta)$ покриває невідомий параметр θ дорівнює α.

4.2.1. Довірливі інтервали для математичного сподівання

По вибірковим даним можна побудувати таку випадкову величину T, що має розподіл, який не залежить від невідомих параметрів a і σ, а саме:

$$T=\frac{\overline{x}-a}{S_X\big/\sqrt{n}}, \tag{22}$$

де: n – об′єм вибірки; $\overline{x}$ – вибіркове середнє; a – генеральне середнє; S_X – стандартне відхилення; σ – середнє квадратичне відхилення.

Функція щільності випадкової величини T має вид []:

$$S(t,n)=\frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)}\Gamma\left(\frac{n-1}{2}\right)}\left(1+\frac{t^2}{n-1}\right)^{-\frac{n}{2}}; \tag{23}$$

$$\Gamma(x)=\int\limits_0^\infty z^{x-1}\mathrm{e}^{-z}dz, \tag{24}$$

де: t – значення величини T; $\Gamma(x)$ – гамма-функція.

Розподіл (23) називають розподілом Стьюдента.[1]

Очевидно, що функція (23) парна, тому:

$$P\left(\left|T\right|<t_\alpha\right)=P\left(\left|\frac{\overline{x}-a}{S_X\big/\sqrt{n}}\right|<t_\alpha\right)=\int\limits_{-t_\alpha}^{t_\alpha}S(t,n)dt=2\int\limits_0^{t_\alpha}S(t,n)dt=\alpha \tag{25}$$

[1] Псевдонім англійського статистика Вільяма Госсета

або
$$P\left(\overline{x}-t_\alpha\frac{S_X}{\sqrt{n}}<a<\overline{x}+t_\alpha\frac{S_X}{\sqrt{n}}\right)=\alpha. \qquad (26)$$

Отже, шуканий довірливий інтервал, що покриває невідоме математичне сподівання (генеральне середнє) з ймовірністю α, має вигляд:

$$\boxed{\overline{x}-t_\alpha\frac{S_X}{\sqrt{n}}<a<\overline{x}+t_\alpha\frac{S_X}{\sqrt{n}}}. \qquad (27)$$

Величину t_α називають квантилем розподілу Стьюдента. Його можна знайти з рівняння (25), яке приводиться до виду:

$$\int_0^{t_\alpha} S(t,n)\,dt=\frac{\alpha}{2}. \qquad (28)$$

Розв′язання рівняння (28) є досить непростим із-за складності підінтегральної функції, тому для практичної зручності з використанням цього рівняння та співвідношень (23), (24) складені таблиці, з яких по заданим значенням надійності α та об′єму вибірки n знаходять величину t_α з додатку 2 (стор. 53).

4.2.2. Довірливі інтервали для середнього квадратичного відхилення

Будемо вимагати, щоб виконувалось співвідношення –

$$P\left(\left|\sigma-S_X\right|<\delta\right)=\alpha, \qquad (29)$$

або:

$$P\left(S_X-\delta<\sigma<S_X+\delta\right)=\alpha, \qquad (30)$$

де: σ – середнє квадратичне відхилення генеральної сукупності; S_X – стандартне відхилення.

Якщо позначити $q=\delta/S_X$, то нерівність (30) (в дужках) можна записати так:

$$S_X\left(1-q\right)<\sigma<S_X\left(1+q\right). \qquad (31)$$

В цій нерівності невідомим залишається параметр q. Для його знаходження розглядають допоміжну випадкову величину:

$$\chi = \frac{S_X}{\sigma}\sqrt{n-1}\,, \tag{32}$$

де: n – об′єм вибірки.

Щільність величини χ має вид [1]:

$$R(\chi, n) = \frac{\chi^{n-2} e^{-\frac{\chi^2}{2}}}{2^{\frac{n-3}{2}}\,\Gamma\left(\frac{n-1}{2}\right)}\,, \tag{33}$$

де: $\Gamma(x)$ – гамма-функція (24).

Цей розподіл не залежить від параметра σ, а тільки від об′єму вибірки n.

Перетворимо далі нерівність (31) так, щоб вона мала вигляд $\chi_1 < \chi < \chi_2$. Тоді ймовірність такого співвідношення буде виражатися інтегралом:

$$\int_{\chi_1}^{\chi_2} R(\chi, n)\,d\chi = \alpha\,. \tag{34}$$

Спочатку запишемо нерівність (31) у вигляді:

$$\frac{1}{S_X(1+q)} < \frac{1}{\sigma} < \frac{1}{S_X(1-q)}\,, \tag{35}$$

після чого помножимо одержаний вираз на $S_X\sqrt{n-1}$. Отже, маємо:

$$\frac{\sqrt{n-1}}{1+q} < \frac{S_X\sqrt{n-1}}{\sigma} < \frac{\sqrt{n-1}}{1-q} \tag{36}$$

або

$$\frac{\sqrt{n-1}}{1+q} < \chi < \frac{\sqrt{n-1}}{1-q} \tag{37}$$

Тепер співвідношення (34) можна переписати так:

$$\int_{\frac{\sqrt{n-1}}{1+q}}^{\frac{\sqrt{n-1}}{1-q}} R(\chi, n)\,d\chi = \alpha\,. \tag{38}$$

Разом з виразами (33) і (24) воно відображає рівняння відносно невідомого параметра q. За допомогою цього рівняння складено таблиці значень параметра q в залежності від об′єму вибірки n і надійності α. Отже, інтервал довіри, який покриває генеральне середнє квадратичне відхилення σ задається нерівністю (31), де значення величини q вибирають з указаної таблиці (додаток 3, стор.53). Якщо ліва частина цієї нерівності виявиться від′ємною – її заміняють нулем. Остаточно для вказаного довірливого інтервалу будемо мати:

$$\boxed{\begin{array}{ll} S_X\left(1-q\right)<\sigma<S_X\left(1+q\right) & при \quad q<1, \\ \quad 0<\sigma<S_X\left(1+q\right) & при \quad q\geq 1. \end{array}} \tag{39}$$

За одержаними залежностями (27) і (39) знайдемо інтервальні оцінки для математичного сподівання – a та середнього квадратичного відхилення – σ випадкової ознаки X, заданої таблицею 2, з використаням, знайдених вище, точечних оцінок. Приймемо для цього рівень надійності $\alpha=0,95$, який найчастіше використовують при вирішенні технічних задач. З таблиці додатку 2 (стор.53) по відомому рівню надійності та об′єму вибірки $n=100$ знаходимо квантиль розподілу Стьюдента: $t_\alpha=1,984$. Далі застосовуємо залежність (27)

$$\bar{x}-t_\alpha\frac{S_X}{\sqrt{n}}<a<\bar{x}+t_\alpha\frac{S_X}{\sqrt{n}}$$

та значення параметрів $\bar{x}=20,174,\ S_X=0,0640,$ що дає:

$$20,174-1,984\cdot\frac{0,0640}{\sqrt{100}}<a<20,174-1,984\cdot\frac{0,0640}{\sqrt{100}}$$

Обчислюючи ліву і праву частину записаної нерівності, остаточно знаходимо інтервал, який з імовірністю, не меншою 0,95, покриває невідоме математичне сподівання (генеральне середнє):

$$\boxed{20{,}161 < a < 20{,}187}\,.$$

Відмітимо, що при $n \geq 30$ розподіл Стьюдента близький до нормального розподілу, тому квантиль t_α можна з деякою незачною похибкою замінити квантилем нормального розподілу z_α, який знаходять з рівняння:

$$\Phi\left(z_\alpha\right) = \frac{\alpha}{2}\,. \qquad (40)$$

Для цього використовують таблицю функції Лапласа $\Phi(z)$, яку читають ‘‘навпаки’’. Тобто, керуючись ‘‘шапкою” таблиці, знаходять ймовірності $\alpha/2$ в полі таблиці, а квантиль z_α виписують, як відповідне значення функції в колонці зліва. В нашому випадку $\alpha/2 = 0{,}95/2 = 0{,}475$ і тоді $z_{0,95} = 1{,}96$ (додаток 1, стор.52). Легко перевірити, що використання знайденого значення для обчислення інтервалу (27) дає практично той же результат, так як об′єм вибірки набагато перевищує (більш, ніж втричі) число 30. Тому різниця в розрахунках буде відчуватися тільки в четвертому знаку після коми. Взагалі, особливо при невеликих об′ємах вибірок, близьких до 30, щоб одержувати більш точні результати досліджень, доцільніше використувати для такого виду задач квантиль Стьюдента.

По об′єму вибірки $n = 100$ і надійності $\alpha = 0{,}95$ з таблиці додатку 3 (стор. 53) знаходимо значення: $q = 0{,}143$. Тоді, враховуючи, що $S_X = 0{,}0640$, за співвідношенням (39)

$$S_X\left(1-q\right) < \sigma < S_X\left(1+q\right)$$

одержимо: $0{,}0640\left(1-0{,}143\right) < \sigma < 0{,}0640\left(1+0{,}143\right)$.

