ÔN TẬP

XÁC SUẤT THỐNG KÊ

HCMUS

Ngày 17 tháng 11 năm 2025

LỜI NÓI ĐẦU

Tài liệu này được soạn dựa trên slide, bài giảng trên lớp và một số tài liệu khác trong quá trình học tập môn Xác suất thống kê của tác giả. Nội dung bao gồm tóm tắt lý thuyết cơ bản và giải một số đề thi giữa kỳ, cuối kỳ.

Trong phần tóm tắt lý thuyết, tài liệu chỉ ghi lại công thức, không bao gồm giải thích ý nghĩa, cách xây dựng hay cách áp dụng. Trong phần giải đề, mặc dù đã cố gắng kiểm tra cẩn thận, song không thể tránh khỏi sai sót trong quá trình tính toán và trình bày. Do đó, người đọc nên kết hợp với bài giảng trên lớp và cách trình bày của giảng viên để đạt được kết quả tốt nhất.

Lưu ý rằng, tài liệu này được chia sẻ miễn phí để hỗ trợ các bạn sinh viên học tập, vui lòng không sử dụng cho mục đích thương mại dưới bất kỳ hình thức nào. Lời cuối, hi vọng rằng tài liệu sẽ có ích với bạn trong quá trình học tập cũng như ôn thi giữa kỳ và cuối kỳ.

MỤC LỤC

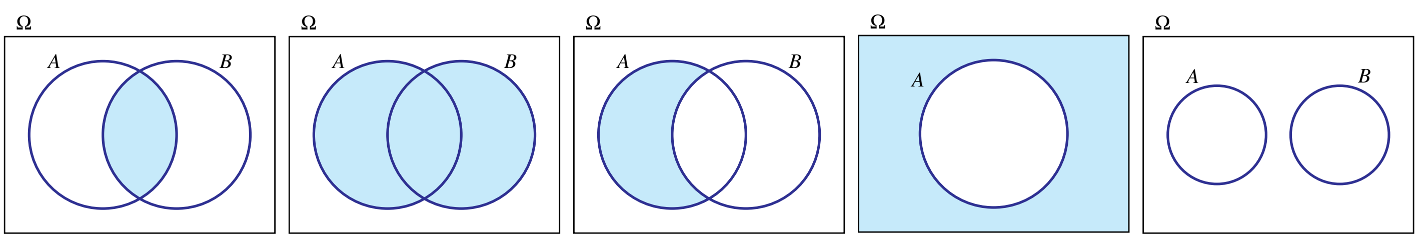
A TÓM TẮT LÝ THUYẾT PHẦN XÁC SUẤT

1 Một số khái niệm cơ bản

- **Phép thử ngẫu nhiên (phép thử):** là sự thực hiện một số điều kiện xác định (một thí nghiệm cụ thể hay quan sát một hiện tượng nào đó), có thể lặp lại nhiều lần và kết quả không xác định trước được.
- **Không gian mẫu**: là tập hợp tất cả các kết quả có thể xảy ra của một phép thử ngẫu nhiên. Không gian mẫu còn được gọi là không gian các biến cố sơ cấp.
Kí hiệu: Ω
- **Biến cố sơ cấp**: là một kết quả của phép thử ngẫu nhiên.
Kí hiệu: ω (ta có: $\omega \in \Omega$)
- **Biến cố ngẫu nhiên (biến cố)**: là một tập con của không gian mẫu.
Kí hiệu: chữ cái in hoa (thường ở đầu bảng chữ cái) như $A, B, C, \ldots$

2 Phép toán trên biến cố

Biểu diễn thông qua biểu đồ Venn:



Biến cố tích của A và B Kí hiệu: $A \cap B$ hoặc AB	Biến cố tổng của A và B Kí hiệu: $A \cup B$ hoặc $A + B$	Biến cố hiệu của A và B Kí hiệu: $A \setminus B$	Biến cố đối của A Kí hiệu: $\overline{A}$ hoặc A^c	A xung khắc với B Kí hiệu: $AB = \varnothing$

Lưu ý: $\overline{A+B} = \overline{A} \cdot \overline{B}$ và $\overline{AB} = \overline{A} + \overline{B}$ (**công thức De Morgan**)

3 Xác suất của biến cố

3.1 Tính chất

- $0 \leq P(A) \leq 1$
- $P(\varnothing) = 0$ và $P(\Omega) = 1$
- $\mathrm{P}(\overline{A}) = 1 - P(A)$
- $P(A + B) = P(A) + P(B) - P(AB)$ (**công thức cộng xác suất**)

3.2 Xác suất có điều kiện

- Xác suất xảy ra biến cố A **khi biết** biến cố B đã xảy ra là:

$$P(A \mid B) = \frac{P(AB)}{P(B)}, \quad \text{với } P(B) > 0$$

- Xác suất xảy ra biến cố B **khi biết** biến cố A đã xảy ra là:

$$P(B \mid A) = \frac{P(AB)}{P(A)}, \quad \text{với } P(A) > 0$$

- Tính chất:
  - $0 \leq P(A \mid B) \leq 1$
  - $P(A \mid A) = 1$
  - $AC = \varnothing \quad \Rightarrow \quad P((A + C) \mid B) = P(A \mid B) + P(C \mid B)$
  - $P(\overline{A} \mid B) = 1 - P(A \mid B)$
- Từ định nghĩa, ta suy ra **công thức nhân xác suất**:

$$P(AB) = P(A \mid B) \cdot P(B) = P(B \mid A) \cdot P(A)$$

3.3 Hai biến cố độc lập

- Định nghĩa:

$$\text{Biến cố } A \text{ độc lập với biến cố } B \quad \Leftrightarrow \quad P(AB) = P(A) \cdot P(B)$$

- Tính chất:
  - $P(A \mid B) = P(A)$
  - $P(B \mid A) = P(B)$
  - Nếu A độc lập với B thì A cũng độc lập với $\overline{B}$

3.4 Hệ đầy đủ các biến cố

Dãy n các biến cố $A_1, A_2, \ldots, A_n$ được gọi là một hệ đầy đủ các biến cố nếu thỏa đồng thời:

- $A_i \cap A_j = \varnothing, \quad \forall i \neq j$
- $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$

Dãy $\{A_1, A_2, \ldots, A_n\}$ còn được gọi là một phân hoạch của không gian mẫu.

3.5 Công thức xác suất đầy đủ - Công thức Bayes

Cho $\{A_1, A_2, \ldots, A_n\}$ là một **hệ đầy đủ** các biến cố và B là một biến cố nào đó liên quan (trong cùng phép thử) thì ta có:

- **Công thức xác suất đầy đủ (toàn phần):**

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i) \cdot P(A_i)$$

  - Với $n = 2$: $P(B) = P(B \mid A_1) \cdot P(A_1) + P(B \mid A_2) \cdot P(A_2)$
  - Với $n = 3$: $P(B) = P(B \mid A_1) \cdot P(A_1) + P(B \mid A_2) \cdot P(A_2) + P(B \mid A_3) \cdot P(A_3)$

- **Công thức Bayes:**

$$P(A_k \mid B) = \frac{P(B \mid A_k) \cdot P(A_k)}{P(B)} = \frac{P(B \mid A_k) \cdot P(A_k)}{\sum_{i=1}^{n} P(B \mid A_i) \cdot P(A_i)}$$

4 Biến ngẫu nhiên

- **Biến ngẫu nhiên** là một hàm số gán một số thực cho mỗi kết quả có thể xảy ra trong không gian mẫu của một phép thử ngẫu nhiên.

- **Kí hiệu:** chữ cái in hoa (thường ở cuối bảng chữ cái) như $X, Y, X, \ldots$ Người ta còn dùng các chữ cái thường tương ứng để kí hiệu **giá trị cụ thể** của biến ngẫu nhiên: $x, y, z, \ldots$

  Ta có:

$$P(X = x) = P(\{\omega \in \Omega \mid X(\omega) = x\})$$
$$P(X \leq x) = P(\{\omega \in \Omega \mid X(\omega) \leq x\})$$

- **Biến ngẫu nhiên rời rạc** là biến ngẫu nhiên có tập giá trị là một tập hữu hạn hoặc vô hạn đếm được.

- **Biến ngẫu nhiên liên tục** là biến ngẫu nhiên có tập giá trị là một khoảng (hoặc một đoạn) trên tập số thực.

5 Biến ngẫu nhiên rời rạc

5.1 Hàm khối xác suất (hàm xác suất)

Cho biến ngẫu nhiên rời rạc X có thể nhận các giá trị $x_1, x_2, \ldots, x_n$, **hàm khối xác suất** của X, kí hiệu p_X, là hàm số thỏa:

- $p_X(x_i) \geq 0, \quad \forall x_i$

- $\sum_{i=1}^{n} p_X(x_i) = 1$

- $p_X(x_i) = P(X = x_i)$

5.2 Hàm phân phối xác suất (hàm phân phối tích lũy)

Cho biến ngẫu nhiên rời rạc X có thể nhận các giá trị $x_1, x_2, \ldots, x_n$, **hàm phân phối xác suất** của X, kí hiệu là F, được xác định bởi:

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p_X(x_i)$$

5.3 Kỳ vọng

$$\mathbb{E}(X) = \mu = \sum_x x \cdot p_X(x)$$

$$\mathbb{E}\left[h(X)\right] = \sum_x h(x) \cdot p_X(x)$$

5.4 Phương sai

$$\mathrm{Var}(X) = \sigma^2 = \sum_x (x-\mu)^2 \cdot p_X(x)$$

$$\mathrm{Var}\left[h(X)\right] = \sum_x \left[h(x) - \mathbb{E}(h(x))\right]^2 \cdot p_X(x)$$

6 Biến ngẫu nhiên liên tục

6.1 Hàm mật độ xác suất

Hàm mật độ xác suất của biến ngẫu nhiên liên tục X, kí hiệu f, là hàm số thỏa:

- $f(x) \geq 0, \quad \forall x \in \mathbb{R}$
- $\displaystyle\int_{-\infty}^{+\infty} f(x)\, dx = 1$
- $P(a \leq X \leq b) = \displaystyle\int_a^b f(x)\, dx, \quad \forall a,b \in \mathbb{R}$

6.2 Hàm phân phối xác suất (hàm phân phối tích lũy)

Hàm phân phối xác suất của biến ngẫu nhiên liên tục X, kí hiệu là F, được xác định bởi:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\, du$$

