

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАВЧАЛЬНО-НАУКОВИЙ КОМПЛЕКС
“ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ”
НАЦІОНАЛЬНОГО ТЕХНІЧНОГО УНІВЕРСИТЕТУ УКРАЇНИ
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ
СІКОРСЬКОГО”
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

РОЗРАХУНКОВО-ГРАФІЧНА РОБОТА №2

з математичної статистики

Виконав студент 2 курсу
групи КА-06
Вергелюк Олександр
Андрійович
Перевірив:
Ільєнко Андрій
Борисович

ЗМІСТ

Вступ	3
1 Первинний аналіз вибірки	5
2 Описові статисти	8
3 Гіпотеза про розподіл генеральної сукупності	9
4 Оцінка параметрів розподілу	10
4.1 Метод моментів	10
4.2 Метод максимальної правдоподібності	10
5 Перевірка параметрів на незміщеність, консистентність та ефективність	12
6 Довірчі інтервали для параметрів розподілу	14
7 Перевірка висунотої гіпотези за критерієм χ^2	16
Висновки	18

ВСТУП

У файлі `svyato.txt` знайти свій набір із 100 чисел. Вони імітують вибірку, отриману із генеральної сукупності.

Дана вибірка

-3.47	0.06	-1.11	-3.77	1.13	2.23	-3.51	-3.2	-0.64	-1.61
0.06	-1.11	-3.77	1.13	2.23	-3.51	-3.2	-0.64	-1.61	-2.44
-1.11	-3.77	1.13	2.23	-3.51	-3.2	-0.64	-1.61	-2.44	-5.44
-3.77	1.13	2.23	-3.51	-3.2	-0.64	-1.61	-2.44	-5.44	-0.6
1.13	2.23	-3.51	-3.2	-0.64	-1.61	-2.44	-5.44	-0.6	1.94
2.23	-3.51	-3.2	-0.64	-1.61	-2.44	-5.44	-0.6	1.94	-2.46
-3.51	-3.2	-0.64	-1.61	-2.44	-5.44	-0.6	1.94	-2.46	-1.12
-3.2	-0.64	-1.61	-2.44	-5.44	-0.6	1.94	-2.46	-1.12	-3.85
-0.64	-1.61	-2.44	-5.44	-0.6	1.94	-2.46	-1.12	-3.85	-1.0
-1.61	-2.44	-5.44	-0.6	1.94	-2.46	-1.12	-3.85	-1.0	-1.18

Відсортована вибірка

-6.78	-6.57	-5.48	-5.44	-5.21	-5.08	-4.49	-4.37	-4.36	-4.11
-3.98	-3.97	-3.92	-3.85	-3.77	-3.6	-3.51	-3.49	-3.47	-3.44
-3.43	-3.34	-3.22	-3.2	-3.18	-3.09	-3.05	-2.99	-2.96	-2.84
-2.75	-2.73	-2.48	-2.46	-2.45	-2.44	-2.34	-2.1	-1.99	-1.96
-1.95	-1.94	-1.91	-1.91	-1.91	-1.78	-1.74	-1.64	-1.63	-1.61
-1.55	-1.5	-1.48	-1.39	-1.34	-1.29	-1.18	-1.15	-1.12	-1.11
-1.08	-1.02	-1.0	-0.91	-0.85	-0.76	-0.71	-0.64	-0.6	-0.49
-0.48	-0.46	-0.04	-0.03	0.06	0.11	0.12	0.16	0.21	0.31
0.32	0.4	0.44	0.87	0.99	1.05	1.13	1.16	1.18	1.24
1.28	1.32	1.86	1.94	2.1	2.23	2.26	2.48	3.28	3.52

Постановка задачі:

1. Проведіть первинний аналіз вибірки. Ци включає статистичний ряд (для неперервних розподілів — інтервальний), емпіричну функцію розподілу (для неперервних розподілів — інтервальну), її графік,

полігон частот (для дискретних розподілів), гістограму (для неперервних розподілів), box-and-whisker plot.

