Л. В. Васильєва,
О. А. Гончаров

ПРАКТИКУМ

із курсу «Теорія ймовірностей та математична статистика»

Суми
Сумський державний університет
2016

УДК 519.2
ББК 22.17
В 19

Васильєва Л. В.
В 19 Практикум із курсу «Теорія ймовірностей та математична статистика»: для студентів вищих навчальних закладів / Л. В. Васильєва, О. А. Гончаров. – Суми : Сумський державний університет, 2016. – 103 с.

Практикум присвячений комп'ютерній обробці даних фізичного та інженерного експерименту. Запропонований ряд робіт, виконуючи які під керівництвом викладача, студенти опановують методологію обробки статистичних даних.

Практикум може бути використаний також магістрами й аспірантами.

УДК 519.2
ББК 22.17

Зміст

Вступ

Запропонований практикум містить 4 самостійні роботи, на виконання кожної з яких відводиться 5 годин: 1 година на підготовку групи та 4 години на виконання і захист.

Вибір тем, винесених на опрацювання, визначається таким:

1) повинне бути засвоєне й апробоване на практиці статистичне поняття ймовірності випадкової події, оскільки під час практичної роботи це, як правило, єдиний спосіб приписати випадковій події конкретне значення ймовірності;

2) повинне бути одержане уявлення про швидкість статистичної збіжності, щоб у подальшій практичній діяльності була можливість передбачити вплив обсягу вибірки на точність одержуваних оцінок;

3) повинне бути набуте вміння обробляти одновимірний статистичний масив: побудова гістограми, знаходження числових характеристик та одержання практичних висновків, виходячи з вигляду гістограми і значень характеристик;

4) повинне бути набуте вміння знаходити статистичну залежність між невипадковим фактором x і випадковим відгуком на нього y, вміння робити за знайденою залежністю прогноз для відгуку y й оцінювати точність цього прогнозу. Оцінка точності прогнозу особливо важлива, тому що наявні статистичні дані, як правило, є сильно зашумленими.

Теми з пунктів 1 і 2 винесені в першу самостійну роботу, теми з пункту 3 – в другу, з пункту 4 – у третю роботу.

Оскільки обробка статистичних даних поєднана з громіздкими обчисленнями, доцільно використовувати математичний пакет, спеціально призначений для їх обробки.

У цьому практикуму наведені приклади розрахунків у системах STATISTICA, MathCad та Excel for Windows.

Теоретичні відомості до самостійних робіт подані у стислому вигляді.

Самостійна робота 1
ГЕОМЕТРИЧНА ЙМОВІРНІСТЬ. СТАТИСТИЧНЕ ВИЗНАЧЕННЯ ЙМОВІРНОСТІ

1.1 Короткі теоретичні відомості

У цій лабораторній роботі використовуються одночасно два визначення ймовірності випадкової події – геометричне й статистичне.

1.1.1 Геометричне визначення ймовірності

Розглянемо геометричне визначення для двовимірного простору. Нехай на площині є деяка область D, площа якої дорівнює S(D) і в ній міститься інша область d із площею S(d). В область D довільно кидають точку (рис. 1.1).

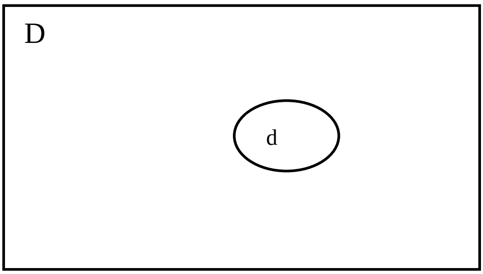


Рисунок 1.1

Яка ймовірність того, що точка потрапляє в область d? Тут передбачається, що ймовірність потрапляння в будь-яку частину області D пропорційна площі цієї частини і не залежить від її розміщення та форми. У такому разі ймовірність потрапляння в область d дорівнює

$$P = \frac{S(d)}{S(D)}. \tag{1.1}$$

У разі одновимірної і тривимірної областей D замість площі мають на увазі відповідно довжину та об'єм.

1.1.2 Статистичне визначення ймовірності

Виконують N випробувань, у кожному з яких може відбутися випадкова подія A. Підраховують число випробувань N_A, в яких подія A дійсно відбулася. Ймовірність події A наближено дорівнює відношенню

$$P(A) \approx \frac{N_A}{N}. \tag{1.2}$$

Інтуїтивно зрозуміло, що чим більше буде N, тим більша точність цієї наближеної оцінки.

1.1.3 Метод Монте-Карло обчислення геометричної ймовірності

Припустимо, що потрібно обчислити площу плоскої фігури A (рис. 1.2), розміщеної в прямокутнику

$$xn \le x \le xk,$$
$$yn \le y \le yn,$$

площа якого дорівнює S_I.

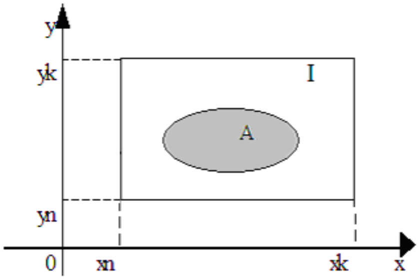


Рисунок 1.2

Усередині прямокутника генеруються N випадкових точок. Кількість точок, що потрапили в область A, позначимо N_A. Тоді

$$S(A) \approx S(I) \cdot \frac{N_A}{N}. \tag{1.3}$$

Метод Монте-Карло називають також ***методом статистичних випробувань.***

Зі збільшенням кількості точок N точність формули (1.3) зростає, але повільно. Доведено, що похибка (1.2) має порядок, близький до $\frac{1}{\sqrt{N}}$. При $N \approx 1000$ це дає точність порядку 5-10%. Тому метод Монте-Карло доцільно використовувати під час розв`язання тих задач, де результат потрібен із невеликою точністю.

Мета самостійної роботи

Повинні бути набуті такі *вміння*: знаходження ймовірності суми та добутку подій, оцінка швидкості статистичної збіжності за результатами статистичного експерименту.

Мають бути засвоєні такі *поняття*: геометричне визначення ймовірності, сума подій, добуток подій, швидкість статистичної збіжності.

Завдання до самостійної роботи

Використовуючи дані зі свого індивідуального завдання, виконати таке:

1) зобразити на рисунку задані області;

2) на окремому рисунку зобразити області AB, A + B, заштрихувати їх;

3) за допомогою рисунка знайти точні значення ймовірностей P(A), P(B), P(AB), P(A+B);

4) знайти значення цих ймовірностей (P'(A), P'(B), P'(AB), P'(A+B)) за методом Монте-Карло (файл MonteKarlo.exe). Оцінити відносну похибку обчислень імовірності за методом Монте-Карло за формулою

$$\Delta = \frac{(P - P')}{P} \cdot 100\% ; \qquad (1.4)$$

5) оцінити швидкість збіжності ймовірності, обчисленої за методом Монте-Карло, до істинного значення, збільшивши число випадкових точок у 100 разів. Для цього усереднити відносні похибки $\Delta_{P(A)}$ для N = 1000 та N = 100 000 і за усередненими результатами зробити висновок про швидкість збіжності.

Приклад виконання самостійної роботи за допомогою програми MonteCarlo

Вихідні дані наведені в таблиці 1.1.

Таблиця 1.1

Варіант	I	A	B
0	$0 \le x \le 10$; $0 \le y \le 10$	$1 \le x \le 8$; $1 \le y \le 8$	$(x-5)^2 +$ $+(y-5)^2 \le 4$

Тут область I – прямокутник, B – прямокутник, B – коло.

Запускаємо файл MonteCarlo.exe. У вікні (рис. 1.3) зліва вказуємо типи заданих областей (у поданому прикладі – прямокутник і коло), межі області I та попередньо розраховані за формулою (1.1) значення ймовірностей P(A), P(B), P(AB), P(A+B); N = 1000.

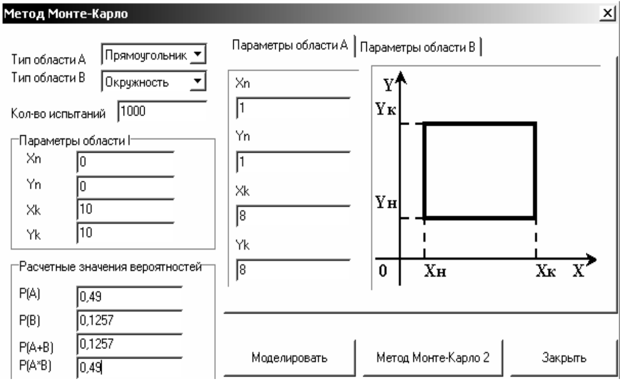


Рисунок 1.3

У правій частині вікна задаємо параметри області А (рис. 1.3) та області В (рис. 1.4). Натискаємо кнопку «Моделювати».

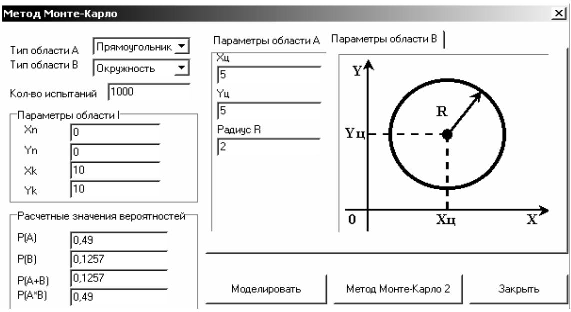


Рисунок 1.4

Звіряємо вигляд раніше намальованих областей з тими, що будуть зображені у вікні (рис. 1.5).

Натискаємо кнопку «Почати моделювання». У результаті розрахунку одержимо значення, як на рис. 1.5.

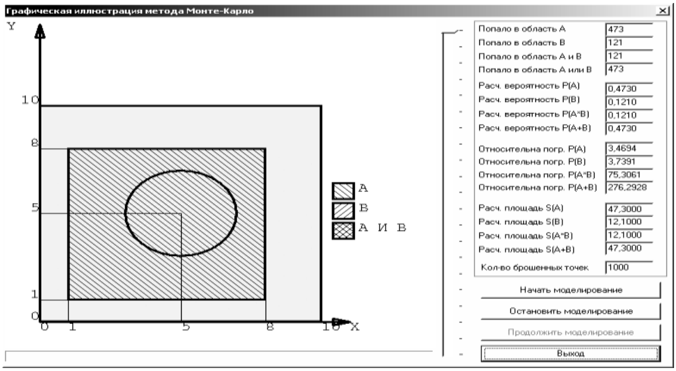


Рисунок 1.5

Розрахункову ймовірність P(A) обчислюємо за формулою

$$P'(A) \approx \frac{N_A}{N} = \frac{473}{1000}. \qquad (1.5)$$

Відносну похибку (у відсотках) для P(A) розраховуємо за формулою

$$\Delta_{P(A)} = \left| \frac{\text{Đ}(\text{À}) - \text{Đ}'(\text{À})}{\text{Đ}(\text{À})} \right| \cdot 100\ \%. \qquad (1.6)$$

Інші розрахункові ймовірності та відносні похибки обчислюються аналогічно.

Якщо збільшити кількість точок, то результати розрахунків уточняться (табл. 1.2).

Таблиця 1.2

N = 1 000		N = 100 000	
P(A)	$\Delta_{P(A)}$	P(A)	$\Delta_{P(A)}$
0,510	4,082	0,491	0,1327
0,498	1,633	0,488	0,3531
0,521	6,327	0,491	0,1306
0,487	0,612	0,492	0,4878
0,474	3,265	0,494	0,8041

Знаходимо середню відносну похибку для події А при N = 1000 і при N = 100 000. За одержаними результатами визначаємо, на скільки порядків зменшиться відносна похибка $\Delta_{P(A)}$, якщо кількість точок N зросте на два порядки з N = 1000 до N = 100000.

Індивідуальні завдання до самостійної роботи

Вар.	I	A	B
1	$2 \le x \le 10;\ 1 \le y \le 7$	$(x-4)^2+(y-4)^2 \le 4$	$2 \le x \le 8;\ 4 \le y \le 6$
2	$0 \le x \le 8;\ 0 \le y \le 6$	$3 \le x \le 5,\ 1 \le y \le 5$	$0 \le x \le 6;\ 3 \le y \le 4$
3	$4 \le x \le 11;\ 2 \le y \le 11$	$4 \le x \le 10,\ 6 \le y \le 10$	$(x-6)^2/1+(y-6)^2/4 \le 1$
4	$0 \le x \le 8,\ 0 \le y \le 6$	$(x-4)^2/9+(y-3)^2/4 \le 1$	$(x-5)^2+(y-3)^2 \le 1$
5	$2 \le x \le 9;\ 0 \le y \le 6$	$4 \le x \le 5;\ 0 \le y \le 3$	$(x-7)^2+(y-4)^2 \le 1$
6	$0 \le x \le 7,\ 0 \le y \le 5$	$(x-4)^2+(y-2)^2 \le 4$	$(x-4)^2/4+(y-2)^2/1 \le 1$
7	$0 \le x \le 8;\ 0 \le y \le 7$	$(x-4)^2+(y-3)^2 \le 4$	$(x-4)^2/9+(y-3)^2/4 \le 1$
8	$2 \le x \le 7,\ 0 \le y \le 4$	$3 \le x \le 6,\ 1 \le y \le 4$	$(x-3)^2+(y-3)^2 \le 1$
9	$0 \le x \le 8;\ 0 \le y \le 8$	$(x-3)^2/9+(y-4)^2/4 \le 1$	$(x-7)^2+(y-4)^2 \le 1$
10	$0 \le x \le 5,\ 0 \le y \le 10$	$(x-2)^2+(y-7)^2 \le 1$	$1 \le x \le 3;\ 0 \le y \le 7$
11	$1 \le x \le 10;\ 0 \le y \le 8$	$(x-3)^2+(y-5)^2 \le 4$	$(x-3)^2+(y-4)^2 \le 1$
12	$1 \le x \le 11,\ 2 \le y \le 12$	$(x-6)^2/9+(y-6)^2/16 \le 1$	$5 \le x \le 8,\ 5 \le y \le 6$
13	$0 \le x \le 12\ \ 2 \le y \le 10$	$1 \le x \le 8;\ 2 \le y \le 8$	$(x-8)^2+(y-5)^2 \le 4$
14	$1 \le x \le 9,\ 0 \le y \le 6$	$(x-4)^2/9+(y-3)^2/4 \le 1$	$(x-4)^2+(y-2)^2 \le 1$
15	$0 \le x \le 9;\ 0 \le y \le 8$	$1 \le x \le 7;\ 1 \le y \le 5$	$(x-7)^2/1+(y-5)^2/4 \le 1$
16	$0 \le x \le 8,\ 1 \le y \le 8$	$3 \le x \le 5,\ 1 \le y \le 5$	$(x-4)^2+(y-4)^2 \le 1$
17	$1 \le x \le 9;\ 1 \le y \le 10$	$(x-5)^2+(y-5)^2 \le 1$	$1 \le x \le 4;\ 1 \le y \le 3$
18	$0 \le x \le 8,\ 0 \le y \le 8$	$1 \le x \le 4,\ 1 \le y \le 5$	$(x-4)^2+(y-3)^2 \le 4$
19	$0 \le x \le 10;\ 0 \le y \le 10$	$(x-5)^2/4+(y-3)^2/9 \le 1$	$(x-5)^2+(y-3)^2 \le 4$
20	$1 \le x \le 10,\ 0 \le y \le 8$	$2 \le x \le 8,\ 2 \le y \le 6$	$4 \le x \le 6,\ 2 \le y \le 4$
21	$2 \le x \le 10;\ 0 \le y \le 8$	$(x-7)^2+(y-3)^2 \le 4$	$(x-4)^2+(y-3)^2 \le 1$
22	$0 \le x \le 10,\ 1 \le y \le 10$	$6 \le x \le 8,\ 3 \le y \le 4$	$(x-7)^2+(y-3)^2 \le 4$
23	$1 \le x \le 9;\ 1 \le y \le 11$	$(x-4)^2+(y-3)^2 \le 1$	$(x-4)^2/4+(y-3)^2/1 \le 1$
24	$2 \le x \le 6,\ 0 \le y \le 4$	$2 \le x \le 5,\ 1 \le y \le 4$	$(x-3)^2+(y-2)^2 \le 1$

Самостійна робота 2
СТАТИСТИЧНА ОБРОБКА ОДНОВИМІРНОГО ВИПАДКОВОГО МАСИВУ

2.1 Короткі теоретичні відомості

Більшість контрольованих параметрів виробів належать до нормально розподілених випадкових величин: розміри деталей, вага виливків, відсотковий вміст хімічних елементів у сплавах, електроємність і опір електротехнічних виробів тощо. При відлагодженому обладнанні й правильно відрегульованому технологічному процесі розподіл контрольованого параметра має бути нормальним, а його середнє значення повинне збігатися зі значенням, заданим у технічній документації. Можливі відхилення від цієї вимоги і передбачувані причини цих відхилень перелічені в таблиці 2.1.

Таблиця 2.1

№	Порушення	Причина
1	Розподіл контрольованого параметра близький до нормального, але вибіркове середнє не збігається зі значенням, заданим технічною документацією	Неправильно відрегульований технологічний процес. Потрібне регулювання
2	Розподіл контрольованого параметра одномодальний, але сильно відрізняється від нормального	Серйозні несправності в обладнанні. Потрібен ремонт
3	Розподіл багатомодальний	Неякісна вибірка, дані взяті з різних генеральних сукупностей. Повторити вибірку

Щоб установити, чи є одне з перелічених порушень, роблять статистичний контроль потрібного параметра. Для цього роблять n випадкових його вимірів $x_1,\ x_2,\ldots,x_n$ (вибірка обсягу n). За вибіркою знаходять такі числові характеристики:

- математичне сподівання –

$$\overline{x}^* = \frac{1}{n}\sum_i x_i\ , \tag{2.1}$$

- дисперсію

$$D^* = \frac{1}{n}\sum_i (x_i - \overline{x}^*)^2\ , \tag{2.2}$$

- середньоквадратичне відхилення

$$\sigma^* = \sqrt{D^*}\ , \tag{2.3}$$

- асиметрію

$$Sk = \frac{1}{(\sigma^*)^3}\cdot\frac{1}{n}\sum_i (x_i - \overline{x}^*)^3\ , \tag{2.4}$$

- ексцес

$$Ks = \frac{1}{(\sigma^*)^4}\cdot\frac{1}{n}\sum_i (x_i - \overline{x}^*)^4 - 3\,. \tag{2.5}$$

Мономодальність або багатомодальність вибіркового розподілу визначають за виглядом гістограми. Близькість закону розподілу до нормального визначають за значеннями асиметрії та ексцесу.

Порівняння вибіркового середнього зі значенням контрольованого параметра, заданого в технічній документації, дозволяє встановити, чи правильно відрегульований технологічний процес.

Мета самостійної роботи

Повинні бути набуті такі *вміння*: будувати гістограми; обчислювати ймовірності потрапляння випадкової величини в заданий проміжок; обчислювати ймовірність відхилення випадкової величини від математичного сподівання не більше ніж на задану величину.

Мають бути засвоєні такі *поняття*: обсяг вибірки, математичне сподівання, розмах вибірки, середньоквадратичне відхилення, асиметрія, ексцес, теоретична й емпірична щільність нормального розподілу.