Lưu ý: Với biến ngẫu nhiên liên tục X, ta có

- $f(x) = \dfrac{dF(x)}{dx}$
- $P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$ vì $P(X = \text{hằng số}) = 0$

6.3 Kỳ vọng

$$\mathbb{E}(X) = \mu = \int_{-\infty}^{+\infty} x f(x)\, dx$$
$$\mathbb{E}\big[h(X)\big] = \int_{-\infty}^{+\infty} h(x) f(x)\, dx$$

6.4 Phương sai

$$\mathrm{Var}(X) = \sigma^2 = \mathbb{E}(X^2) - \big[\mathbb{E}(X)\big]^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\, dx$$
$$\mathrm{Var}\big[h(X)\big] = \mathbb{E}\big[h(x)^2\big] - \mathbb{E}\big[h(x)\big]^2$$

7 Tính chất của kỳ vọng và phương sai

Cho hai biến ngẫu nhiên X, Y và hằng số c, ta có:

7.1 Tính chất của kỳ vọng

- $\mathbb{E}(c) = c$
- $\mathbb{E}(cX) = c\mathbb{E}(X)$
- $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
- X độc lập với $Y \;\Rightarrow\; \mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

7.2 Tính chất của phương sai

- $\mathrm{Var}(c) = 0$
- $\mathrm{Var}(cX) = c^2 \cdot \mathrm{Var}(X)$
- $\mathrm{Var}(X+c) = \mathrm{Var}(X)$
- X độc lập với $Y \;\Rightarrow\; \mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

8 Các phân phối xác suất thường gặp

8.1 Phân phối Bernoulli

- **Kí hiệu:** $X \sim B(1\,;\ p)$
- **Kỳ vọng:** $\mathbb{E}(X) = p$
- **Phương sai:** $\mathrm{Var}(X) = p(1-p)$

8.2 Phân phối nhị thức

- **Kí hiệu:** $X \sim B\,(n\,;\;p)$
- **Hàm khối xác suất:** $P(X = k) = C_n^k \cdot p^k \cdot (1-p)^{n-k}$
- **Kỳ vọng:** $\mathbb{E}(X) = np$
- **Phương sai:** $\mathrm{Var}(X) = np(1-p)$

8.3 Phân phối Poisson

- **Kí hiệu:** $X \sim \mathcal{P}(\lambda)$ (λ: trung bình số sự kiện xảy ra trong một đơn vị thời gian hoặc không gian)
- **Hàm khối xác suất:** $P(X = k) = \dfrac{\lambda^k \cdot e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots$
- **Kỳ vọng:** $\mathbb{E}(X) = \lambda$
- **Phương sai:** $\mathrm{Var}(X) = \lambda$

8.4 Phân phối đều

- **Kí hiệu:** $X \sim \mathcal{U}([a,b])$ (biến ngẫu nhiên liên tục X có phân phối đều trên đoạn $[a, b]$)
- **Hàm mật độ xác suất:**

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{nếu } x \in [a,b] \\ 0, & \text{nếu } x \notin [a,b] \end{cases}$$

- **Hàm phân phối xác suất:**

$$F(x) = P(X \le x) = \begin{cases} 0, & \text{nếu } x < a \\ \dfrac{x-a}{b-a}, & \text{nếu } a \le x \le b \\ 1, & \text{nếu } x > b \end{cases}$$

- **Kỳ vọng:** $\mathbb{E}(X) = \dfrac{1}{2}(a+b)$
- **Phương sai:** $\mathrm{Var}(X) = \dfrac{1}{12}(b-a)^2$

8.5 Phân phối mũ

- **Kí hiệu:** $T \sim \mathrm{Exp}(\lambda)$ (biến ngẫu nhiên liên tục T có phân phối mũ với tham số λ)
- **Hàm mật độ xác suất:**

$$f(t) = \lambda e^{-\lambda t}, \quad t \ge 0$$

- **Hàm phân phối xác suất:**
$$F(t) = P(T \le t) = 1 - e^{-\lambda t}, \quad t \ge 0$$

- **Kỳ vọng:** $\mathbb{E}(T) = \dfrac{1}{\lambda}$

- **Phương sai:** $\text{Var}(T) = \dfrac{1}{\lambda^2}$

8.6 Phân phối chuẩn

- **Kí hiệu:** $X \sim \mathcal{N}\left(\mu\ ;\ \sigma^2\right)$ (biến ngẫu nhiên liên tục X có phân phối chuẩn với tham số μ và σ^2)
- **Kỳ vọng:** $\mathbb{E}(X) = \mu$
- **Phương sai:** $\text{Var}(X) = \sigma^2$
- **Định lý 1:** Nếu $X \sim \mathcal{N}\left(\mu\ ;\ \sigma^2\right)$ thì $Y = aX + b \sim \mathcal{N}\left(a\mu + b\ ;\ a^2\sigma^2\right)$
- **Định lý 2:** Nếu các biến ngẫu nhiên $X_1, X_2, \ldots, X_n$ độc lập với nhau và $X_i \sim \mathcal{N}\left(\mu_i\ ;\ \sigma_i^2\right)$ thì $S = \sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\mu_1 + \ldots + \mu_n\ ;\ \sigma_1^2 + \ldots + \sigma_n^2\right)$
- **Phân vị** mức α của biến ngẫu nhiên liên tục X, kí hiệu x_α, là giá trị của X thỏa:
$$P(X \le x_\alpha) = \alpha \quad \text{(lưu ý có nơi quy ước vế phải là } 1 - \alpha\text{)}$$

8.7 Phân phối chuẩn tắc

- Biến ngẫu nhiên Z có phân phối chuẩn tắc nếu nó có phân phối chuẩn với $\mu = 0$ và $\sigma^2 = 1$.
- **Kí hiệu:** $Z \sim \mathcal{N}(0\ ;\ 1)$
- **Hàm phân phối xác suất:** $\Phi(z) = P(Z \le z)$
(giá trị này có thể tra bảng phân phối chuẩn tắc thay vì tính tích phân hàm mật độ xác suất)
- **Tính chất**:
  - $\Phi(-z) = 1 - \Phi(z)$
  - $P(Z \le b) = \Phi(b)$
  - $P(a < Z \le b) = \Phi(b) - \Phi(a)$
- **Lưu ý:** Nếu $X \sim \mathcal{N}\left(\mu\ ;\ \sigma^2\right)$ thì $Z = \dfrac{X - \mu}{\sigma} \sim \mathcal{N}(0\ ;\ 1)$

Note: Người ta thường dùng kí hiệu Z cho biến ngẫu nhiên có phân phối chuẩn tắc và kí hiệu $\Phi(\cdot)$ cho hàm phân phối xác suất của Z.

9 Định lý giới hạn trung tâm

9.1 Sử dụng định lý

Xét

- $X_1, X_2, \ldots, X_n$ là dãy biến ngẫu nhiên độc lập và có cùng phân phối, với kỳ vọng $\mathbb{E}(X) = \mu$ và phương sai $\mathrm{Var}(X) = \sigma^2$
- $S_n = X_1 + X_2 + \ldots + X_n$

thì khi n đủ lớn, ta có:

$$Z_n = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}} = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \sim \mathcal{N}(0\,;\,1)$$

Ngoài ra, nếu đặt $\overline{X} = \dfrac{X_1 + X_2 + \ldots + X_n}{n} = \dfrac{S_n}{n}$ thì ta có:

$$Z_n = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0\,;\,1)$$

9.2 Hiệu chỉnh liên tục

Khi thực hiện xấp xỉ các phân phối rời rạc bởi phân phối chuẩn, để tăng độ chính xác cho kết quả, ta luôn sử dụng **hiệu chỉnh liên tục** như sau:

- $P(X \leq a) \approx \Phi\left(\dfrac{a - \mu_X + \mathbf{0{,}5}}{\sigma_X}\right)$
- $P(X < a) \approx \Phi\left(\dfrac{a - \mu_X - \mathbf{0{,}5}}{\sigma_X}\right)$
- $P(a < X \leq b) = P(X \leq b) - P(X \leq a) \approx \Phi\left(\dfrac{b - \mu_X + \mathbf{0{,}5}}{\sigma_X}\right) - \Phi\left(\dfrac{a - \mu_X + \mathbf{0{,}5}}{\sigma_X}\right)$

A1 GIỮA KỲ HÈ 2024-2025

Câu 1 (2 điểm)

Có hai lô sản phẩm, lô thứ nhất có 10 sản phẩm loại I và 2 sản phẩm loại II. Lô thứ hai có 16 sản phẩm loại I và 4 sản phẩm loại II. Từ mỗi lô ta lấy ngẫu nhiên một sản phẩm. Sau đó, từ 2 sản phẩm thu được lấy ngẫu nhiên ra một sản phẩm. Tìm xác suất để sản phẩm lấy ra sau cùng là sản phẩm loại I.

Lời giải.

Phép thử: Chọn ngẫu nhiên một sản phẩm từ hai sản phẩm đã lấy ra trước đó

Đặt các biến cố:

A_i: Sản phẩm được chọn thuộc lô thứ i $(i = 1,2)$

B: Sản phẩm được chọn là sản phẩm loại I

Theo đề bài:

$P(A_1) = 0{,}5$ $\qquad$ $P(A_2) = 0{,}5$

$P(B \mid A_1) = \dfrac{10}{10+2} = \dfrac{5}{6}$ $\qquad$ $P(B \mid A_2) = \dfrac{16}{16+4} = \dfrac{4}{5}$

Vì $\{A_1, A_2\}$ là một **hệ đầy đủ** nên xác suất để sản phẩm được chọn là sản phẩm loại I:

$$\begin{aligned} P(B) &= P(B \mid A_1) \cdot P(A_1) + P(B \mid A_2) \cdot P(A_2) \\ &= \frac{5}{6} \cdot 0{,}5 + \frac{4}{5} \cdot 0{,}5 \\ &\approx 0{,}8167 \end{aligned}$$

■

Câu 2 (4 điểm)

Giả sử thời gian tự học của sinh viên (đơn vị: giờ/ngày) là biến ngẫu nhiên X có hàm phân phối xác suất như sau:

$$F(x) = \begin{cases} 0, & \text{khi } x < 0 \\ \dfrac{x^2}{c}, & \text{khi } 0 \le x \le 8 \\ 1, & \text{khi } x > 8 \end{cases}$$

Giả sử rằng thời gian tự học của sinh viên là độc lập với nhau.