Остаточно інтервал, який з імовірністю, не меншою 0,95, покриває невідоме середнє квадратичне відхилення, запишеться так:

$$\boxed{0{,}0548 < \sigma < 0{,}0732}\,.$$

5. Перевірка статистичних гіпотез

Під **статистичною гіпотезою** розуміють будь-яке висловлювання про генеральну сукупність, яке перевіряється за результатами аналізу вибірки.

Статистичні гіпотези підрозділяють на гіпотези про закони розподілу та гіпотези про параметри розподілу. Далі будемо розглядати лише перші з них. Гіпотези про параметри розподілу викладені у більш повних виданнях [2, 3, 5, 11].

Нехай треба перевірити гіпотезу H_0 про те, що випадкова величина X має закон розподілу $F(x)$. По вибірковим даним можна побудувати емпіричний розподіл $F^*(x)$ цієї випадкової величини. Порівняння $F(x)$ та $F^*(x)$ проводиться за допомогою спеціально підібраної випадкової величини – статистичного критерію узгодженості, який характеризує міру відхилення вибіркового розподілу від теоретичного. Існує кілька таких критеріїв: "хі-квадрат" Пірсона (χ^2), Колмогорова, Смирнова та ін.

5.1. Критерій χ^2 Пірсона

Розіб'ємо всю область зміни випадкової величини X на l інтервалів $\Delta_i = x_{i-1} \ldots x_i,\ i = 1, 2, \ldots, l$ (x_i – межі інтервалів) і підрахуємо кількість n_i значень випадкової величини, які попали в кожний із інтервалів Δ_i.

Вважаючи теоретичний закон розподілу відомим, завжди можна визначити ймовірність попадання значень випадкової величини в інтервали Δ_i:

$$p_i = \int_{x_{i-1}}^{x_i} f(x)dx = F(x_i) - F(x_{i-1}), \qquad (41)$$

де $f(x)$ – щільність розподілу. Тепер теоретичну кількість m_i значень випадкової величини X, які попали в кожний із інтервалів Δ_i знаходимо за формулою: $m_i = np_i$ (n – об'єм вибірки). Результати обчислень заносимо до таблиці.

Таблиця 3

Номер інтервалу $-\ i$	1	2	$\cdots$	i	$\cdots$	l
Емпіричні частоти – n_i	n_1	n_2	$\cdots$	n_i	$\cdots$	n_l
Теоретичні частоти – m_i	m_1	m_2	$\cdots$	m_i	$\cdots$	m_l

За критерій, що характеризує ступінь розходження між теоретичними та емпіричними розподілами приймають:

$$\chi^2 = \sum_{i=1}^{l} \frac{(n_i - np_i)^2}{np_i} = \sum_{i=1}^{l} \frac{(n_i - m_i)^2}{m_i}. \qquad (42)$$

Очевидно, що чим більше χ^2, тим більше розходження і тим більше підстав відкинути висунуту гіпотезу.

Статистика (42) має розподіл, який називають "хі-квадрат" з $\kappa = l - r - 1$ ступенями вільності. Тут l – число інтервалів, а r – число параметрів розподілу $F(x)$, який приймається гіпотезою. Наприклад, для показникового закону розподілу[2] маємо один параметр – λ ($r = 1$), а для нормального[3] – два параметри a і σ ($r = 2$). Для зручності використання розподіл "хі-квадрат" затабульовано – додаток 4 (стор. 54).

Правило застосування χ^2 – критерію.

Розрахувавши значення χ^2 за формулою (42) та ступінь вільності κ і задавшись рівнем значущості $\gamma = 1 - \alpha$, з таблиці χ^2-розподілу по значенням κ і γ знаходять критичне значення $\chi^2_{\kappa,\gamma}$ критерію. Якщо $\chi^2 < \chi^2_{\kappa,\gamma}$, то гіпотезу H_0 приймають, в противному разі – відхиляють.

---

[2] $f(x) = \begin{cases} \lambda e^{-\lambda x} & при\ x > 0 \\ 0 & при\ x < 0 \end{cases}$ – щільність показникового розподілу.

[3] $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$ – щільність нормального розподілу.

П р и м і т к а. Умовою успішного застосування χ^2 - критерію є наявність не менше 5 значень випадкової величини в кожному інтервалі таблиці 3. Тому, якщо деякі частоти n_i виявляться меншими ніж 5, то такі інтервали (стовпці таблиці) об'єднують так, щоб зазначена умова була виконана.

Застосуємо тепер критерій Пірсона до випадкової величини, заданої таблицею 2. По вигляду гістограми та полігону розподілу (рис.2) приймаємо гіпотезу H_0: "**Випадкова величина розподілена за нормальним законом**". Тоді щільність розподілу матиме вигляд:

$$f(x) = \frac{1}{S_X\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2S_X^2}} . \qquad (43)$$

Враховуючи, знайдені раніше, числові значення величин $\bar{x}$ та S_X ($\bar{x} = 20,174$; $S_X = 0,064$), одержимо:

$$f(x) = \frac{1}{0,064\sqrt{2\pi}} e^{-\frac{(x-20,174)^2}{2\cdot 0,064^2}} = 6,273 e^{-123,61(x-20,174)^2} . \quad (44)$$

За прийнятою залежністю на рис.5 побудовано графік щільності ймовірностей для досліджуваної випадкової величини. Там же за емпіричними даними зображено полігон щільності відносних частот ($n_i/(nh) = W_i/h$, $n = 100$ – об'єм вибірки, $h = 0,05$ – крок інтервалу).

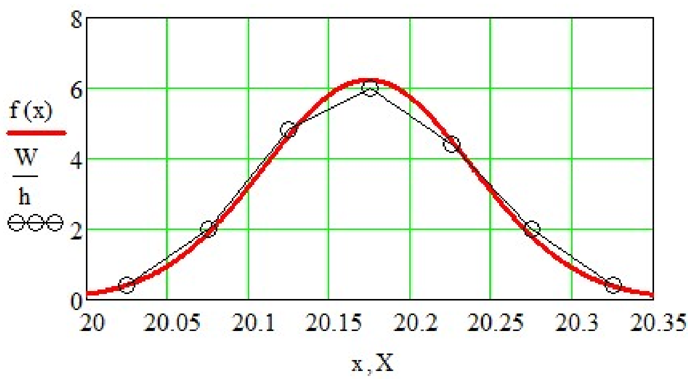


Рис.5

Візуально спостерігається достатньо хороша збіжність теоретичних і емпіричних даних, але впевнено про це можна стверджувати лише після вивчення питання на основі критерію узгодженості.

Для використання критерію Пірсона складаємо таблицю емпіричних та теоретичних частот. Емпіричні частоти приведені в таблиці 2. Теоретичні частоти знаходимо наближено за формулою:

$$np_i \approx nhf(x_i) = 100 \cdot 0,05 \cdot f(x_i) = 5 \cdot f(x_i), \qquad (45),$$

де $x_i, i = 1, 2, \ldots, 7$ – середини інтервалів.

Нижче наведені приклади знаходження теоретичних частот і таблиця 4 результатів обчислень. Одержані значення заносимо в таблицю після округлення до цілого.

$$np_1 = 5 \cdot f(20,025) = 5 \cdot 6,273 e^{-123,61(20,025-20,174)^2} = 5 \cdot 0,403 = 2,02;$$

$$np_2 = 5 \cdot f(20,075) = 5 \cdot 6,273 e^{-123,61(20,075-20,174)^2} = 5 \cdot 1,868 = 9,34;$$

. . . .

$$np_7 = 5 \cdot f(20,325) = 5 \cdot 6,273 e^{-123,61(20,325-20,174)^2} = 5 \cdot 0,374 = 1,87.$$

Таблиця 4

Емпіричні частоти - n_i	2	10	24	30	22	10	2
Теоретичні частоти - np_i	2	9	23	31	23	9	2

Для успішного застосування критерію необхідно, щоб на кожний інтервал припадало не менше 5 спостережень. Ця вимога буде виконана, якщо об’єднати окремо два лівих і два правих стовпці таблиці 4. Одержимо наступну таблицю, яка включає $l = 5$ стовпців числових даних.

Таблиця 5

Емпіричні частоти - n_i	12	24	30	22	12
Теоретичні частоти - np_i	11	23	31	23	11

Відповідно до таблиці 5 за формулою (42) обчислюємо критерій χ^2 Пірсона:

$$\chi^2 = \sum_{i=1}^{l} \frac{\left(n_i - np_i\right)^2}{np_i} = \frac{\left(12-11\right)^2}{11} + \frac{\left(24-23\right)^2}{23} +$$

$$+ \frac{\left(30-31\right)^2}{31} + \frac{\left(22-23\right)^2}{23} + \frac{\left(12-11\right)^2}{11} = 0,28$$

При рівні значущості $\gamma = 0,05$ і ступені вільності $k = l - r - 1 = 5 - 2 - 1 = 2$ ($r = 2$) з таблиці (додаток 4, стор.54) знаходимо критичне значення критерію: $\chi^2_{\gamma,k} = 6,0$.

