Розрахункова Робота 2
з математичної статистики
Варіант 133

Воробйов Георгій

20 травня 2021 р.

Зміст

0 Завдання

1. Проведіть первинний аналіз вибірки. Це включає статистичний ряд (для розподілів — інтервальний), емпіричну функцію розподілу (для неперервних розподілів інтервальну), її графік, полігон частот (для дискретних розподілів), гістограму (неперервнихрозподілів), box-and-whisker plot.

2. Знайдіть вибіркове середнє, вибіркову дисперсію, виправлену вибіркову дисперсію, вибіркову медіану, вибіркову моду, вибіркові коефіцієнти асиметрії та ексцесу.

3. Обґрунтуйтета 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 параметра N біноміальної генеральної сукупності). Це чудовий спосіб проявити креативність та/або вміння користуватися Google.

5. Для кожного параметра кращу з цих двох оцінок перевірте на (асимптотичну)незміщеність, консистентність та ефективність.
6. Побудуйте довірчі інтервали надійністю0:95для параметрів розподілу.
7. Нарешті, перевірте висунуту гіпотезу про розподіл генеральної сукупності за допомогою критерію χ^2
8. Проявіть всі свої літературні здібності та напишіть висновки

Задана вибірка:
2 1 1 4 4 3 4 3 2 7 6 1 5 3 3 1 4 3 2 3 2 3 2 3 4 5 3 5 5 1 2 3 6 3 5 5 2 5 2 2 0 3 0 2 6 2 3 4 3 2 4 1 4 3 4 2 4 1 4 5 5 3 3 3 2 4 3 2 4 3 3 3 3 4 2 3 6 1 2 3 3 4 0 3 5 1 4 4 3 3 1 3 3 6 2 2 3 2 5 3

1 Розв'язок

1.1 Вступ

Запишемо відсортовану вибірку:
0 0 0 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 7

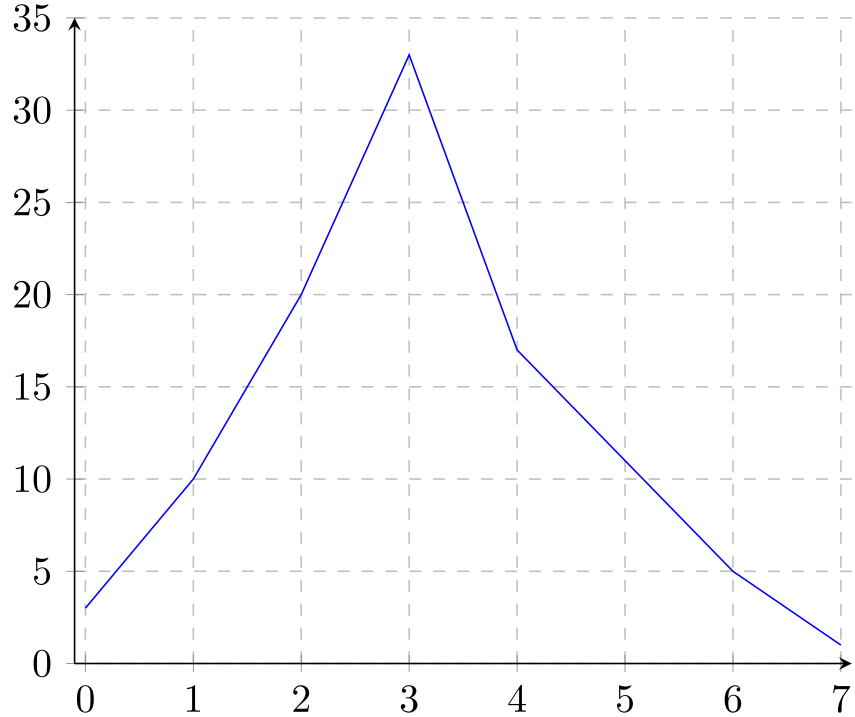
1.2 Аналіз вибірки

Побудуємо статистичний ряд даної вибірки

елементи	Частота n_i	Кумулятивна частота n_i^*	Відносна частота ν_i	Відносна кумулятивна частота ν_i^*
0	3	3	0.03	0.03
1	10	13	0.1	0.13
2	20	33	0.2	0.33
3	33	66	0.33	0.66
4	17	83	0.17	0.83
5	11	94	0.11	0.94
6	5	99	0.05	0.99
7	1	100	1	1