(a) Tìm c để F là hàm phân phối xác suất.

(b) Chọn ngẫu nhiên một sinh viên. Tính xác suất sinh viên đó có thời gian tự học ít nhất 5 giờ/ngày.

(c) Chọn ngẫu nhiên 50 sinh viên. Tính xác suất có không quá 10 sinh viên có thời gian tự học ít hơn 3 giờ/ngày.

(d) Tính trung bình và phương sai của X.

(e) Khảo sát 500 sinh viên, tính xác suất thời gian tự học trung bình của 500 sinh viên này là hơn 5,3 giờ/ngày.

Lời giải.

(a) Vì F là hàm phân phối xác suất nên F liên tục tại mọi điểm, suy ra:

$$\lim_{x\to 8^-} F(x) = \lim_{x\to 8^+} F(x) \quad\Leftrightarrow\quad \lim_{x\to 8^-} \frac{x^2}{c} = \lim_{x\to 8^+} 1 \quad\Leftrightarrow\quad \frac{64}{c} = 1 \quad\Leftrightarrow\quad c = 64$$

Vậy hàm phân phối xác suất của X là:

$$F(x) = \begin{cases} 0, & \text{khi } x < 0 \\ \dfrac{x^2}{64}, & \text{khi } 0 \le x \le 8 \\ 1, & \text{khi } x > 8 \end{cases}$$

(b) Xác suất sinh viên có thời gian tự học ít nhất 5 giờ/ngày:

$$P(X \ge 5) = 1 - P(X < 5) = 1 - F(5) = 1 - \frac{5^2}{64} = 0{,}609375$$

(c) Ta có:

$$P(X < 3) = F(3) = \frac{3^2}{64} = 0{,}140625$$

Gọi Y là số sinh viên có thời gian tự học ít hơn 3 giờ/ngày trong 50 sinh viên.
Suy ra: $Y \sim B(50\,;\,0{,}140625)$
Xác suất có không quá 10 sinh viên có thời gian tự học ít hơn 3 giờ/ngày:

$$P(Y \le 10) = \sum_{k=0}^{10} C_k^{50} \cdot 0{,}140625^k \cdot (1 - 0{,}140625)^{50-k} \approx 0{,}9155$$

(d) Hàm mật độ xác suất của X:

$$f(x) = \frac{dF}{dx} = \begin{cases} \dfrac{x}{32}, & \text{khi } x \in [0;8] \\ 0, & \text{khi } x \notin [0;8] \end{cases}$$

Ta có:

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x)\,dx = \int_0^8 x \cdot \frac{x}{32}\,dx = \frac{16}{3} \approx 5{,}3333$$

$$\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)\,dx = \int_0^8 x^2 \cdot \frac{x}{32}\,dx = 32$$

$$Var(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = 32 - \left(\frac{16}{3}\right)^2 = \frac{32}{9} \approx 3{,}5556$$

(e) Gọi T (giờ/ngày) là thời gian tự học trung bình của 500 sinh viên.

Ta có: $T = \frac{1}{n}\sum_{i=1}^{n} X_i$ với $n = 500,\ \mu_X = \frac{16}{3},\ \sigma_X^2 = \frac{32}{9}$

Theo định lý giới hạn trung tâm, ta có $Z = \dfrac{T - \mu_T}{\sigma_T} \sim \mathcal{N}(0\,;\,1)$ với:

- $\mu_T = \frac{1}{n} \cdot n\mu_X = \mu_X = \frac{16}{3}$
- $\sigma_T = \sqrt{\frac{1}{n^2} \cdot n\sigma_X^2} = \sqrt{\frac{\sigma_X^2}{n}} = \sqrt{\frac{\frac{32}{9}}{500}} \approx 0{,}0843$

Xác suất để thời gian tự học trung bình của 500 sinh viên lớn hơn 5,3 giờ/ngày:

$$\begin{aligned}
P(T > 5{,}3) &= 1 - P(T \le 5{,}3) \\
&= 1 - P\left(\frac{T - \frac{16}{3}}{0{,}0843} \le \frac{5{,}3 - \frac{16}{3}}{0{,}0843}\right) \\
&\approx 1 - P(Z \le -0{,}40) \\
&= 1 - \Phi(-0{,}4) \;=\; \Phi(0{,}4) \;=\; 0{,}6554
\end{aligned}$$

■

Câu 3 (4 điểm)

Một công ty sản xuất trò chơi điện tử thực hiện một khảo sát về thời gian người chơi hoàn thành hai cấp độ trong một trò chơi vừa phát hành. Họ ghi nhận rằng thời gian để hoàn thành cấp độ I là một biến ngẫu nhiên có phân phối chuẩn với trung bình 54 phút và độ lệch chuẩn 11,5 phút.

(a) Tính xác suất người chơi cần ít hơn 55 phút để hoàn thành cấp độ I của trò chơi này.

(b) Tìm ngưỡng thời gian t_0 sao cho 89,8% người chơi sẽ cần ít hơn t_0 phút để hoàn thành cấp độ I của trò chơi.

(c) Biết rằng có 74% người chơi hoàn thành cấp độ I, 72% người chơi đã hoàn thành cấp độ

I chọn tiếp tục chơi cấp độ II, và cũng có 19% người chơi đã không hoàn thành cấp độ I nhưng họ chọn trả phí để chơi tiếp cấp độ II. Trong các khách hàng của trò chơi điện tử này, tính tỷ lệ người chơi cấp độ II.

(d) Giả sử rằng thời gian để hoàn thành cấp độ II của trò chơi này là một biến ngẫu nhiên có phân phối chuẩn với trung bình 97 phút và độ lệch chuẩn 14,6 phút. Chọn ngẫu nhiên một khách hàng của trò chơi điện tử này, tính xác suất để người này đã chơi cấp độ II và cần ít hơn 99,1 phút để hoàn thành cấp độ II.

Lời giải.

(a) Gọi X (phút) là thời gian người chơi hoàn thành cấp độ I của trò chơi.
Theo đề: $X \sim \mathcal{N}\left(54\ ;\ 11{,}5^2\right)$
Xác suất người chơi cần ít hơn 55 phút để hoàn thành cấp độ I:

$$\begin{aligned} P(X < 55) &= P\left(\frac{X-54}{11{,}5} < \frac{55-54}{11{,}5}\right) \\ &\approx P(Z < 0{,}09) = \Phi(0{,}09) = 0{,}5359 \end{aligned}$$

(b) Ta có:

$$\begin{aligned} & P(X < t_0) &&= 0{,}898 \\ \Leftrightarrow\ & P\left(\frac{X-54}{11{,}5} < \frac{t_0-54}{11{,}5}\right) &&= 0{,}898 \\ \Leftrightarrow\ & \Phi\left(\frac{t_0-54}{11{,}5}\right) &&= \Phi(1{,}27) \\ \Leftrightarrow\ & \frac{t_0-54}{11{,}5} &&= 1{,}27 \\ \Leftrightarrow\ & t_0 &&= 68{,}605 \end{aligned}$$

Vậy 89,8% người chơi sẽ cần ít hơn 68,605 phút để hoàn thành cấp độ I của trò chơi.

(c) Phép thử: Chọn ngẫu nhiên một khách hàng của trò chơi điện tử này
Đặt các biến cố:

A: Người này đã hoàn thành cấp độ I

B: Người này đã chơi cấp độ II

Theo đề bài:

$P(A) = 0{,}74$ $\qquad$ $P(\overline{A}) = 1 - 0{,}74 = 0{,}26$

$P(B \mid A) = 0{,}72$ $\qquad$ $P(B \mid \overline{A}) = 0{,}19$

Vì $\{A, \overline{A}\}$ là một **hệ đầy đủ** nên theo công thức xác suất đầy đủ, ta có:

$$\begin{aligned} P(B) &= P(B \mid A) \cdot P(A) + P(B \mid \overline{A}) \cdot P(\overline{A}) \\ &= 0{,}72 \cdot 0{,}74 + 0{,}19 \cdot 0{,}26 \\ &= 0{,}5822 \end{aligned}$$

Vậy có 58,22% người chơi đã chơi cấp độ II.

(d) ...

■

A2 GIỮA KỲ II 2024-2025

Câu 1 (4 điểm)

Một phòng thí nghiệm công nghệ sinh học nhập khẩu các loại enzyme dùng trong phản ứng sinh học từ ba nhà cung cấp khác nhau. Trong đó, 30% số enzyme được nhập từ nhà cung cấp X, 45% từ nhà cung cấp Y, và phần còn lại từ nhà cung cấp Z. Tỷ lệ enzyme kém chất lượng (không đạt chuẩn phản ứng) từ ba nhà cung cấp X, Y, Z lần lượt là 4%, 8% và 5%. Chọn ngẫu nhiên một enzyme từ kho của phòng thí nghiệm này.

(a) Tính xác suất để enzyme được chọn là enzyme kém chất lượng và đến từ nhà cung cấp X.

(b) Tính xác suất để enzyme được chọn là enzyme kém chất lượng.

(c) Tính xác suất để enzyme được chọn là kém chất lượng hoặc đến từ nhà cung cấp Y.

(d) Biết rằng enzyme được chọn là kém chất lượng, tính xác suất nó đến từ nhà cung cấp Z.

Lời giải.