Одержані результати дозволяють сформулювати наступний **висновок**: *оскільки* $\chi^2 < \chi^2_{\gamma,k}$ *(*$0,28 < 6,0$*), емпіричні дані з надійністю не меншою 0,95 не суперечать теоретичним, тому відкидати гіпотезу немає підстав.*

5.2. Критерій Колмогорова

В цьому критерії, для виявлення ступені відповідності емпіричних і теоретичних даних, порівнюються значення теоретичної функції розподілу $F(x)$ (прийнятою нульовою гіпотезою H_0) і емпіричної – $F^*(x)$ (функції накоплення відносних частот).

Для застосування критерію слід знайти різниці вказаних функцій $D_i = F^*(x_i) - F(x_i)$, $i = 0, 1, 2, \ldots, l$ і визначити, яка з цих різниць найбільша за абсолютною величиною, тобто:

$$D_{\max} = \max\left\{\left|F^*(x_i) - F(x_i)\right|\right\},\ i = 0, 1, 2, \ldots, l, \tag{46}$$

де l – число інтервалів (1), x_i – межі інтервалів.

Мірою розходження теоретичного і емпіричного розподілів обрана величина:

$$\lambda = D_{\max}\sqrt{n}, \tag{47}$$

де n – об'єм вибірки.

А. Н. Колмогоров показав, що ймовірність $P(\lambda)$ нерівності $D_{\max}\sqrt{n} \ge \lambda$ при $n \to \infty$ прямує до границі:

$$P(\lambda) = 1 - \sum_{k=-\infty}^{\infty} (-1)^k \, \mathrm{e}^{-2k^2\lambda^2} . \tag{48}$$

При цьому, якщо ця ймовірність мала, то це означає, що розходження між теоретичними і емпіричними даними надто великі для того, щоб їх можна було вважати випадковими і в цьому разі гіпотезу H_0 про закон розподілу слід відкинути. Якщо ж ймовірність $P(\lambda)$ досить велика, то це доводить випадковість розбіжності вказаних даних і також – обґрунтованість висунутої гіпотези. Значення ймовірності (48) приведені в таблиці додатку 5 (стор. 55).

Правило застосування критерію Колмогорова.

Визначають різниці теоретичного і емпіричного законів розподілу та вибирають найбільшу з них – $D_{\max}$ (46).

Обчислюють величину λ за формулою (47) та знаходять табличне значення ймовірності $P(\lambda)$ (додаток 5, стор. 55).

Висунуту гіпотезу H_0 про закон розподілу приймають, якщо знайдена ймовірність $P(\lambda)$ виявиться достатньо великою, в противному разі гіпотезу відкидають.

П р и м і т к а. При малих об'ємах вибірки цей критерій дає дещо завищені оцінки правдоподібності гіпотези. В цих випадках існує ризик вважати правдоподібною гіпотезу, яка незадовільно погоджується з дослідними даними.

Тепер, для спрощення викладок і можливості більш наглядного порівняння методики впровадження різних критеріїв узгодженості, розглянемо застосування критерію Колмогорова на прикладі тієї ж випадкової величини (див. табл.2), для якої раніше був застосований критерій Пірсона. Як уже відомо, за виглядом гістограми та полігону розподілу (рис.2) можна висунути гіпотезу $\boldsymbol{H_0}$: "**Випадкова величина розподілена за нормальним законом**". Функцію розподілу для цього закону можна представити у вигляді [6]:

$$F(x) = 0,5 + \Phi\left(\frac{x - \overline{x}}{S_X}\right). \tag{49}$$

Тут $\Phi(t)$ – функція Лапласа, значення якої можна вибрати по аргументу t з таблиці додатку 1 (стор. 52). При цьому слід враховувати, що ця функція непарна $\Phi(-t) = -\Phi(t)$, тому в таблиці наведені її значення лише для додатних значень аргумента.

За формулою (49) підраховуємо значення теоретичної функції розподілу в точках x_i, $i = 0, 1, 2, \ldots, 7$:

$$F(x_0) = 0{,}5 + \Phi\left(\frac{20{,}00 - 20{,}174}{0{,}064}\right) = 0{,}5 + \Phi\left(-2{,}72\right) = 0{,}5 - 0{,}49674 = 0{,}003;$$

$$F(x_1) = 0{,}5 + \Phi\left(\frac{20{,}05 - 20{,}174}{0{,}064}\right) = 0{,}5 + \Phi\left(-1{,}94\right) = 0{,}5 - 0{,}47381 = 0{,}026;$$

. . . .

$$F(x_7) = 0{,}5 + \Phi\left(\frac{20{,}35 - 20{,}174}{0{,}064}\right) = 0{,}5 + \Phi\left(2{,}75\right) = 0{,}5 + 0{,}49702 = 0{,}997;$$

Одержані результати заносимо в таблицю (четвертий рядок). Продублюємо також в третьому рядку таблиці значення емпіричної функції розподілу, що спрощує знаходження різниць емпіричної та теоретичної функцій за абсолютною величиною (п'ятий рядок).

П р и м і т к а. На початку діапазону при $x_0 = 20{,}00$ емпірична функція розподілу має нульове значення $F^*(x_0) = 0$, що витікає з принципу побудови цієї функції (стор.14). Це може привести до нехарактерного (збільшеного) значення різниці функцій (46) та неадекватного застосування критерію, коли гіпотеза H_0 не признається тоді, коли насправді вона є правдоподібною. Тому результати розрахунків у першому стовпчику слід відкинути, якщо вони є підозрілими.

Таблиця 6

i	0	1	2	3	4	5	6	7
x_i	20,00	20,05	20,10	20,15	20,20	20,25	20,30	20,35
$F^*(x_i)$	0	0.02	0,12	0,36	0,66	0,88	0,98	1
$F(x_i)$	0,003	0,026	0,123	0,352	0,659	0,883	0,976	0,997
$\lvert D_i \rvert$	0,003	0,006	0,003	0,008	0,001	0,003	0,004	0,003

З п'ятого рядка таблиці 6 знаходимо $D_{\max} = 0,008$, що дає:

$$\lambda = D_{\max}\sqrt{n} = 0,008\sqrt{100} = 0,08$$

З таблиці додатку 5 (стор. 55) одержуємо: $P(0,08) = 1$. Отже, маємо наступний **висновок**: *ймовірність $P(\lambda)$ (48) є максимально можливою, тому прийняття висунутої гіпотези про нормальний закон розподілу є правомірним.*

Далі для закріплення викладеного матеріалу наводимо індивідуальні завдання з прикладом їх виконання. Допоміжно дається зразок виконання завдання в середовищі «Mathcad». Перевага використання системи «Mathcad» полягає в суттєвому спрощенні і розширенні можливостей обчислень, а також набування студентами досвіду застосування автоматизованих обчислювальних систем в процесі навчання.

6. Варіанти індивідуальних завдань

За даними вибіркової сукупності складено інтервальний ряд. Користуючись цими даними потрібно:

1) знайти вибіркові оцінки числових характеристик випадкової величини (вибіркове середнє значення – $\bar{x}$, вибіркову дисперсію – S_X^2 та стандартне відхилення – S_X, вибірковий коефіцієнт варіації – V_X, вибіркові асиметрію – A_X та ексцес – E_X);

2) встановити інтервали довіри для генерального середнього а також середнього квадратичного відхилення;

3) побудувати гістограму щільності відносних частот;

4) по вигляду гістограми відносних частот прийняти нульову гіпотезу про закон розподілу та перевірити її за допомогою критерію узгодженості Пірсона або Колмогорова.

Варіанти числових даних взяти з таблиці 7, де варіант вказано в першому стовпчику таблиці.

Таблиця 7

1	2	3	4	5	6	7	8	9
1	*Інтервал*	-4÷-2	-2÷0	0÷2	2÷4	4÷6	6÷8	8÷10
	Частота n_i	8	14	26	22	15	10	5
2	*Інтервал*	1÷4	4÷7	7÷10	10÷13	13÷16	16÷19	19÷22
	Частота n_i	6	15	28	21	16	9	5
3	*Інтервал*	-2÷0	0÷2	2÷4	4÷6	6÷8	8÷10	10÷12
	Частота n_i	4	10	27	24	18	10	7
4	*Інтервал*	0÷3	3÷6	6÷9	9÷12	12÷15	15÷18	18÷21
	Частота n_i	9	11	29	25	13	8	5
5	*Інтервал*	-1÷1	1÷3	3÷5	5÷7	7÷9	9÷11	11÷13
	Частота n_i	7	15	28	20	15	10	5
6	*Інтервал*	-3÷-1	-1÷1	1÷3	3÷5	5÷7	7÷9	9÷11
	Частота n_i	8	17	28	20	14	8	5
7	*Інтервал*	4÷6	6÷8	8÷10	10÷12	12÷14	14÷16	16÷18
	Частота n_i	7	14	21	26	17	10	5
8	*Інтервал*	-1,5÷0	0÷1,5	1,5÷3	3÷4,5	4,5÷6	6÷7,5	7,5÷9
	Частота n_i	6	13	23	27	15	11	5
9	*Інтервал*	-3÷0	0÷3	3÷6	6÷9	9÷12	12÷15	15÷18
	Частота n_i	8	15	22	27	16	7	5
10	*Інтервал*	2÷4	4÷6	6÷8	8÷10	10÷12	12÷14	14÷16
	Частота n_i	6	12	21	25	18	11	7
11	*Інтервал*	6÷8	8÷10	10÷12	12÷14	14÷16	16÷18	18÷20
	Частота n_i	5	14	28	19	16	10	8
12	*Інтервал*	5÷8	8÷11	11÷14	14÷17	17÷20	20÷23	23÷26
	Частота n_i	7	14	22	27	15	9	6
13	*Інтервал*	2,5÷4	4÷5,5	5,5÷7	7÷8,5	8,5÷10	10÷11,5	11,5÷13
	Частота n_i	6	12	28	21	17	11	5
14	*Інтервал*	-4÷0	0÷4	4÷8	8÷12	12÷16	16÷20	20÷24
	Частота n_i	7	18	30	24	9	7	5
15	*Інтервал*	-1÷1	1÷3	3÷5	5÷7	7÷9	9÷11	11÷13
	Частота n_i	8	15	23	24	15	10	5