2. Знайдіть вибіркове середнє, вибірккову дисперсію, виправлену вибірккову дисперсію, вибірккову медіану, вибірккову моду, вибірккові коефіцієнти асиметрії та ексцесу.
3. Обґрунтуйте та висуньте (нову) гіпотезу про розподіл генеральної сукупності.
4. Методом моментів та методом максимальної вірогідності знайдіть оцінки параметрів розподілу. В деяких випадках це може бути не дуже просто (як, наприклад, для параметру N біноміальної генеральної сукупності). Це чудовий спосіб проявити креативність та/або вміння користуватися Google.
5. Для кожного параметру кращу у цих двох оцінок перевірте на (асимптотичну) незміщеність, консистентність та ефективність. Тут також має сенс зауваження до попереднього пункту. У випадку нездоланих труднощів — а це відноситься виключно до перевірки ефективності a та b в $U(a, b)$, a в $\text{Exp}(y, a)$ та N в $\text{Bin}(N, p)$ — відповідну перевірку можна пропустити.
6. Побудуйте довірчі інтервали надійністю 0.95 для параметрів розподілу. (The above notes still apply!)
7. Нарешті перевірте висунуту гіпотезу про розподіл генеральної сукупності за допомогою критерію χ^2 . Якщо гіпотеза суперечить вибіркковим даним, перейдіть до п.3.
8. Проявіть всі свої літературні здібності та напишіть висновки.

Інструменти, що я використовував під час роботи над цією РГР:

- Jupyter notebook

1 ПЕРВИННИЙ АНАЛІЗ ВИБІРКИ

З загального виду вибірки видно, що її переважну більшість становлять раціональні числа, серед яких лише одне значення повторюється. Отже, можна висунути припущення, що надана вибірка з неперервного розподілу генеральної сукупності. Подальший аналіз вибірки буде наподитися для випадку неперервного розподілу.

Побудуємо інтервальний ряд для вибірки. Для цього за формулою Стерджеса визначимо кількість інтервалів.

$$N_{intervals} = 1 + 3.322 * \lg n \approx 8$$

Тепер знайдемо розмах вибірки, щоб обрати довжини інтервалів.

$$R = x_{(n)} - x_{(1)} = 3.52 - (-6.78) = 10.3$$

Для зручності в подальшому оберемо однакові довжини для всіх інтервалів причому такі, щоб точки, які визначають межі інтервалів мали лише один знак після коми.

$$h = \frac{R}{N_{intervals}} = \frac{10.3}{8} \approx 1.3$$

Тепер можемо побудувати статистичний ряд, табл. 1.1.

Таблиця 1.1 — Інтервальний статистичний ряд

Δ_k	x_k^*	n	n^*	ν	ν^*
[-6.8 ; -5.5)	-6.15	2	2	0.02	0.02
[-5.5 ; -4.2)	-4.85	7	9	0.07	0.09
[-4.2 ; -2.9)	-3.55	20	29	0.2	0.29
[-2.9 ; -1.6)	-2.25	21	50	0.21	0.5
[-1.6 ; -0.3)	-0.95	22	72	0.22	0.72
[-0.3 ; 1.0)	0.35	13	85	0.13	0.85
[1.0 ; 2.3)	1.65	12	97	0.12	0.97
[2.3 ; 3.6]	2.95	3	100	0.03	1

Тепер за статистичним рядом можна побудувати емпіричну інтервальну функцію розподілу.

$$F_n^*(x) = \begin{cases} 0 & x \leq -6.8 \\ 0.02 & -6.8 < x \leq -5.5 \\ 0.09 & -5.5 < x \leq -4.2 \\ 0.29 & -4.2 < x \leq -2.9 \\ 0.5 & -2.9 < x \leq -1.6 \\ 0.72 & -1.6 < x \leq -0.3 \\ 0.85 & -0.3 < x \leq 1 \\ 0.97 & 1 < x \leq 2.3 \\ 1 & 2.3 < x \leq 3.5 \end{cases}$$

Графік цієї функції наведено на рис. 1.1.