Phép thử: Chọn ngẫu nhiên một enzyme từ kho của phòng thí nghiệm

Đặt các biến cố:

A_1, A_2, A_3: Enzyme được chọn đến từ nhà cung cấp lần lượt là X, Y, Z

B: Enzyme được chọn là enzyme kém chất lượng

Theo đề bài:

$P(A_1) = 0{,}3$	$P(A_2) = 0{,}45$	$P(A_3) = 1 - 0{,}3 - 0{,}45 = 0{,}25$
$P(B \mid A_1) = 0{,}04$	$P(B \mid A_2) = 0{,}08$	$P(B \mid A_3) = 0{,}05$

(a) Xác suất để enzyme được chọn là enzyme kém chất lượng **và** đến từ nhà cung cấp X:

$$P(BA_1) = P(B \mid A_1) \cdot P(A_1) = 0{,}04 \cdot 0{,}3 = 0{,}012$$

(b) Vì $\{A_1, A_2, A_3\}$ là một **hệ đầy đủ** nên xác suất để enzyme được chọn là enzyme kém chất lượng:

$$\begin{aligned} P(B) &= P(B \mid A_1) \cdot P(A_1) + P(B \mid A_2) \cdot P(A_2) + P(B \mid A_3) \cdot P(A_3) \\ &= 0{,}04 \cdot 0{,}3 + 0{,}08 \cdot 0{,}45 + 0{,}05 \cdot 0{,}25 \\ &= 0{,}0605 \end{aligned}$$

(c) Ta có:

$$P(BA_2) = P(B \mid A_2) \cdot P(A_2) = 0{,}08 \cdot 0{,}45 = 0{,}036$$

Xác suất để enzyme được chọn là kém chất lượng **hoặc** đến từ nhà cung cấp Y:

$$P(B + A_2) = P(B) + P(A_2) - P(BA_2) = 0{,}0605 + 0{,}45 - 0{,}036 = 0{,}4745$$

(d) Xác suất enzyme đến từ nhà cung cấp Z **biết rằng** nó là enzyme kém chất lượng:

$$P(A_3 \mid B) = \frac{P(B \mid A_3) \cdot P(A_3)}{P(B)} = \frac{0{,}05 \cdot 0{,}25}{0{,}0605} \approx 0{,}2066$$

■

Câu 2 (6 điểm)

Một nhà máy sản xuất viên tẩy rửa công nghiệp cho biết khối lượng hoạt chất trong mỗi viên có phân phối chuẩn, với trung bình là 1,2 gram và độ lệch chuẩn là 0,03 gram. Bất kỳ viên nào chứa ít hơn 1,14 gram hoạt chất được xem là viên không đạt chuẩn.

(a) Chọn ngẫu nhiên một viên tẩy rửa. Tính xác suất viên đó không đạt chuẩn.

(b) Viên tẩy rửa được đóng gói 20 viên mỗi gói. Một gói có ít nhất một viên không đạt chuẩn được gọi là gói không đạt chuẩn.

- **(i)** Tính xác suất để một gói là không đạt chuẩn.
- **(ii)** Tính xác suất để trong một gói có không quá 5 viên không đạt chuẩn.

(c) Một thùng hàng chứa 500 gói tẩy rửa. Sử dụng xấp xỉ chuẩn có hiệu chỉnh hãy tính xác suất để có tối đa 200 gói không đạt chuẩn trong thùng hàng.

Lời giải.

(a) Gọi X (gram) là khối lượng hoạt chất trong mỗi viên tẩy rửa.
Theo đề: $X \sim \mathcal{N}\left(1{,}2\ ;\ 0{,}03^2\right)$
Xác suất viên tẩy rửa được chọn ngẫu nhiên là không đạt chuẩn:

$$\begin{aligned} P(X < 1{,}14) &= P\left(\frac{X - 1{,}2}{0{,}03} < \frac{1{,}14 - 1{,}2}{0{,}03}\right) \\ &= P(Z < -2) \\ &= 1 - \Phi(2) = 1 - 0{,}9772 = 0{,}0228 \end{aligned}$$

(b) Gọi Y là số viên tẩy rửa không đạt chuẩn trong mỗi gói 20 viên.
Ta có: $Y \sim B\,(20\ ;\ 0{,}0228)$

(i) Xác suất để một gói là không đạt chuẩn:

$$P(Y \geq 1) = 1 - P(Y < 1) = 1 - C_{20}^{0} \cdot 0{,}0228^{0} \cdot (1 - 0{,}0228)^{20} \approx 0{,}3695$$

(ii) Xác suất để trong một gói có không quá 5 viên không đạt chuẩn:

$$P(Y \leq 5) = \sum_{i=0}^{5} C_{20}^{i} \cdot 0{,}0228^{i} \cdot (1 - 0{,}0228)^{20-i} \approx 1$$

(c) Gọi S là số gói tẩy rửa không đạt chuẩn trong mỗi thùng hàng 500 gói.

Ta có: $S \sim B\,(500\,;\ 0{,}3695)$

Xấp xỉ phân phối nhị thức bằng phân phối chuẩn, ta được $Z = \dfrac{S - \mu_S}{\sigma_S} \sim \mathcal{N}\,(0\,;\ 1)$ với:

- $\mu_S = 500 \cdot 0{,}3695 = 184{,}75$
- $\sigma_S = \sqrt{500 \cdot 0{,}3695 \cdot (1 - 0{,}3695)} \approx 10{,}7928$

Xác suất để có tối đa 200 gói không đạt chuẩn trong thùng hàng:

$$P(S \leq 200) \approx \Phi\left(\frac{200 - 184{,}75 + \mathbf{0{,}5}}{10{,}7928}\right) \approx \Phi(1{,}48) = 0{,}9306$$

■

A3 GIỮA KỲ II 2024-2025 (CNTT-23CLC01)

Câu 1 (3 điểm)

Trong một trận đá bóng, có 3 cầu thủ A, B và C đang ngồi chờ dự bị. Khả năng 3 cầu thủ này được chọn để vào sân lần lượt là 40%, 35%, 25%. Với mỗi cầu thủ, xác suất cầu thủ này ghi bàn trong 15 phút tiếp theo lần lượt là 15%, 10% và 8%.

(a) Tính xác suất một trong 3 cầu thủ sẽ ghi bàn trong 15 phút tiếp theo.

(b) Nếu biết một trong 3 cầu thủ đã ghi bàn trong 15 phút tiếp theo, xác suất cầu thủ nào ghi bàn là cao nhất?

(c) Biết rằng không có bàn thắng nào được ghi trong 15 phút sau khi thay người, xác suất cầu thủ C là người được thay vào sân là bao nhiêu?

Lời giải.

Phép thử: Chọn ngẫu nhiên một cầu thủ dự bị vào sân trong trận đá bóng

Đặt các biến cố:

A, B, C: Cầu thủ được chọn lần lượt là A, B, C

D: Cầu thủ được chọn ghi bàn trong 15 phút tiếp theo

Theo đề bài:

$$P(A) = 0{,}4 \qquad P(B) = 0{,}35 \qquad P(C) = 0{,}25$$
$$P(D \mid A) = 0{,}15 \qquad P(D \mid B) = 0{,}1 \qquad P(D \mid C) = 0{,}08$$

(a) Vì $\{A, B, C\}$ là một **hệ đầy đủ** nên xác suất để một trong ba cầu thủ ghi bàn trong 15 phút tiếp theo là:

$$\begin{aligned} P(D) &= P(D \mid A) \cdot P(A) + P(D \mid B) \cdot P(B) + P(D \mid C) \cdot P(C) \\ &= 0{,}15 \cdot 0{,}4 + 0{,}1 \cdot 0{,}35 + 0{,}08 \cdot 0{,}25 \\ &= 0{,}115 \end{aligned}$$

(b) Xác suất mỗi cầu thủ ghi bàn **nếu biết** một trong ba cầu thủ đã ghi bàn trong 15 phút tiếp theo:

$$P(A \mid D) = \frac{P(D \mid A) \cdot P(A)}{P(D)} = \frac{0{,}15 \cdot 0{,}4}{0{,}115} \approx 0{,}5217$$

$$P(B \mid D) = \frac{P(D \mid B) \cdot P(B)}{P(D)} = \frac{0{,}1 \cdot 0{,}35}{0{,}115} \approx 0{,}3043$$

$$P(C \mid D) = \frac{P(D \mid C) \cdot P(C)}{P(D)} = \frac{0{,}08 \cdot 0{,}25}{0{,}115} \approx 0{,}1739$$

Vậy xác suất cầu thủ A ghi bàn là cao nhất nếu biết một trong ba cầu thủ đã ghi bàn trong 15 phút tiếp theo.

(c) Xác suất cầu thủ C là người được thay vào sân **biết rằng** không có bàn thắng nào được ghi trong 15 phút sau khi thay người:

$$P(C \mid \overline{D}) = \frac{P(\overline{D} \mid C) \cdot P(C)}{P(\overline{D})} = \frac{\left[1 - P(D \mid C)\right] \cdot P(C)}{1 - P(D)} = \frac{(1 - 0{,}08) \cdot 0{,}25}{1 - 0{,}115} \approx 0{,}2599$$

■

Câu 2 (3 điểm)

Tuổi thọ của một bóng đèn (đơn vị: giờ) là biến ngẫu nhiên $Y = 100 \cdot X$ với X là biến ngẫu nhiên có hàm mật độ được cho bởi:

$$f(x) = \begin{cases} \dfrac{C}{x^4}, & \text{nếu } x \geq 1 \\ 0, & \text{ngược lại} \end{cases}$$

với C là hằng số chưa biết.

(a) Tìm hằng số C.

(b) Tính $\mathbb{E}(Y), Var(Y)$.

(c) Xét 2 biến cố sau:

- A : Tuổi thọ của bóng đèn ít nhất 150 giờ.
- B : Tuổi thọ của bóng đèn không quá 200 giờ.

Hai biến cố này có độc lập hay không? Vì sao?

Lời giải.

(a) Ta có:

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_{1}^{+\infty} \frac{C}{x^4}\,dx = \lim_{t \to +\infty} \int_{1}^{t} \frac{C}{x^4}\,dx = \lim_{t \to +\infty} \left[\frac{C}{3}\left(1 - \frac{1}{t^3}\right)\right] = \frac{C}{3}$$

Để $f(x)$ là hàm mật độ xác suất thì:

(i) $\displaystyle\int_{-\infty}^{+\infty} f(x)\,dx = 1 \quad \Leftrightarrow \quad \frac{C}{3} = 1 \quad \Leftrightarrow \quad C = 3$ (1)

(ii) $f(x) \geq 0, \quad \forall x \in \mathbb{R} \quad \Leftrightarrow \quad \dfrac{C}{x^4} \geq 0, \quad \forall x \geq 1 \quad \Leftrightarrow \quad C \geq 0$ (2)

Từ (1) và (2), suy ra $C = 3$.