1	2	3	4	5	6	7	8	9
16	*Інтервал*	2÷5	5÷8	8÷11	11÷14	14÷17	17÷20	20÷23
	Частота n_i	9	11	24	26	14	9	7
17	*Інтервал*	4÷6	6÷8	8÷10	10÷12	12÷14	14÷16	16÷18
	Частота n_i	8	15	21	25	16	10	5
18	*Інтервал*	0÷3	3÷6	6÷9	9÷12	12÷15	15÷18	18÷21
	Частота n_i	5	14	24	28	17	7	5
19	*Інтервал*	1÷3	3÷5	5÷7	7÷9	9÷11	11÷13	13÷15
	Частота n_i	5	14	22	27	17	10	5
20	*Інтервал*	-4÷-1	-1÷2	2÷5	5÷8	8÷11	11÷14	14÷17
	Частота n_i	7	17	25	23	15	8	5
21	*Інтервал*	2÷4	4÷6	6÷8	8÷10	10÷12	12÷14	14÷16
	Частота n_i	5	11	32	24	16	7	5
22	*Інтервал*	4÷6	6÷8	8÷10	10÷12	12÷14	14÷16	16÷18
	Частота n_i	6	17	28	21	14	8	6
23	*Інтервал*	6÷8	8÷10	10÷12	12÷14	14÷16	16÷18	18÷20
	Частота n_i	6	17	28	21	12	9	7
24	*Інтервал*	8÷10	10÷12	12÷14	14÷16	16÷18	18÷20	20÷22
	Частота n_i	5	16	27	23	16	8	5
25	*Інтервал*	-4÷-2	-2÷0	0÷2	2÷4	4÷6	6÷8	8÷10
	Частота n_i	7	15	25	28	13	7	5
26	*Інтервал*	-2÷1	1÷4	4÷7	7÷10	10÷13	13÷16	16÷19
	Частота n_i	6	14	28	23	13	10	6
27	*Інтервал*	-4÷0	0÷4	4÷8	8÷12	12÷16	16÷20	20÷24
	Частота n_i	6	15	25	20	16	11	7
28	*Інтервал*	-1÷1	1÷3	3÷5	5÷7	7÷9	9÷11	11÷13
	Частота n_i	8	17	29	21	14	6	5
29	*Інтервал*	8÷10	10÷12	12÷14	14÷16	16÷18	18÷20	20÷22
	Частота n_i	5	14	28	23	13	10	7
30	*Інтервал*	-3÷0	0÷3	3÷6	6÷9	9÷12	12÷15	15÷18
	Частота n_i	5	16	28	21	14	10	6

1	2	3	4	5	6	7	8	9
31	*Інтервал*	-3÷-1	-1÷1	1÷3	3÷5	5÷7	7÷9	9÷11
	Частота n_i	8	14	27	21	15	9	6
32	*Інтервал*	0÷2	2÷4	4÷6	6÷8	8÷10	10÷12	12÷14
	Частота n_i	6	15	27	22	16	10	4
33	*Інтервал*	-2÷0	0÷2	2÷4	4÷6	6÷8	8÷10	10÷12
	Частота n_i	5	9	28	23	18	11	6
34	*Інтервал*	1÷3	3÷5	5÷7	7÷9	9÷11	11÷13	13÷15
	Частота n_i	7	13	28	26	13	8	5
35	*Інтервал*	-1÷2	2÷5	5÷8	8÷11	11÷14	14÷17	17÷20
	Частота n_i	7	14	27	21	16	9	6
36	*Інтервал*	-3÷0	0÷3	3÷6	6÷9	9÷12	12÷15	15÷18
	Частота n_i	7	18	29	19	15	7	5
37	*Інтервал*	4÷6	6÷8	8÷10	10÷12	12÷14	14÷16	16÷18
	Частота n_i	5	13	22	28	17	9	6
38	*Інтервал*	-1,5÷0	0÷1,5	1,5÷3	3÷4,5	4,5÷6	6÷7,5	7,5÷9
	Частота n_i	5	12	23	30	18	7	5
39	*Інтервал*	-2÷0	0÷2	2÷4	4÷6	6÷8	8÷10	10÷12
	Частота n_i	8	15	26	20	16	8	5
40	*Інтервал*	1÷3	3÷5	5÷7	7÷9	9÷11	11÷13	13÷15
	Частота n_i	6	12	24	28	15	9	6
41	*Інтервал*	6÷8	8÷10	10÷12	12÷14	14÷16	16÷18	18÷20
	Частота n_i	6	15	27	19	16	11	6
42	*Інтервал*	4÷7	7÷10	10÷13	13÷16	16÷19	19÷22	22÷25
	Частота n_i	7	17	24	28	12	7	5
43	*Інтервал*	2÷4	4÷6	6÷8	8÷10	10÷12	12÷14	14÷16
	Частота n_i	6	15	28	21	15	10	5
44	*Інтервал*	-4÷0	0÷4	4÷8	8÷12	12÷16	16÷20	20÷24
	Частота n_i	7	14	28	24	15	7	5
45	*Інтервал*	-1÷1	1÷3	3÷5	5÷7	7÷9	9÷11	11÷13
	Частота n_i	8	15	27	24	12	9	5

1	2	3	4	5	6	7	8	9
46	*Інтервал*	2÷5	5÷8	8÷11	11÷14	14÷17	17÷20	20÷23
	Частота n_i	9	11	22	30	12	9	7
47	*Інтервал*	4÷6	6÷8	8÷10	10÷12	12÷14	14÷16	16÷18
	Частота n_i	5	11	21	28	19	10	6
48	*Інтервал*	0÷2	2÷4	4÷6	6÷8	8÷10	10÷12	12÷14
	Частота n_i	9	13	21	29	12	9	7
49	*Інтервал*	-1÷1	1÷3	3÷5	5÷7	7÷9	9÷11	11÷13
	Частота n_i	5	12	23	30	15	9	6
50	*Інтервал*	-4÷-2	-2÷0	0÷2	2÷4	4÷6	6÷8	8÷10
	Частота n_i	7	15	21	28	16	8	5
51	*Інтервал*	-3÷-1	-1÷1	1÷3	3÷5	5÷7	7÷9	9÷11
	Частота n_i	5	12	20	28	19	10	6
52	*Інтервал*	-1÷1	1÷3	3÷5	5÷7	7÷9	9÷11	11÷13
	Частота n_i	7	15	28	21	15	8	6
53	*Інтервал*	0÷2	2÷4	4÷6	6÷8	8÷10	10÷12	12÷14
	Частота n_i	5	14	21	28	16	11	5
54	*Інтервал*	3÷5	5÷7	7÷9	9÷11	11÷13	13÷15	15÷17
	Частота n_i	7	14	21	26	18	9	5
55	*Інтервал*	-2÷0	0÷2	2÷4	4÷6	6÷8	8÷10	10÷12
	Частота n_i	5	12	20	29	22	8	4
56	*Інтервал*	-4÷-2	-2÷0	0÷2	2÷4	4÷6	6÷8	8÷10
	Частота n_i	7	11	22	27	20	8	5
57	*Інтервал*	3÷5	5÷7	7÷9	9÷11	11÷13	13÷15	15÷17
	Частота n_i	5	13	21	28	19	8	6
58	*Інтервал*	6÷8	8÷10	10÷12	12÷14	14÷16	16÷18	18÷20
	Частота n_i	7	15	24	28	12	9	5
59	*Інтервал*	0÷3	3÷6	6÷9	9÷12	12÷15	15÷18	18÷21
	Частота n_i	4	11	20	31	18	10	6
60	*Інтервал*	5÷7	7÷9	9÷11	11÷13	13÷15	15÷17	17÷19
	Частота n_i	5	12	28	21	17	11	6

6.1. Приклад виконання індивідуального завдання

Інтервал	$5\div7$	$7\div9$	$9\div11$	$11\div13$	$13\div15$	$15\div17$	$17\div19$
Частота n_i	4	11	20	31	18	10	3

Добудуємо задану таблицю, добавивши рядки для середини інтервалів – x_i, відносних частот – $W_i = n_i / n$ та щільності – $p_w = W_i / h$ відносних частот. Для цього знайдемо об'єм вибірки.

$$n = \sum_{n=1}^{7} n_i = 4 + 11 + 20 + 31 + 18 + 10 + 3 = 100$$

Крок інтервалу складає: $h = 2$.