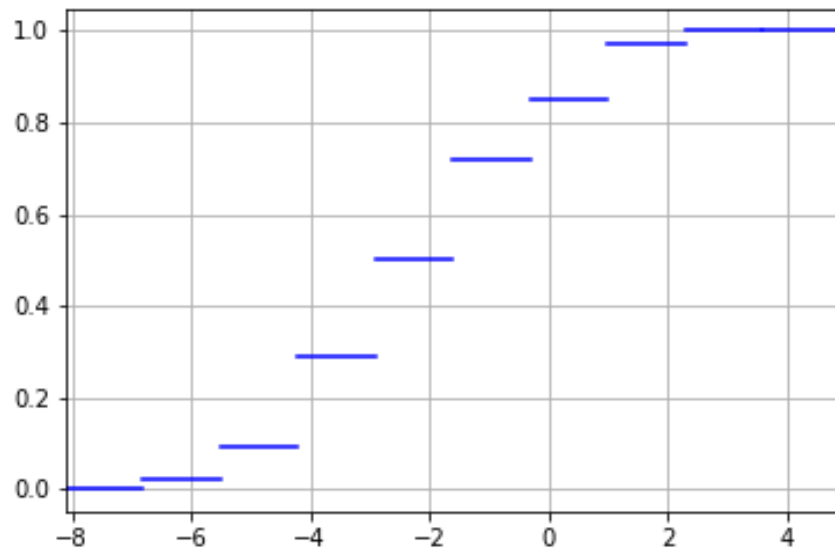


Рисунок 1.1 — Графік емпіричної інтервальної функції розподілу

Гістограма для даної вибірки наведена на рис. 1.2

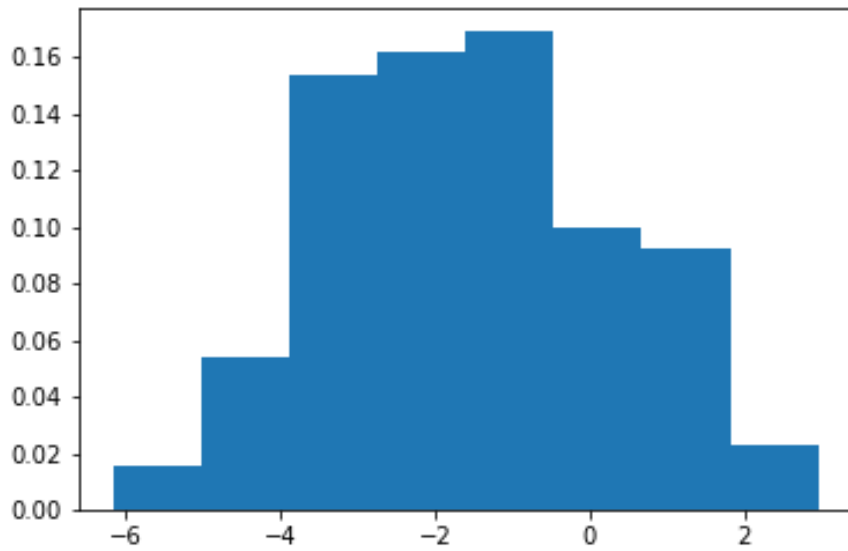


Рисунок 1.2 — Гістограма даної вибірки

Тепер побудуємо діаграму box-and-whisker (ящик з вусами) для даної вибірки (рис. 1.3).

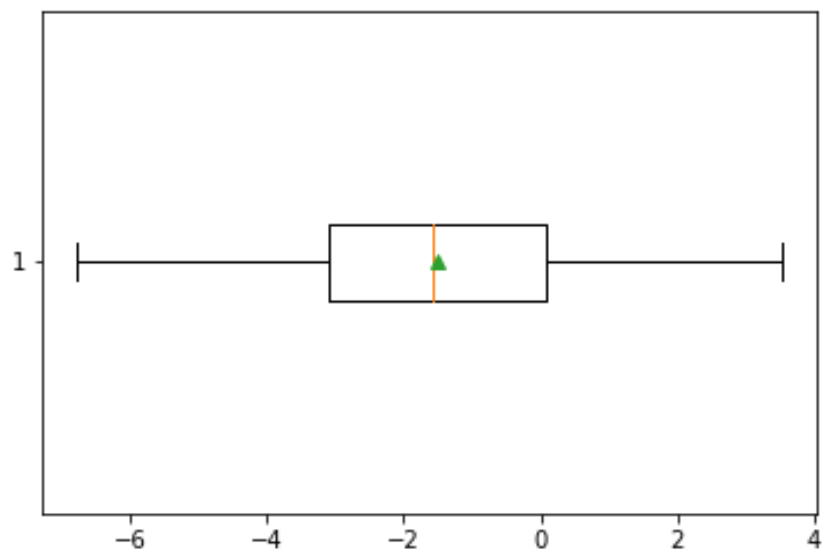


Рисунок 1.3 — Box-and-whisker

2 ОПИСОВІ СТАСТИКИ

Тепер знайдемо описові статистики вибірки: вибіркове середнє, вибірккову дисперсію, виправлену вибірккову десперсію, вибірккову медіану, мибірккову моду, вибірккові коефіцієнти асиметрії та ексцесу.