За даними таблиці можемо побудувати полігон частот та функцію розподілу

Рис. 1: Полігон частот



Маємо наступну емпіричну функцію розподілу

$$F_n^*(x) = \begin{cases} 0 & x \le 0 \\ 0.03 & 0 < x \le 1 \\ 0.13 & 1 < x \le 2 \\ 0.33 & 2 < x \le 3 \\ 0.66 & 3 < x \le 4 \\ 0.83 & 4 < x \le 5 \\ 0.94 & 5 < x \le 6 \\ 0.99 & 6 < x \le 7 \\ 1 & x > 7 \end{cases}$$

Відповідний їй графік:

Рис. 2: Епірична функція розподілу

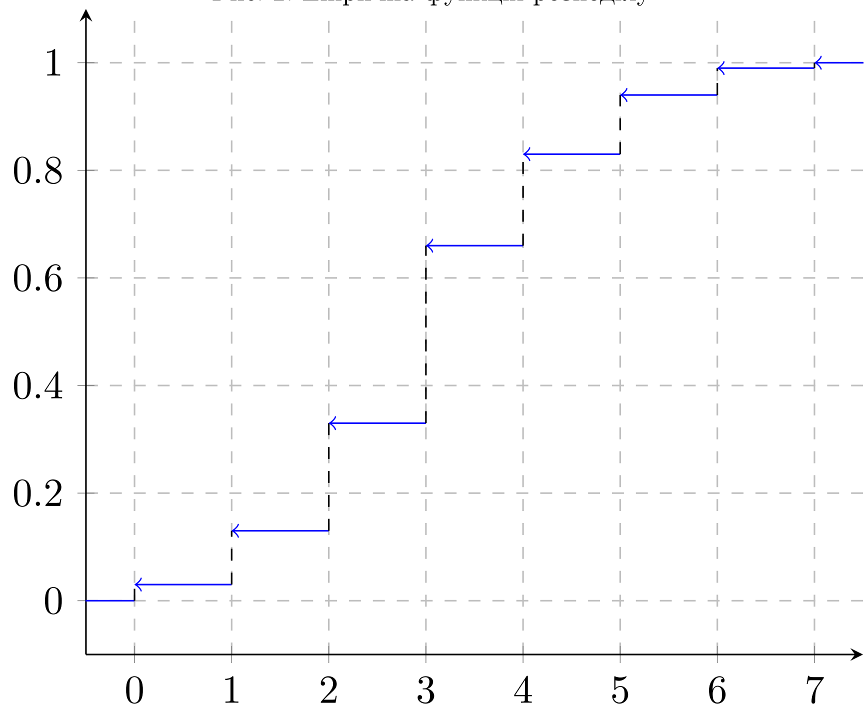
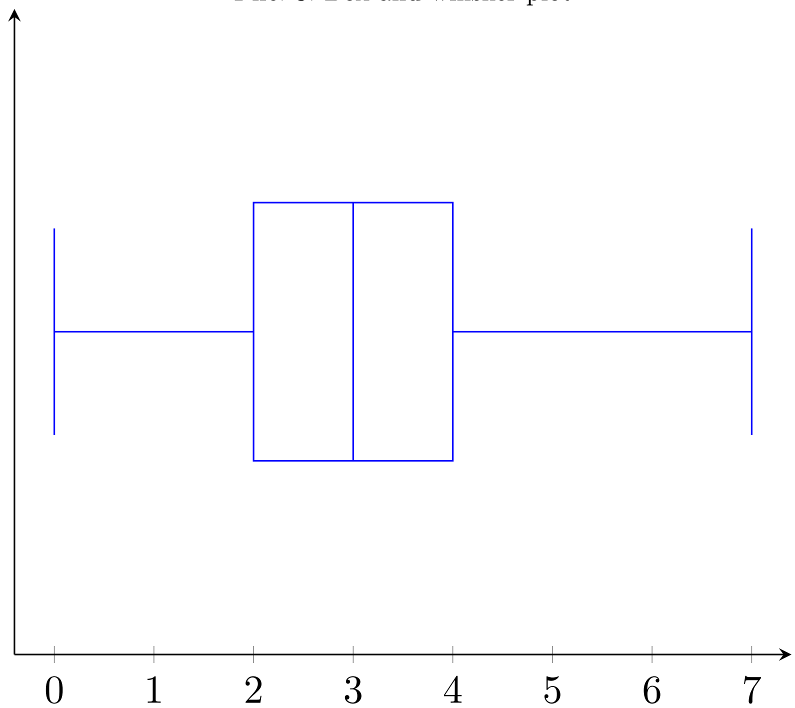