Таблиця а

Інтервал	$5\div7$	$7\div9$	$9\div11$	$11\div13$	$13\div15$	$15\div17$	$17\div19$
Частота n_i	4	11	20	31	18	10	6
Середини інтервалів: x_i	6	8	10	12	14	16	18
Відносна частота: W_i	0,04	0,11	0,20	0,31	0,18	0,10	0,06
Щільність відносних частот: p_w	0,02	0,055	0,10	0,155	0,09	0,05	0,03

1. Знаходимо вибіркові оцінки числових характеристик випадкової величини:

– середнє значення

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \frac{1}{100}\big(6\cdot4 + 8\cdot11 + 10\cdot20 + 12\cdot31 + 14\cdot18 +$$

$$+16\cdot10 + 18\cdot6\big) = 12,04$$

– вибіркова дисперсія

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \cdot n_i = \frac{1}{99}\Big[(6-12,04)^2\cdot4 + (8-12,04)^2\cdot11 + (10-12,04)^2\cdot20 +$$

$$+(12-12,04)^2\cdot31 + (14-12,04)^2\cdot18 + (16-12,04)^2\cdot10 + (18-12,04)^2\cdot6\Big] = 8,564$$

– стандартне відхилення

$$S_X=\sqrt{S_X^2}=\sqrt{8,564}=2,926$$

– вибірковий коефіцієнт варіації

$$V_X=\frac{S_X}{\overline{x}}=\frac{2,926}{12,04}=0,243$$

– вибіркова асиметрія

$$A_X=\frac{\frac{1}{n}\sum_{i=1}^{k}\left(x_i-\overline{x}\right)^3 n_i}{S_X^3}=\frac{1}{100\cdot 2,926^3}\Big[\left(6-12,04\right)^3\cdot 4+\left(8-12,04\right)^3\cdot 11+$$

$$+\left(10-12,04\right)^3\cdot 20+\left(12-12,04\right)^3\cdot 31+\left(14-12,04\right)^3\cdot 18+$$

$$+\left(16-12,04\right)^3\cdot 10+\left(18-12,04\right)^3\cdot 6\Big]=0,100$$

– вибірковий ексцес

$$E_X=\frac{\frac{1}{n}\sum_{i=1}^{k}\left(x_i-\overline{x}\right)^4 n_i}{S_X^4}-3=\frac{1}{100\cdot 2,926^4}\Big[\left(6-12,04\right)^4\cdot 4+\left(8-12,04\right)^4\cdot 11+$$

$$+\left(10-12,04\right)^4\cdot 20+\left(12-12,04\right)^4\cdot 31+\left(14-12,04\right)^4\cdot 18+$$

$$+\left(16-12,04\right)^4\cdot 10+\left(18-12,04\right)^4\cdot 6\Big]-3=-0,422$$

2. Встановлюємо інтервал довіри для генерального середнього (математичного сподівання). Для цього приймаємо надійність $\alpha=0,95$. З таблиці додатку 2 (стор. 53) по об'єму вибірки ($n=100$) і значенню надійності знаходимо квантиль розподілу Стьюдента $t_\alpha=1,984$. Далі обчислюємо:

– ліва межа інтервалу

$$\overline{x}-\frac{t_\alpha S_X}{\sqrt{n}}=12,04-\frac{1,984\cdot 2,926}{\sqrt{100}}=11,46$$

– права межа інтервалу

$$\overline{x}+\frac{t_\alpha S_X}{\sqrt{n}}=12,04+\frac{1,984\cdot 2,926}{\sqrt{100}}=12,62$$

Отже, інтервал довіри для генерального середнього має вид: $\boxed{11,46<a<12,62}$.

Встановлюємо тепер інтервал довіри для середнього квадратичного відхилення. По, вказаним вище, об'єму вибірки і значенню надійності з таблиці додатку 3 (стор. 53) знайдемо величину $q = 0{,}143$. Тоді:

– ліва межа інтервалу

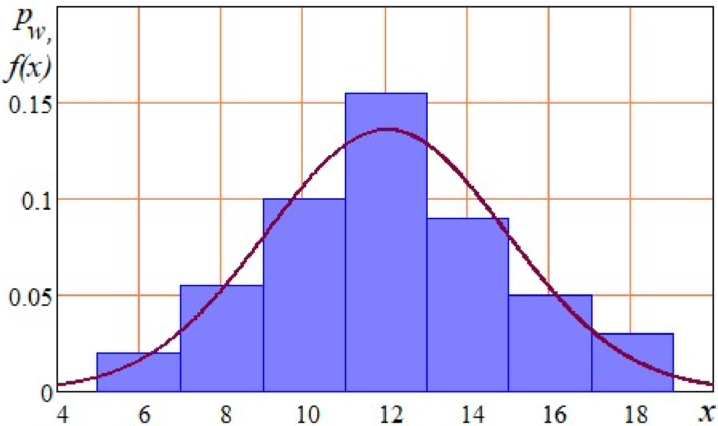
$$S_X(1-q) = 2{,}926(1-0{,}143) = 2{,}51$$

– права межа інтервалу

$$S_X(1+q) = 2{,}926(1+0{,}143) = 3{,}35$$

Інтервал довіри для середнього квадратичного відхилення: $\boxed{2{,}51 < q < 3{,}35}$.

3. За даними *таблиці а* будуємо гістограму щільності відносних частот.



4. Визначені вище, вибіркові оцінки асиметрії та ексцесу близькі до нуля. Це значить, що розподіл може бути симетричним, а його гостровершинність – відподати нормальному закону. Крім того, сама гістограма по формі нагадує нормальний закон. Спираючись на це, висуваємо робочу гіпотезу $\boldsymbol{H_0}$ **– "випадкова величина розподілена за нормальним законом".** В нашому випадку цей закон можна записати так:

$$f(x) = \frac{1}{\sqrt{2\pi}S_X} e^{-\frac{(x-\bar{x})^2}{2S_X^2}} = \frac{1}{2{,}926\sqrt{2\pi}} e^{-\frac{(x-12{,}04)^2}{2\cdot 2{,}926^2}} = 0{,}136\, e^{-\frac{(x-12{,}04)^2}{17{,}123}}$$

Для побудови графіка підрахуємо значення функції в середині кожного інтервалу.

$$f\left(x_1\right)=0{,}136\,\mathrm{e}^{-\frac{(6-12{,}04)^2}{17{,}123}}=0{,}016;\quad f\left(x_2\right)=0{,}136\,\mathrm{e}^{-\frac{(8-12{,}04)^2}{17{,}123}}=0{,}052;$$

$$f\left(x_3\right)=0{,}136\,\mathrm{e}^{-\frac{(10-12{,}04)^2}{17{,}123}}=0{,}107;\quad f\left(x_4\right)=0{,}136\,\mathrm{e}^{-\frac{(12-12{,}04)^2}{17{,}123}}=0{,}136;$$

$$f\left(x_5\right)=0{,}136\,\mathrm{e}^{-\frac{(14-12{,}04)^2}{17{,}123}}=0{,}109;\quad f\left(x_6\right)=0{,}136\,\mathrm{e}^{-\frac{(16-12{,}04)^2}{17{,}123}}=0{,}054;$$

$$f\left(x_7\right)=0{,}136\,\mathrm{e}^{-\frac{(18-12{,}04)^2}{17{,}123}}=0{,}017.$$

Знайдені значення використаємо при зображенні нормального закону розподілу, суміщеного з гістограмою. Сумісне порівняння вказаних графіків також говорить на користь прийнятої гіпотези.

Для подальшого обґрунтування гіпотези застосуємо критерій Пірсона. Обчислимо теоретичні значення частот попадання значень випадкової величини на кожний інтервал за формулою: $np_i \approx nhf(x_i)=200f(x_i)$. Тут ймовірність p_i попадання значень випадкової величини на кожний інтервал обчислюється наближено, як площа стовпчика шириною h і висотою $f(x_i)$. Значення функції $f(x_i)$ одержано вище. Маємо:

$$np_1=200\cdot 0{,}016=3;\quad np_2=200\cdot 0{,}052=10;\quad np_3=200\cdot 0{,}107=21;$$

$$np_4=200\cdot 0{,}136=27;\quad np_5=200\cdot 0{,}109=22;\quad np_6=200\cdot 0{,}054=11;$$

$$np_7=200\cdot 0{,}017=3.$$

Результати обчислень, округлені до цілого, заносим в *Таблицю б.*.

Таблиця б

Емпірична частота n_i	4	11	20	31	18	10	6
Теоретична частота np_i	3	10	21	27	22	11	3

Для того, щоб частоти в стовпчиках даних були не меншими п'яти, слід об'єднати два лівих і два правих стовпці. Тоді одержимо *таблицю в*.