Знайдемо вибірккове середнє:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = -1.5207$$

Тепер обчислимо вибірккову дисперсію

$$\mathbb{D}\xi^{**} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = 4.5888$$

Знайдемо виправлену вибірккову дисперсію

$$\mathbb{D}\xi^{***} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 4.6352$$

Оскільки в нашій вибірці $n = 100$, то медіана вибірки:

$$M_e^* = \frac{x_{(50)} + x_{(51)}}{2} = \frac{-1.61 - 1.55}{2} = -1.58$$

Оскільки наша вибірка з неперервного розподілу генеральної сукупності, то вибіркова мода визначається за формулою:

$$\begin{aligned} M_o^* &= y_{mo-1} + h \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} = \\ &= -1.6 + 1.3 \frac{22 - 21}{(22 - 21) + (22 - 13)} = -1.47 \end{aligned}$$

Обчислимо вибірковий коефіцієнт асиметрії:

$$A_s = \frac{\overline{M_3}}{s_0^3} = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{ns_0^3} = 0.0318$$

Знайдемо вибірковий коефіцієнт ексцесу:

$$E_k = \frac{\overline{M_4}}{s_0^4} - 3 = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{ns_0^4} - 3 = -0.3145$$

3 ГІПОТЕЗА ПРО РОЗПОДІЛ ГЕНЕРАЛЬНОЇ СУКУПНОСТІ

Оскільки з самого початку було прийнято припущення, що генеральна сукупність має неперервний розподіл, то логічно визначати можливий розподіл із неперервних. За умовою, це можуть бути 3 варіанти:

- гауссівський;
- рівномірний;
- експоненціальний із зсувом.

Візьмемо до уваги кілька фактів, які ми отримали при первинному аналізі вибірки.

1. Гістограма за виглядом дещо нагадує графік щільності розподілу гауссівського розподілу ("дзвін").
2. Емпірична функція розподілу лише трохи нагадує функцію розподілу гауссівського розподілу, але зовсім трохи.
3. За графіком box-and-whisker видно, що половина усіх значень вибірки лежить у відносно вузькому проміжку, який трохи менше за $\frac{1}{3}$ розмаху вибірки. Це схоже із правилом для гауссівського розподілу, коли в межах одного стандартного відхилення від математичного сподівання лежить 68.26% усіх значень. Це ще один аргумент на користь гауссівського розподілу.
4. Вибіркове середнє (-1.52), вибірова мода (-1.47) та вибірова медіана (-1.58) ці значення доволі близькі один до одного, що схоже на гауссівський розподіл, у якого ці значення мають співпадати.

За сукупністю вище зазначених факторів приймемо гіпотезу H_0 — вибірка з генеральної сукупності із гауссівським розподілом.

4 ОЦІНКА ПАРАМЕТРІВ РОЗПОДІЛУ

Відповідно до висунутої гіпотези H_0 знайдемо оцінки гауссівського розподілу $N(a, \sigma^2)$ знайдемо оцінки a^* і $(\sigma^2)^*$ методом моментів та методом максимальної ймовірності.

4.1 Метод моментів

За визначенням для гауссівського розподілу:

$$a = \mathbb{E}\xi$$

$$\sigma^2 = \mathbb{D}\xi$$

Тоді оцінки розподілу будуть:

$$a_{MM}^* = \bar{\xi} \Rightarrow a_{MM}^* = -1.52$$

$$(\sigma^2)_{MM}^* = \mathbb{D}\xi^{**} \Rightarrow (\sigma^2)_{MM}^* = 4.6352$$

Отже, маємо гауссівський розподіл $N(-1.52, 4.6352)$.

