

# Xử lý số liệu thống kê

## Assignment 1

Nhóm E:

Trần Tiến Đạt  
Nguyễn Thị Ngọc Anh  
Nguyễn Thái Hưng Thịnh

Ngày nộp: November 10, 2025

## **Contents**

Bài 4	2
Bài 5	3

## Bài 4

Xét hai bộ dữ liệu  $x_1 < x_2 < x_3 < \dots < x_n$  và  $y_1 < y_2 < y_3 < \dots < y_n$ , có trung bình mẫu tương ứng là  $\bar{x}$ ,  $\bar{y}$  và trung vị lần lượt là  $m_x$  và  $m_y$ . Đặt  $w_i = x_i + y_i$ .

(a) Chứng minh hoặc đưa ra phản chứng rằng:  $\bar{x} + \bar{y}$  là trung bình mẫu của  $w_1, w_2, \dots, w_n$

**Lời giải :**

Trước hết, ta xét trung bình mẫu của  $w_1, w_2, \dots, w_n$ :

$$\begin{aligned}\bar{w} &= \frac{1}{n} \sum_{i=1}^n w_i \\ &= \frac{1}{n} \sum_{i=1}^n (x_i + y_i) \\ &= \frac{1}{n} \left( \sum_{i=1}^n x_i + \sum_{i=1}^n y_i \right) \\ &= \frac{1}{n} \sum_{i=1}^n x_i + \frac{1}{n} \sum_{i=1}^n y_i \\ &= \bar{x} + \bar{y}\end{aligned}$$

Vậy suy ra  $\bar{w} = \bar{x} + \bar{y}$  là trung bình mẫu của  $w_1, w_2, \dots, w_n$ .

(b) Chứng minh hoặc đưa ra phản chứng rằng:  $m_x + m_y$  là trung vị của  $w_1, w_2, \dots, w_n$

**Lời giải :**

Ta sẽ chứng minh mệnh đề trên là đúng, trước hết ta sẽ xét với trường hợp  $n$  là số lẻ, tức rằng ta có trung vị của  $\{x\}_n$  và  $\{y\}_n$  lần lượt là:

$$m_x = x_{\frac{n+1}{2}}$$

$$m_y = y_{\frac{n+1}{2}}$$

Khi đó với dãy  $w$  ta có:

$$\begin{aligned}m_w &= w_{\frac{n+1}{2}} \\ &= x_{\frac{n+1}{2}} + y_{\frac{n+1}{2}} \\ &= m_x + m_y\end{aligned}\tag{1}$$

Với trường hợp  $n$  là số chẵn, ta có trung vị của  $\{x\}_n$  và  $\{y\}_n$  lần lượt là:

$$m_x = \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1})$$

$$m_y = \frac{1}{2}(y_{\frac{n}{2}} + y_{\frac{n}{2}+1})$$

Khi đó với dãy  $w$  ta có:

$$\begin{aligned}
m_w &= \frac{1}{2}(w_{\frac{n}{2}} + w_{\frac{n}{2}+1}) \\
&= \frac{1}{2}((x_{\frac{n}{2}} + y_{\frac{n}{2}}) + (x_{\frac{n}{2}+1} + y_{\frac{n}{2}+1})) \\
&= \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) + \frac{1}{2}(y_{\frac{n}{2}} + y_{\frac{n}{2}+1}) \\
&= m_x + m_y
\end{aligned} \tag{2}$$

Từ (2) và (1) ta suy ra  $m_x + m_y$  là trung vị của  $w_1, w_2, \dots, w_n$  với mọi  $n$ .

## Bài 5

Giả sử rằng ta có dữ liệu ngẫu nhiên  $X_1, X_2, \dots, X_n$  là độc lập cùng phân phối Poisson  $P(\lambda)$ . Ta biết rằng  $\bar{X}$  là ước lượng không chêch của  $\lambda$ .

(a) Sử dụng định lý giới hạn trung tâm, hãy xác định công thức cho khoảng tin cậy 95% cho  $\lambda$ .

**Lời giải :**

Với  $X_i \sim \text{Poisson}(\lambda)$ , ta có

$$\mathbb{E}[X_i] = \lambda, \quad \text{Var}(X_i) = \lambda.$$

Khi đó, trung bình mẫu

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

là ước lượng không chêch của  $\lambda$ .

Từ định lý giới hạn trung tâm, ta có:

$$\frac{\bar{X} - \lambda}{\sqrt{\lambda/n}} \xrightarrow{d} \mathcal{N}(0, 1),$$

tức là khi  $n$  đủ lớn, phân phối của  $\bar{X}$  xấp xỉ chuẩn với trung bình  $\lambda$  và phương sai  $\frac{\lambda}{n}$ :

$$\bar{X} \approx \mathcal{N}\left(\lambda, \frac{\lambda}{n}\right).$$

Do đó,

$$P\left(-1.96 \leq \frac{\bar{X} - \lambda}{\sqrt{\lambda/n}} \leq 1.96\right) \approx 0.95,$$

hay tương đương,

$$P\left(\bar{X} - 1.96\sqrt{\frac{\lambda}{n}} \leq \lambda \leq \bar{X} + 1.96\sqrt{\frac{\lambda}{n}}\right) \approx 0.95.$$

Do tham số  $\lambda$  chưa biết nên trong thực tế ta không thể tính chính xác sai số chuẩn  $\sqrt{\lambda/n}$ . Vì  $\bar{X}$  là ước lượng hợp lý và không chênh của  $\lambda$ , ta có thể thay  $\lambda$  bằng  $\bar{X}$  trong biểu thức này để thu được ước lượng xấp xỉ của sai số chuẩn. Khi đó, khoảng tin cậy 95% (theo xấp xỉ Wald) cho  $\lambda$  được viết là:

$$\lambda \in \left[ \bar{X} - 1.96\sqrt{\frac{\bar{X}}{n}}, \bar{X} + 1.96\sqrt{\frac{\bar{X}}{n}} \right].$$

**(b)** Áp dụng công thức trong ý (a), hãy xác định khoảng tin cậy cho  $\lambda$  theo dữ liệu 4, 6, 7, 9, 10, 13.

**Lời giải :**

Với dữ liệu đã cho, ta có  $n = 6$  và

$$\bar{x} = \frac{4 + 6 + 7 + 9 + 10 + 13}{6} = \frac{49}{6} \approx 8.167.$$

Độ tin cậy:

$$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow 1 - \frac{\alpha}{2} = 0.975 \Rightarrow z_{1-\alpha/2} = z_{0.975} = 1.96.$$

Biên độ tin cậy:

$$\Delta = z_{0.975} \sqrt{\frac{\bar{x}}{n}} = 1.96 \sqrt{\frac{49/6}{6}} = 1.96 \sqrt{\frac{49}{36}} = 1.96 \cdot \frac{7}{6} = \frac{343}{150} \approx 2.287.$$

Khi đó, khoảng tin cậy 95% cho  $\lambda$  là

$$\lambda \in [8.167 - 2.287, 8.167 + 2.287] = [5.880, 10.454].$$

**Kết luận:** Khoảng tin cậy 95% cho tham số  $\lambda$  của phân phối Poisson dựa trên mẫu đã cho là

$$\lambda \in (5.880, 10.454).$$