Завдання до самостійної роботи

Використовуючи дані зі свого індивідуального завдання, виконати таке:

1) створити в STATISTICA файл вихідних даних (* .sta);

2) створити файл автозвіту * .rtf;

3) створити статистичну таблицю з такими даними: обсяг вибірки (Valid N), математичне сподівання (Mean), розмах вибірки (Minimum – Maximum), середнє квадратичне відхилення (Std.Dev.), асиметрія (Skewness), ексцес (Kurtosis);

4) побудувати гістограму, обравши число часткових інтервалів, що дорівнює 10. Відзначити, чи задовольняє гістограма поставленим до неї вимогам. Якщо не задовольняє, зменшити, наскільки це допустимо, число часткових інтервалів;

5) за значеннями асиметрії та ексцесу і вигляду гістограми зробити висновок, чи значно відрізняється розподіл випадкової величини від нормального;

6) за допомогою ймовірнісного калькулятора знайти ймовірність потрапляння випадкової величини в проміжок, заданий у варіанті завдання;

7) вважаючи, що технологічний процес відрегульований правильно, а допуск становить 10 % від значення контрольованого параметра, знайти випуск придатної продукції у відсотках.

Приклад виконання самостійної роботи в пакеті Statistica

Працюємо в модулі Basic Statistics and Tables. Порядок роботи в системі STATISTICA описаний у додатку А.

1 Створити в STATISTICA файл вихідних даних під ім'ям свого варіанта (наприклад, fio2.sta).

2 Створити файл автозвіту * .rtf.

3 Створити статистичну таблицю з такими даними: обсяг вибірки (Valid N), математичне сподівання (Mean), розмах вибірки (Minimum – Maximum), середнє квадратичне відхилення (Std.Dev.), асиметрія (Skewness), ексцес (Kurtosis).

Для цього виділити таблицю з даними – Statistics – Basic Statistics and Tables – Descriptive statistics – Ok – кнопка Variables – виділити потрібну змінну (в нашому випадку X) – OK – кнопка Advanced – активувати опції Valid N, Mean, Standard Deviation, Minimum і Maximum, Skewness, Kurtosis – кнопка Summary.

Потрібна таблиця (табл. 2.2) вставляється у звіт із самостійної роботи з автозвіту:

Таблиця 2.2

Variable	Descriptive Statistics (Vertical.sta)						
	Valid N	Mean	Minimum	Maximum	Std.Dev.	Skewness	Kurtosis
VERTICAL	80	2,45775	0,3100(	4,1700(	0,75182	292,0000(	0,11862

4 Побудувати гістограму. Для цього потрібно активувати таблицю fio2.sta – пункт верхнього меню Graphs – 2D Graphs – Histogramm – кнопка Variables – виділити потрібну змінну (в нашому випадку X) – OK – кнопка Advanced. Вибрати опції Graph type : Regular, Fit type : Off і

число часткових інтервалів (Categories), що дорівнює 10. Одержимо гістограму (рис. 2.1).

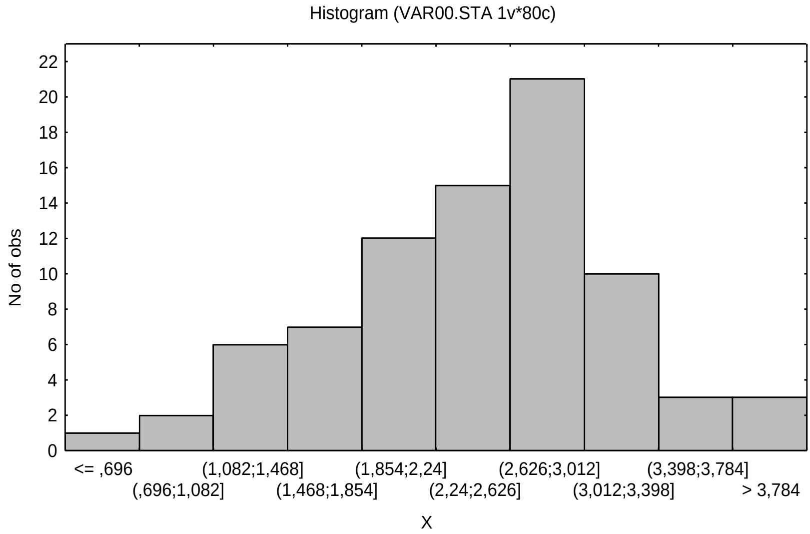


Рисунок 2.1

Під кожним стовпцем гістограми записаний інтервал зміни змінної, що відповідає даному стовпцю. Висота стовпців дорівнює кількості точок, що потрапили у відповідний частковий інтервал. Ця гістограма вимоги, поставлені до гістограми, не задовольняє (пояснити самостійно). Зменшити число часткових інтервалів до 5 і побудувати нову гістограму (рис. 2.2).

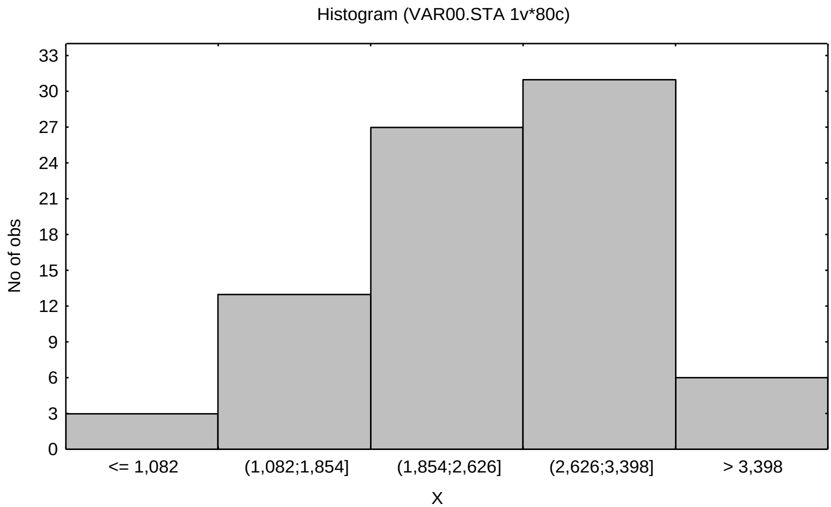


Рисунок 2.2

Дайте відповідь на запитання: чи задовольняє ця гістограма вимоги, пропонованим до гістограми? Чи можна ще зменшити число часткових інтервалів?

5 За значеннями асиметрії та ексцесу й вигляду гістограми зробити висновок, чи значно відрізняється розподіл випадкової величини від нормального.

6 За допомогою ймовірнісного калькулятора знайти ймовірність потрапляння випадкової величини в проміжок, заданий у варіанті завдання. Ймовірність потрапляння випадкової величини в проміжок (a, b) розраховують за формулою $P(a < X < b) = F(b) - F(a)$, де $F(x)$ – функція розподілу випадкової величини. Значення функції розподілу знаходимо за допомогою ймовірнісного калькулятора: Basic Statistic / Tables – Probability Calculator – Distribution Z (Normal) (рис. 2.3). У поля mean, st.dev. і Z вписуємо відповідні значення, натискаємо на кнопку Compute й одержуємо потрібну ймовірність: F (2,1) = 0,3170. Аналогічно знаходимо F (3,2):

$$P(2,1 < X < 3,2) = F(3,2) - F(2,1) = 0,8381 - 03170 = 0,5211.$$

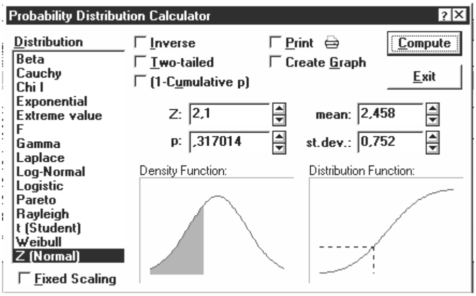


Рисунок 2.3

7 Оскільки технологічний процес відрегульований неправильно, то вибіркове середнє $\overline{x}^*$ можна взяти за значення параметра, заданого в технічній документації. Десятивідсоткове відхилення знаходимо за формулою $\delta = 0{,}1 \cdot \overline{x} = 0{,}1 \cdot 2{,}458.$ Потім обчислюємо ймовірність $P(|x - \overline{x}| < \delta) = P(|x - 2{,}458| < 0{,}245) = 0{,}2554$ за допомогою ймовірнісного калькулятора, як показано на рис. 2.4. Одержуємо відповідь: вихід придатної продукції при заданому допуску $0{,}1\overline{x}$ складає 25,5 % від усієї продукції.

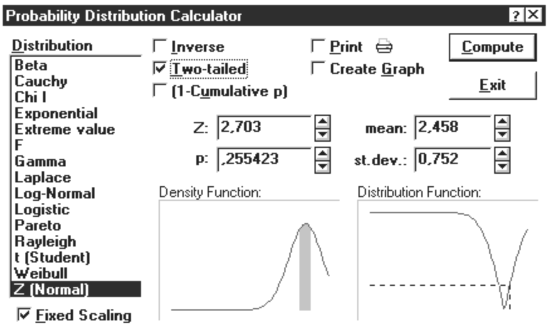


Рисунок 2.4

Примітка. Перед друкуванням звіту рис. 2.3, 2.4 з тексту звіту можна видалити.

Індивідуальні завдання до самостійної роботи

Варіант 1

1,67	2,41	0,79	1,41	2,50	2,29	2,58	1,32
3,75	1,94	0,95	3,48	2,39	1,17	1,92	1,04
2,13	1,58	2,18	2,30	3,03	1,50	2,53	1,91
1,31	3,62	1,49	1,98	2,14	3,35	2,89	2,51
2,31	2,34	1,00	2,03	0,64	2,67	0,09	1,78
3,24	1,91	1,20	1,61	2,35	1,73	2,93	2,32
2,84	1,29	2,28	2,54	1,85	2,40	2,22	2,90
2,37	2,68	2,00	2,70	2,33	2,86	0,36	1,98
2,53	0,80	2,89	0,73	1,01	1,85	2,05	1,16
1,76	2,78	2,43	1,85	1,21	1,53	1,54	2,43

$P(0{,}93 < X < 1{,}52) = ?$

Варіант 2

2,46	1,70	2,44	0,82	1,50	2,53	2,32	2,61
1,35	3,78	1,97	0,98	3,51	2,42	1,20	1,95
1,07	2,16	1,61	2,21	2,33	3,06	1,53	2,56
1,94	1,34	3,63	1,52	2,01	2,17	3,38	2,92
2,54	2,34	2,37	1,03	2,06	0,67	2,70	1,12
1,81	3,27	1,94	1,23	1,64	2,38	1,76	2,96
2,35	2,87	1,32	2,31	2,57	1,88	2,43	1,88
2,93	2,40	2,71	2,03	2,76	2,36	2,89	0,39
2,01	2,56	0,83	2,92	0,76	1,04	1,88	2,08
1,19	1,79	2,81	2,46	1,88	1,24	1,56	1,57

$P(0{,}92 < X < 1{,}54) = ?$

Варіант 3

1,60	2,49	1,73	2,47	0,85	1,53	2,56	2,35
2,64	1,38	3,81	2,00	1,01	3,54	2,45	1,23
1,98	1,10	2,19	1,64	2,24	2,36	3,09	1,56
2,59	1,97	1,37	3,68	1,55	2,04	2,20	3,41
2,95	2,57	2,37	2,40	1,06	2,09	0,70	2,73
0,45	1,84	3,30	1,97	1,26	1,67	2,41	1,79
2,99	2,38	2,90	1,35	2,34	2,60	1,91	2,46
2,28	2,96	2,43	2,74	2,06	2,76	2,39	2,92
0,42	2,04	2,59	0,86	2,95	0,79	1,07	1,91
2,11	1,22	1,82	2,84	2,49	1,91	1,27	1,59

$$P(0{,}91 < X < 1{,}55) = ?$$

Варіант 4

1,62	1,63	2,52	1,76	2,50	0,88	1,56	2,59
2,38	2,67	1,14	3,84	2,03	1,04	3,57	2,48
1,86	2,01	1,13	2,22	1,67	2,27	2,38	3,12
1,59	2,62	2,00	1,40	3,71	1,58	2,07	2,23
3,44	2,98	2,60	2,40	2,43	1,09	2,12	0,73
2,76	0,18	1,87	3,32	2,00	1,29	1,70	2,44
1,82	3,02	2,41	2,93	1,38	2,37	2,63	1,94
2,49	2,31	2,99	2,46	2,77	2,09	2,79	2,42
2,95	0,45	2,07	2,62	0,89	2,98	0,82	1,10
1,94	2,14	1,25	1,83	2,87	2,52	1,94	1,30

$$P(0{,}90 < X < 1{,}56) = ?$$

Варіант 5

3,31	1,15	1,43	2,27	2,47	1,58	2,18	3,20
2,85	2,27	1,63	1,95	1,96	2,85	2,09	2,83
1,21	1,89	2,92	2,71	3,00	1,74	4,17	2,36
1,37	3,90	2,81	1,59	2,34	1,46	2,55	2,00
2,60	2,78	3,45	1,92	2,95	2,33	1,73	4,04
1,91	2,40	2,56	3,77	3,31	2,93	2,73	2,76
1,42	2,45	1,06	3,09	0,31	2,20	3,66	2,33
1,62	2,03	2,77	2,15	3,35	2,74	3,26	1,71
2,70	2,96	2,27	2,82	2,64	3,32	2,79	3,10
2,42	3,12	2,75	3,28	0,78	2,40	2,95	1,22

$P(0{,}89 < X < 1{,}57) = ?$

Варіант 6

2,46	1,70	2,44	0,82	1,50	2,53	2,32	2,61
1,35	3,78	1,97	0,98	3,51	2,42	1,20	1,95
1,07	2,16	1,61	2,21	2,33	3,06	1,53	2,56
1,94	1,34	3,63	1,52	2,01	2,17	3,38	2,92
2,54	2,34	2,37	1,03	2,06	0,67	2,70	1,12
3,24	1,91	1,20	1,61	2,35	1,73	2,93	2,32
2,84	1,29	2,28	2,54	1,85	2,40	2,22	2,90
2,37	2,68	2,00	2,70	2,33	2,86	0,36	1,98
2,53	0,80	2,89	0,73	1,01	1,85	2,05	1,16
1,76	2,78	2,43	1,85	1,21	1,53	1,54	2,43

$P(1{,}08 < X < 1{,}68) = ?$

Варіант 7

1,67	2,41	0,79	1,41	2,50	2,29	2,58	1,32
3,75	1,94	0,95	3,48	2,39	1,17	1,92	1,04
2,13	1,58	2,18	2,30	3,03	1,50	2,53	1,91
1,31	3,62	1,49	1,98	2,14	3,35	2,89	2,51
2,31	2,34	1,00	2,03	0,64	2,67	0,09	1,78
1,81	3,27	1,94	1,23	1,64	2,38	1,76	2,96
2,35	2,87	1,32	2,31	2,57	1,88	2,43	1,88
2,93	2,40	2,71	2,03	2,76	2,36	2,89	0,39
2,01	2,56	0,83	2,92	0,76	1,04	1,88	2,08
1,19	1,79	2,81	2,46	1,88	1,24	1,56	1,57

$$P(1{,}07 < X < 1{,}69) = ?$$

Варіант 8

1,62	1,63	2,52	1,76	2,50	0,88	1,56	2,59
2,38	2,67	1,14	3,84	2,03	1,04	3,57	2,48
1,86	2,01	1,13	2,22	1,67	2,27	2,38	3,12
1,59	2,62	2,00	1,40	3,71	1,58	2,07	2,23
3,44	2,98	2,60	2,40	2,43	1,09	2,12	0,73
0,45	1,84	3,30	1,97	1,26	1,67	2,41	1,79
2,99	2,38	2,90	1,35	2,34	2,60	1,91	2,46
2,28	2,96	2,43	2,74	2,06	2,76	2,39	2,92
0,42	2,04	2,59	0,86	2,95	0,79	1,07	1,91
2,11	1,22	1,82	2,84	2,49	1,91	1,27	1,59

$$P(1{,}06 < X < 1{,}70) = ?$$

Варіант 9

1,60	2,49	1,73	2,47	0,85	1,53	2,56	2,35
2,64	1,38	3,81	2,00	1,01	3,54	2,45	1,23
1,98	1,10	2,19	1,64	2,24	2,36	3,09	1,56
2,59	1,97	1,37	3,68	1,55	2,04	2,20	3,41
2,95	2,57	2,37	2,40	1,06	2,09	0,70	2,73
2,76	0,18	1,87	3,32	2,00	1,29	1,70	2,44
1,82	3,02	2,41	2,93	1,38	2,37	2,63	1,94
2,49	2,31	2,99	2,46	2,77	2,09	2,79	2,42
2,95	0,45	2,07	2,62	0,89	2,98	0,82	1,10
1,94	2,14	1,25	1,83	2,87	2,52	1,94	1,30

$$P(1{,}05 < X < 1{,}71) = ?$$

Варіант 10

1,43	2,03	3,05	2,70	2,13	1,48	1,80	1,81
2,70	1,94	3,63	1,06	1,74	2,77	2,56	2,85
1,59	3,08	2,21	1,22	3,75	2,66	1,44	3,19
1,81	2,40	1,85	2,45	2,57	3,30	1,77	2,80
3,18	1,58	2,89	1,76	2,25	2,41	3,62	2,13
1,95	0,45	2,07	2,62	0,89	2,98	0,82	1,10
1,94	2,14	1,25	1,83	2,87	2,52	1,94	1,30
1,62	1,03	2,77	2,15	2,35	2,74	3,26	1,71
2,70	2,96	2,27	1,82	2,64	3,32	2,79	3,10
2,42	1,12	2,75	1,28	0,78	2,40	2,95	1,22

$$P(1{,}04 < X < 1{,}72) = ?$$

Варіант 11

2,46	1,70	1,43	2,27	2,47	1,58	2,32	2,61
1,35	3,78	1,63	1,95	1,96	2,85	1,20	1,95
1,07	2,16	1,87	3,32	2,00	1,29	1,53	2,56
1,94	1,34	2,41	2,93	1,38	2,37	3,38	2,92
2,54	2,34	2,99	2,46	2,77	2,09	2,70	1,12
3,24	1,91	2,07	2,62	0,89	2,98	2,93	2,32
2,84	1,29	1,25	1,83	2,87	2,52	2,22	2,90
2,37	2,68	2,77	2,15	3,35	2,74	0,36	1,98
2,53	0,80	2,27	2,82	2,64	3,32	2,05	1,16
1,76	2,78	2,75	3,28	0,78	2,40	1,54	2,43

$P(1{,}23 < X < 1{,}83) = ?$

Варіант 12

3,31	1,15	2,44	0,82	1,50	2,53	2,18	3,20
2,85	2,27	1,97	0,98	3,51	2,42	2,09	2,83
2,76	0,18	1,61	2,21	2,33	3,06	1,70	2,44
1,82	3,02	3,63	1,52	2,01	2,17	2,63	1,94
2,49	2,31	2,37	1,03	2,06	0,67	2,79	2,42
2,95	0,45	1,20	1,61	2,35	1,73	0,82	1,10
1,94	2,14	2,28	2,54	1,85	2,40	1,94	1,30
1,62	2,03	2,00	2,70	2,33	2,86	3,26	1,71
2,70	2,96	2,89	0,73	1,01	1,85	2,79	3,10
2,42	3,12	2,43	1,85	1,21	1,53	2,95	1,22