Таблиця в

Емпірична частота n_i	15	20	31	18	16
Теоретична частота np_i	13	21	27	22	14

Розраховуємо фактичне значення критерію Пірсона за даними *таблиці в*.

$$\chi^2_{факт} = \sum_{i=1}^{5} \frac{(n_i - np_i)^2}{np_i} = \frac{(15-13)^2}{13} + \frac{(20-21)^2}{21} + \frac{(31-27)^2}{27} +$$

$$+ \frac{(18-22)^2}{22} + \frac{(16-14)^2}{14} = 1,96$$

Приймаємо рівень значущості $\gamma = 0,05$ та знаходимо ступінь вільності $k = \ell - r - 1 = 5 - 2 - 1 = 2$. Тут $\ell = 5$ – число стовпчиків даних таблиці 3, а $r = 2$ – число параметрів нормального закону розподілу (параметрами є математичне сподівання "a" та середнє квадратичне відхилення "σ").

За значеннями $\gamma = 0,05$ та $k = 2$ з таблиці додатку 4 (стор. 54) знаходимо критичне значення критерію: $\chi^2_{крит} = 6,0$.

Порівнюючи знайдені значення χ^2 критерію, приходимо до такого **висновку:** *оскільки* $\chi^2_{факт} < \chi^2_{крит}$ *відкидати гіпотезу* $\boldsymbol{H_0}$ *немає підстав.*

Застосуємо тепер для обґрунтування прийнятої гіпотези **критерій узгодженості Колмогорова**. Для цього обчислимо різниці значень емпіричної і теоретичної функцій розподілу в на кінцях інтервалів x_i, $i = 0, 1, 2, \ldots, 7$. Результати розрахунків будем записувати в наступну таблицю.

Таблиця г

Кінці інтервалів x_i	5	7	9	11	13	15	17	19
Емпірична функція розподілу $F^*(x_i)$	0	0,04	0,15	0,35	0,66	0,84	0,94	1
Теоретична функція розподілу $F(x)$	0,008	0,043	0,149	0,359	0,629	0,844	0,954	0,991
Параметр D	0,008	0,003	0,001	0,009	0,031	0,004	0,014	0,009

Значення емпіричної функції розподілу в лівій точці $x_0 = 5$ дорівнює нулю: властивість 3 (стор.12). Значення в інших точках можна знайти за формулою, яка витікає зі смислу цієї функції, як функції накоплених відносних частот:

$$F^*(x_i) = \sum_{k=1}^{i} W_k, \; i = 1, 2, \ldots, 7.$$

Маємо: $F^*(7) = 0,04$; $F^*(9) = 0,04 + 0,11 = 0,15$;

$F^*(11) = 0,15 + 0,20 = 0,35$; $F^*(13) = 0,35 + 0,31 = 0,66$;

$F^*(15) = 0,66 + 0,18 = 0,84$; $F^*(17) = 0,84 + 0,10 = 0,94$;

$F^*(19) = 0,94 + 0,06 = 1$.

Значення теоретичної функції розподілу визначимо за формулою (49). При цьому значення функції Лапласа знайдемо з таблиці додатку 1 (стор. 52).

$$F(5) = 0,5 + \Phi\left(\frac{5 - 12,04}{2,926}\right) = 0,5 + \Phi(-2,41) = 0,5 - 0,492 = 0,008;$$

$$F(7) = 0,5 + \Phi\left(\frac{7 - 12,04}{2,926}\right) = 0,5 + \Phi(-1,72) = 0,5 - 0,457 = 0,043;$$

$$F(9) = 0,5 + \Phi\left(\frac{9 - 12,04}{2,926}\right) = 0,5 + \Phi(-1,04) = 0,5 - 0,351 = 0,149;$$

$$F(11) = 0,5 + \Phi\left(\frac{11 - 12,04}{2,926}\right) = 0,5 + \Phi(-0,36) = 0,5 - 0,141 = 0,359;$$

$$F(13) = 0,5 + \Phi\left(\frac{13 - 12,04}{2,926}\right) = 0,5 + \Phi(0,33) = 0,5 + 0,129 = 0,629;$$

$$F(15) = 0,5 + \Phi\left(\frac{15 - 12,04}{2,926}\right) = 0,5 + \Phi(1.01) = 0,5 + 0,344 = 0,844;$$

$$F(17) = 0,5 + \Phi\left(\frac{17 - 12,04}{2,926}\right) = 0,5 + \Phi(1,69) = 0,5 + 0,454 = 0,954;$$

$$F(19) = 0,5 + \Phi\left(\frac{19 - 12,04}{2,926}\right) = 0,5 + \Phi(2,38) = 0,5 + 0,491 = 0,991.$$

Знайденими числами заповнимо третій рядок *Таблиці г.*

Різниці теоретичної і емпіричної функцій тепер можна знайти, як різниці другого і третього рядка *таблиці г.* При обчисленнях вказані різниці (46) записуємо, за абсолютним значенням, в останній рядок цієї таблиці.

З записаного рядка безпосередньо випливає: $D_{\max} = 0,031$. Отже, параметр λ дорівнює (47):

$$\lambda = D_{\max}\sqrt{n} = 0,031\sqrt{100} = 0,31 .$$

З таблиці додатку 5 (стор. 55) за знайденим параметром знаходимо ймовірність $P(\lambda) = 0,99998$. Оскільки ця ймовірність виявилася достатньо великою – *гіпотеза H_0 про нормальний закон розподілу приймається.*

6.2. Статистичні обчислення засобами «Mathcad»

6.2.1. Умови задач і створення варіантів індивідуальних завдань

Для контролю якості ознаки X при виготовленні деталей взята вибірка (таблиця X).

1. Знайти статистичні оцінки точечних характеристик генеральної сукупності ознаки X: середнє значення, виправлену дисперсію і стандартне відхилення, коєфіцієнт варіації, асиметрію та ексцес.
2. Знайти інтервальні оцінки для параметрів розподілу генеральної сукупності: математичного сподівання та середнього квадратичного відхилення.
3. Побудувати полігон та гістограму щільності відносних частот.
4. Висунути гіпотезу про закон розподілу і виконати перевірку узгодженості емпіричного і теоретичного розподілів за допомогою критерію Пірсона або Колмогорова.

Для знаходження даних використовуються можливості «Mathcad», як показано на листі 1.

ORIGIN := 1

M := rnorm(80 ,50 + Variant ,10)

$i := 1..10 \quad j := 1..8 \quad X_{i,j} := M_{(i-1)\cdot 8+j}$

Номер варіанта: Variant ≡ 39

X :=

	1	2	3	4	5	6	7	8
1	82.4	67.8	96.4	62.7	70.8	75.8	91.6	77.1
2	46.0	72.7	69.7	24.5	70.0	76.8	83.5	72.8
3	109.4	124.6	85.4	70.8	85.6	82.3	75.2	79.8
4	28.7	79.8	72.0	102.4	90.0	86.8	57.8	85.7
5	124.3	68.5	63.0	106.0	80.9	57.8	107.5	61.8
6	61.6	114.5	34.3	71.3	105.5	103.0	60.9	49.2
7	92.8	102.3	72.1	42.5	23.0	55.2	47.5	148.1
8	80.1	86.2	46.5	49.0	107.3	76.6	62.0	45.3
9	43.7	63.6	42.8	115.5	88.5	87.1	86.1	90.5
10	56.6	99.0	40.1	101.2	33.5	101.1	89.6	105.4

Лист 1

Для отримання вибірки у вигляді таблиці X (лист 1) треба на розрахунковому листі «Mathcad» набрати 5 рядків.

Перший рядок “ORIGIN = 1”– системна змінна, яка показує з якого номеру розпочинається відлік індексів індексованих даних.

В другому рядку функція “rnorm(m,a,σ)” формує масив $m = 80$ випадкових чисел, розподілених за нормальним законом з математичним сподіванням – a і середнім квадратичним відхиленням – σ, який задається у вигляді вектора M. Функція “rnorm(m,a,σ)” вибирається з категорії “випадкові числа” набору влаштованих функцій $\boldsymbol{f(x)}$.

Третій рядок перетворює вектор M в таблицю X даних, яка має 10 рядків і 8 стовпців.

В четвертому рядку враховується варіант даних, для чого змінній “Variant” присвоюється номер варіанту індивідуального завдання (в нашому прикладі цей номер: 39). При цьому слід застосовувати знак глобального присвоювання “≡”.

В п'ятому рядку відбувається вивід даних у вигляді таблиці X. Для цього достатньо набрати “X = ”.