4.2 Метод максимальної правдоподібності

За методом максимальної правдоподібності:

$$\mathcal{L}(\vec{x}, a, \sigma^2) = \prod_{k=1}^n \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_k - a)^2}{2\sigma^2}}$$

$$\ln \mathcal{L}(\vec{x}, a, \sigma^2) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{k=1}^n (x_k - a)^2$$

Для знаходження максимуму функції від двох змінних запишемо систему:

$$\begin{cases} \frac{\partial \ln \mathcal{L}}{\partial a} = \frac{1}{\sigma^2} \sum_{k=1}^n (x_k - a) = 0 \\ \frac{\partial \ln \mathcal{L}}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{k=1}^n (x_k - a)^2 = 0 \end{cases}$$
$$\begin{cases} \frac{1}{\sigma^2} n(\bar{x} - a) = 0 \\ \frac{1}{2\sigma^4} \sum_{k=1}^n (x_k - a)^2 = \frac{n}{2\sigma^2} \end{cases}$$

$$\begin{cases} \frac{n\bar{x}}{\sigma^2} = \frac{na}{\sigma^2} \\ \frac{n}{2\sigma^4} \mathbb{D}x^{**} = \frac{n}{2\sigma^2} \end{cases} \Rightarrow \begin{cases} a^* = \bar{x} \\ (\sigma^2)^* = \mathbb{D}x^{**} \end{cases}$$

Тепер знайдемо матрицю других похідних:

$$\begin{pmatrix} \frac{\partial^2 \ln \mathcal{L}}{\partial a^2} & \frac{\partial^2 \ln \mathcal{L}}{\partial a \partial \sigma^2} \\ \frac{\partial^2 \ln \mathcal{L}}{\partial \sigma^2 \partial a} & \frac{\partial^2 \ln \mathcal{L}}{\partial (\sigma^2)^2} \end{pmatrix} = \begin{pmatrix} -\frac{n}{\sigma^2} & -\frac{n}{\sigma^4}(\bar{x} - a) \\ -\frac{n}{\sigma^4}(\bar{x} - a) & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{k=1}^n (x_k - a)^2 \end{pmatrix}$$

При $a = a^*$ і $\sigma^2 = (\sigma^2)^*$:

$$\begin{pmatrix} -\frac{n}{(\sigma^2)^*} & 0 \\ 0 & \frac{n}{((\sigma^2)^*)^3} \left(\frac{(\sigma^2)^*}{2} - \frac{1}{n} \sum_{k=1}^n (x_k - a^*)^2 \right) \end{pmatrix} = \begin{pmatrix} -\frac{n}{(\sigma^2)^*} & 0 \\ 0 & -\frac{n}{2((\sigma^2)^*)^2} \end{pmatrix}$$

Ця матриця є від'ємно визначеною, тому $\ln \mathcal{L}(\vec{x}, a, \sigma^2)$ дійсно досягає максимуму при $a = \bar{x}$ і $\sigma^2 = \mathbb{D}\xi^{**}$. Отже, отримали оцінки $a_{MMP}^* = \bar{\xi}$ та $(\sigma^2)_{MMP}^* = \mathbb{D}\xi^{**}$.

Як бачимо, обом методами отримано однакові оцінки для параметрів даного розподілу: $a^* = \bar{\xi} = -1.52$ та $(\sigma^2)^* = \mathbb{D}\xi^{**} = 4.6352$.

5 ПЕРЕВІРКА ПАРАМЕТРІВ НА НЕЗМІЩЕНІСТЬ, КОНСИСТЕНТНІСТЬ ТА ЕФЕКТИВНІСТЬ

Перевіримо отримані оцінки для кожного параметра на (асимптотичну) незміщеність, консистентність та ефективність. Нехай θ^* оцінка параметру θ , тоді за означеннями:

- Незміщеність:

$$\mathbb{E}\theta_n^* = \theta \quad \forall n \in \mathbb{N}$$

Асимптотично незміщеність:

$$\lim_{n \rightarrow \infty} \mathbb{E}\theta_n^* = \theta$$

- Консистентність:

$$\forall \varepsilon > 0 : \lim_{n \rightarrow \infty} \mathbb{P}\{|\theta_n^* - \theta| \geq \varepsilon\} = 0 \Leftrightarrow \theta_n^* \xrightarrow{\mathbb{P}} \theta, n \rightarrow \infty$$

- Ефективність. Нехай Θ_n — множина усіх незміщених оцінок параметру θ за вибірками фіксованого обсягу n . Оцінка $\theta_{\text{еф}}^*$ називається ефективною, якщо

$$\mathbb{D}\theta_n^* = \inf_{\theta^* \in \Theta_n} \mathbb{D}^*$$

Зручно використовувати наступний критерій ефективності незміщеної оцінки (наслідок з нерівності Крамера-Рао):

$$\frac{\partial \ln \mathcal{L}(\vec{\xi}, \theta)}{\partial \theta} = C(n, \theta) \cdot (\theta^*(\vec{\xi} - \theta))$$

Як відомо, вибіркове середнє та вибіркова дисперсія консистентні оцінки, тому обидва параметра розподілу також консистентні, оскільки дорівнюють їм.