$P(1{,}22 < X < 1{,}84) = ?$

Варіант 13

1,60	2,49	0,79	1,41	2,50	2,29	2,56	2,35
2,64	1,38	0,95	3,48	2,39	1,17	2,45	1,23
1,98	1,10	2,18	2,30	3,03	1,50	3,09	1,56
2,59	1,97	1,49	1,98	2,14	3,35	2,20	3,41
2,95	2,57	1,00	2,03	0,64	2,67	0,70	2,73
2,76	0,18	1,94	1,23	1,64	2,38	1,70	2,44
1,82	3,02	1,32	2,31	2,57	1,88	2,63	1,94
2,49	2,31	2,71	2,03	2,76	2,36	2,79	2,42
2,95	0,45	0,83	2,92	0,76	1,04	0,82	1,10
1,94	2,14	2,81	2,46	1,88	1,24	1,94	1,30

$$P(1{,}21 < X < 1{,}85) = ?$$

Варіант 14

0,67	2,41	1,73	2,47	0,85	1,53	1,58	1,32
0,75	1,94	3,81	2,00	1,01	3,54	1,92	1,04
2,13	1,58	2,19	1,64	2,24	2,36	2,53	1,91
1,31	3,62	1,37	3,68	1,55	2,04	2,89	2,51
2,31	0,34	2,37	2,40	1,06	2,09	0,09	1,78
1,81	2,27	1,87	3,32	2,00	1,29	1,76	2,96
2,35	2,87	2,41	2,93	1,38	2,37	2,43	1,88
2,93	2,40	2,99	2,46	1,77	2,09	2,89	0,39
2,01	2,56	2,07	2,62	0,89	2,98	1,88	2,08
1,19	1,79	1,25	1,83	2,87	2,52	1,56	1,57

$$P(1{,}20 < X < 1{,}86) = ?$$

Варіант 15

1,62	1,63	2,52	1,76	2,50	0,88	1,56	2,59
2,38	2,67	3,14	3,84	2,03	1,04	3,57	2,48
2,86	2,01	3,13	2,22	1,67	2,27	2,38	3,12
1,59	2,62	2,00	1,40	3,71	1,58	2,07	2,23
0,42	2,04	2,59	0,86	2,95	0,79	1,07	1,91
2,11	3,22	1,82	2,84	2,49	2,91	3,27	1,59
1,76	2,50	1,88	3,08	2,47	2,99	1,44	2,43
2,69	2,00	2,55	2,37	3,05	2,52	1,83	2,15
2,85	3,48	3,01	0,51	2,13	2,68	0,95	3,04
0,88	1,16	2,00	3,20	1,31	1,91	2,93	2,58

$P(1{,}19 < X < 1{,}87) = ?$

Самостійна робота 3
ПРОГНОЗ НА ОСНОВІ ЛІНІЙНОЇ РЕГРЕСІЇ. ТОЧНІСТЬ ПРОГНОЗУ. ЩІЛЬНІСТЬ ЛІНІЙНОГО ЗВ'ЯЗКУ

3.1 Короткі теоретичні відомості

Рівняння лінійної регресії $y = b_0 + b_1 x$ знаходять вибіркою методом найменших квадратів. На рис. 3.1 це похила пряма, зображена суцільною лінією.

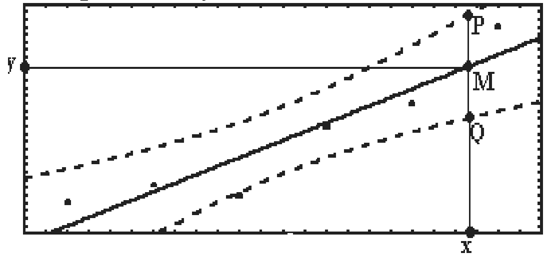


Рисунок 3.1

Точна лінійна залежність між x і y: $y = \beta_0 + \beta_1 x + \varepsilon$ (ε – випадковий член) невідома. Можна лише стверджувати, що вона з імовірністю γ розміщена в довірчій області, обмеженої лініями гіперболи (рис. 3.2).

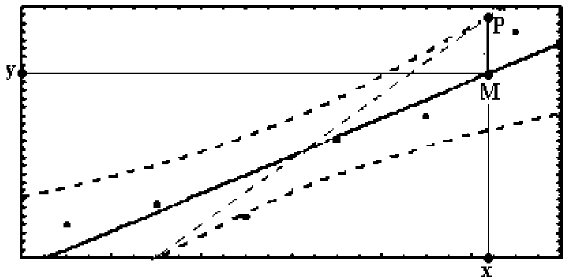


Рисунок 3.2

Імовірність γ називається рівнем довіри. Зазвичай

γ = 0,95 або $\gamma = 0{,}99$ (95 , 99 %). Точна лінія регресії $y = \beta_0 + \beta_1 x + \varepsilon$ зображена на рис. 3.2 пунктирною прямою. Прогноз y в точці x роблять за рівнянням $y = b_0 + b_1 x$. Точне значення прогнозу може з імовірністю γ відповідати будь-якій точці довірчого інтервалу. PQ (див. рис. 3.1).

У найбільш несприятливому випадку точний прогноз потрапляє на край довірчої області. У цьому разі абсолютна похибка прогнозу дорівнює напівширині довірчого інтервалу δ = MP = MQ. Відносну похибку прогнозу у відсотках обчислюємо за формулою

$$\Delta = \left| \frac{\delta}{y_{\text{ïðîãíîç}}} \right| \cdot 100\,\% . \tag{3.1}$$

Мета самостійної роботи

Повинні бути набуті такі *вміння*:

1) знаходження графіка та рівняння лінійної регресії в довірчій області при заданому рівні довіри;

2) знаходження за графіком прогнозу відгуку при заданому значенні фактора;

3) знаходження за графіком максимальної абсолютної похибки прогнозу;

4) розрахунок максимальної відносної похибки прогнозу у відсотках;

5) оцінювання тісноти лінійного зв'язку за значенням коефіцієнта кореляції.

Мають бути засвоєні такі *поняття*: кореляційне поле, область прогнозів, прогноз, довірча область, довірчий інтервал, рівень довіри, напівширина довірчого інтервалу, абсолютна і відносна похибки прогнозу, коефіцієнт кореляції, тіснота лінійного зв'язку.

Завдання до самостійної роботи

Використовуючи дані зі свого індивідуального завдання, виконати таке:

1 Побудувати графіки лінії регресії з 80, 95 і 99 % довірчими областями.

2 Нанести вручну на лінію регресії центр розсіювання.

3 Знайти за графіком прогноз в точці, що відповідає центру розсіювання для всіх трьох значень рівня довіри (80, 95, 99 %), а також прогноз в будь-якій довільній точці з області прогнозів.

4 Знайти за графіком напівширину довірчого інтервалу в точці, що відповідає центру розсіювання для всіх трьох значень рівня довіри (80 , 95 , 99 %): δ_{80}, δ_{95}, δ_{99}.

5 Оцінити максимальну відносну помилку прогнозу (у відсотках) для всіх трьох значень рівня довіри (80, 95, 99 %) за формулою $\left|\dfrac{\delta_\gamma}{y_{\text{ïðîãíîç}}}\right| \cdot 100\,\%$ (δ_γ і $y_{прогноз}$ знаходимо за кресленням).

6 Зробити висновок про взаємозв'язок рівня довіри γ та відносної похибки прогнозу.

7 Знайти коефіцієнт кореляції й оцінити за ним щільність лінійного зв'язку.

Приклад виконання самостійної роботи в пакеті Statistica

1 Створити таблицю даних. Таблиця матиме дві змінні: X – фондовіддача і Y – рентабельність. Потім створити автозвіт.

2 Внести до автозвіту створену таблицю даних.
З'явиться таблиця даних, подібна до таблиці 3.1.

Таблиця 3.1

	1 X	2 Y
1	1,033	1,83
2	0,012	0,58
3	0,045	1,34
4	0,243	1,34
5	0,266	1,64
6	0,302	1,65
7	0,451	1,91
8	1,041	1,96
9	1,423	2,08
10	1,914	2,18

3 Побудувати графіки лінійної регресії з 80, 95 і 99 % довірчими областями.

Для 80 %: Graphs – 2D Graphs – Scatterplots – кнопка Variables (вибрати X, Y – Ok), кнопка Advanced – Graph type : Regular, Fit type : Linear, Confidence level – 0,80 – OK. Для 95 і 99 % аналогічно.

Одержимо три графіки: для рівня довіри 80 % (рис. 3.3), рівня довіри 95 % (рис. 3.4), рівня довіри 99 % (рис. 3.5).

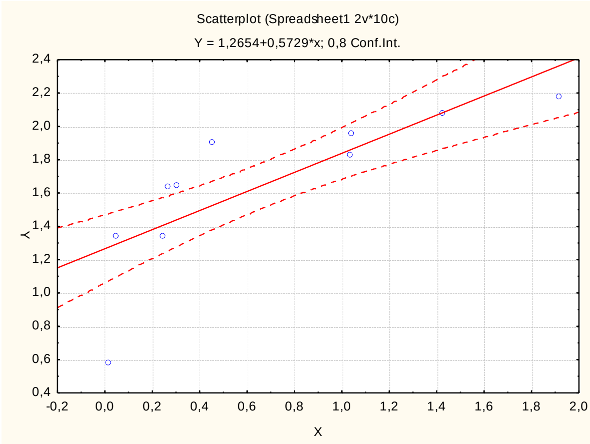


Рисунок 3.3

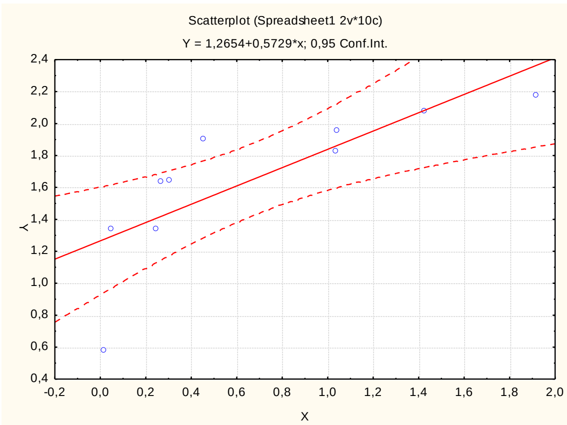


Рисунок 3.4

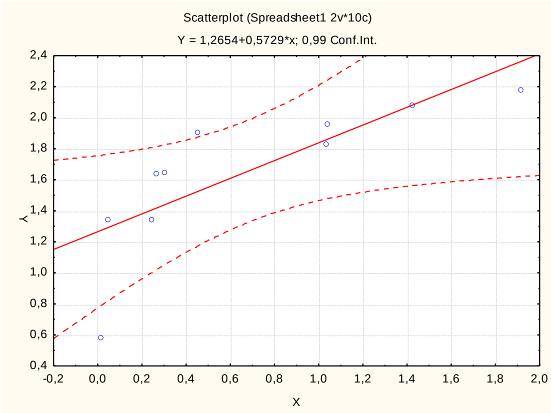


Рисунок 3.5

4 Знайти координати центру розсіювання та область прогнозів.

Для цього виділити таблицю з даними – Statistics – Basic Statistics and Tables – Descriptive statistics – Ok – кнопка Variables – виділити потрібну змінну (в даному випадку X) – OK – кнопка Advanced – активувати опції Mean, Minimum і Maximum – кнопка Summary.

Необхідні статистики виводяться в автозвіт у вигляді таблиці 3.2:

Таблиця 3.2

	Descriptive Statistics (Spreadsheet1)			
Variable	Valid N	Mean	Minimum	Maximum
X	10	0,67300(	0,01200(	1,91400(
Y	10	1,65100(	0,58000(	2,18000(

Середні значення (Mean) дають координати центру розсіювання X = 0,673 і Y = 1,651. Виконання пунктів завдання здійснюють вручну.

Область прогнозів задають інтервалом (Xmin; Xmax), у цьому прикладі (0,012; 1,914).

5 Знайти коефіцієнт кореляції. Для цього створюємо кореляційну таблицю: активуємо таблицю даних – Statistics – Basic Statistics and Tables – Correlation matrices – One variables list (X, Y) – Summary: Correlation matrix. Результат одержуємо у вигляді таблиці 3.3.

Таблиця 3.3

	Correlations (Spreadsheet1) Marked correlations are significant at p < ,05 N=10 (Casewise deletion of missing data)	
Variable	X	Y
X	1,00	0,78
Y	0,78	1,00

Коефіцієнт кореляції для змінних X, Y дорівнює 0,78. Оскільки 0,6 < 0,78 < 0,9, то лінійний зв'язок між цими змінними достатній.

Індивідуальні завдання до самостійної роботи

Варіант 1

Продуктивність праці та рівень рентабельності плодоовочевих консервних заводів області за рік характеризуються такими даними.

Номер заводу	Фактор X	Показник Y
	Продуктивність праці, грн.	**Рівень рентабельності, %**
1	8540	38,34
2	2911	44,69
3	6630	39,4
4	8492	38,93
5	2901	46,96
6	5410	39,48
7	1920	46,07
8	2569	43,5
9	3520	50,11
10	2340	42,79
11	6921	40,15
12	7671	40,44
13	1586	60,76
14	3223	42,99
15	7224	40,69

Варіант 2

Відомі такі дані про збитковість виробництва яловичини у КСП адміністративних районів області за рік.

Номер району	**Фактор X**	**Показник Y**
	Середньодобовий приріст, грн.	**Рівень збитковості, %**
1	249	37,7
2	231	29,7
3	245	26,8
4	242	28,4
5	250	43,2
6	190	48
7	283	33,9
8	273	29,1
9	290	29,8
10	150	66
11	294	19,6
12	196	48,8
13	241	27,4
14	214	53,6
15	188	62,1

Варіант 3

Фондовіддача і рівень рентабельності плодоконсервних заводів області за рік характеризуються такими даними.

Номер заводу	**Фактор X_1**	**Фактор X_2**	**Показник Y**
	Фондовіддача, грн.	**Продуктивність праці, грн.**	**Рівень рентабельності, %**
1	1,12	7343	20,1
2	1,05	3991	20
3	0,99	5760	18
4	0,7	3000	11,7
5	0,98	5241	17,9
6	1,04	4500	16,8
7	1,03	4300	15,6
8	1,35	7500	24,3
9	1,03	6743	18,1
10	0,89	5234	17,8
11	0,78	2500	13
12	0,87	3930	14,2
13	1,43	7433	24,2
14	1,03	6980	20
15	1,05	6740	19,3

Варіант 4

Фондовіддача і рівень рентабельності хлібозаводів області за рік характеризуються такими даними.

Номер заводу	Фактор X	Показник Y
	Фондовіддача, грн.	Рівень рентабель-ності, %
1	33,4	14,3
2	29,1	14,7
3	25,3	11,9
4	27,1	12,1
5	43,3	22,3
6	47,2	23,1
7	49,3	24,3
8	35,7	18,3
9	45,8	27,6
10	52,4	25,3
11	42,1	25,1
12	40,1	20,2
13	33,3	15,7
14	41,2	19,9
15	39,7	17,2

Варіант 5

Наведені дані про питому вагу робітників із спеціальною технічною підготовкою і продуктивність праці плодоовочевих заводів області за рік.

Номер заводу	**Фактор X**	**Показник Y**
	Питома вага робітників із технічною підготовкою, %	**Продуктивність праці, грн.**
1	64	4300
2	61	4150
3	49	3000
4	52	3300
5	53	3300
6	54	4300
7	57	4280
8	61	4100
9	56	3700
10	52	3500
11	60	4000
12	59	4450
13	63	4270
14	50	3300
15	65	4500

Варіант 6

Наведені дані про відносний рівень витрат обігу і рівень рентабельності магазинів промислових товарів за рік.

Номер магазину	**Фактор X**	**Показник Y**
	Відносний рівень витрат обігу, %	**Рівень рентабельності, %**
1	***2***	***4***
1	9,17	6,9
2	6,5	11,1
3	6,81	10,2
4	7,89	8,9
5	7,01	8,3
6	8,91	7,8
7	6,17	13,1
8	10,11	4,9
9	5,98	13,3
10	6,1	10,7
11	5,9	13,4
12	6,13	10,8
13	9,01	4,7
14	10,41	3,9
15	8,13	7,6

Варіант 7

Наведені дані про рівень технічної підготовки робітників і рівень заробітної платні на цукрових заводах області за рік.

Номер заводу	**Фактор X**	**Показник Y**
	Питома вага підготовки робітників із технічною підготовкою, %	**Заробітна плата за місяць, грн.**
1	35	152,2
2	33	180,33
3	37	204,2
4	39	229,95
5	37	204,37
6	41	199,8
7	49	220,11
8	38	218,33
9	58	263,3
10	43	222,72
11	56	239,39
12	47	217,01
13	44	223,4
14	55	237,87
15	54	234,2

Варіант 8

Фондовіддача і рівень рентабельності плодоконсервних заводів області за рік характеризуються такими даними.

Номер заводу	**Фактор X**	**Показник Y**
	Фондовіддача, грн.	**Рівень рентабельності, %**
1	5,46	37,6
2	5,53	37,9
3	7,05	32,1
4	7,29	32,1
5	7,4	31,9
6	7,1	33,4
7	6,25	31,3
8	8,64	39,3
9	5,18	29,8
10	1,81	20
11	2,3	25,5
12	5,53	37,6
13	2,22	20,3
14	3,54	29,1
15	3,23	27,7

Варіант 9

Наведені дані про питому вагу рілля в сільськогосподарських угіддях і рівень збитковості продукції тваринництва по районах області за рік.

Номер району	**Фактор X**	**Показник Y**
	Питома вага ріллі в сільськогосподарських угіддях, %	**Рівень збитковості продукції тваринництва, %**
1	*2*	*4*
1	80	20
2	87,2	37,5
3	90,8	43,4
4	94,7	45,6
5	81,4	23,4
6	79,2	25
7	71,3	17,2
8	86,2	33,3
9	71,4	15
10	77,7	18,7
11	75,4	24,8
12	77,9	34,5
13	87,2	33,1
14	68,1	19,2
15	86,2	31,8

Варіант 10

Наведені дані про питому вагу в товарообігу споживчої кооперації продукції власного виробництва і рівень рентабельності підприємств області за рік.

Номер підприємства	**Фактор X**	**Показник Y**
	Питома вага продукції власного виробництва, %	**Рівень рентабельності, %**
1	25,2	11,8
2	58,2	19,8
3	42,2	14,8
4	46,8	19,4
5	60,5	21,4
6	66,1	20,4
7	26,5	15,4
8	59,9	20,7
9	43,2	16,4
10	47,8	18,4
11	61,8	19,7
12	68,1	22,4
13	32	13,7
14	60,2	22,4
15	44,2	16,7

Варіант 11

Фондовіддача та рівень рентабельності м'ясокомбінатів області за рік характеризуються такими даними.

Номер заводу	**Фактор X**	**Показник Y**
	Фондовіддача, грн.	**Рівень рентабельності, %**
1	1,25	9,2
2	2,32	13,7
3	1,71	11,3
4	1,64	10
5	1,38	6,1
6	1,18	9,1
7	1,44	9,8
8	1,17	6,4
9	1,72	14,2
10	2,21	13,8
11	1,64	13,2
12	1,73	11,4
13	1,17	8,1
14	1,39	9
15	2,07	11,1

Варіант 12

Збитковість вирощування овочів у сільськогосподарських підприємствах і збір овочів з 1 га характеризуються такими даними за рік.