(b) Ta có:

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf(x)\,dx = \int_{1}^{+\infty} x \cdot \frac{3}{x^4}\,dx = \lim_{t\to+\infty}\int_{1}^{t} \frac{3}{x^3}\,dx = \lim_{t\to+\infty}\left[\frac{3}{2}\left(1-\frac{1}{t^2}\right)\right] = 1{,}5$$

$$\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)\,dx = \int_{1}^{+\infty} x^2 \cdot \frac{3}{x^4}\,dx = \lim_{t\to+\infty}\int_{1}^{t} \frac{3}{x^2}\,dx = \lim_{t\to+\infty}\left[3\left(1-\frac{1}{t}\right)\right] = 3$$

$$Var(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = 3 - 1{,}5^2 = 0{,}75$$

Suy ra:

$$\mathbb{E}(Y) = \mathbb{E}(100\cdot X) = 100\cdot\mathbb{E}(X) = 100\cdot 1{,}5 = 150$$
$$Var(Y) = Var(100\cdot X) = 100^2\cdot Var(X) = 100^2\cdot 0{,}75 = 7500$$

(c) Ta có:

- $Y \geq 150 \quad\Leftrightarrow\quad 100\cdot X \geq 150 \quad\Leftrightarrow\quad X \geq 1{,}5$. Do đó:

$$P(A) = P(Y\geq 150) = P(X\geq 1{,}5) = \int_{1,5}^{+\infty}\frac{3}{x^4}\,dx = \lim_{t\to+\infty}\int_{1,5}^{t}\frac{3}{x^4}\,dx = \lim_{t\to+\infty}\left(\frac{1}{1{,}5^3}-\frac{1}{t^3}\right) = \frac{8}{27}$$

- $Y \leq 200 \quad\Leftrightarrow\quad 100\cdot X \leq 200 \quad\Leftrightarrow\quad X \leq 2$. Do đó:

$$P(B) = P(Y\leq 200) = P(X\leq 2) = \int_{1}^{2}\frac{3}{x^4}\,dx = \frac{7}{8}$$

- Mặt khác:

$$P(AB) = P(150\leq X\leq 200) = P(1{,}5\leq X\leq 2) = \int_{1,5}^{2}\frac{3}{x^4}\,dx = \frac{37}{216}$$

Vì $P(A)\cdot P(B) = \frac{8}{27}\cdot\frac{7}{8} = \frac{7}{27} \neq P(AB) = \frac{37}{216}$ nên A, B không phải là hai biến cố độc lập.

■

Câu 3 (4 điểm)

Trọng lượng của một túi thức ăn cho gia súc có phân phối chuẩn với trung bình là 1,5 kg và độ lệch chuẩn 250 g. Một nông trại mua 500 túi thức ăn cho đàn gia súc của mình.

(a) Tính xác suất một túi thức ăn có trọng lượng lớn hơn 2 kg.

(b) Mỗi tuần, đàn gia súc cần 175 kg thức ăn. Tính xác suất nông trại còn dư ít nhất 55 kg thức ăn vào cuối tháng. Giả sử một tháng có 4 tuần, và lượng thức ăn ở mỗi tuần là như nhau.

(c) Trước khi sử dụng, chủ nông trại cân túi thức ăn để đảm bảo đủ lượng thức ăn cho gia súc. Nếu trọng lượng túi thức ăn dưới 1 kg thì chủ nông trại sẽ phải yêu cầu đổi trả túi thức ăn mới. Tính xác suất chủ nông trại phải yêu cầu đổi ít nhất 10 túi thức ăn.

Lời giải.

(a) Gọi X (kg) là khối lượng của một túi thức ăn được chọn ngẫu nhiên.
Theo đề: $X \sim \mathcal{N}\left(1{,}5\ ;\ 0{,}25^2\right)$
Xác suất một túi thức ăn có trọng lượng lớn hơn 2 kg:

$$\begin{aligned} P(X > 2) &= 1 - P(X \le 2) \\ &= 1 - P\left(\frac{X - 1{,}5}{0{,}25} \le \frac{2 - 1{,}5}{0{,}25}\right) \\ &= 1 - \Phi(2) = 1 - 0{,}9772 = 0{,}0228 \end{aligned}$$

(b) Gọi Y (kg) là khối lượng thức ăn còn dư vào cuối tháng của nông trại.
Ta có: $Y = 500 \cdot X - 175 \cdot 4 = 500 \cdot X - 700$
Vì $X \sim \mathcal{N}\left(\mu_X\ ;\ \sigma_X^2\right)$ nên $Y \sim \mathcal{N}\left(\mu_Y\ ;\ \sigma_Y^2\right)$ với:

- $\mu_Y = 500 \cdot \mu_X - 700 = 500 \cdot 1{,}5 - 700 = 50$
- $\sigma_Y = \sqrt{500^2 \cdot \sigma_X^2} = \sqrt{500^2 \cdot 0{,}25^2} = 125$

Xác suất nông trại còn dư ít nhất 55 kg thức ăn vào cuối tháng:

$$P(Y \ge 55) = 1 - P(Y < 55) = 1 - P\left(\frac{Y - 50}{125} < \frac{55 - 50}{125}\right) = 1 - \Phi(0{,}04) = 1 - 0{,}5160 = 0{,}484$$

(c) Xác suất một túi thức ăn bị yêu cầu đổi (có trọng lượng dưới 1 kg):

$$P(X < 1) = P\left(\frac{X - 1{,}5}{0{,}25} < \frac{1 - 1{,}5}{0{,}25}\right) = \Phi(-2) = 1 - \Phi(2) = 1 - 0{,}9772 = 0{,}0228$$

Gọi T là số túi thức ăn bị yêu cầu đổi trong 500 túi mà nông trại đã mua.
Ta có: $T \sim B\,(500\ ;\ 0{,}0228)$
Xác suất để chủ nông trại phải yêu cầu đổi ít nhất 10 túi thức ăn:

$$P(T \ge 10) = 1 - P(T < 10) = 1 - \sum_{i=0}^{9} C_{500}^{i} \cdot 0{,}0228^{i} \cdot (1 - 0{,}0228)^{500-i} \approx 1 - 0{,}2959 = 0{,}7041$$

■

A4 GIỮA KỲ I 2024-2025 (CA 1)

Câu 1 (4 điểm)

Một chương trình máy tính bao gồm hai khối mã được viết một cách độc lập nhau bởi hai lập trình viên. Khối mã đầu tiên có lỗi với xác suất 0,2. Khối mã thứ hai có lỗi với xác suất 0,3. Chương trình sẽ hoạt động nếu cả hai khối mã đều không có lỗi. Ngược lại, nếu có một khối mã có lỗi thì chương trình sẽ báo lỗi.

(a) Tính xác suất để cả hai khối mã đều có lỗi.

(b) Tính xác suất để chương trình báo lỗi.

(c) Biết rằng chương trình báo lỗi, hỏi xác suất để có lỗi ở cả hai khối mã là bao nhiêu?

(d) Biết rằng chương trình báo lỗi, hỏi xác suất để khối mã đầu tiên không có lỗi là bao nhiêu?

Lời giải.

Phép thử: Chọn ngẫu nhiên một chương trình máy tính

Đặt các biến cố:

A: Khối mã thứ nhất bị lỗi

B: Khối mã thứ hai bị lỗi

Theo đề bài: $P(A) = 0{,}2 \quad ; \quad P(B) = 0{,}3$

(a) Vì A, B **độc lập** với nhau nên xác suất để cả hai khối mã đều có lỗi:

$$P(AB) = P(A) \cdot P(B) = 0{,}2 \cdot 0{,}3 = 0{,}06$$

(b) Xác suất để chương trình báo lỗi:

$$P(A+B) = P(A) + P(B) - P(AB) = 0{,}2 + 0{,}3 - 0{,}06 = 0{,}44$$

(c) Xác suất để có lỗi ở cả hai khối mã **biết rằng** chương trình báo lỗi:

$$P(AB \mid A+B) = \frac{P((AB) \cap (A+B))}{P(A+B)} = \frac{P(AB)}{P(A+B)} = \frac{0{,}06}{0{,}44} \approx 0{,}1364$$

(d) Vì A độc lập với B nên $\overline{A}$ cũng độc lập với B, do đó:

$$P(\overline{A}B) = P(\overline{A}) \cdot P(B) = (1 - 0{,}2) \cdot 0{,}3 = 0{,}24$$

Xác suất để khối mã đầu tiên không có lỗi **biết rằng** chương trình báo lỗi:

$$P(\overline{A} \mid A+B) = \frac{P(\overline{A} \cap (A+B))}{P(A+B)} = \frac{P(\overline{A}B)}{P(A+B)} = \frac{0{,}24}{0{,}44} \approx 0{,}5455$$

Câu 2 (3 điểm)

Tuổi thọ (đơn vị: năm) của một loại linh kiện điện tử là một biến ngẫu nhiên với hàm mật độ xác suất như sau:

$$f(x) = \begin{cases} C - \dfrac{x}{50} & \text{nếu } 0 < x < 10 \\ 0 & \text{nếu khác} \end{cases}$$

(a) Tìm hằng số C.

(b) Tính xác suất để tuổi thọ của loại linh kiện điện tử này không quá 5 năm.

(c) Tính kỳ vọng tuổi thọ của loại linh kiện điện tử này.

Lời giải.

(a) Ta có:

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_0^{10} \left(C - \frac{x}{50}\right) dx = 10C - 1$$

Để $f(x)$ là hàm mật độ xác suất thì:

(i) $$\int_{-\infty}^{+\infty} f(x)\,dx = 1 \quad \Leftrightarrow \quad 10C - 1 = 1 \quad \Leftrightarrow \quad C = \frac{1}{5} \tag{1}$$

(ii) $$f(x) \geq 0, \quad \forall x \in \mathbb{R} \quad \Leftrightarrow \quad C - \frac{x}{50} \geq 0, \quad \forall x \in (0{,}10) \quad \Leftrightarrow \quad C \geq \frac{1}{5} \tag{2}$$

Từ (1) và (2), suy ra $C = \dfrac{1}{5}$.

(b) Gọi X (năm) là tuổi thọ của loại linh kiện điện tử này.

Xác suất để tuổi thọ của loại linh kiện điện tử này không quá 5 năm:

$$P(X \leq 5) = \int_{-\infty}^{5} f(x)\,dx = \int_0^5 \left(\frac{1}{5} - \frac{x}{50}\right) dx = 0{,}75$$

(c) Kỳ vọng tuổi thọ của loại linh kiện điện tử này:

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x)\,dx = \int_0^{10} x\left(\frac{1}{5} - \frac{x}{50}\right) dx = \frac{10}{3} \approx 3{,}3333 \text{ (năm)}$$

Câu 3 (3 điểm)

Một nhà phân tích thị trường cho rằng giá màn hình vi tính tinh thể lỏng (LCD) tuân theo phân phối chuẩn với kỳ vọng 130$ và độ lệch chuẩn 32$.