6.2.2. Вибіркові оцінки числових характеристик ознаки X

ORIGIN := 1

X :=

	1	2	3	4	5	6	7	8
1	82.4	67.8	96.4	62.7	70.8	75.8	91.6	77.1
2	46.0	72.7	69.7	24.5	70.0	76.8	83.5	72.8
3	109.4	124.6	85.4	70.8	85.6	82.3	75.2	79.8
4	28.7	79.8	72.0	102.4	90.0	86.8	57.8	85.7
5	124.3	68.5	63.0	106.0	80.9	57.8	107.5	61.8
6	61.6	114.5	34.3	71.3	105.5	103.0	60.9	49.2
7	92.8	102.3	72.1	42.5	23.0	55.2	47.5	148.1
8	80.1	86.2	46.5	49.0	107.3	76.6	62.0	45.3
9	43.7	63.6	42.8	115.5	88.5	87.1	86.1	90.5
10	56.6	99.0	40.1	101.2	33.5	101.1	89.6	105.4

Статистичні оцінки точечних характеристик ознаки X

$x_{cp} := \text{mean}(X) = 76.3$ - середнє значення

$\min(X) = 23.02$ - мінімальне значення

$\max(X) = 148.07$ - максимальне значення

$Rx := \max(X) - \min(X) = 125.0$ - розмах значень

$SS := \text{Var}(X) = 621.3$ - незміщена (виправлена) вибіркова дисперсія

$S := \text{Stdev}(X) = 24.9$ - незміщене стандартне відхилення

$V := \dfrac{S}{x_{cp}} = 0.327$ - вибірковий коефіцієнт варіації

$A := \text{skew}(X) = 0.080$ - вибіркова асиметрія

$E := \text{kurt}(X) = -0.021$ - вибірковий ексцес

Інтервальні оцінки точечних характеристик ознаки X

Об'єм вибірки: $n := \text{rows}(X)\cdot\text{cols}(X) = 80$

Рівень надійності: $\alpha := 0.95$ Рівень значущості: $\gamma := 1 - \alpha = 0.05$

Квантиль розподілу Стьюдента: $t := qt\left(1 - \dfrac{\gamma}{2}, n - 1\right) = 1.99$

Для заданого ріівня надійності α і обєму вибірки- n знаходимо табличне значення:

$q := 0.161$

Довірливий інтервал для генерального середнього:

$$x_{cp} - t\cdot\frac{S}{\sqrt{n}} < M_x < x_{cp} + t\cdot\frac{S}{\sqrt{n}} \;\Bigg|\; \begin{array}{l}\text{float}, 3\\ \text{simplify}\end{array} \rightarrow 70.8 < M_x < 81.9 \qquad \boxed{70.8 < M_x < 81.9}$$

Довірливий інтервал для генерального стандартного відхилення:

$$S\cdot(1 - q) < \sigma_x < S\cdot(1 + q) \text{ float}, 3 \;\rightarrow 20.9 < \sigma_x < 28.9 \qquad \boxed{20.9 < \sigma_x < 28.9}$$

Лист 2

6.2.3. Перевірка статистичної гіпотези про закон розподілу ознаки X за критерієм Пірсона

$ORIGIN := 1$ $X :=$

	1	2	3	4	5	6	7	8
1	82.4	67.8	96.4	62.7	70.8	75.8	91.6	77.1
2	46.0	72.7	69.7	24.5	70.0	76.8	83.5	72.8
3	109.4	124.6	85.4	70.8	85.6	82.3	75.2	...

Об'єм вибірки: $n := rows(X) \cdot cols(X) = 80$

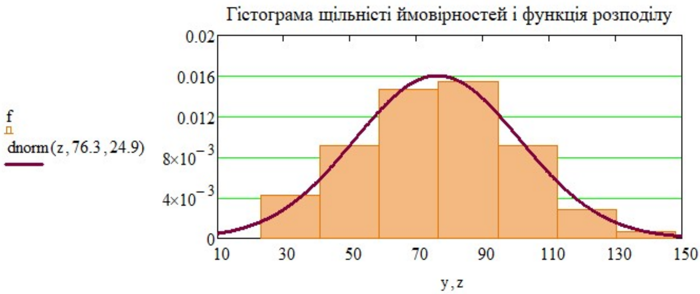
Число інтервалів за формулою Стерджеса: $Nint := round(1 + 3.322 \cdot \log(n)) = 7$

Крок інтервалу: $h := \dfrac{\max(X) - \min(X)}{Nint} = 17.9$

Межі інтервалів: $i := 1 .. Nint + 1$ $\quad x_i := \min(X) + (i-1) \cdot h$

Щільність відносних частот: $H := hist(x, X)$ $\quad H^T = (6\ \ 13\ \ 21\ \ 22\ \ 13\ \ 4\ \ 1)$ $\quad f := \dfrac{H}{n \cdot h}$

Середини інтервалів: $y := x + \dfrac{h}{2}$



Виходячи з виду гістограми, висуваємо гіпотезу H_0: ***"Ознака X розподілена за нормальним законом"***

Функція щільності ймовірностей: $x_{cp} := mean(X) = 76.3$ $\quad S_x := Stdev(X) = 24.9$ $\quad fn(z) := dnorm(z, x_{cp}, S_x)$

Теоретичні частоти і їх порівняння з емпіричними:

$j := 1 .. Nint$ $\quad p_j := pnorm(x_{j+1}, x_{cp}, S_x) - pnorm(x_j, x_{cp}, S_x)$ $\quad Nt_j := n \cdot p_j$

$A := augment(H, Nt)$ $\quad A^T = \begin{pmatrix} 6 & 13 & 21 & 22 & 13 & 4 & 1 \\ 5 & 13 & 21 & 21 & 13 & 5 & 1 \end{pmatrix}$ $\quad B := \begin{pmatrix} 6 & 13 & 21 & 22 & 13 & 5 \\ 5 & 13 & 21 & 21 & 13 & 6 \end{pmatrix}$

Об'єднуємо два правих стовпці так, щоб частота значень в них була не нижчою п'яти (табл. B).

Знаходимо фактичне значення χ^2- критерію: $Ne := \left(B^T\right)^{\langle 1 \rangle}$ $\quad Nt := \left(B^T\right)^{\langle 2 \rangle}$ $\quad \chi_{kv} := \sum_{j=1}^{cols(B)} \dfrac{(Ne_j - Nt_j)^2}{Nt_j} = 0.4$

Критичне значення χ^2- критерію при рівні надійності $\alpha := 0.95$ дорівнює:

$\chi_{kr} := qchisq(\alpha, cols(B) - 3) = 7.8$

ВИСНОВОК: оскільки фактичне значення критерію менше критичного, гіпотеза H_0 приймається

Лист 3

6.2.4. Перевірка статистичної гіпотези про закон розподілу ознаки X за критерієм Колмогорова

$ORIGIN := 1$ $X :=$

	1	2	3	4	5	6	7	8
1	82.4	67.8	96.4	62.7	70.8	75.8	91.6	77.1
2	46.0	72.7	69.7	24.5	70.0	76.8	83.5	72.8
3	109.4	124.6	85.4	70.8	85.6	82.3	75.2	...

Об'єм вибірки: $n := rows(X)\cdot cols(X) = 80$

Число інтервалів за формулою Стерджеса: $Nint := round(1 + 3.322\cdot\log(n)) = 7$

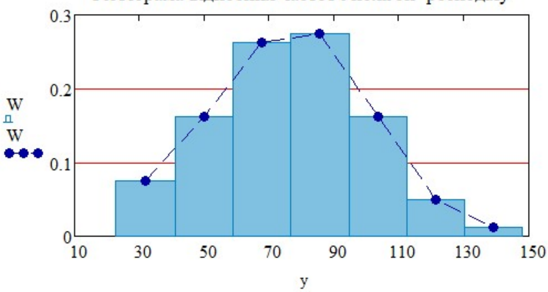
Крок інтервалу: $h := \dfrac{\max(X) - \min(X)}{Nint} = 17.9$

Межі інтервалів: $i := 1 .. Nint + 1$ $x_i := \min(X) + (i-1)\cdot h$

Відносні частоти: $H := hist(x, X)$ $H^T = (6\ \ 13\ \ 21\ \ 22\ \ 13\ \ 4\ \ 1)$ $W := \dfrac{H}{n}$

Середини інтервалів: $y_i := x_i$ $y := y + \dfrac{h}{2}$

Гістограма відносних частот і полігон розподілу



Виходячи з виду гістограми і полігону розподілу, висуваємо гіпотезу H_0:

"Ознака X розподілена за нормальним законом"

Емпірична функція розподілу: $k := 1 .. Nint$ $Fe_k := \sum_{j=1}^{k} W_j$

$Fe^T = (0.075\ \ 0.238\ \ 0.5\ \ 0.775\ \ 0.938\ \ 0.988\ \ 1)$

Теоретична функція розподілу: $x_{cp} := mean(X) = 76.3$ $S_x := Stdev(X) = 24.9$ $F(z) := pnorm(z, x_{cp}, S_x)$

Різниця значень емпіричної і теоретичної функцій розподілу: $D_k := \left|Fe_k - F(x_{k+1})\right|$

$D^T = (0.002\ \ 0.003\ \ 0.004\ \ 0.008\ \ 0.012\ \ 0.003\ \ 0.002)$ $D\max = 0.012$

Критерій Колмогорова: $\lambda := D\max\cdot\sqrt{n} = 0.11$ $P(\lambda) := 2\cdot\left[\sum_{k=1}^{\infty}\left[(-1)^{k+1}\cdot e^{-2\cdot k^2\cdot\lambda^2}\right]\right]$

$P(0.11)\ float, 5\ \rightarrow 1.0$

ВИСНОВОК: ***оскільки ймовірність $P(\lambda)$ достатннььо велика, гіпотеза H_0 приймається.***

Лист 4

6.2.5. Коротка довідка про використання влаштованих функцій «Mathcad»

Влаштовані функції в програмі «Mathcad» розподілені за категоріями і розміщені в списку функцій $\boldsymbol{f(x)}$. При активації піктограми $\boldsymbol{f(x)}$ відкривається дворівневий список, в першій частині якого за допомогою курсора вказують категорію, а в другій – необхідну функцію.