Номер району	**Фактор X**	**Показник Y**
	Збір овочів з 1га, ц	**Рівень збитковості,%**
1	*2*	*4*
1	52,8	31,4
2	72,6	30,9
3	50,4	37,1
4	33,4	45,7
5	31,5	57,7
6	54,6	46,7
7	54,3	33,3
8	36,6	63,8
9	15,6	68,8
10	73,2	29,8
11	65,9	39,4
12	44,6	46,2
13	23,7	68,8
14	64,6	34
15	25,6	47,6

Варіант 13

Збитковість вирощування овочів у сільськогосподарських підприємствах і витрати праці (людино-годин на 1 ц) характеризуються такими даними за рік.

Номер району	**Фактор X**	**Показник Y**
	Витрати праці на 1 ц, людино-годин	**Рівень збитковості, %**
1	2,3	8,8
2	26,8	39,4
3	22,8	26,2
4	56,6	78,8
5	16,4	34
6	26,5	47,6
7	26	43,7
8	12,4	23,6
9	10	19,9
10	41,7	50
11	47,9	63,1
12	32,4	44,2
13	20,2	11,2
14	39,6	52,8
15	18,4	20,2

Варіант 14

Збитковість вирощування овочів у сільськогосподарських підприємствах і **собівартість 1 ц** характеризуються такими даними за рік.

Номер району	**Фактор X**	**Показник Y**
	Собівартість 1 ц, грн.	**Рівень збитковості,%**
1	31,84	31,4
2	32,3	20,9
3	32,21	37,1
4	48,95	45,7
5	42,48	57,7
6	35,38	46,7
7	29,11	33,3
8	67,06	63,8
9	65,52	68,8
10	21,26	12,8
11	31,29	39,4
12	33,63	26,2
13	73,35	68,8
14	40,12	34
15	43,63	47,6

Варіант 15

Рівень рентабельності й один із показників господарської діяльності торгових підприємств характеризуються такими даними за рік.

Номер підприємства	**Фактор X**	**Показник Y**
	Товарообіг на одну людину, грн.	**Рівень рентабельності, %**
1	27	3,62
2	29	2,02
3	24	2,77
4	21	2,01
5	33	4,33
6	28	4,01
7	23	2,12
8	28	3,73
9	30	3,92
10	25	2,87
11	22	2,11
12	34	4,39
13	31	4,11
14	22	2,13
15	29	3,2

Самостійна робота 4
КРИТЕРІЙ УЗГОДЖЕНОСТІ ПІРСОНА, КОЛМОГОРОВА – СМІРНОВА

4.1 Короткі теоретичні відомості

Дані, одержані під час контролювання технологічного процесу, для подальшої обробки бажано подати у вигляді теоретичного розподілу, що максимально відповідає експериментальному розподілу. Перевірку гіпотези про вигляд функції розподілу проводять за критеріями згоди – Пірсона, Колмогорова та іншими.

Критеріями згоди називають спеціальні випадкові величини, призначення яких – перевіряти гіпотезу про вигляд розподілу.

Критерій узгодженості Пірсона є універсальним. Для його застосування потрібні вибірки великих обсягів $(n \geq 100)$.

Нехай дана вибірка: $x_1, x_2, \dots x_n$. За нею розраховують спостережуване значення критерію $\chi^2_{набл}$, за статистичними таблицями знаходять критичне значення критерію $\chi^2_{кр}$. Залежно від того, як співвідносяться ці дві величини, гіпотезу про вигляд розподілу або приймають, або відкидають.

4.1.1 Послідовність дій

1 Розмах вибірки від $x_{\min}$ до $x_{\max}$ розбивають на s непересічних частин (рис. 4.1), які можуть бути різної довжини: $m_1 + m_2 + \dots m_s = n$. Величини $m_1, m_2, \dots m_s$ називають *вибірковими частотами*.

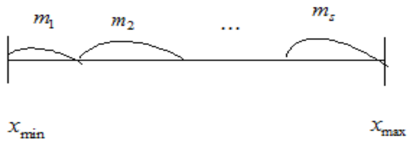


Рисунок 4.1

2 Обчислюють теоретичні частоти m_i' :

$$m_i' = np_i \,, \qquad (4.1)$$

де p_i – імовірність потрапляння випадкових величин x в інтервал s_i . Ця величина залежить від того, який теоретичний закон розподілу перевіряється.

Очевидно, що якщо гіпотеза про вигляд закону розподілу вірна, то $m_i \approx m_i'$.

3 Обчислюють статистику

$$\chi^2_{\text{í àáë}} = \sum_{i=1}^{s} \frac{\left(m_i - m_i'\right)^2}{m_i'} \qquad (4.2)$$

яка при $n \to \infty$ прагне до розподілу χ^2 с $\nu = s - r - 1$ ступенями свободи, де r – число параметрів теоретичного розподілу $F(x)$, оцінюваних на основі вибірки, а s – число інтервалів, на які розбито всю безліч спостережених даних (табл. 4.1).

Таблиця 4.1

Вигляд розподілу	Число ступенів свободи
Нормальний	$s-1-2=s-3$
Експоненціальний	$s-1-1=s-2$
Рівномірний	$s-1-0=s-1$

4 Для знаходження критичного значення $\chi^2_{кр}$ використовують статистичні таблиці $\chi^2(\alpha, n)$ або ймовірнісний калькулятор у програмі Statistica, де α – рівень значущості гіпотези.

5 Якщо $\chi^2_{набл} < \chi^2_{кр}$, то гіпотеза про вигляд розподілу приймається. Якщо $\chi^2_{набл} > \chi^2_{кр}$ – відкидається.

Якщо жодна гіпотеза не підтвердилася при використанні критерію Пірсона, то можна змінити рівень значущості гіпотези або ж виконати перевірку за критерієм Колмогорова – Смирнова.

Цей критерій дозволяє оцінити істотність відмінностей між двома вибірками, можливе його застосування для порівняння емпіричного розподілу з теоретичним. Критерій дозволяє знайти точку, в якій сума накопичених частот розбіжностей між двома розподілами є найбільшою, й оцінити достовірність цієї розбіжності. Нульова гіпотеза H_0 = {відмінності між двома розподілами недостовірні (роблячи висновки з точки максимальної накопиченої розбіжності між ними)}.

Алгоритм застосування критерію Колмогорова – Смирнова можна подати так:

1. Записати варіаційні ряди експериментальної та контрольної групи.

2. Обчислити відносні частоти f для цих двох вибірок.

3. Знайти модулі різниць $\left| f_{експ} - f_{контр} \right|$.

4. Знайти найбільший модуль $d_{\max}$.

5. Обчислити $\lambda_{експ} = d_{\max} \cdot \sqrt{\dfrac{n_1 \cdot n_2}{n_1 + n_2}}$, де n_1– сума емпіричних частот (експериментальної групи), n_1– сума теоретичних частот.

6. Із заданим рівнем значущості гіпотези α обчислити $\lambda_{кр(\alpha)}$. Якщо $\lambda_{експ} > \lambda_{кр(\alpha)}$, то гіпотеза H_0 відкидається при заданому рівні значущості.

Мета самостійної роботи

Повинні бути набуті такі *вміння*: побудова гістограм у парі з кривими розподілу; розбиття розмаху вибірки на часткові інтервали довільної довжини; знаходження критичного значення $\chi^2_{кр}$ для заданого рівня значущості; розрахунок теоретичних частот розподілу.

Повинні бути освоєні такі *поняття*: критерій згоди, обрані (емпіричні) частоти, теоретичні частоти, статистична гіпотеза, рівень значущості гіпотези, число ступенів свободи.

Робота розрахована на 4 години.

Завдання до самостійної роботи

Використовуючи дані зі свого індивідуального завдання, виконати наступне:

1 Створити в STATISTICA файл початкових даних.

2 Створити файл автозвіту.

3 На основі файла даних створити статистичну таблицю: обсяг вибірки (Valid N), математичне сподівання (Mean), середнє квадратичне відхилення (Std.Dev) , дисперсію (Variance).

4 За файлом даних побудувати 3 гістограми в парі з такими кривими розподілу: нормального, експоненціального, рівномірного. За виглядом рисунків зробити висновок про те, чи можна однозначно стверджувати, який вигляд розподілу відповідає вибірці.

5 Перевірити відповідність вигляду розподілу вибірці за допомогою тестів Колмогорова – Смирнова і χ^2.

Приклад виконання самостійної роботи в пакеті Statistica

1 Створюємо в STATISTICA файл вихідних даних.

2 Створюємо файл автозвіту.

3 Створюємо статистичну таблицю.

Для цього виділити таблицю з даними – Statistics – Basic Statistics and Tables – Descriptive statistics – Ok – кнопка Variables – виділити потрібну змінну (в даному випадку X) – OK – кнопка Advanced – активувати опції Valid N, Mean, Standard Deviation, Variance – кнопка Summary (табл. 4.2).

Таблиця 4.2

	Descriptive Statistics (Spreadsheet2)			
Variable	Valid N	Mean	Variance	Std.Dev.
X	80	2,00750	0,63284	0,79551

4 *Будуємо гістограму.*

Для цього потрібно активувати таблицю *.sta – пункт верхнього меню Graphs – 2D Graphs – Histogramm – кнопка Variables – виділити потрібну змінну (в нашому випадку X) – OK – кнопка Advanced (рис. 4.2). Вибрати опції Graph type : Regular, Fit type : Normal і число часткових інтервалів (Categories) що дорівнює 7 (вибрати самим). Одержимо гістограму (рис. 4.3).

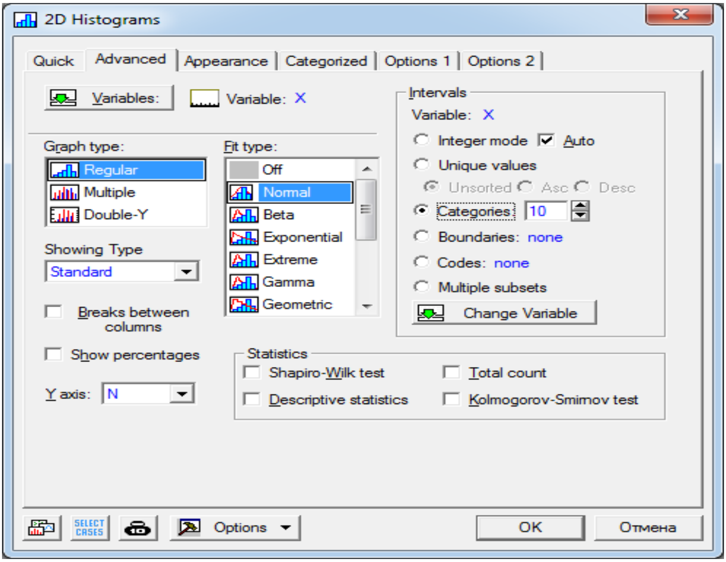


Рисунок 4.2

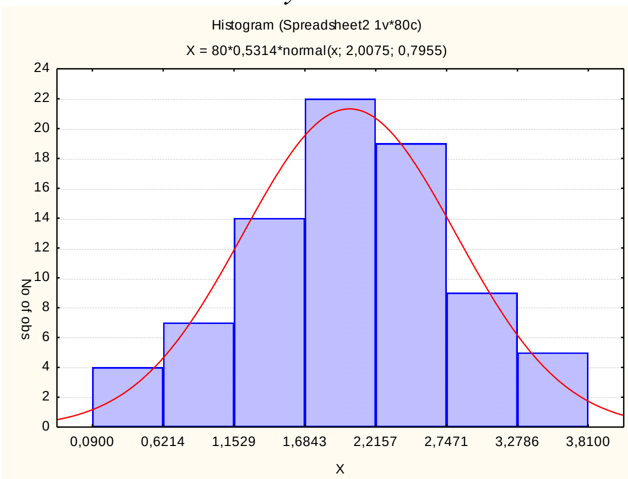


Рисунок 4.3

Аналогічно будуємо гістограми з кривими для експоненціального (Exponential) і нормального (Lognormal) розподілів (рис. 4.4, 4.5).

За виглядом графіків можна попередньо припустити, що найбільш відповідним є нормальний розподіл.

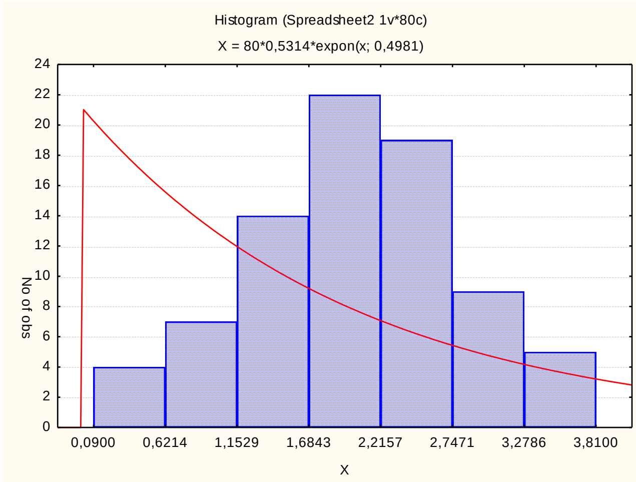


Рисунок 4.4

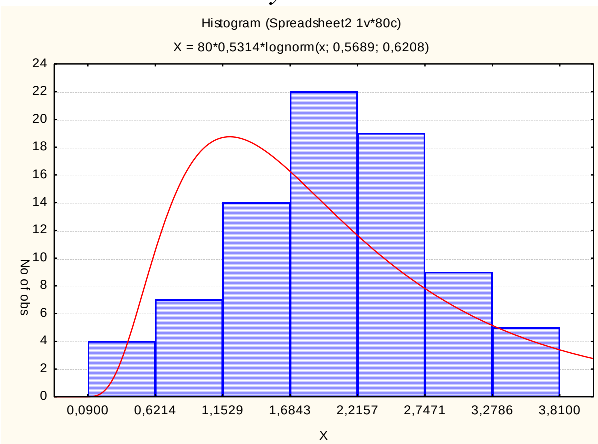


Рисунок 4.5

5 Підключаємо модуль Distribution Fitting (Підгонка розподілів) (рис. 4.6).

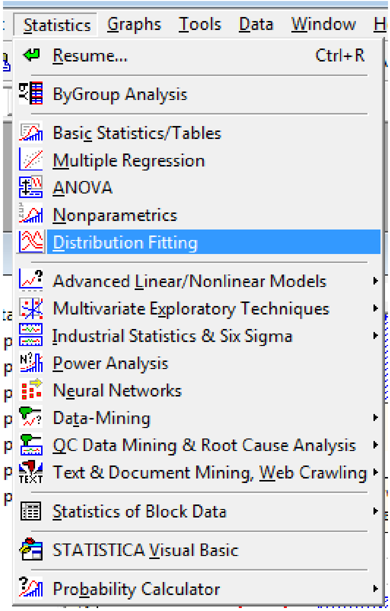


Рисунок 4.6

У вікні ставимо перемикач на Continuous Distribution (Безперервні розподіли) (рис. 4.7), OK.

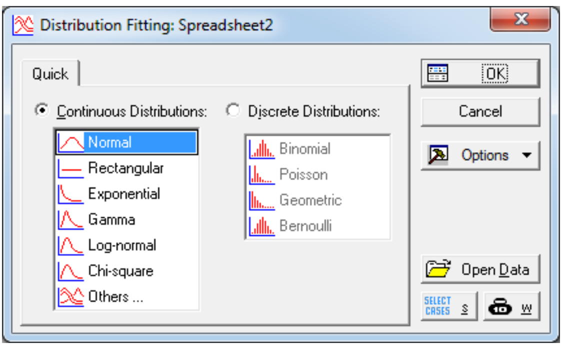


Рисунок 4.7

У вікні Fitting Continuous Distributions вибираємо

Variables – X (змінна – X), Distribution: Normal (Розподіл – нормальний), кнопка Parameters (рис. 4.8); кнопка Options (рис. 4.9) – Summary.

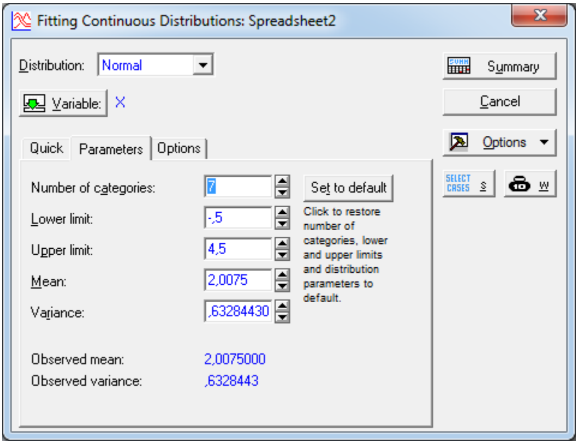


Рисунок 4.8

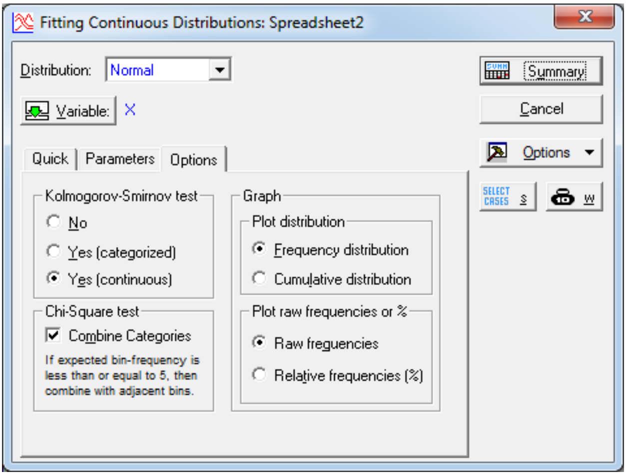


Рисунок 4.9

Результат одержимо у вигляді таблиці 4.3. Значення статистики хі-квадрат невелике – Chi-Square = 2,34730, що свідчить на користь гіпотези. Питання про те, що таке велике чи невелике значення статистики, знімається поняттям рівня значущості. У заголовку таблиці бачимо, що рівень значущості p = 0,30924, що означає, що гіпотезу про згоду даних із нормальним розподілом можна відкинути на рівні 0, 30924. Іншими словами, відкидаючи гіпотезу про нормальний розподіл, ми ризикуємо помилитися з імовірністю 30,9 %.

Аналогічним чином отримуємо таблиці для перевірки гіпотез про згоду даних з експоненціальним і логнормальним розподілами (табл. 4.4, 4.5).

Висновки із цих таблиць зробіть самостійно.

Індивідуальні завдання до самостійної роботи

Індивідуальні завдання до самостійної роботи взяти з самостійної роботи 2.