Рис. 3: Box and whisker plot



1.3 Вибіркові статистики

Порахуємо значення вибіркового середнього:

$$\overline{\xi} = \frac{1}{n}\sum_{i=1}^{100}\xi_i$$

$$\overline{\xi}_{\text{Знач}} = 3.09$$

Порахуємо значення вибіркової дисперсії:

$$\mathbb{D}^{**} = \frac{1}{n}\sum_{i=1}^{100}(\xi_i - \overline{\xi})^2$$

$$\mathbb{D}^{**}\xi_{\text{Знач}} = 2.08$$

Порахуємо значення виправленої вибіркової дисперсії:

$$\mathbb{D}^{***} = \frac{n}{n-1}\mathbb{D}^{**}$$

$$\mathbb{D}^{***}\xi_{\text{Знач}} = 2.06$$

Порахуємо вибіркову медіану:

$$Me^{*}\xi_{\text{Знач}} = \langle \text{середина вибірки} \rangle = 3$$

Порахуємо вибіркову моду - значення вибірки, що зустрічається найчастіше:

$$Mo^{*}\xi = 3$$

Для розрахунку вибіркових коефіцієнтів асиметрії та ексцесу розрахуємо потрібні початкові та центральні вибіркові моменти

$$As^{*}\xi = \frac{\mu_3^{*}}{(\sigma^{*})^3}$$

$$Ex^{*}\xi = \frac{\mu_4^{*}}{(\sigma^{*})^4} - 3$$

$$\mu_3^{*} = \mathbb{E}\left[\xi - \overline{\xi}\right]^3 = 0.667$$

$$\mu_4^{*} = \mathbb{E}\left[\xi - \overline{\xi}\right]^4 = 12.429$$

$$\sigma^{*} = \sqrt{\mathbb{D}^{**}} = 1.443$$

$$As^{*}\xi_{\text{Знач}} = 0.221$$

$$Ex^{*}\xi_{\text{Знач}} = -0.14$$

1.4 Висунення Гіпотези

1. Маємо дискретну ГС

2. Значення дисперсії менше ніж математичного сподівання(Що схоже на npq та np відповідно)

3. Має вибіркову медіану рівну вибірковій моді

Тож можемо припустити, що данна ГС є розподіленою біноміально
$\xi \sim Bin(N, p)$

1.5 Оцінки параметра розподілу

1.5.1 Метод моментів

Маємо відомі нам значення матсподівання та дисперсії:

$$\begin{cases} \mathbb{E}\xi = Np \\ \mathbb{D}\xi = Np(1-p) \end{cases}$$

Звідси, підставивши значення $\mathbb{E}^*\xi = \overline{\xi}$ та D^{***} можемо виразити значення N* та p* :

$$\begin{cases} 1 - p^* = \frac{\mathbb{D}^{***}\xi}{\overline{\xi}} \\ \overline{\xi} = N^* p^* \end{cases}$$

$$\begin{cases} p^* = \frac{\overline{\xi} - \mathbb{D}^{***}\xi}{\overline{\xi}} \\ \overline{\xi} = N^* p^* \end{cases}$$

$$\begin{cases} p^* = \frac{\overline{\xi} - \mathbb{D}^{***}\xi}{\overline{\xi}} \\ N^* = \frac{\overline{\xi}^2}{\overline{\xi} - \mathbb{D}^{***}\xi} \end{cases}$$

Для значення нашої вибірки отримуємо значення p^* та N^*

$$\begin{cases} p^* = \frac{3.09 - 2.06}{3.09} \\ N^* = \frac{3.09^2}{3.09 - 2.06} \end{cases}$$

$$\begin{cases} p^* = \frac{1}{3} \\ N^* = 9.27 \end{cases}$$

1.5.2 Метод максимальної правдоподібності

Запишемо функцію Правдоподібності

$$\mathcal{L}(\overrightarrow{x}, N, p) = \prod_{i=1}^{n} \mathbb{P}\{\xi = x_i\} = \prod_{i=1}^{n} C_N^{x_i} p^{x_i} (1-p)^{N - x_i}$$

$$\ln \mathcal{L}(\overrightarrow{x}, N, p) = \sum_{i=1}^{n} ln C_N^{x_k} + \ln p \sum_{i=1}^{n} x_k + \ln(1-p) \left(nN - \sum_{i=1}^{n} x_k \right)$$