

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 статистики	8
3 Гіпотеза про розподіл генеральної сукупності	9
4 Оцінка параметрів розподілу	10
5 Перевірка параметрів на незміщеність, консистентність та ефективність	11
6 Довірчі інтервали для параметрів розподілу	12
7 Перевірка висунотої гіпотези за критерієм χ^2	13
Висновки	14

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, вибіркиму дисперсію, виправлену вибіркиму дисперсію, вибіркиму медіану, вибіркиму моду, вибіркові 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 вибірким 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.

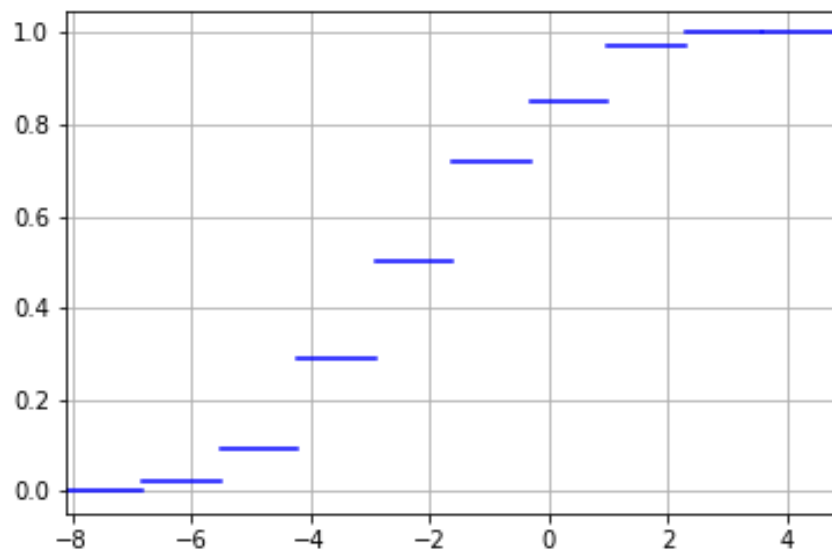


Рисунок 1.1 — Графік емпіричної інтервальної функції розподілу

Гістограма для даної вибірки наведена на рис. 1.2

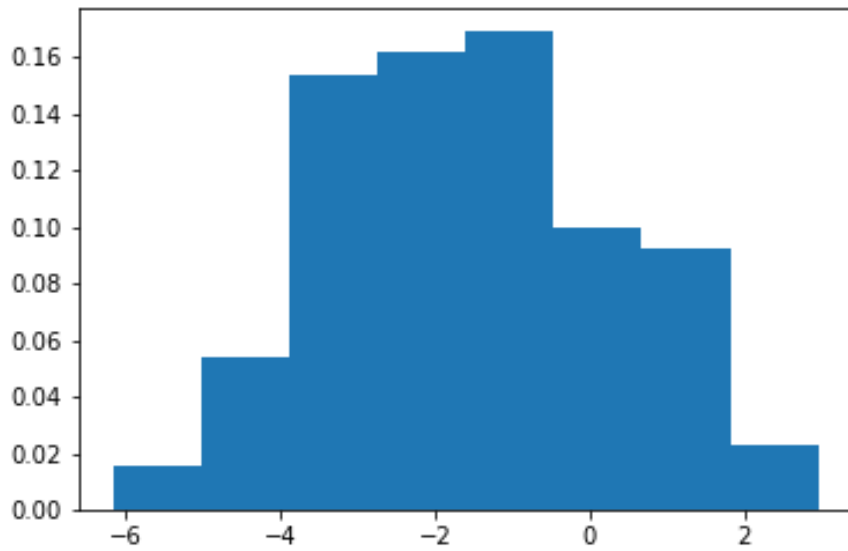


Рисунок 1.2 — Гістограма даної вибірки

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

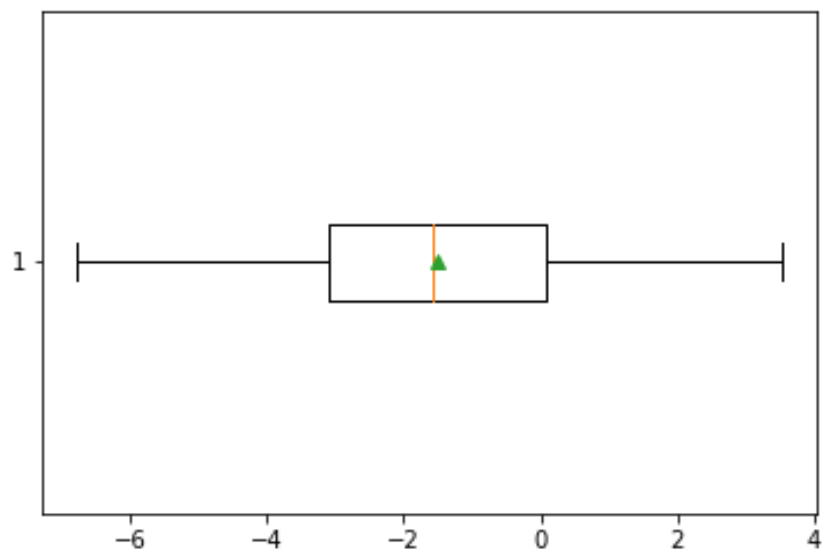


Рисунок 1.3 — Box-and-whisker

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 вибіркове 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

5 ПЕРЕВІРКА ПАРАМЕТРІВ НА НЕЗМІЩЕНІСТЬ, КОНСИСТЕНТНІСТЬ ТА ЕФЕКТИВНІСТЬ

7 ПЕРЕВІРКА ВИСУНОТОЇ ГІПОТЕЗИ ЗА КРИТЕРІЄМ χ^2

ВИСНОВКИ