Таблиця 4.3

Number of valid cases:80
Observed mean = 2,007500, Observed variance = 0,632844
Distribution: Normal
Parameters: Mean = 2,007500, Variance = ,6328443

Upper Boundary	Variable:X, Distribution: Normal (Spreadsheet2.sta) Kolmogorov-Smirnov d = 0,07014, p = n.s., Lilliefors p = n.s. Chi-Square = 2,34730, df = 2 (adjusted) , p = 0,30924								
	Observed Frequency	Cumulative Observed	Percent Observed	Cumul. % Observed	Expected Frequency	Cumulative Expected	Percent Expected	Cumul. % Expected	Observed-Expected
<= 0,21429	2	2	2,50000	2,5000	0,96746	0,96746	1,20932	1,2093	1,0325
0,92857	6	8	7,50000	10,0000	6,03315	7,00061	7,54144	8,7508	-0,0331
1,64286	17	25	21,25000	31,2500	18,86677	25,86737	23,58340	32,3342	-1,8667
2,35714	26	51	32,50000	63,7500	27,72117	53,58854	34,65140	66,9857	-1,7211
3,07143	24	75	30,00000	93,7500	19,16789	72,75641	23,95980	90,9455	4,8321
3,78571	4	79	5,00000	98,7500	6,22767	78,98409	7,78458	98,7301	-2,2276
< Infinity	1	80	1,25000	100,0000	1,01591	80,00000	1,26989	100,0000	-0,0159

Таблиця 4.4

Number of valid cases:80
Observed mean = 2,007500, Observed variance = 0,632844
Distribution: Exponential
Parameters: Lambda = ,4981320

	Variable:X, Distribution: Exponential (Spreadsheet2.sta) Kolmogorov-Smirnov d = 0,31349, p < 0,01 Chi-Square = 86,28110, df = 5, p = 0,00000								
Upper Boundary	Observed Frequency	Cumulative Observed	Percent Observed	Cumul. % Observed	Expected Frequency	Cumulative Expected	Percent Expected	Cumul. % Expected	Observed-Expected
<= 0,21429	2	2	2,5000	2,500	8,0994	8,0994	12,6553	10,124	-6,0994
0,92857	6	8	7,5000	10,000	21,5265	29,6259	33,6352	37,032	-15,526
1,64286	17	25	21,2500	31,250	15,0816	44,7076	23,5650	55,884	1,9184
2,35714	26	51	32,5000	63,750	10,5663	55,2739	16,5098	69,092	15,433
3,07143	24	75	30,0000	93,750	7,4028	62,6767	11,5669	78,345	16,597
3,78571	4	79	5,0000	98,750	5,1864	67,8632	8,1038	84,829	-1,1865
< Infinity	1	80	1,2500	100,000	12,1367	80,0000	18,9637	100,000	-11,136

Таблиця 4.5

Number of valid cases:80
Observed mean = 2,007500, Observed variance = 0,632844
Distribution: Log-normal
Parameters: Mean = ,5689392, Variance = ,3854039

Variable:X, Distribution: Log-normal (Spreadsheet2.sta)
Kolmogorov-Smirnov d = 0,18881, p < 0,01
Chi-Square = 30,92500, df = 3 (adjusted) , p = 0,00000

Upper Boundary	Observed Frequency	Cumulative Observed	Percent Observed	Cumul. % Observed	Expected Frequency	Cumulative Expected	Percent Expected	Cumul. % Expected	Observed-Expected
<= 0,21429	2	2	2,5000	2,500	0,0271	0,0271	0,0339	0,034	1,9728
0,92857	6	8	7,5000	10,000	11,9842	12,0114	14,9803	15,014	-5,9842
1,64286	17	25	21,2500	31,250	24,2697	36,2811	30,3371	45,351	-7,2697
2,35714	26	51	32,5000	63,750	18,0339	54,3151	22,5424	67,893	7,9660
3,07143	24	75	30,0000	93,750	10,7698	65,0850	13,4623	81,356	13,2301
3,78571	4	79	5,0000	98,750	6,1357	71,2207	7,6696	89,025	-2,1357
< Infinity	1	80	1,2500	100,000	8,7792	80,0000	10,9741	100,000	-7,7792

Додаток А
Стислі відомості про систему statistica v6.0

А.1 Структура пакета STATISTICA V6.0

Програма STATISTICA містить декілька незалежно працюючих модулів, що відкриваються за допомогою пункту меню **Statistics** (рис. А.1). У кожному модулі зібрані логічно пов'язані між собою статистичні процедури. Завантажити можна відразу кілька модулів. Кнопки цих модулів знаходяться у нижній частині екрана. Переходити між ними можна стандартним способом, натиснувши лівою клавішею мишки по відповідну кнопку.

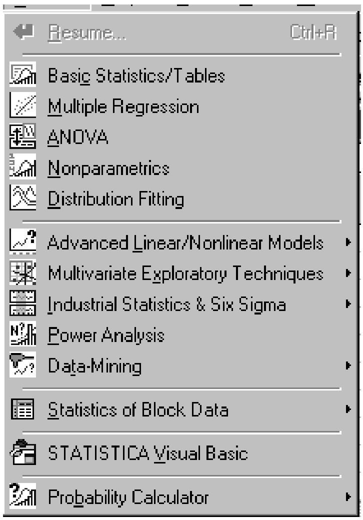


Рисунок А.1

А.2 Створення нової таблиці даних

Для цього потрібно вибрати пункт меню **File – New – Ok.** Відкриється порожня електронна таблиця розміром 10х10 (рис. А.2). У стовпчиках розміщені змінні (Vars), в рядках – випадки (Cases).

А.3 Вилучення та додання нових змінних і випадків

Виконуються командами Delete (вилучити) і Add (додати). Після виділення рядка або стовпця натиснути кнопку Vars, якщо вилучаються або додаються змінні, й зазначити, скільки елементів вилучається чи додається, Ok. Для випадків – аналогічно, але з кнопкою Cases.

	1 Var1	2 Var2	3 Var3	4 Var4	5 Var5	6 Var6	7 Var7	8 Var8	9 Var9	10 Var10
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										

Рисунок А.2

А.4 Коригування таблиці

Коректування таблиці зводиться до коректування назв змінних, вмісту стовпців цілком й окремих кліток. Для цього клацнути по імені стовпця, потім – по кнопці Vars (змінні), за пунктом меню **Current Specs** (поточні специфікації). Після цього можна задавати нове ім'я стовпцю в рядку Name (ім'я) і нові значення випадків за допомогою формули у віконці Long name (повне ім'я) (рис. А.3). Також можна змінити подання числа (автоматично – 8 позицій (Column width), число позицій після коми (Decimals) – 3).

Коригування числових значень в окремих осередках виконується, як в EXCEL.

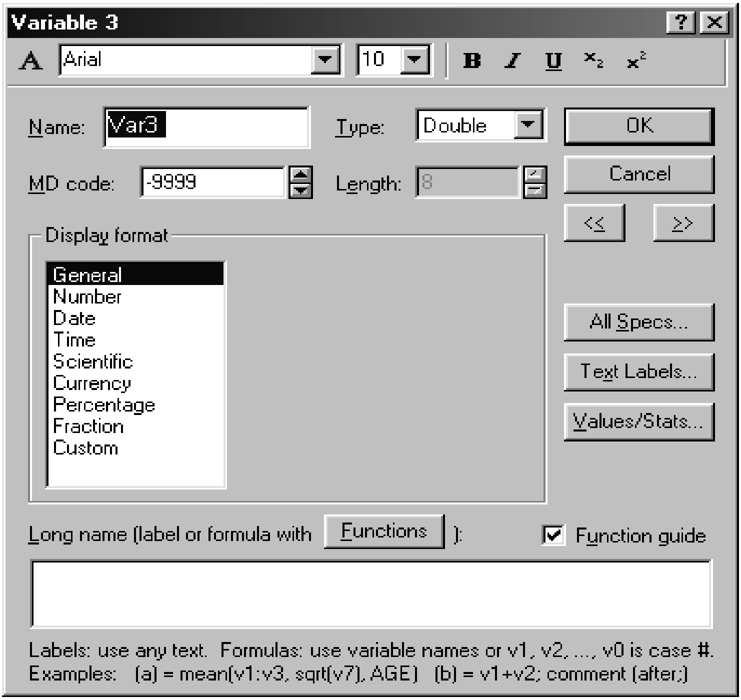


Рисунок А.3

А.5 Обчислення статистичних характеристик для значень змінних

(наприклад, максимальне, мінімальне, середнє значення, дисперсія і т. ін.).

Активізувати таблицю даних, потім активізувати пункти меню **Statistics – Basic Statistics/Tables – Descriptive Statistics** – вкладка **Advanced** (аналіз – описові статистики – усі статистики) (рис. А.4), виділити змінні, для яких шукають характеристики (кнопка Variables), Ok, вибрати зі списку потрібні статистичні характеристики (наприклад, Min – мінімальне значення, Max – максимальне значення, Valid N – обсяг вибірки, Mean – середнє значення, Standard Deviation – середнє квадратичне відхилення, Variance –

дисперсія), кнопка Summary. На екрані з’явиться таблиця з потрібними характеристиками.

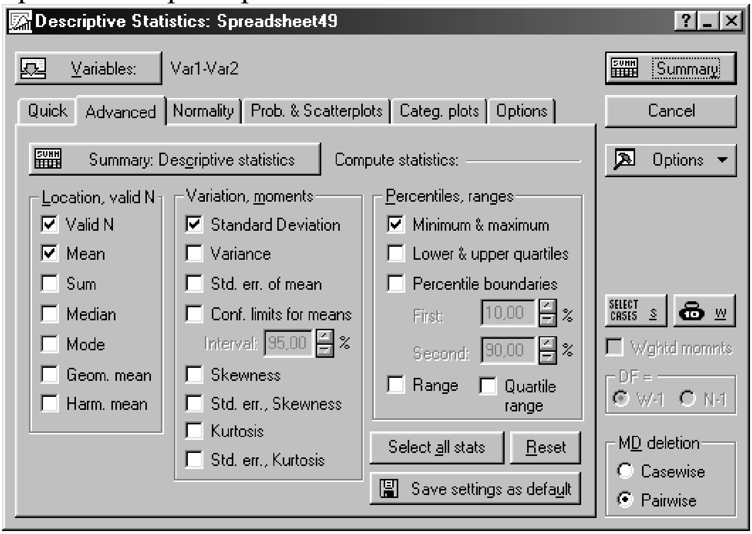


Рисунок А.4

А.6 Одержання графіка і рівняння лінійної регресії

Активізувати таблицю. Вибрати пункт меню **Craphs – 2D Craphs – Scatterplots** – вкладка **Advanced** (графіки, статистичні двомірні графіки, точковий графік), вибрати змінні Variables (для аргументу – x і функції – y), Ok, вибрати опції Regular, Linear (регулярний, лінійний), Ok (рис. А.5).

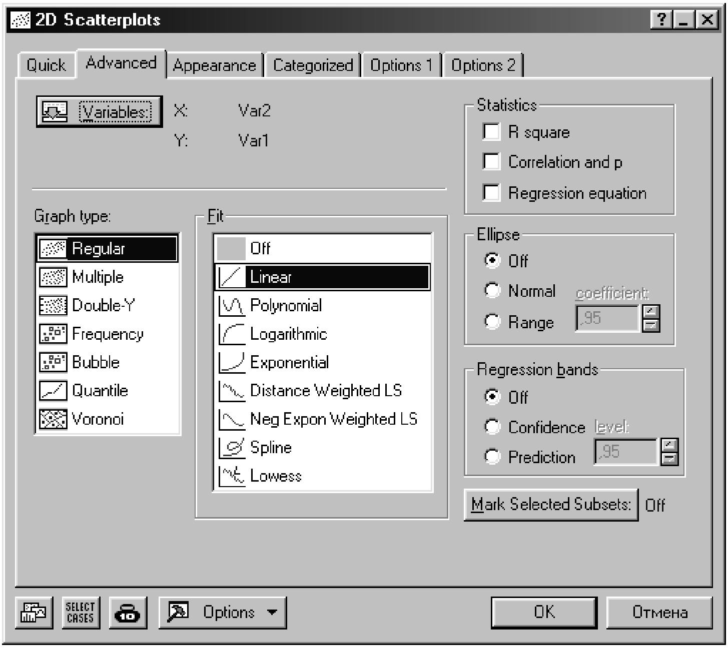


Рисунок А.5

З'явиться графік лінійної регресії, над яким записане рівняння регресії (рис. А.6).

А.7 Типи файлів у системі Statistica

Типи файлів:

*.sta – початкові дані;

*.stw – результати обробки даних, Workbook;

*.str, або *.rtf – звіт.

Довжина імені файла, як і будь-якого іншого ідентифікатора, не більше 8 символів.

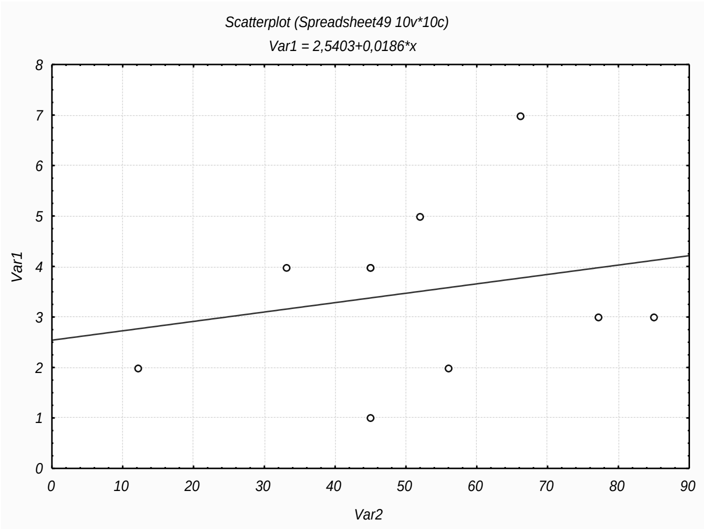


Рисунок А.6

А.8 Створення автозвіту

Автозвіт бажано створювати при кожному сеансі роботи з пакетом для того, щоб усі результати роботи (таблиці та графіки) запам’ятовувалися в автозвіті.

Далі визначений шлях для створення стислого автозвіту: File, Output Manager. З'явиться багатосторінкове меню.

На сторінці Output Manager потрібно відмітити опції: Single Workbook, Place results in Workbook automatically, Also send to Report Window, Single Report (рис. А.7). На сторінці Workbook, крім уже заданих опцій, у полі Add to Workbook performs відмітити опцію Copy. На сторінці Report, крім уже заданих опцій, у полі Add to Report performs відмітити опцію Copy. Після цього натиснути Ok. Тепер усі розрахункові таблиці та графіки автоматично заносяться до звіту.

Файл звіту потрібно зберегти з розширенням *.rtf. Тепер його можна редагувати у WORD.

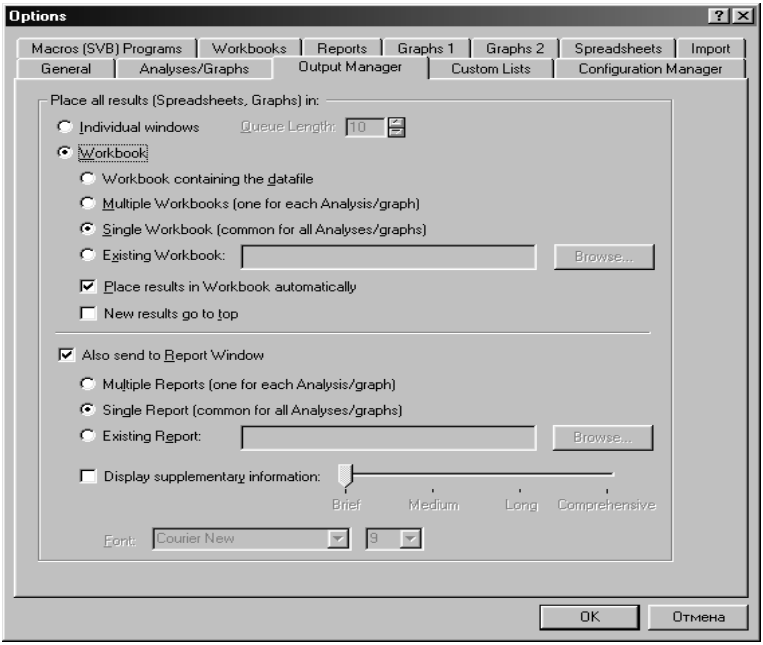

Options
Macros (SVB) Programs
Workbooks
Reports
Graphs 1
Graphs 2
Spreadsheets
Import
General
Analyses/Graphs
Output Manager
Custom Lists
Configuration Manager
Place all results (Spreadsheets, Graphs) in:
Individual windows
Queue Length:
10
Workbook
Workbook containing the datafile
Multiple Workbooks (one for each Analysis/graph)
Single Workbook (common for all Analyses/graphs)
Existing Workbook:
Browse...
Place results in Workbook automatically
New results go to top
Also send to Report Window
Multiple Reports (one for each Analysis/graph)
Single Report (common for all Analyses/graphs)
Existing Report:
Browse...
Display supplementary information:
Brief
Medium
Long
Comprehensive
Font:
Courier New
9
OK
Отмена

Рисунок А.7

Додаток Б
Приклад виконання самостійної роботи 2 в середовищі MathCad

1 Засобами стандартного редактора Блокнот створюємо файл даних dan.dat.

2 Розраховуємо такі значення: обсяг вибірки n, математичне сподівання mean, розмах вибірки R = xmin - xmax, середньоквадратичне відхилення σ, асиметрію Sk, ексцес Ex.

$ORIGIN := 1$

$i := 1 .. 80$

$\xi_i := READ("dan.dat")$

$xmax := max(\xi)$ $xmin := min(\xi)$

$xmax = 3.75$ $xmin = 0.09$

$\xi := sort(\xi)$ $n := length(\xi)$ $n = 80$

$R := xmax - xmin$

$mean := mean(\xi)$ $mean = 2.03$

$disp := var(\xi)\cdot\frac{n}{n-1}$ $disp = 0.574$

$\sigma := \sqrt{disp}$ $\sigma = 0.758$

$\mu_3 := \left(\frac{1}{n}\right)\cdot\sum_{i=1}^{n}\left(\xi_i - mean\right)^3$ $\mu_4 := \frac{1}{n}\cdot\sum_{i=1}^{n}\left(\xi_i - mean\right)^4$

$Sk := \frac{\mu_3}{\sigma^3}$ $Sk = -0.173$

$Ex := \left(\frac{\mu_4}{\sigma^4}\right) - 3$ $Ex = -0.288$

За значеннями асиметрії та ексцесу робимо висновок, чи значно відрізняється розподіл випадкової величини від нормального.

3 Будуємо гістограму (рис. Б.1), задавши число часткових інтервалів, що дорівнює 10.

$m := 10 \qquad h := \frac{R}{m} \qquad h = 0.366$

$j := 1 .. m \qquad k := 1 .. m - 1$

$x_j := xmin + \left(\frac{h}{2}\right) \cdot (2 \cdot j - 1)$

$f := hist(x, \xi)$

Висота стовпців дорівнює кількості точок, що потрапили у відповідний частковий інтервал. Ця гістограма вимог, що ставляться до гістограм, не задовольняє (пояснити).

Зменшуємо число часткових інтервалів до 7 і будуємо нову гістограму (рис. Б.2).

Дайте відповідь на питання: чи задовольняє ця гістограма вимогам, що пред'являються до гістограм? Чи можна ще зменшити число часткових інтервалів?