Наведемо список категорій та функцій, які задіяні в розрахунках, на листах 2-4.

<u>Статистика</u>:

mean(X) – середнє значення елементів масиву X;

Var(X) – незміщена дисперсія елементів масиву X;

Stdev(X) – стандартне відхилення елементів масиву X;

skev(X) – асиметрія елементів масиву X;

kurt(X) – ексцес елементів масиву X;

hist(x,X) – вектор, елементами якого є частоти масиву X, які приходяться на інтервали з кінцями, визначеними вектором x;

<u>Вектори і матриці</u>:

min(X) – мінімальне значення елементів масиву X;

max(X) – максимальне значення елементів масиву X;

cols(X) – число стовпців матриці X;

rows(X) – число рядків матриці X;

augment(A,B) – послідовне об'єднання матриць A і B з однаковою кількістю рядків;

<u>Розподіли ймовірностей</u>:

pnorm(x,a,σ) – функція нормального розподілу з математичним сподіванням a і середнім квадратичним відхиленням σ;

$\mathrm{qt}\left(1-\frac{\gamma}{2}, n\text{-}1\right)$ – квантиль розподілу Стьюдента при рівні значушості γ і об'ємі вибірки n;

Щільність ймовірностей:

dnorm(x,a,σ) – щільність нормального розподілу з математичним сподіванням a і середнім квадратичним відхиленням σ;

qchisq(α,k) – критичне значення χ^2_{kr} при надійності α і ступені вільності k;

В розрахунках використані деякі оператори програми «Mathcad», які вносяться на розрахунковий лист з відповідних панелей. З панелі "Символьні операції" вводились наступні оператори:

float, m – вивід результату обчислення в десятичній формі з точністю m значущих цифр;

simplify – спрощення результату обчислень.

З панелі інструментів "Матриці" вводились оператори:

M^{T} – транспонування матриці М;

$\mathrm{M}^{\langle j \rangle}$ – j-тий стовпець матриці М.

Досить застосовним є оператор присвоєння змінній величині i значень з інтервалу від 1 до m з кроком, рівним одиниці. Він виглядає так: $i := 1..m$, тобто це означає, що $i = 1, 2, 3, \ldots, m$. Оператор набирають з клавіатури, при цьому двокрапку обов'язково треба вносити клавішею $\langle ж \rangle$.

Набір індексів $i = 1, 2, \ldots$ масиву даних x_i виконують за допомогою клавіші $\langle [\rangle$. Клавіша $\langle . \rangle$ (крапка) зміщує, записані після неї літери або цифри в нижній регістр, але це не перетворює запис в елемент масиву. Наприклад, $x_{\text{ср}}$, S_X не є елементами масиву (проставлена точка після x та S не спостерігається і спливає лише після наведення на них курсору).

Відмітимо, що, за виключенням текстових регіонів, в «Mathcad» використовується тільки латинський шрифт.

Далі приводимо ДОДАТКИ, в яких розміщені статистичні таблиці, необхідні для проведення розрахунків розглянутого типу.

ДОДАТОК 1

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

x	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0,0	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	08317	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	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
0,7	25804	26115	25424	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	0,34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
1,2	38433	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	46080	46164	46246	47327
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2,0	0,47725	47778	47831	47882	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

x	**Φ(x)**	**x**	**Φ(x)**	**x**	**Φ(x)**	**x**	**Φ(x)**	**x**	**Φ(x)**
3,0	0,49865	3,1	49903	3,2	49931	3,3	49952	3,4	49966
3,5	49977	3,6	49984	3,7	49989	3.8	49993	3,9	49995
4,0	499968								
4,5	499997								
5,0	49999997								

ДОДАТОК 2

Таблиця значень квантилю розподілу Стьюдента $t_\alpha = t(\alpha, n)$

n	α			n	α		
	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	2,54	4,03	6,86	25	2,064	2,894	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	2,648	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 3

Таблиця значень функції $q = q(\alpha, n)$

n	α			n	α		
	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	0,226	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	0,089	0,120	0,162

ДОДАТОК 4

Критичні точки розподілу χ^2

Число ступенів вільності, k	Рівень значущості γ					
	0,01	0,025	0,05	0,95	0,975	0,89
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	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	9,36	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	9,56	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 5

Значення ймовірності $P(\lambda) = 1 - \sum\limits_{k=-\infty}^{\infty} (-1)^k e^{-2k^2\lambda^2}$

λ	0	1	2	3	4	5	6	7	8	9
0,3	0,99999	99998	99995	99991	99983	99970	99949	99917	99872	99807
0,4	99719	99603	99452	99262	99027	98741	98400	97998	97532	96998
0,5	96394	95719	94969	94147	93250	92282	91242	90134	88960	87724
0,6	86428	85077	83678	82225	80732	79201	77636	76042	74422	72781
0,7	71124	69453	67774	66089	64402	62717	61036	59363	57700	56050
0,8	54414	52796	51197	49619	48063	46532	45026	43545	42093	40668
0,9	39273	37907	36571	35266	33992	32748	31536	30356	29206	28087
1,0	27000	25943	24917	23922	22957	22021	21114	20236	19387	18556
1,1	17772	17005	16264	15550	14861	14196	13556	12939	12345	11774
1,2	11225	10697	10190	09703	09235	08787	08357	07944	07550	07171
1,3	06809	06463	06132	05815	05513	05224	04949	04686	04435	04196
1,4	03968	03751	03545	03348	03162	02984	02815	02655	02503	02359
1,5	02222	02092	01969	01852	01742	01638	01539	01446	01357	01274
1,6	01195	01121	01051	00985	00922	00864	00808	00756	00707	00661
1,7	00618	00577	00539	00503	00469	00438	00408	00380	00354	00330
1,8	00307	00285	00265	00247	00229	00213	00198	00186	00170	00158
1,9	00146	00136	00126	00116	00108	00100	00092	00085	00079	00073
2,0	00067	00062	00057	00053	00048	00045	00041	00038	00035	00032
2,1	00030	00027	00025	00023	00021	00019	00018	00016	00015	00014
2,2	00013	00011	00010	00010	00009	00008	00007	00007	00006	00006
2,3	00005	00005	00004	00004	00004	00003	00003	00003	00002	00002
2,4	00002	00002	00002	00001	00001	00001	00001	00001	00001	00001

ЛІТЕРАТУРА

1. Васильків І.М. Основи теорії ймовірностей і математичної статистики: навч. посіб.– Львів: ЛНУ імені Івана Франка, 2020.– 184 с.
2. Руденко В.М. Математична статистика: навч. посіб.– Київ: Центр учбової літератури, 2012.– 304 с.
3. Барковський В. В., Барковська Н. В., Лопатін О. К. Теорія ймовірностей та математична статистика.– Київ: Центр учбової літератури, 2010.– 424 с.
4. Сметанкін В.О., Завгородній О.І., Мазнєва Г.Г., Сметанкіна Н.В. Випадкові величини та математична статистика: навчально-методичний посібник.– Харків: КП Міськдрук, 2009.– 100 с.
5. Жлуктенко В. І., Наконечний С. І. Теорія ймовірностей і математична статистика. Ч. ІІ. Математична статистика.– Київ: КНЕУ, 2007.– 368 с.
6. Макаров Е.Г. Инженерные расчеты в Mathcad 14. – СПб.: Питер, 2007.– 592с.
7. Завгородній О.І., Сметанкін В.О., Мазнєва Г.Г., Сметанкіна Н.В. Теорія ймовірностей і математична статистика.– Харків: ХНТУСГ, 2005.– 275 с.
8. Гмурман В.Е. Теория вероятностей и математическая статистика.– М.: Высшая школа, 2002.– 479 с.
9. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике.– М.: Высшая школа, 2002.– 405 с.
10. Білущак Т.І., Чабанюк Я.М. Теорія ймовірностей і математична статистика.– Практикум.– Львів: 2001.– 418 с.
11. Турчин В.М. Математична статистика.– Київ: Видавничий центр "Академія", 1999.– 511 с.

НАВЧАЛЬНЕ ВИДАННЯ

МАТЕМАТИЧНА СТАТИСТИКА

Основи теорії та методика розв'язування задач
з варіантами індивідуальних завдань

ЗАВГОРОДНІЙ Олексій Іванович
СОЛОВИЧЕНКО Ольга Володимирівна
СТОРОЖЕНКО Ігор Петрович
ЛЕВКІН Дмитро Артурович
СИЧОВА Тетяна Олександрівна

Формат 60х84 1/16. Гарнітура Times New Roman
Папір для цифрового друку. Друк ризографічний.
Ум. друк. арк. 2,6
Наклад 100 пр.
Державний біотехнологічний університет
61002, м. Харків, вул. Алчевських, 44

Для нотаток