(a) Tính xác suất một màn hình LCD được chọn ngẫu nhiên có giá thấp hơn 150$.

(b) Chọn ngẫu nhiên 10 màn hình LCD.

- **(i)** Tính xác suất có tối đa 3 màn hình LCD có giá thấp hơn 150$.
- **(ii)** Tính xác suất giá trung bình của 10 màn hình LCD được chọn thấp hơn 150$.

(c) Chọn ngẫu nhiên 500 màn hình LCD, tính xác suất có tối thiểu 200 màn hình LCD có giá thấp hơn 150$.

Lời giải.

(a) Gọi X ($) là giá của một màn hình LCD được chọn ngẫu nhiên.
Theo đề: $X \sim \mathcal{N}\left(130\ ;\ 32^2\right)$
Xác suất một màn hình LCD được chọn ngẫu nhiên có giá thấp hơn 150$:

$$P(X < 150) = P\left(\frac{X-130}{32} < \frac{150-130}{32}\right) = P(Z < 0{,}625) \approx \Phi(0{,}63) = 0{,}7357$$

(b) Gọi Y là số màn hình LCD có giá thấp hơn 150$ trong 10 màn hình LCD được chọn ngẫu nhiên.
Ta có: $Y \sim B\,(10\ ;\ 0{,}7357)$

(i) Xác suất có tối đa 3 màn hình LCD có giá thấp hơn 150$:

$$P(Y \leq 3) = \sum_{i=0}^{3} C_{10}^{i} \cdot 0{,}7357^{i} \cdot (1-0{,}7357)^{10-i} \approx 0{,}0049$$

(ii) Gọi S ($) là giá tiền trung bình của 10 màn hình LCD được chọn.
Ta có: $S = \dfrac{1}{n}(X_1 + X_2 + \ldots + X_n)$, trong đó $X_i \sim \mathcal{N}\left(\mu\ ;\ \sigma^2\right)$, $\quad n = 10, \mu = 130, \sigma = 32$
Theo định lý giới hạn trung tâm, ta có $Z = \dfrac{S - \mu_S}{\sigma_S} \sim \mathcal{N}\,(0\ ;\ 1)$ với:

- $\mu_S = \dfrac{1}{n} \cdot n\mu = \mu = 130$
- $\sigma_S = \sqrt{\dfrac{1}{n^2} \cdot n\sigma^2} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{32}{\sqrt{10}} \approx 10{,}1193$

Xác suất giá trung bình của 10 màn hình LCD được chọn thấp hơn 150$:

$$P(S < 150) = P\left(\frac{S-130}{10{,}1193} < \frac{150-130}{10{,}1193}\right) \approx \Phi(1{,}98) = 0{,}9761$$

(c) Gọi T là số màn hình LCD có giá thấp hơn 150\$ trong 500 màn hình được chọn ngẫu nhiên.
Ta có: $T \sim B\,(500\ ;\ 0{,}7357)$

Xấp xỉ phân phối nhị thức bằng phân phối chuẩn, ta được $Z = \dfrac{T - \mu_T}{\sigma_T} \sim \mathcal{N}\,(0\ ;\ 1)$ với:

- $\mu_T = 500 \cdot 0{,}7357 = 367{,}85$
- $\sigma_T = \sqrt{500 \cdot 0{,}7357 \cdot (1 - 0{,}7357)} \approx 9{,}8602$

Xác suất có tối thiểu 200 màn hình LCD có giá thấp hơn 150\$:

$$\begin{aligned} P(T \geq 200) &= 1 - P(T < 200) \\ &\approx 1 - \Phi\left(\frac{200 - 367{,}85 - \mathbf{0{,}5}}{9{,}8602}\right) \\ &\approx 1 - (1 - \Phi(17{,}07)) \approx 1 - (1 - 1) = 1 \end{aligned}$$

■

A5 GIỮA KỲ I 2024-2025 (CA 2)

Câu 1 (4 điểm)

Một công ty lắp ráp máy tính nhận 24% số linh kiện từ nhà cung cấp A, 36% số linh kiện từ nhà cung cấp B, và còn lại 40% số linh kiện từ nhà cung cấp C. Phần trăm số linh kiện bị lỗi từ ba nhà cung cấp A, B, C lần lượt là 5%, 10%, và 6%. Chọn ngẫu nhiên một linh kiện.

(a) Tính xác suất để linh kiện bị lỗi và đến từ nhà cung cấp C.

(b) Tính xác suất để linh kiện bị lỗi.

(c) Tính xác suất để linh kiện bị lỗi hoặc đến từ nhà cung cấp C.

(d) Biết rằng linh kiện bị lỗi, hỏi xác suất để linh kiện này đến từ nhà cung cấp C là bao nhiêu?

Lời giải.

Phép thử: Chọn ngẫu nhiên một linh kiện

Đặt các biến cố:

A, B, C: Linh kiện được chọn đến từ nhà cung cấp lần lượt là A, B, C

D: Linh kiện được chọn bị lỗi

Theo đề bài:

$$P(A) = 0{,}24 \qquad P(B) = 0{,}36 \qquad P(C) = 0{,}4$$
$$P(D \mid A) = 0{,}05 \qquad P(D \mid B) = 0{,}1 \qquad P(D \mid C) = 0{,}06$$

(a) Xác suất để linh kiện bị lỗi **và** đến từ nhà cung cấp C:

$$P(DC) = P(D \mid C) \cdot P(C) = 0{,}06 \cdot 0{,}4 = 0{,}024$$

(b) Vì $\{A, B, C\}$ là một **hệ đầy đủ** nên xác suất để linh kiên được chọn bị lỗi:

$$\begin{aligned} P(D) &= P(D \mid A) \cdot P(A) + P(D \mid B) \cdot P(B) + P(D \mid C) \cdot P(C) \\ &= 0{,}05 \cdot 0{,}24 + 0{,}1 \cdot 0{,}36 + 0{,}06 \cdot 0{,}4 \\ &= 0{,}072 \end{aligned}$$

(c) Xác suất để linh kiện bị lỗi **hoặc** đến từ nhà cung cấp C:

$$P(D + C) = P(D) + P(C) - P(DC) = 0{,}072 + 0{,}4 - 0{,}024 = 0{,}448$$

(d) Xác suất linh kiện đến từ nhà cung cấp C **biết rằng** linh kiện bị lỗi:

$$P(C \mid D) = \frac{P(CD)}{P(D)} = \frac{0{,}024}{0{,}072} \approx 0{,}3333$$

■

Câu 2 (3 điểm)

Thời gian (đơn vị: phút) cần để một hệ thống máy tính khởi động lại là một biến ngẫu nhiên liên tục có hàm mật độ xác suất là:

$$f(x) = \begin{cases} C(10-x)^2, & 0 < x < 10 \\ 0, & \text{nơi khác} \end{cases}$$

(a) Tìm hằng số C.

(b) Tính xác suất để thời gian cần để khởi động lại hệ thống nằm giữa một và hai phút.

(c) Tính kỳ vọng thời gian cần để khởi động lại hệ thống.

Lời giải.

(a) Ta có:

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_0^{10} C(10-x)^2\,dx = \frac{1000}{3}C$$

Để $f(x)$ là hàm mật độ xác suất thì:

(i) $\displaystyle\int_{-\infty}^{+\infty} f(x)\,dx = 1 \quad \Leftrightarrow \quad \frac{1000}{3}C = 1 \quad \Leftrightarrow \quad C = \frac{3}{1000}$ (1)

(ii) $f(x) \geq 0, \quad \forall x \in \mathbb{R} \quad \Leftrightarrow \quad C(10-x)^2 \geq 0, \quad \forall x \in (0,10) \quad \Leftrightarrow \quad C \geq 0$ (2)

Từ (1) và (2), suy ra $C = \dfrac{3}{1000}$.

(b) Gọi X (phút) là thời gian cần thiết để khởi động lại hệ thống.
Xác suất để thời gian cần để khởi động lại hệ thống nằm giữa một và hai phút:

$$P(1 < X < 2) = \int_1^2 f(x)\,dx = \int_1^2 \frac{3}{1000}(10-x)^2\,dx = 0{,}217$$

(c) Kỳ vọng thời gian cần để khởi động lại hệ thống:

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x)\,dx = \int_0^{10} x \cdot \frac{3}{1000}(10-x)^2\,dx = 2{,}5 \text{ (phút)}$$

■

Câu 3 (3 điểm)

Một nhà sản xuất bút huỳnh quang cho biết lượng mực phát quang chứa trong mỗi cây bút có phân phối chuẩn với trung bình là 1,2 gram và độ lệch chuẩn là 0,03 gram. Bất kỳ bút huỳnh quang nào có ít hơn 1,14 gram mực phát quang được gọi là "bút không đạt chuẩn".

(a) Chọn ngẫu nhiên một cây bút. Tính xác suất được cây bút không đạt chuẩn.

(b) Người ta đóng gói 20 cây bút mỗi hộp. Hộp có chứa bút không đạt chuẩn được gọi là "hộp không đạt chuẩn".

- **(i)** Tính xác suất để hộp không đạt chuẩn.
- **(ii)** Tính xác suất để trong hộp có không quá năm bút không đạt chuẩn.

(c) Người ta đóng gói 500 hộp bút vào một thùng hàng. Tính xác suất để trong thùng hàng có tối đa 200 hộp bút không đạt chuẩn.

Lời giải.