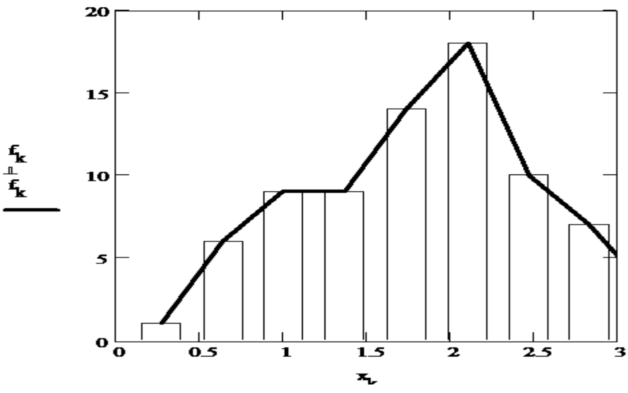


Рисунок Б.1

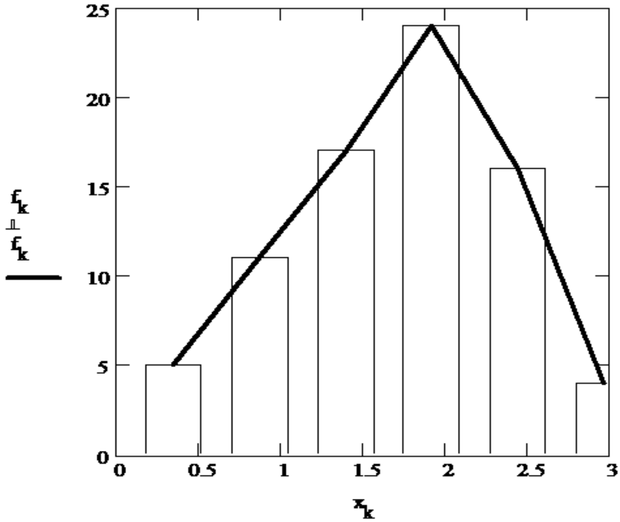


Рисунок Б.2

4 Імовірність потрапляння випадкової величини в проміжок (a, b) розраховуємо за формулою Р (a < X < b) = F (b) – F (a), де F (x) – функція розподілу випадкової величини.

Значення функції розподілу знаходимо за допомогою вбудованої функції pnorm(x, mean, σ):

pnorm(3.2, mean, σ) – pnorm(2.1, mean, σ) = 0.402

$P(2{,}1 < X < 3{,}2) = F(3{,}2) - F(2{,}1) = 0{,}402.$

5 Оскільки технологічний процес відрегульований правильно, то вибіркове середнє можна прийняти за значення параметра, заданого в технічній документації. Десятивідсоткове відхилення знаходимо за формулою $\delta = 0{,}1 \cdot \overline{x} = 0{,}1 \cdot 2{,}03.$ Потім обчислюємо імовірність:

$$\mathrm{P}\left(\left|\mathrm{x} - \overline{\mathrm{x}}\right| < \delta\right) = 2\mathrm{F}\left(\frac{\delta}{\sigma}\right).$$

$$\delta := 0.1 \cdot \text{mean} \qquad\qquad \delta = 0.203$$

$$2 \cdot \text{pnorm}\left(\frac{\delta}{\sigma}, \text{mean}, \sigma\right) = 0.02$$

Одержуємо відповідь: вихід придатної продукції при заданому допуску $0{,}1\overline{x}$ становить 2 % від усієї продукції.

Додаток В
Приклад виконання самостійної роботи 3 в середовищі MathCad.

Засобами редактора Блокнот створюємо файли даних dan _ x.dat і dan _ y.dat.

Дані будуть подані у вигляді рис. В.1, В.2.

$ORIGIN := 1 \qquad N := 15 \qquad i := 1..N$

$x_i := READ("dan\_x.dat")$

$x^T =$

	1	2	3	4	5	6	7	8	9	10
1	7.89	14.41	6.01	9.17	6.78	8.91	6.17	10.11	5.98	6.1

Рисунок В.1

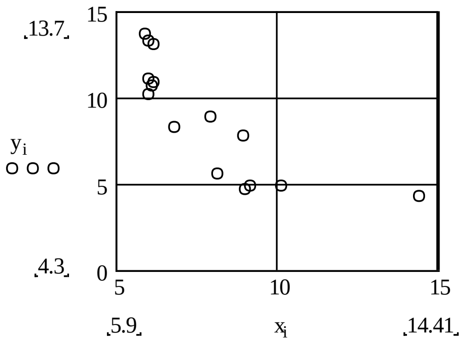
$y_i := READ("dan\_y.dat")$

$y^T =$

	1	2	3	4	5	6	7	8	9	10
1	8.9	4.3	10.2	4.9	8.3	7.8	13.1	4.9	13.3	10.7

Рисунок В.2

Будуємо кореляційне поле (рис. В.3) і знаходимо коефіцієнт кореляції.



$corr(x, y) = -0.808$

Рисунок В.3

Із вигляду кореляційного поля можна припустити, що залежність між X і Y існує. Для визначення тісноти лінійного зв'язку знаходимо коефіцієнт кореляції.

Коефіцієнт кореляції для змінних X, Y дорівнює -0,808. Оскільки 0,6 < |- 0,808|< 0,9, то лінійний зв'язок між цими змінними достатній.

Методом найменших квадратів знаходимо коефіцієнти моделі y = b0 + b1x:

$$b0 := \operatorname{intercept}(x, y) \qquad b0 = 17{,}818$$

$$b1 := slope(x, y) \qquad b1 = -1{,}157$$

$$yr_i := b0 + b1 \cdot x_i$$

Знаходимо координати центра розсіяння та область прогнозів.

Середнє значення фактора X :

$$Xmean := mean(x) \qquad Xmean = 7{,}78$$

Середнє значення фактора Y:

$$Ymean := mean(y) \qquad Ymean = 8{,}82$$

Середні значення (Mean) дають координати центра розсіювання:

$$(\bar{x}, \bar{y}) = (7.78, 8.82)$$

Розраховуємо значення дисперсії :

$$S2 := \left(\frac{1}{N-1}\right) \cdot \sum_{k=1}^{N} \left(y_k - yr_k\right)^2 \qquad S2 = 3.88$$

Область прогнозів задаємо у вигляді Xmin ≤ X ≤ ≤ Xmax, де мінімальне і максимальне значення знаходимо таким чином:

$X\min := \min(x) \qquad X\min = 5.9$

$X\max := \max(x) \qquad X\max = 14.41$

Область прогнозів задаємо інтервалом (Xmin; Xmax), у цьому прикладі (5,9; 14,41).
Будуємо графіки лінійної регресії з 80, 95 та 99 % довірчими областями.
Одержимо три графіки (рис. В.4, В.5, В.6).

Рівень довіри γ = 80 .

$$\alpha := 0.20 \qquad t := qt\left[1-\left(\frac{\alpha}{2}\right), N-2\right]$$

$$\delta_i := t\cdot\sqrt{S2}\cdot\sqrt{\left(\frac{1}{N}\right)+\frac{\left(x_i - Xmean\right)^2}{\sqrt{\sum_{k=1}^{N}\left(x_k - Xmean\right)^2}}}$$

$$yleft_i := yr_i - \delta_i \qquad yright_i := yr_i + \delta_i$$

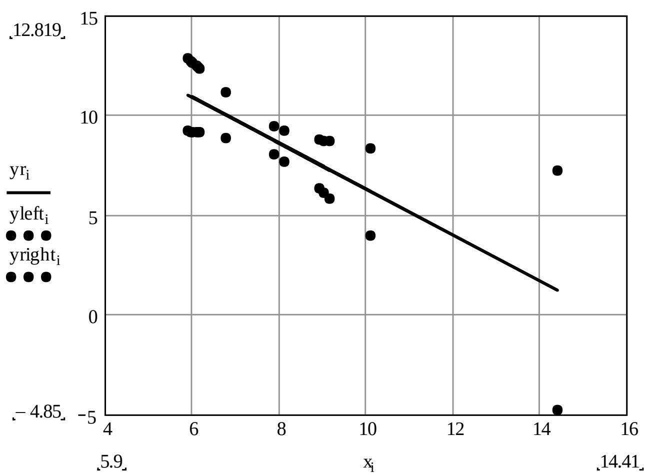


Рисунок В.4

Рівень довіри γ = 95 %.

$$\alpha := 0.05 \qquad t := qt\left[1 - \left(\frac{\alpha}{2}\right), N - 2\right]$$

$$\delta_i := t \cdot \sqrt{S2} \cdot \sqrt{\left(\frac{1}{N}\right) + \frac{(x_i - Xmean)^2}{\sqrt{\sum_{k=1}^{N} (x_k - Xmean)^2}}}$$

$$yleft_i := yr_i - \delta_i \qquad yright_i := yr_i + \delta_i$$

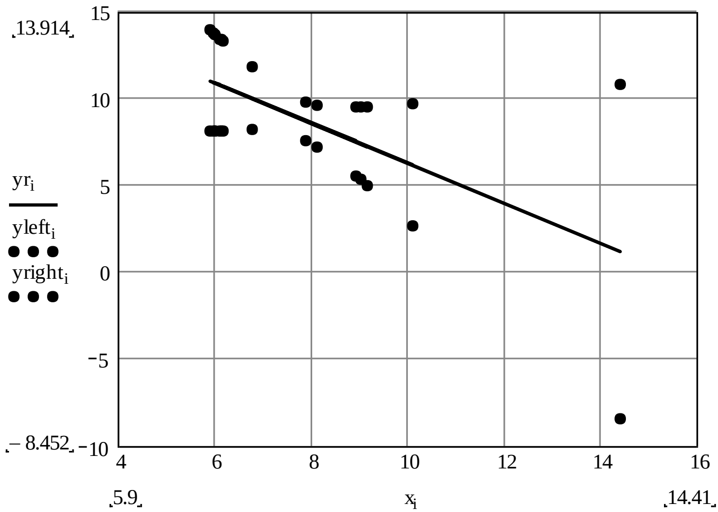


Рисунок В.5

Рівень довіри $\gamma = 99\ \%$.

$$\alpha := 0.01 \qquad t := qt\left[1-\left(\frac{\alpha}{2}\right), N-2\right]$$

$$\delta_i := t\cdot\sqrt{S2}\cdot\sqrt{\left(\frac{1}{N}\right)+\frac{(x_i - Xmean)^2}{\sqrt{\sum_{k=1}^{N}(x_k - Xmean)^2}}}$$

$$yleft_i := yr_i - \delta_i \qquad yright_i := yr_i + \delta_i$$

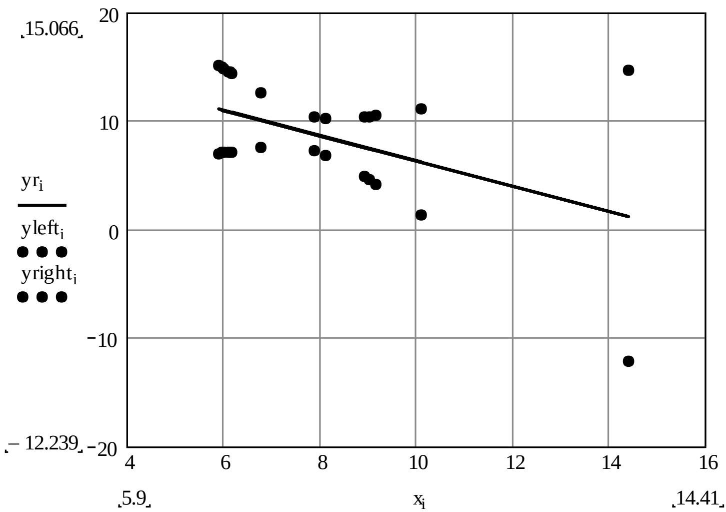


Рисунок В.6

Додаток Г
Приклад виконання самостійної роботи 2 в Excel

1 Створити файл початкових даних під своїм ім'ям на основі файла «Шаблон_лаб 2.xls».

2 Розрахувати основні статистики: обсяг вибірки (функція СЧЕТЗ () або COUNTA()), математичне сподівання (функція СРЗНАЧ () або AVERAGE()), розмах вибірки (функції МИН (), МАКС () або MIN (), MAX()), середнє квадратичне відхилення (функція СТАНДОТКЛОН () або STDEV()), асиметрію (функція СКОС () або SKEW()), ексцес (функція ЭКСЦЕСС () або KURT()).

Одержимо (табл. Г.2).

Таблиця Г.2

Обсяг вибірки	80
Математичне сподівання	2,0295
Мінімум	0,09
Максимум	3,75
Дисперсія	0,574169
Середнє відхилення	0,75774
Асиметрія	-0,17991
Ексцес	-0,15484

3 Побудувати гістограми. Взяти число часткових інтервалів, що дорівнює 10 (потім 9, 8, 7). Одержимо гістограми (рис. Г.1).

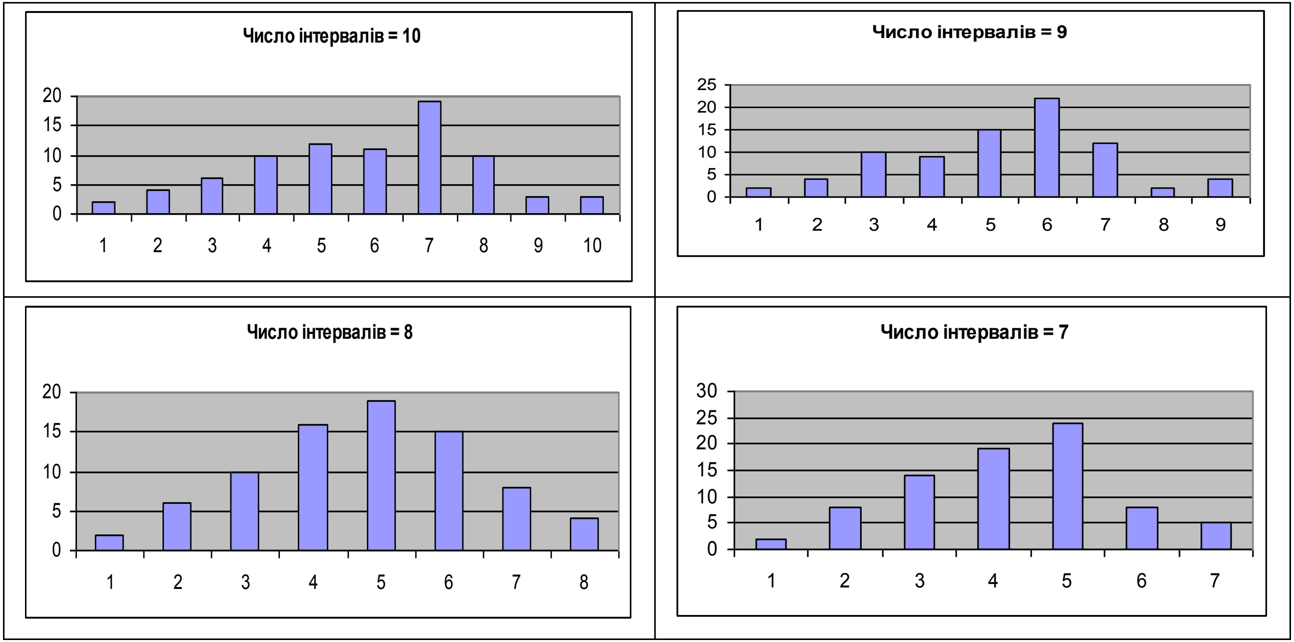

Число інтервалів = 10
20
15
10
5
0
1 2 3 4 5 6 7 8 9 10
Число інтервалів = 9
25
20
15
10
5
0
1 2 3 4 5 6 7 8 9
Число інтервалів = 8
20
15
10
5
0
1 2 3 4 5 6 7 8
Число інтервалів = 7
30
25
20
15
10
5
0
1 2 3 4 5 6 7

Рисунок Г.1

Інтервали зміни змінної, що відповідають кожному стовпцю, розраховані над гістограмою (див. файл «Шаблон_лаб 2.xls»). Висота стовпців дорівнює кількості точок, що потрапили у відповідний частковий інтервал. Визначити, яка з одержаних гістограм задовольняє вимоги, поставлені до гістограм, пояснити чому. Зменшити число часткових інтервалів до 6 або 5 і побудувати нову гістограму. Дайте відповідь на запитання: чи можна ще зменшити число часткових інтервалів?

3 За значеннями асиметрії та ексцесу й виглядом гістограми зробити висновок, чи значно відрізняється розподіл випадкової величини від нормального.

4 Знайти ймовірність потрапляння випадкової величини у проміжок, заданий у варіанті завдання.

Імовірність потрапляння випадкової величини в проміжок (a, b) розраховуємо за формулою $P (a < X < b) = F (b) – F (a)$, де $F (x)$ – функція розподілу випадкової величини. Значення функції розподілу знаходимо за допомогою стандартної функції НОРМРАСП () (чи NORMDIST()). Одержимо (наприклад, для a = 1,2; b = 2,2) :

F(2,2) =	0,5890148
F(1,2) =	0,1368233
P(1,2 < X < 2,2) = F(2,2) - F(1,2) =	0,452191

5 Вважаючи, що технологічний процес відрегульований правильно, а допуск становить 10 % від значення контрольованого параметра, знайти випуск придатної продукції у відсотках.

Оскільки технологічний процес відрегульований правильно, то вибіркове середнє можна прийняти за значення параметра, заданого в технічній документації. Десятивідсоткове відхилення знаходимо за формулою

$\delta = 0{,}1 \cdot \bar{x} = 0{,}1 \cdot 2{,}0295 = 0{,}20295$. Потім обчислюємо ймовірність.

$$P(\left|x - \bar{x}\right| < \delta) = P(2{,}0295 - 0{,}02095 < x < 2{,}0295 + 0{,}02095) =$$
$$= P(1{,}82655 < x < 2{,}23245) = 0{,}605587 - 0{,}394413 = 0{,}211174$$

за допомогою стандартної функції НОРМРАСП (), як у п. 5.

Переводимо результат у відсотки: $0{,}211174 \cdot 100\,\% = 21{,}1174\,\%$.

Одержимо відповідь: вихід придатної продукції при заданому допуску становить 21,1 % від усієї продукції.

Додаток Д
Приклад виконання самостійної роботи 3 в Excel

Хід виконання роботи відповідає п. 3.1. – 3.5.

Створити файл початкових даних під своїм ім'ям можна на основі файла «Шаблон_лаб_3.xls».

Нижче наведені робочі листи Excel – у вигляді розрахунків та у вигляді вікна формул.

У таблиці кольором виділені діапазони, за якими побудовані довірчі області (колір виділення відповідає кольору заголовка діаграми).