Вибіркове середнє — незміщена оцінка, тоді оцінка параметру розподілу $a^* = \bar{\xi}$ — незміщена оцінка. Вибіркова дисперсія — асимптотично незміщена оцінка, тоді оцінка параметру розподілу $(\sigma^2)^* = \mathbb{D}\xi^{**}$ — асимптотично незміщена.

Перевіримо оцінки на ефективність. Для оцінки параметру $a^* = \bar{\xi}$:

$$\frac{\partial \ln \mathcal{L}(\vec{\xi}, a, \sigma^2)}{\partial a} = \frac{1}{\sigma^2} \sum_{k=1}^n (x_k - a) = \frac{n}{\sigma^2} (\bar{\xi} - a) = C(n) \cdot (\bar{\xi} - a)$$

Отже, можна зробити висновок, що оцінка параметру $a^* = \bar{\xi}$ — ефективна.

Оцінка параметру $(\sigma^2)^* = \mathbb{D}\xi^{**}$ лише асимптотично незміщена, тому не може бути ефективною.

6 ДОВІРЧІ ІНТЕРВАЛИ ДЛЯ ПАРАМЕТРІВ РОЗПОДІЛУ

Побудуємо довірчі інтервали надійністю 0.95 для параметрів розподілу.

Для параметра a : $\mathbb{P}\{|a - a^*| < \varepsilon\} = 0.95$. Маємо:

$$\sqrt{n-1} \frac{\bar{\xi} - a}{\sqrt{\mathbb{D}^{**}\xi}} \sim St_{n-1}$$

З рівності

$$\mathbb{P}\left\{\frac{\sqrt{n-1}|\bar{\xi} - a|}{\sqrt{\mathbb{D}^{**}\xi}} < t_\gamma\right\} = 0.95$$

знайдемо t_γ .

$$\frac{\sqrt{n-1}|\bar{\xi} - a|}{\sqrt{\mathbb{D}^{**}\xi}} < t_\gamma$$

$$|\bar{\xi} - a| < \frac{t_\gamma \sqrt{\mathbb{D}^{**}\xi}}{\sqrt{n-1}}$$

$$\bar{\xi} - \frac{t_\gamma \sqrt{\mathbb{D}^{**}\xi}}{\sqrt{n-1}} < a < \frac{t_\gamma \sqrt{\mathbb{D}^{**}\xi}}{\sqrt{n-1}} + \bar{\xi}$$

$$\int_{t_{\gamma;n}}^{\infty} f(x)dx = \frac{1-\gamma}{2} = \frac{1-0.95}{2} = 0.025$$

Отже, треба знайти значення коефіцієнта $t_{0.025;99}$. В таблицях цей коефіцієнт не наводиться, найближчим є $t_{0.025;100} = 1.984$. Візьмемо це значення за коефіцієнт $t_{0.025;99} \approx 1.984$. Оскільки вибіркві статистики $\bar{\xi} = -1.5207$, а $\mathbb{D}^{**}\xi = 4.5888$. Тоді:

$$\frac{t_\gamma \sqrt{\mathbb{D}^{**}\xi}}{\sqrt{n-1}} = \frac{1.984 \cdot \sqrt{4.5888}}{\sqrt{99}} = 0.4271$$

Отже, $a \in (-1.9478; -1.0936)$.

Для параметра $\mathbb{D}^* * \xi$:

$$\frac{n\mathbb{D}^* * \xi}{\sigma^2} \in \chi_{n-1}^2$$

Маємо:

$$\frac{n\mathbb{D}^{**}\xi}{t_{\frac{1-\gamma}{2}}} < \sigma^2 < \frac{n\mathbb{D}^{**}\xi}{t_{\frac{1+\gamma}{2}}}$$

$$\frac{1-\gamma}{2} = 0.025$$

$$\frac{1 + \gamma}{2} = 0.975$$

Тоді беремо коефіцієнти $\chi^2_{0.025;99} \approx 129.6$ і $\chi^2_{0.975;99} \approx 74.22$.

Межі довірчого інтервалу:

$$\frac{n\mathbb{D}^{**}\xi}{t_{0.025;99}} = 3.5407$$

$$\frac{n\mathbb{D}^{**}\xi}{t_{0.975;99}} = 6.1827$$

Отже, $\sigma^2 \in (3.5407; 6.1827)$.