(a) Gọi X (gram) là khối lượng mực phát quang trong một cây bút được chọn ngẫu nhiên.
Theo đề: $X \sim \mathcal{N}\left(1{,}2\ ;\ 0{,}03^2\right)$
Xác suất để cây bút được chọn ngẫu nhiên không đạt chuẩn:

$$\begin{aligned} P(X < 1{,}14) &= P\left(\frac{X-1{,}2}{0{,}03} < \frac{1{,}14-1{,}2}{0{,}03}\right) \\ &= P(Z < -2) \\ &= 1 - \Phi(2) = 1 - 0{,}9772 = 0{,}0228 \end{aligned}$$

(b) Gọi Y là số cây bút không đạt chuẩn trong một hộp được chọn ngẫu nhiên gồm 20 bút.
Ta có: $Y \sim B\,(20\ ;\ 0{,}0228)$

(i) Xác suất để hộp được chọn không đạt chuẩn:

$$P(Y \geq 1) = 1 - P(Y < 1) = 1 - C_{20}^{0} \cdot 0{,}0228^{0} \cdot (1-0{,}0228)^{20} \approx 0{,}3695$$

(ii) Xác suất để trong hộp có không quá 5 cây bút không đạt chuẩn:

$$P(Y \leq 5) = \sum_{i=0}^{5} C_{20}^{i} \cdot 0{,}0228^{i} \cdot (1-0{,}0228)^{20-i} \approx 1$$

(c) Gọi S là số hộp bút không đạt chuẩn trong thùng hàng 500 hộp.
Ta có: $S \sim B\,(500\ ;\ 0{,}3695)$
Xấp xỉ phân phối nhị thức bằng phân phối chuẩn, ta được $Z = \dfrac{S-\mu_S}{\sigma_S} \sim \mathcal{N}\,(0\ ;\ 1)$ với:

- $\mu_S = 500 \cdot 0{,}3695 = 184{,}75$
- $\sigma_S = \sqrt{500 \cdot 0{,}3695 \cdot (1 - 0{,}3695)} \approx 10{,}7928$

Xác suất để có tối đa 200 hộp bút không đạt chuẩn trong thùng hàng:

$$P(S \leq 200) \approx \Phi\left(\frac{200 - 184{,}75 + 0{,}5}{10{,}7928}\right) \approx \Phi(1{,}46) = 0{,}9279$$

■

A6 GIỮA KỲ II 2023-2024

Câu 1 (3 điểm)

Một công ty bảo hiểm thực hiện khảo sát trên diện rộng về tình trạng sức khỏe và việc mua bảo hiểm sức khỏe của một cộng đồng dân cư. Kết quả được tổng hợp ở bảng dưới đây:

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="5">Tình trạng sức khỏe</td></tr>
<tr><td>Hoàn hảo</td><td>Rất tốt</td><td>Tốt</td><td>Tương đối</td><td>Yếu</td></tr>
<tr><td rowspan="2">Bảo hiểm sức khỏe</td><td>Không có</td><td>459</td><td>727</td><td>854</td><td>385</td><td>99</td></tr>
<tr><td>Có</td><td>4198</td><td>6245</td><td>4821</td><td>1634</td><td>578</td></tr>
<tr><td colspan="2">Tổng cộng</td><td>4657</td><td>6972</td><td>5675</td><td>2019</td><td>677</td></tr>
</table>

Dựa vào bảng tổng hợp trên, hãy tính xác suất trong các trường hợp sau:

(a) Gặp một người có sức khỏe yếu?

(b) Gặp một người không mua bảo hiểm sức khỏe?

(c) Gặp một người có sức khỏe yếu khi biết người đó không mua bảo hiểm sức khỏe?

Lời giải.

Phép thử: Chọn ngẫu nhiên một người được khảo sát trong cộng đồng dân cư

Đặt các biến cố:

A: Người đó có sức khỏe yếu

B: Người đó không mua bảo hiểm sức khỏe

Số người được khảo sát: $n(\Omega) = 4657 + 6972 + 5675 + 2019 + 677 = 20000$

(a) Xác suất gặp một người có sức khỏe yếu:

$$P(A) = \frac{n(A)}{n(\Omega)} = \frac{677}{20000} = 0{,}03385$$

(b) Số người không mua bảo hiểm sức khỏe:

$$n(B) = 459 + 727 + 854 + 385 + 99 = 2524$$

Xác suất gặp một người không mua bảo hiểm sức khỏe:

$$P(B) = \frac{n(B)}{n(\Omega)} = \frac{2524}{20000} = 0{,}1262$$

(c) Ta có:

$$P(AB) = \frac{n(AB)}{n(\Omega)} = \frac{99}{20000} = 0{,}00495$$

Xác suất gặp một người có sức khỏe yếu **khi biết** người đó không mua bảo hiểm sức khỏe:

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{0{,}00495}{0{,}1262} \approx 0{,}0392$$

■

Câu 2 (3 điểm)

Cho X là một biến ngẫu nhiên liên tục, có hàm mật độ xác suất như sau:

$$f(x) = \begin{cases} C\dfrac{1}{x^2} & \text{nếu } 1 < x < 70 \\ 0 & \text{nếu } x \le 1 \text{ hoặc } x \ge 70 \end{cases}$$

(a) Tìm hằng số C.

(b) Tính xác suất $P(X > 35)$.

(c) Đặt $Y = X^2 + 1$. Hãy tính trung bình và phương sai của Y.

Lời giải.

(a) Ta có:

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_1^{70} C\frac{1}{x^2}\,dx = \frac{69}{70}C$$

Để $f(x)$ là hàm mật độ xác suất thì:

(i) $$\int_{-\infty}^{+\infty} f(x)\,dx = 1 \quad \Leftrightarrow \quad \frac{69}{70}C = 1 \quad \Leftrightarrow \quad C = \frac{70}{69} \qquad (1)$$

(ii) $$f(x) \ge 0, \quad \forall x \in \mathbb{R} \quad \Leftrightarrow \quad C\frac{1}{x^2} \ge 0, \quad \forall x \in (1{,}70) \quad \Leftrightarrow \quad C \ge 0 \qquad (2)$$

Từ (1) và (2), suy ra $C = \dfrac{70}{69}$.

(b) Ta có:

$$P(X > 35) = \int_{35}^{+\infty} f(x)\,dx = \int_{35}^{70} \frac{70}{69}\frac{1}{x^2}\,dx = \frac{1}{69} \approx 0{,}0145$$

(c) Ta có:

$$\mathbb{E}(Y) = \mathbb{E}(X^2+1) = \int_{-\infty}^{+\infty} (x^2+1)f(x)\,dx = \int_1^{70} (x^2+1)\frac{70}{69}\frac{1}{x^2}\,dx = 71$$

$$\mathbb{E}(Y^2) = \mathbb{E}\left[(X^2+1)^2\right] = \int_{-\infty}^{+\infty} (x^2+1)^2 f(x)\, dx = \int_1^{70} (x^2+1)^2 \frac{70}{69}\frac{1}{x^2}\, dx \approx 116131$$

$$Var(Y) = \mathbb{E}(Y^2) - [\mathbb{E}(Y)]^2 = 116131 - 71^2 = 111090$$

■

Câu 3 (2 điểm)

Giả sử trọng lượng hành lý ký gửi của hành khách đi máy bay tuân theo phân phối chuẩn với trung bình 20,41 kg và độ lệch chuẩn 1,45 kg. Hầu hết các hãng hàng không đều tính phí hành lý nặng hơn 22,67 kg.

(a) Xác định bao nhiêu phần trăm hành khách đi máy bay phải chịu khoản phí này.

(b) Giả sử rằng trong một chuyến bay có 200 hành khách. Cân nặng hành lý của từng hành khách là độc lập nhau. Gọi Y là số lượng hành khách bị phạt vì có hành lý nặng hơn 22,67 kg. Hãy xác định phân phối xác suất của Y, và tính xác suất cho việc có từ 10 tới 20 hành khách bị phạt, tức là $P(10 \le Y \le 20)$.

Lời giải.

(a) Gọi X (kg) là khối lượng hành lý ký gửi của hành khách đi máy bay.
Theo đề: $X \sim \mathcal{N}\left(20{,}41\ ;\ 1{,}45^2\right)$
Xác suất hành khách có hành lý ký gửi nặng hơn 22,67 kg:

$$\begin{aligned} P(X > 22{,}67) &= 1 - P(X \le 22{,}67) \\ &= 1 - P\left(\frac{X - 20{,}41}{1{,}45} \le \frac{22{,}67 - 20{,}41}{1{,}45}\right) \\ &\approx 1 - \Phi(1{,}56) = 1 - 0{,}9406 = 0{,}0594 \end{aligned}$$

Vậy có khoảng 5,94% hành khách đi máy bay chịu khoản phí hành lý nặng hơn 22,67 kg.

(b) Theo đề, Y là số lượng hành khách bị phạt vì có hành lý nặng hơn 22,67 kg trong 200 hành khách trên chuyến bay.
Phân phối xác suất của Y: $Y \sim B\,(200\ ;\ 0{,}0594)$
Xác suất để có từ 10 tới 20 hành khách bị phạt:

$$P(10 \le Y \le 20) = \sum_{i=10}^{20} C_{200}^{i} \cdot 0{,}0594^{i} \cdot (1 - 0{,}0594)^{200-i} \approx 0{,}7465$$

■

Câu 4 (2 điểm)

Một quán cà phê phục vụ trung bình 75 khách hàng mỗi giờ vào giờ cao điểm buổi sáng.

(a) Phân phối xác suất nào là thích hợp nhất để tính xác suất mà một số lượng khách hàng nhất định đến trong vòng một giờ vào thời điểm này trong ngày?

(b) Giá trị trung bình và độ lệch chuẩn của số lượng khách hàng mà quán cà phê này phục vụ trong một giờ vào thời gian này trong ngày là bao nhiêu?
Tính xác suất mà quán cà phê này phục vụ 70 khách hàng trong một giờ vào thời điểm này trong ngày?

(c) Một sự kiện được coi là bất thường nếu xác suất xảy ra sự kiện đó là nhỏ hơn 1%, nhưng sự kiện đó vẫn xảy ra trong thực tế. Giả sử rằng, có nhiều nhất là 55 khách hàng đến quán cà phê này trong một giờ vào giờ cao điểm buổi sáng. Liệu số lượng khách hàng này có được coi là thấp bất thường?

Lời giải.

(a) Gọi X là số lượng khách hàng đến quán cà phê trong vòng một giờ vào giờ cao điểm buổi sáng. Phân phối xác suất thích hợp cho X là phân phối Poisson.