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		x	y	y(x)	δ при 80%	Y-δ	Y+δ	δ при 95%	Y-δ	Y+δ	δ при 99%	Y-δ	Y+δ
2	1	1,033	1,830	1,857257	0,233749	1,623508	2,091006	0,38589716	1,47136	2,243154	0,56150569	1,295751	2,418763
3	2	0,012	0,580	1,272289	0,242073	1,030216	1,514362	0,399638912	0,87265	1,671928	0,58150084	0,690788	1,85379
4	3	0,045	1,340	1,291196	0,240938	1,050258	1,532134	0,397765193	0,893431	1,688961	0,57877446	0,712422	1,86997
5	4	0,243	1,340	1,404637	0,235269	1,169368	1,639906	0,388405754	1,016232	1,793043	0,56515586	0,839482	1,969793
6	5	0,266	1,640	1,417815	0,234741	1,183074	1,652556	0,387534306	1,030281	1,805349	0,56388784	0,853927	1,981703
7	6	0,302	1,650	1,438441	0,233971	1,20447	1,672412	0,386262939	1,052178	1,824704	0,56203792	0,876403	2,000479
8	7	0,451	1,910	1,523808	0,231525	1,292283	1,755333	0,382224438	1,141584	1,906033	0,55616164	0,967647	2,07997
9	8	1,041	1,960	1,861841	0,23391	1,627931	2,09575	0,386162124	1,475678	2,248003	0,56189123	1,299949	2,423732
10	9	1,423	2,080	2,080702	0,245393	1,835309	2,326095	0,405120005	1,675582	2,485822	0,58947619	1,491226	2,670178
11	10	1,914	2,180	2,362014	0,269845	2,092168	2,631859	0,445488186	1,916526	2,807502	0,64821454	1,713799	3,010228

	A	B	C	D
12				
13				
14	Знаходимо координати центра розсіювання			
15	Xmean=	0,673		
16	Ymean=	1,651		
17				
18	Знаходимо область прогнозів			
19	Xmax=	1,914	Ymax=	2,180
20	Xmin=	0,012	Ymin=	0,580
21				
22	Розраховуємо значения дисперсій			
23	Dx=	0,415153778		
24	Dy=	0,222121111		
25				
26	Обчислюємо значения середньоквадратичних відхилень			
27	σx=	0,64432428		
28	σy=	0,471297264		
29				
30	Знаходимо коеффіцієнт кореляції			
31	CORR(x,y)=	0,783277971		
32				

Лист з розрахунками – рядки1-31

	A	B	C	D	E	F	G	H	I	J	K	L
1		x	y	y(x)	δ при 80%	Y-δ	Y+δ	δ при 95%	Y-δ	Y+δ	δ при 99%	Y-δ
2	1	1,033	1,83	=$B$34+$B$35*B2	=$B$38*$B$54*КОРЕНЬ((	=D2-E2	=D2+E2	=$B$54*$B$39*КОРЕНЬ((1+(1	=D2-H2	=D2+H2	=$B$54*$B$40*КОРЕНЬ((	=D2-K2
3	2	0,012	0,58	=$B$34+$B$35*B3	=$B$38*$B$54*КОРЕНЬ((	=D3-E3	=D3+E3	=$B$54*$B$39*КОРЕНЬ((1+(1	=D3-H3	=D3+H3	=$B$54*$B$40*КОРЕНЬ((	=D3-K3
4	3	0,045	1,34	=$B$34+$B$35*B4	=$B$38*$B$54*КОРЕНЬ((	=D4-E4	=D4+E4	=$B$54*$B$39*КОРЕНЬ((1+(1	=D4-H4	=D4+H4	=$B$54*$B$40*КОРЕНЬ((	=D4-K4
5	4	0,243	1,34	=$B$34+$B$35*B5	=$B$38*$B$54*КОРЕНЬ((	=D5-E5	=D5+E5	=$B$54*$B$39*КОРЕНЬ((1+(1	=D5-H5	=D5+H5	=$B$54*$B$40*КОРЕНЬ((	=D5-K5
6	5	0,266	1,64	=$B$34+$B$35*B6	=$B$38*$B$54*КОРЕНЬ((	=D6-E6	=D6+E6	=$B$54*$B$39*КОРЕНЬ((1+(1	=D6-H6	=D6+H6	=$B$54*$B$40*КОРЕНЬ((	=D6-K6
7	6	0,302	1,65	=$B$34+$B$35*B7	=$B$38*$B$54*КОРЕНЬ((	=D7-E7	=D7+E7	=$B$54*$B$39*КОРЕНЬ((1+(1	=D7-H7	=D7+H7	=$B$54*$B$40*КОРЕНЬ((	=D7-K7
8	7	0,451	1,91	=$B$34+$B$35*B8	=$B$38*$B$54*КОРЕНЬ((	=D8-E8	=D8+E8	=$B$54*$B$39*КОРЕНЬ((1+(1	=D8-H8	=D8+H8	=$B$54*$B$40*КОРЕНЬ((	=D8-K8
9	8	1,041	1,96	=$B$34+$B$35*B9	=$B$38*$B$54*КОРЕНЬ((	=D9-E9	=D9+E9	=$B$54*$B$39*КОРЕНЬ((1+(1	=D9-H9	=D9+H9	=$B$54*$B$40*КОРЕНЬ((	=D9-K9
10	9	1,423	2,08	=$B$34+$B$35*B10	=$B$38*$B$54*КОРЕНЬ((	=D10-E10	=D10+E10	=$B$54*$B$39*КОРЕНЬ((1+(1	=D10-H10	=D10+H10	=$B$54*$B$40*КОРЕНЬ((	=D10-K
11	10	1,914	2,18	=$B$34+$B$35*B11	=$B$38*$B$54*КОРЕНЬ((	=D11-E11	=D11+E11	=$B$54*$B$39*КОРЕНЬ((1+(1	=D11-H11	=D11+H11	=$B$54*$B$40*КОРЕНЬ((	=D11-K
12												
13												
14	Знаходимо координати центра розсіювання											
15	Xmean=	=СРЗНАЧ(B2:B11)										
16	Ymean=	=СРЗНАЧ(C2:C11)										
17												
18	Знаходимо область прогнозів											
19	Xmax=	=МАКС(B2:B11)	Ymax=	=МАКС(C2:C11)								
20	Xmin=	=МИН(B2:B11)	Ymin=	=МИН(C2:C11)								
21												
22	Розраховуємо значення дисперсій											
23	Dx=	=ДИСП(B2:B11)										
24	Dy=	=ДИСП(C2:C11)										
25												
26	Обчислюємо значення середньоквадратичних відхилень											
27	σx=	=СТАНДОТКЛОН(B2:B11)										
28	σy=	=СТАНДОТКЛОН(C2:C11)										
29												
30	Знаходимо коеффіцієнт кореляції											
31	CORR(x,y)=	=КОРРЕЛ(B2:B11;C2:C11)										

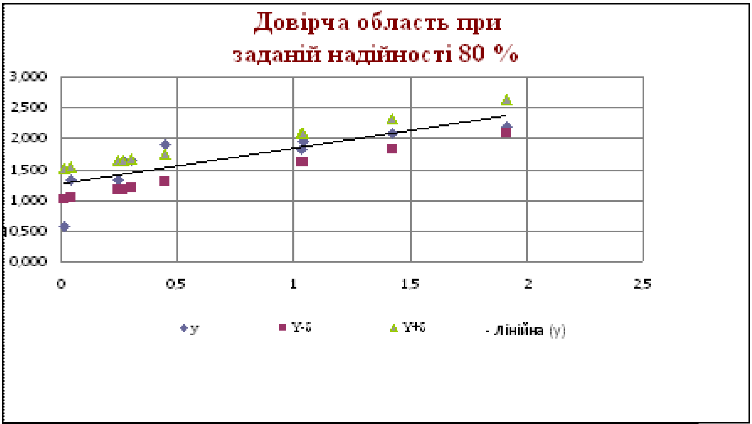
Лист із формулами – рядки 1-31

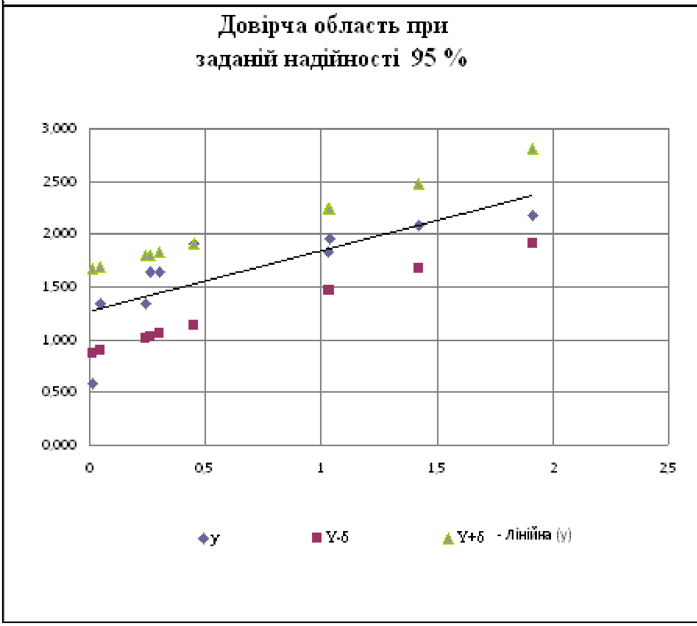
	A	B	C	D	E	F	G	H
33	За методом найменших квадратів знаходимо коефіцієнти моделі y=b0+b1x :							
34	b0=	1,265413874						
35	b1=	0,572936293						
36								
37	Знахождження критичних точок розподілу Стьюдента							
38	для γ=80%	1,39681531						
39	для γ=95%	2,306004135						
40	для γ=99%	3,355387331						
41								
42	Розрахуємо напівширину довірчого інтервалу по формуле:							
43								
44	$\delta = \sigma_e \cdot t_y \cdot \sqrt{1+\frac{1}{n}+\frac{(x_{np}-x_i)^2}{\sum_{i=1}^{n}(x_{np}-\bar{x})^2}}$							
45								
46								
47								
48	Обчислюємо знаменник		$\sum_{i=1}^{n}(x_i-\bar{x}) = S_x = (n-1)D(x)$					
49								
50	Sx=	3,736384						
51								
52								
53	Розрахуємо середньоквадратичне відхилення вибраних точок від лінії регрессії							
54	σe=	0,157099088						
55								

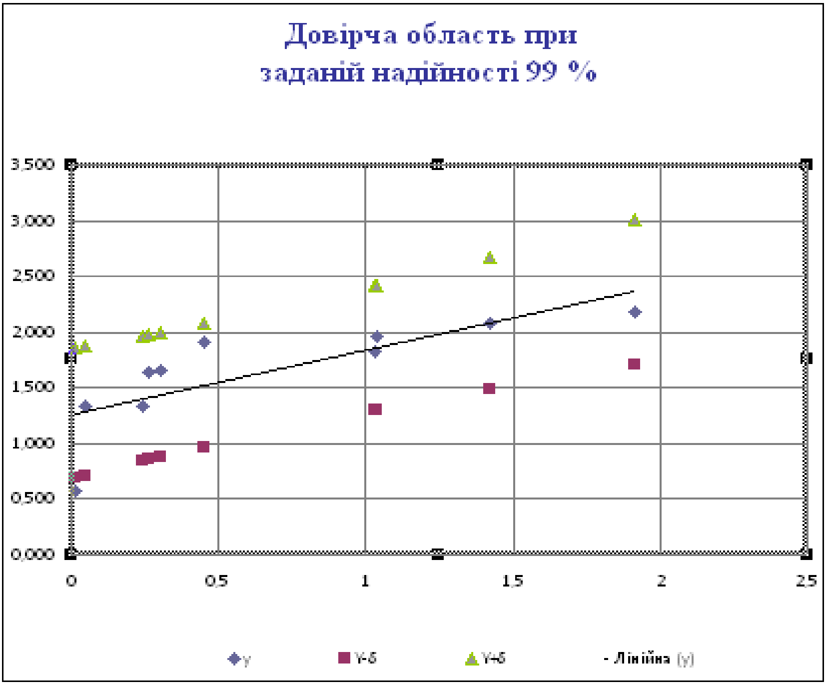
Лист із розрахунками – рядки 32-54

33	За методом найменших квадратів знаходимо коефіцієнти моделі y=b0+b1x :	
34	b0=	=ОТРЕЗОК(C2:C11;B2:B11)
35	b1=	=НАКЛОН(C2:C11;B2:B11)
36		
37	Знаходження критичних точок розподілу Стьюдента	
38	для γ=80%	=СТЬЮДРАСПОБР(0,2;8)
39	для γ=95%	=СТЬЮДРАСПОБР(0,05;8)
40	для γ=99%	=СТЬЮДРАСПОБР(0,01;8)
41		
42	Розраховуємо напівширину довірчого інтервалу за формулою:	
43		
44	$\delta = \sigma_e \cdot t_y \cdot \sqrt{1+\frac{1}{n}+\frac{(x_{np}-x_i)^2}{\sum_{i=1}^{n}(x_{np}-\bar{x})^2}}$	
45		
46		
47		
48	Обчислимо знаменник	$\sum_{i=1}^{n}(x_i-\bar{x}) = S_x = (n-1)D(x)$
49		
50	Sx=	=(10-1)*B23
51		
52		
53	Розрахуємо середньоквадратичне відхилення вибіркових точок від лінії регресії	
54	σe=	=КОРЕНЬ((1/(10-1)*B24))
55		

Лист із формулами – рядки 32-54


Довірча область при
заданій надійності 80 %
3,000
2,500
2,000
1,500
1,000
0,500
0,000
0
0,5
1
1,5
2
2,5
y
Y-δ
Y+δ
Лінійна (y)


Довірча область при
заданій надійності 95 %
3,000
2,500
2,000
1,500
1,000
0,500
0,000
0
0,5
1
1,5
2
2,5
y
Y-δ
Y+δ
Лінійна (y)


Довірча область при
заданій надійності 99 %
3,500
3,000
2,500
2,000
1,500
1,000
0,500
0,000
0
0,5
1
1,5
2
2,5
y
Y-δ
Y+δ
Лінійна (y)

Додаток Е
Приклад виконання самостійної роботи 4 в Excel

Масив даних:

0,75	2,41	0,17	2,47	0,85	1,53	1,58	1,32
0,75	1,94	3,81	2,00	1,01	3,54	1,92	1,04
2,13	1,58	2,19	1,64	2,24	2,36	2,53	1,91
1,31	3,62	1,37	3,68	1,55	2,04	2,89	2,51
2,31	0,34	2,37	2,40	1,06	2,09	0,09	1,78
1,81	2,27	1,87	3,32	2,00	1,29	1,79	2,96
2,35	2,87	2,41	2,93	1,38	2,37	2,43	1,88
2,93	2,40	2,99	2,46	1,77	2,09	2,89	0,39
2,01	2,56	2,07	2,62	0,89	2,98	1,88	2,08
1,19	1,79	1,25	1,83	2,87	2,52	1,56	1,57

Для вибірки знаходимо основні її характеристики:

1. Вибіркове (емпіричне) середнє.
   Формула Excel = СРЗНАЧ (< діапазон вибірки >)
2. Дисперсія.
   Формула Excel = ДИСП.В (< діапазон вибірки >)
3. Середньоквадратичне відхилення.
   Формула Excel = КОРЕНЬ (дисперсії)
4. Обсяг вибірки -
   Формула Excel = СЧЁТ (< діапазон вибірки >)

Одержуємо значення:

x_сер.	2,01
Дисперсія	0,63270486
Сер. кв. відхил.	0,79542747
n	80

Використання критерію Пірсона для визначення вигляду розподілу вимагає розділення усього розмаху вибірки на n часткових інтервалів. Розмахом вибірки є відстанню між максимальним і мінімальним членами варіаційного ряду (масиву даних).

Для знаходження мінімального і максимально елементів вибірки використовуємо формули Excel :

= МИН (< діапазон вибірки >)

= МАКС (< діапазон вибірки >).

Кількість часткових інтервалів беруть 7 – 10. Для визначення кроку для розбиття розмаху вибірки використовуємо формулу:

=(МАКС (<…>) -МИН (<…>))/k,

де k – обрана кількість інтервалів.

x_min	0,09
x_max	3,81
h	0,53142857

Для одержання початкової гіпотези H_0 побудуємо гістограму, відклавши по горизонталі кармани вибірки, по вертикалі – частоти.

Побудову гістограми можна виконати за допомогою налаштування Excel «Анализ Данных» або ручної побудови.

Розрахунки для ручної побудови гістограми:

Карман	*Частота*
Мінімальне елемент вибірки	–
Попередній карман +h	ЧАСТОТА(…)
Попередній карман +h	ЧАСТОТА(…)
Попередній карман +h	ЧАСТОТА(…)
Попередній карман +h	ЧАСТОТА(…)
Попередній карман +h	ЧАСТОТА(…)
Попередній карман +h	ЧАСТОТА(…)
Максимальний елемент вибірки	ЧАСТОТА(…)

Формула Excel = ЧАСТОТА(…) має 2 обов'язкові аргументи:

1) масив_даних – масив або посилання на множину значень, для яких обчислюються частоти. Якщо аргумент «масив _ даних» не містить значень, функція ЧАСТОТА повертає масив нулів;

2) масив_інтервалів – масив або посилання на множину інтервалів, в які групуються значення аргументу «масив _ даних». Якщо аргумент «масив _ інтервалів» не містить значень, функція ЧАСТОТА повертає кількість елементів у аргументі «масив _ даних».

Формула в комірці є формулою масиву. Щоб ця функція повертала значення в комірках, які знаходяться нижче, виділіть діапазон разом з осередком, в якому одержаний результат, натисніть клавішу F2, а потім – клавіші CTRL SHIFT Enter. Інакше буде повернено лише значення виділеної комірки.

Одержуємо

Карман	частота
0,09	
0,62	4
1,15	7
1,68	14
2,22	22
2,75	19
3,28	9
3,81	5

Розраховані значення використовуємо при побудові гістограми за допомогою «Диаграмма – Гистограмма».

Побудову гістограми можна виконати за допомогою налаштування Excel «Анализ Данных».

Для одержання гістограми переходимо в пункт верхнього меню «Данные – Анализ Данных – Гистограмма». У діалоговому вікні (рис. Д.1) потрібно заповнити поля:

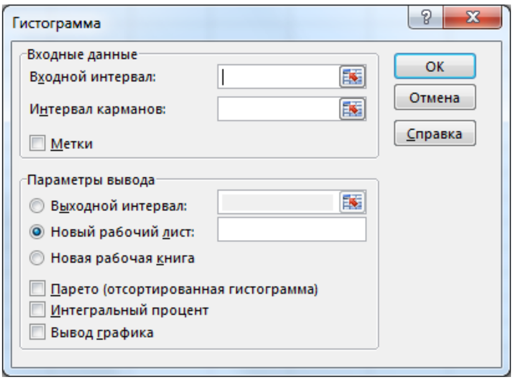


Рисунок Д.1 –Діалогове вікно «Гистограмма»

– вхідний інтервал є інтервалом, у якому розміщена вибірка;

– інтервал карманів – вибираємо осередки, в яких ми заздалегідь порахували кармани;

– вихідний інтервал – місце на листі де буде розміщена таблиця з карманами і частотами й сама гістограма;

– вибираємо «Вывод графика», оскільки без цього ми одержимо лише кармани й частоти.

Обидва варіанти дадуть потрібну гістограму (рис. Д.2), за якою гіпотетично визначити вигляд розподілу.

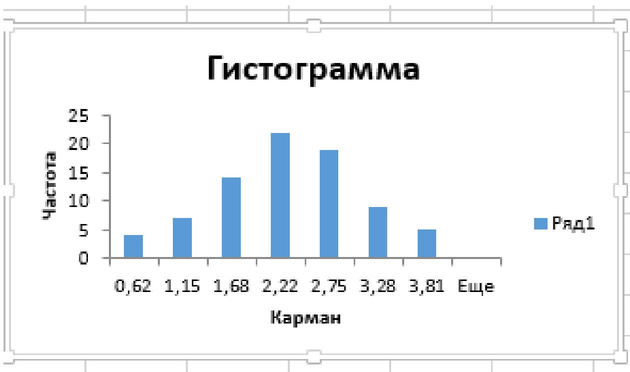


Рисунок Д.2 – Побудована гістограма

Використання критерію Пірсона.