7 ПЕРЕВІРКА ВИСУНОТОЇ ГІПОТЕЗИ ЗА КРИТЕРІЄМ χ^2

Перевіримо висунуту гіпотезу про розподіл генеральної сукупності за допомогою критерію χ^2 .

Нехай $\xi \sim N(-1.5207, 4.5888)$. Тоді побудуємо таблицю 7.1 для перевірки критерію χ^2 .

Таблиця 7.1 — Таблиця для критерію χ^2

Δ_k	n	$p_k = \mathbb{P}\{\xi \in \Delta_k\}$	np_k
$[-6.8 ; -5.5)$	2	0.0252	2.52
$[-5.5 ; -4.2)$	7	0.0742	7.42
$[-4.2 ; -2.9)$	20	0.1555	15.55
$[-2.9 ; -1.6)$	21	0.2229	22.29
$[-1.6 ; -0.3)$	22	0.2317	23.17
$[-0.3 ; 1.0)$	13	0.1653	16.53
$[1.0 ; 2.3)$	12	0.0815	8.15
$[2.3 ; 3.6]$	3	0.0293	2.93

Як видно, умова $np_k \geq 10$ виконується не для всіх інтервалів, тому необхідно об'єднати деякі з них (табл. 7.2).

Таблиця 7.2 — Таблиця для критерію χ^2

Δ_k	n	$p_k = \mathbb{P}\{\xi \in \Delta_k\}$	np_k
$[-6.8 ; -2.9)$	29	0.2549	25.49
$[-2.9 ; -1.6)$	21	0.2229	22.29
$[-1.6 ; -0.3)$	22	0.2317	23.17
$[-0.3 ; 1.0)$	13	0.1653	16.53
$[1.0 ; 3.6]$	15	0.0815	11.08

Тепер дослідимо величину

$$\chi^2(n) = \sum_{k=1}^m \frac{(n_k - np_k)^2}{np_k} \sim \chi_{m-s-1}^2,$$

де m — кількість інтервалів,

s — кількість оцінених параметрів розподілу.

В даному випадку $\chi_{m-s-1}^2 = \chi_2^2$, $\alpha = 0.05$, то треба взяти $\chi_{2;0.05}^2 = 5.99$. Тепер обчислимо

$$\chi^2(n) = \sum_{k=1}^m \frac{(n_k - np_k)^2}{np_k} = \sum_{k=1}^5 \frac{(n_k - np_k)^2}{np_k} = 2.758$$

Як видно, $\chi^2(n) < \chi_{2;0.05}^2$, то наявні дані не суперечать висунутій гіпотезі, тобто немає підстави відхилити запропоновану гіпотезу на рівні значущості 0.05.

ВИСНОВКИ

В ході виконання розрахунково-графічної роботи було проаналізовано надану вибірку за допомогою широкого набору статистичних засобів. Спочатку проведено первинний аналіз даних та застосовано візуальні інструменти аналізу для отримання першої інтуїтивної гіпотези про клас розподілу генеральної сукупності. Гіпотеза полягала в тому, що генеральна сукупність розподілена за гауссівським розподілом.

На основі висунутої гіпотези про розподіл генеральної сукупності висунуто оцінки параметрів розподілу за допомогою методу моментів та методу максимальної правдоподібності. В обох випадках отримано оцінки $a^* = \bar{\xi}$ та $(\sigma^2)^* = \mathbb{D}^{**}\xi$. В подальшому дані оцінки перевірено на консистентність, (асимптотичну) незміщеність та ефективність. Оцінка параметру $a^* = \bar{\xi}$ має усі зазначені властивості, а оцінка $(\sigma^2)^* = \mathbb{D}^{**}\xi$ консистентна, асимптотично незміщена, але не ефективна.

Для кожного параметру розподілу на основі отриманих оцінок побудовано довірчі інтервали з ймовірністю 0.95. В результаті отримано, що $a \in (-1.9478; -1.0936)$, а $\sigma^2 \in (3.5407; 6.1827)$.

В кінці виконано перевірку висунутої гіпотези про гауссівський розподіл генеральної сукупності за допомогою критерію χ^2 . В результаті визначено, що на рівні значущості 0.05 немає підстав відхилити початкову гіпотезу про гауссівський розподіл генеральної сукупності.