Для розрахунку $\chi^2_{набл}$ побудуємо таку таблицю:

Кармани	Емпіричні частоти	Теоретичні частоти	X^2_набл
…	…	…	…
…	…	…	…
…	…	…	….

Кармани, які ми знайшли раніше, переносимо в перший стовпець. Емпіричні частоти – кількість потраплянь у і-й інтервал. Ці частоти ми отримали за допомогою функції = ЧАСТОТА (...) або після використання пакета «Анализ Данных – Гистограмма».

Теоретичні частоти – це ймовірність попадання в і-й інтервал, помножена на обсяг вибірки.

За припущення, що вибірка має нормальний розподіл, теоретичні частоти обчислюватимемо таким чином:

= (НОРМРАСП (X; середнє; стандарт_відхилення; 1) -НОРМРАСП (X - 1; середнє; стандарт_відхилення; 1)) *СЧЁТ(< діапазон вибірки >),

де X- 1 і X – кінці інтервалу [X- 1; X].

За припущення, що вибірка має експоненціальний розподіл, теоретичні частоти обчислюємо таким чином:

= (EXP (-X/середнє) – EXP (- (X - 1) / середнє)) *СЧЁТ (< діапазон вибірки >),

де X-1 і X – кінці інтервалу [X- 1; X].

За припущення, що вибірка має рівномірний розподіл, теоретичні частоти обчислюємо таким чином:

= СЧЁТ (< діапазон вибірки >) * (X - (X - 1)) / (МАКС (< діапазон вибірки >) -МИН (< діапазон вибірки >)),

де X- 1 і X – кінці інтервалу [X- 1; X].

Після знаходження теоретичних та емпіричних частот знаходимо X^2 набл = ((Емп.част – Теор.част.) ^ 2 / Теор. част.) на і-му інтервалі.

Після того як заповнена вся таблиця, будуємо гістограму теоретичних і емпіричних частот, щоб оцінити вибірку (рис. Д.3 – Д.5).

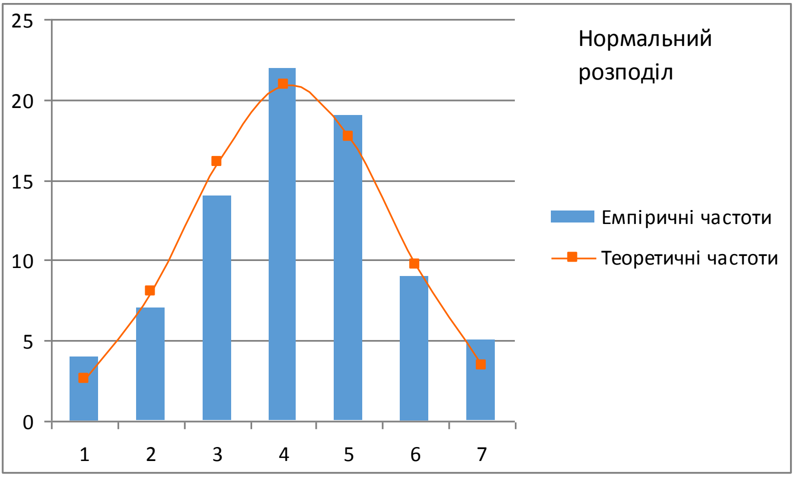


Рисунок Д.3 – Гістограма теоретичних та емпіричних частот (нормальний розподіл)

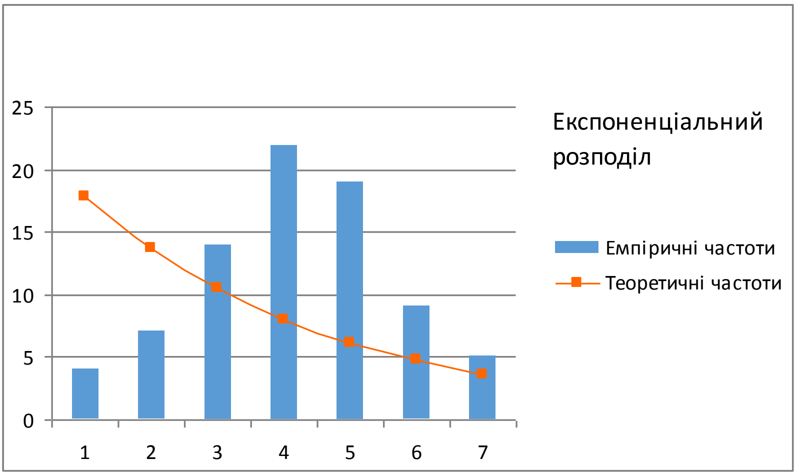


Рисунок Д.4 – Гістограма теоретичних та емпіричних частот (експоненціальний розподіл)

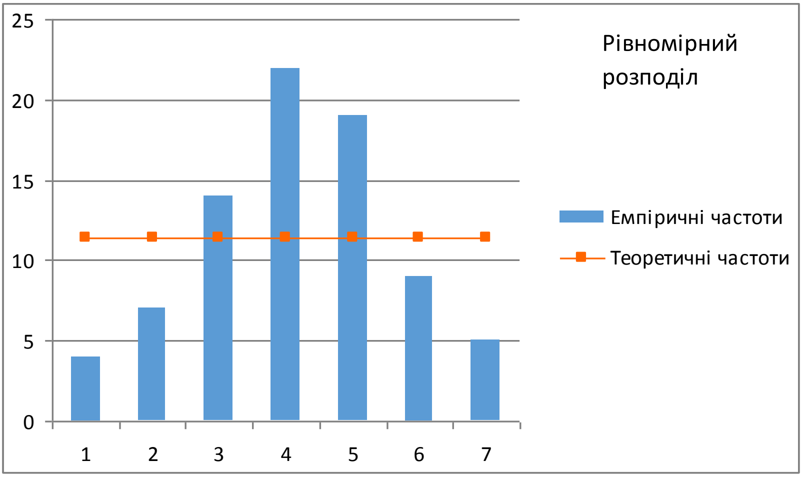


Рисунок Д.3 – Гістограма теоретичних та емпіричних частот (рівномірний розподіл)

Виконуємо оцінювання. Знайдемо X^2 як суму X^2_набл на усіх і-х інтервалах. Критичне значення X^2_кр є табличним. Його значення одержуємо за допомогою функції Excel =ХИ2.ОБР (рівень значущості; кількість ступенів свободи). Рівнем значущості є ймовірність відхилити правильну гіпотезу. Як правило, рівень значущості набуває значень 0,05 і 0,1.

Кількість ступенів свободи визначають таким чином:

Нормальний розподіл	N-1-2
Експоненціальний розподіл	N-1-1
Рівномірний розподіл	N-1-0

де N – обсяг вибірки, параметр 0; 1 або 2 показує кількість характеристик, від яких залежить розподіл.

Якщо X^2_набл < X^2_кр – гіпотезу приймаємо, в іншому разі гіпотезу відхиляємо. Якщо гіпотеза виявилася неправильною, перевіряємо вибірку для іншого вигляду розподілу.

Якщо жодна гіпотеза не підтвердилася при використанні критерію Пірсона, то можна збільшити рівень значущості або ж виконати розрахунок за критерієм Колмогорова – Смирнова.

Для перевірки критерію Колмогорова – Смирнова заповнимо таку таблицю:

Кармани	К-ть потраплянь	Емпіричні частоти	Теоретичні частоти	Модуль різниці частот
…	…	…	…	…

Кармани, знайдені раніше, переносимо в перший стовпчик. Емпіричні частоти –ймовірність потрапляння в і-й інтервал. Ці частоти одержані за допомогою функції = ЧАСТОТА (...) /СЧЁТ (<...>).

Теоретичні частоти – це ймовірність потрапляння в і-й інтервал:

За припущення, що вибірка має нормальний розподіл, теоретичні частоти обчислюють таким чином:
= (НОРМРАСП (X; середнє; стандарт_відхилення; 1) - НОРМРАСП (X - 1; середнє; стандарт_відхилення; 1)), де X- 1 і X – кінці інтервалу [X- 1; X].

За припущення, що вибірка має експоненціальний розподіл, теоретичні частоти обчислюють так:
= (EXP(-X/середнє) - EXP(-(X-1)/ середнє)), де X-1 і X – кінці інтервалу [X - 1; X].

За припущення, що вибірка має рівномірний розподіл, теоретичні частоти обчислюємо таким чином: = (X -

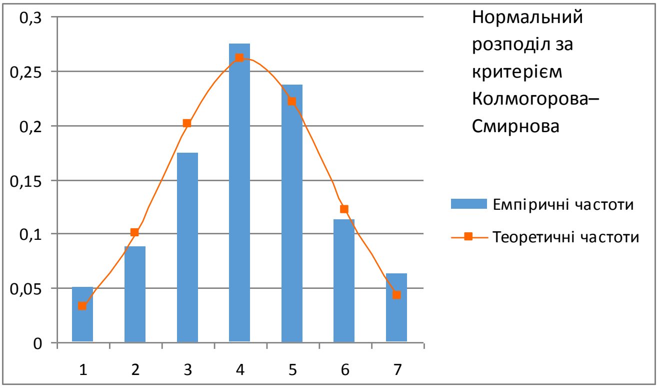
(X - 1)) / (МАКС (< діапазон вибірки >) -МІН (< діапазон вибірки >)), де X- 1 і X - кінці інтервалу [X- 1; X].

Після знаходження теоретичних та емпіричних частот знаходимо: модуль різниці =ABS (Емпір.ч. – Теор.ч.) на і-му інтервалі; n1 = суму емпіричних частот і n2 = суму теоретичних частот. Для розрахунку L^2набл знайдемо таку величину: Z = n1*n2/ (n1+n2), тоді L^2набл = Z*Корінь (dmax), де dmax – максимальне значення з модуля різниці теоретичних і емпіричних частот. L^2_кр – критичне значення критерію Колмогорова – Смирнова для відносного рівня значущості, що знаходиться в таблицях.

Одержимо:

Карман	Частота	Емпірична частота	Теоретична частота	Модуль різниць	Z
0,09					
0,62	4	0,05	0,032741127	0,01725887	
1,15	7	0,0875	0,100599059	0,01309906	
1,68	14	0,175	0,200927808	0,02592781	
2,22	22	0,275	0,261001183	0,01399882	
2,75	19	0,2375	0,220538044	0,01696196	
3,28	9	0,1125	0,121202057	0,00870206	
3,81	5	0,0625	0,043304993	0,01919501	
Сума	80	1	0,980314271		0,49502965
dmax	0,025927808				
L^2_набл	0,079710239				
λкр(0,05)	1,224				

Гістограма частот матиме вигляд



Нижче наведені робочі листи Excel у вигляді вікна формул.

	A	B	C	D	E
28	1) H0: нормальний розподіл				
29					
30	Частоти				
31		Емпір. ч.	Теор. ч.		
32	=B12				
33	=A32+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	4;1)-НОРМРАСП(A32; $F$12;$F$14;1))*$B$41	=(B33-C33)^2/C33	
34	=A33+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=(НОРМРАСП(A34;$F$12;$F$1 4;1)-НОРМРАСП(A33; $F$12;$F$14;1))*$B$41	=(B34-C34)^2/C34	
35	=A34+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=(НОРМРАСП(A35;$F$12;$F$1 4;1)-НОРМРАСП(A34; $F$12;$F$14;1))*$B$41	=(B35-C35)^2/C35	
36	=A35+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=(НОРМРАСП(A36;$F$12;$F$1 4;1)-НОРМРАСП(A35; $F$12;$F$14;1))*$B$41	=(B36-C36)^2/C36	
37	=A36+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=(НОРМРАСП(A37;$F$12;$F$1 4;1)-НОРМРАСП(A36; $F$12;$F$14;1))*$B$41	=(B37-C37)^2/C37	
38	=A37+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=(НОРМРАСП(A38;$F$12;$F$1 4;1)-НОРМРАСП(A37; $F$12;$F$14;1))*$B$41	=(B38-C38)^2/C38	
39	=A38+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=(НОРМРАСП(A39;$F$12;$F$1 4;1)-НОРМРАСП(A38; $F$12;$F$14;1))*$B$41	=(B39-C39)^2/C39	
40	=A39+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=(НОРМРАСП(A40;$F$12;$F$1 4;1)-НОРМРАСП(A39; $F$12;$F$14;1))*$B$41	=(B40-C40)^2/C40	
41		=СУММ(B33:B40)			
42					
43	X^2_набл	=СУММ(D33:D40)			
44					
45	X^2_кр	=ХИ2.ОБР(0,1;8-1-2)			
46					
47	H0 відхилено				

	A	B	C	D	E
49	2) H0: експоненциальний розподіл				
50					
51	Частоти				
52		Емпір. Ч.	Теор. Ч.		
53	=B12				
54	=A53+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=(EXP(-A53/$F$12)-EXP(-A54/$F$12))*$B$62	=(B54-C54)^2/C54	
55	=A54+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=(EXP(-A54/$F$12)-EXP(-A55/$F$12))*$B$62	=(B55-C55)^2/C55	
56	=A55+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=(EXP(-A55/$F$12)-EXP(-A56/$F$12))*$B$62	=(B56-C56)^2/C56	
57	=A56+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=(EXP(-A56/$F$12)-EXP(-A57/$F$12))*$B$62	=(B57-C57)^2/C57	
58	=A57+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=(EXP(-A57/$F$12)-EXP(-A58/$F$12))*$B$62	=(B58-C58)^2/C58	
59	=A58+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=(EXP(-A58/$F$12)-EXP(-A59/$F$12))*$B$62	=(B59-C59)^2/C59	
60	=A59+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=(EXP(-A59/$F$12)-EXP(-A60/$F$12))*$B$62	=(B60-C60)^2/C60	
61	=A60+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=(EXP(-A60/$F$12)-EXP(-A61/$F$12))*$B$62	=(B61-C61)^2/C61	
62		=СУММ(B54:B61)			
63					
64	X^2_набл	=СУММ(D54:D61)			
65					
66	X^2_кр	=ХИ2.ОБР(0,1;8-1-1)			
67					
68	H0 відхилено				

	A	B	C	D	E
70	3)H0: рівномірний розподіл				
71					
72	Частоти				
73		Емпір. Ч.	Теор. Ч.		
74	=B12				
75	=A74+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=$B$83*(A75-A74)/($B$13-$B$1	=(B75-C75)^2/C75	
76	=A75+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=$B$83*(A76-A75)/($B$13-$B$1	=(B76-C76)^2/C76	
77	=A76+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=$B$83*(A77-A76)/($B$13-$B$1	=(B77-C77)^2/C77	
78	=A77+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=$B$83*(A78-A77)/($B$13-$B$1	=(B78-C78)^2/C78	
79	=A78+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=$B$83*(A79-A78)/($B$13-$B$1	=(B79-C79)^2/C79	
80	=A79+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=$B$83*(A80-A79)/($B$13-$B$1	=(B80-C80)^2/C80	
81	=A80+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=$B$83*(A81-A80)/($B$13-$B$1	=(B81-C81)^2/C81	
82	=A81+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=$B$83*(A82-A81)/($B$13-$B$1	=(B82-C82)^2/C82	
83		=СУММ(B75:B82)			
84					
85	X^2_набл	=СУММ(D75:D82)			
86					
87	X^2_кр	=ХИ2.ОБР(0,1;8-1-0)			
88					
89	H0 відхиляемо				

	A	B	C	D	E	F	G	H
97								
98	Критерий Колмогорова-Смирнова							
99								
100	1) H0: нормальний розподіл							
101								
102	Частоти							
103			Емпір. Ч.	Теор. Ч.				
104	=B12							
105	=A104+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=B105/$B$113	=(НОРМРАСП(A105;$F$12;$F$14;1)-НОРМРАСП(A104; $F$12;$F$14;1))	=ABS(C105-D105)			
106	=A105+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=B106/$B$113	=(НОРМРАСП(A106;$F$12;$F$14;1)-НОРМРАСП(A105; $F$12;$F$14;1))	=ABS(C106-D106)			
107	=A106+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=B107/$B$113	=(НОРМРАСП(A107;$F$12;$F$14;1)-НОРМРАСП(A106; $F$12;$F$14;1))	=ABS(C107-D107)			
108	=A107+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=B108/$B$113	=(НОРМРАСП(A108;$F$12;$F$14;1)-НОРМРАСП(A107; $F$12;$F$14;1))	=ABS(C108-D108)			
109	=A108+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=B109/$B$113	=(НОРМРАСП(A109;$F$12;$F$14;1)-НОРМРАСП(A108; $F$12;$F$14;1))	=ABS(C109-D109)			
110	=A109+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=B110/$B$113	=(НОРМРАСП(A110;$F$12;$F$14;1)-НОРМРАСП(A109; $F$12;$F$14;1))	=ABS(C110-D110)			
111	=A110+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=B111/$B$113	=(НОРМРАСП(A111;$F$12;$F$14;1)-НОРМРАСП(A110; $F$12;$F$14;1))	=ABS(C111-D111)			
112	=A111+$B$15	=ЧАСТОТА(A1:H10;A33:A40)	=B112/$B$113	=(НОРМРАСП(A112;$F$12;$F$14;1)-НОРМРАСП(A111; $F$12;$F$14;1))	=ABS(C112-D112)			
113		=СУММ(B105:B112	=СУММ(C105:C112)	=СУММ(D105:D112)		=C113*D113/(C113+D113)		
114	dmax	=МАКС(E105:E112)						
115	L^2_набл	=B114*КОРЕНЬ(F113)						
116								
117	L^2_кр	0,12067						
118								
119	H0 приймаємо							

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. П. С. Сеньо. – Теорія ймовірностей та математична статистика: підручник / Київ : Знання, 2007.
2. В. В. Барковський, Н. В. Барковська, О. К. Лопатін. – Теорія ймовірностей та математична статистика / Київ: ЦУЛ, 2002.
3. Е. С. Вентцель. – Теория вероятностей: навч. посібник / Москва: Наука, 1973.
4. Г. Е. Гмурман. Теория вероятностей и математическая статистика / Москва: Высшая школа, 1977.
5. Гмурман Г. Е. Руководство к решению задач по теории вероятностей и математической статистике. Высшая школа, М., 1975
6. А. М. Зубков, Б. А. Севастьянов, В. П. Чистяков. – Сборник задач по теории вероятностей / Москва: Наука, 1989.

Навчальне видання

Васильєва Людмила Володимирівна,
Гончаров Олександр Андрійович

ПРАКТИКУМ
із курсу «Теорія ймовірностей та математична статистика»

Художнє оформлення обкладинки__________
Редактор Н. А Гавриленко
Комп'ютерне верстання ___________________

Формат 60×84/16. Ум. друк. арк. 6,05. Обл.-вид. арк. 4,55. Тираж 50 пр. Зам. №

Видавець і виготовлювач
Сумський державний університет,
вул. Римського-Корсакова, 2, м. Суми, 40007
Свідоцтво суб'єкта видавничої справи ДК № 3062 від 17.12.2007.