(b) Ta có: $X \sim \mathcal{P}(\lambda)$ với $\lambda = 75$
Trung bình và phương sai của X:

$$\mathbb{E}(X) = Var(X) = \lambda = 75$$

Xác suất để quán cà phê phục vụ 70 khách hàng trong một giờ vào thời điểm này trong ngày:

$$P(X = 70) = e^{-75} \cdot \frac{75^{70}}{70!} = e^{-75} \cdot \prod_{i=1}^{70} \frac{75}{i} \approx 0{,}0402$$

Note: Người giải đưa về dấu $\prod$ để có thể bấm máy tính CASIO.

(c) Theo định lý giới hạn trung tâm, ta có $Z = \dfrac{X-\mu}{\sigma} \sim \mathcal{N}(0\,;\,1)$ với:

- $\mu = \lambda = 75$
- $\sigma = \sqrt{\lambda} = \sqrt{75} \approx 8{,}6603$

Xác suất để có nhiều nhất 55 khách hàng đến quán cà phê này trong một giờ vào giờ cao điểm buổi sáng:

$$P(X \le 55) \approx \Phi\left(\frac{55 - 75 + 0{,}5}{8{,}6603}\right) \approx \Phi(-2{,}25) = 1 - \Phi(2{,}25) = 1 - 0{,}9878 = 0{,}0122$$

Vì $P(X \le 55) = 1{,}22\% \ge 1\%$ nên số lượng khách hàng này không được coi là thấp bất thường.

■

A7 GIỮA KỲ I 2023-2024 (CA 1)

Câu 1 (3 điểm)

Một công ty sản xuất ô tô thực hiện khảo sát ý kiến của khách hàng về các sản phẩm của mình. Kết quả thu được như sau: 90% số sản phẩm loại 1 nhận được đánh giá tích cực, 60% số sản phẩm loại 2 nhận được đánh giá tích cực, và 10% số sản phẩm loại 3 nhận được đánh giá tích cực. Giả sử, công ty sản xuất 40% sản phẩm loại 1, 40% sản phẩm loại 2, và 20% sản phẩm loại 3. Một sản phẩm được chọn ngẫu nhiên để kiểm tra.

(a) Tính xác suất sản phẩm được chọn nhận được đánh giá tích cực.

(b) Tính xác suất sản phẩm được chọn nhận được đánh giá tích cực hoặc thuộc loại 1.

(c) Nếu sản phẩm được chọn nhận được đánh giá tích cực, thì xác suất nó là sản phẩm thuộc loại 1 là bao nhiêu?

Lời giải.

Phép thử: Chọn ngẫu nhiên một sản phẩm để kiếm tra

Đặt các biến cố:

A_i: Sản phẩm được chọn thuộc loại i $(i = 1,2,3)$

B: Sản phẩm được chọn nhận được đánh giá tích cực

Theo đề bài:

$$P(A_1) = 0{,}4 \qquad P(A_2) = 0{,}4 \qquad P(A_3) = 0{,}2$$
$$P(B \mid A_1) = 0{,}9 \qquad P(B \mid A_2) = 0{,}6 \qquad P(B \mid A_3) = 0{,}1$$

(a) Vì $\{A_1, A_2, A_3\}$ là một **hệ đầy đủ** nên xác suất để sản phẩm được chọn nhận đánh giá tích cực:

$$\begin{aligned} P(B) &= P(B \mid A_1) \cdot P(A_1) + P(B \mid A_2) \cdot P(A_2) + P(B \mid A_3) \cdot P(A_3) \\ &= 0{,}9 \cdot 0{,}4 + 0{,}6 \cdot 0{,}4 + 0{,}1 \cdot 0{,}2 \\ &= 0{,}62 \end{aligned}$$

(b) Ta có:

$$P(BA_1) = P(B \mid A_1) \cdot P(A_1) = 0{,}9 \cdot 0{,}4 = 0{,}36$$

Xác suất sản phẩm được chọn nhận được đánh giá tích cực **hoặc** thuộc loại 1:

$$P(B + A_1) = P(B) + P(A_1) - P(BA_1) = 0{,}62 + 0{,}4 - 0{,}36 = 0{,}66$$

(c) Xác suất sản phẩm được chọn thuộc loại 1 **nếu** nó nhận được đánh giá tích cực:

$$P(A_1 \mid B) = \frac{P(A_1B)}{P(B)} = \frac{0{,}36}{0{,}62} \approx 0{,}5806$$

Câu 2 (3 điểm)

Đường kính (X) của một phân tử (đơn vị: μm) được mô hình bởi hàm mật độ xác suất như sau:

$$f(x) = \begin{cases} \dfrac{C}{(x-6)^2}, & \text{nếu } 0 \le x \le 4 \\ 0, & \text{nếu khác} \end{cases}$$

(a) Tìm hằng số C.

(b) Tính giá trị của hàm phân phối xác suất tại 3.

(c) Cho $Y = 2X + 3$. Tính kỳ vọng của Y.

Lời giải.

(a) Ta có:

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_0^4 \frac{C}{(x-6)^2}\,dx = \frac{C}{3}$$

Để $f(x)$ là hàm mật độ xác suất thì:

(i) $$\int_{-\infty}^{+\infty} f(x)\,dx = 1 \quad \Leftrightarrow \quad \frac{C}{3} = 1 \quad \Leftrightarrow \quad C = 3 \tag{1}$$

(ii) $$f(x) \ge 0, \quad \forall x \in \mathbb{R} \quad \Leftrightarrow \quad \frac{C}{(x-6)^2} \ge 0, \quad \forall x \in [0,4] \quad \Leftrightarrow \quad C \ge 0 \tag{2}$$

Từ (1) và (2), suy ra $C = 3$.

(b) Giá trị của hàm phân phối xác suất tại 3:

$$F_X(3) = \int_{-\infty}^{3} f(x)\,dx = \int_0^3 \frac{3}{(x-6)^2}\,dx = 0{,}5$$

(c) Ta có:

$$\mathbb{E}(Y) = \mathbb{E}(2X+3) = \int_{-\infty}^{+\infty} (2x+3) f(x)\,dx = \int_0^4 (2x+3)\frac{3}{(x-6)^2}\,dx \approx 8{,}4083$$

Câu 3 (3 điểm)

Giả sử thời gian hoàn thành đường chạy cự li 100m của các nam sinh trường T là một biến ngẫu nhiên tuân theo phân phối chuẩn với trung bình là 15 giây và độ lệch chuẩn là 1,5 giây. Những nam sinh có thành tích chạy dưới 11,5 giây sẽ được chọn vào đội tuyển của trường.

(a) Hãy tính tỷ lệ nam sinh được chọn vào đội tuyển.

(b) Trong một nhóm gồm 30 nam sinh được chọn ngẫu nhiên, hãy tính xác suất có ít nhất hai nam sinh được chọn vào đội tuyển. Gọi Y là tổng thời gian hoàn thành đường chạy của 30 sinh viên này. Tính $\mu_Y - \sigma_Y$, với μ_Y và σ_Y lần lượt là kỳ vọng và độ lệch chuẩn của Y.

Lời giải.

(a) Gọi X (giây) là thời gian hoàn thành đường chạy cự ly 100m của một nam sinh trường T.
Theo đề: $X \sim \mathcal{N}\left(15\,;\, 1{,}5^2\right)$
Xác suất để nam sinh có thành tích chạy dưới 11,5 giây:

$$\begin{aligned} P(X < 11{,}5) &= P\left(\frac{X-15}{1{,}5} < \frac{11{,}5-15}{1{,}5}\right) \\ &= P(Z < -2{,}33) \\ &\approx \Phi(-2{,}33) = 1 - \Phi(2{,}33) = 1 - 0{,}9901 = 0{,}0099 \end{aligned}$$

Vậy có khoảng 0,99% nam sinh được chọn vào đội tuyển.

(b) Theo đề, Y là tổng thời gian hoàn thành đường chạy 100m của 30 nam sinh.
Suy ra: $Y = X_1 + X_2 + \ldots + X_n$, trong đó $X_i \sim \mathcal{N}\left(\mu\,;\,\sigma^2\right)$, $\quad n = 30, \mu = 15, \sigma = 1{,}5$
Ta có: $Y \sim \mathcal{N}\left(\mu_Y\,;\,\sigma_Y^2\right)$ với:

- $\mu_Y = n\mu = 30 \cdot 15 = 450$
- $\sigma_Y = \sqrt{n\sigma^2} = \sqrt{30 \cdot 1{,}5^2} \approx 8{,}2158$

Vậy: $\quad \mu_Y - \sigma_Y = 450 - 8{,}2158 = 441{,}7842$

■

Câu 4 (1 điểm)

Tuổi trung bình của sinh viên một trường đại học là 22,3 với độ lệch chuẩn là 4. Chọn ngẫu nhiên 64 sinh viên. Tính xác suất tuổi trung bình các sinh viên này lớn hơn 23.

Lời giải.

Gọi X là tuổi của một sinh viên trường đại học được chọn ngẫu nhiên.

Theo đề: $X \sim \mathcal{N}\left(22{,}3\ ;\ 4^2\right)$

Gọi Y là tuổi trung bình của 64 sinh viên được chọn ngẫu nhiên

Ta có: $Y = \dfrac{1}{n}(X_1 + X_2 + \ldots + X_n)$, trong đó $X_i \sim \mathcal{N}\left(\mu\ ;\ \sigma^2\right)$, $\quad n = 64, \mu = 22{,}3, \sigma = 4$

Theo định lý giới hạn trung tâm, ta có $Z = \dfrac{Y - \mu_Y}{\sigma_Y} \sim \mathcal{N}(0\ ;\ 1)$ với:

- $\mu_Y = \dfrac{1}{n} \cdot n\mu = \mu = 22{,}3$
- $\sigma_Y = \sqrt{\dfrac{1}{n^2} \cdot n\sigma^2} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{4}{\sqrt{64}} = 0{,}5$

Xác suất tuổi trung bình các sinh viên này lớn hơn 23:

$$\begin{aligned} P(Y > 23) &= 1 - P(Y \leq 23) \\ &= 1 - P\left(\frac{Y - 22{,}3}{0{,}5} > \frac{23 - 22{,}3}{0{,}5}\right) \\ &= 1 - \Phi(1{,}4) = 1 - 0{,}9192 = 0{,}0808 \end{aligned}$$

■

A8 GIỮA KỲ I 2023-2024 (CA 2)

Câu 1 (3 điểm)

Một nhà máy có bốn ca làm việc. Trung bình mỗi ngày, tỷ lệ phế phẩm của bốn ca lần lượt là 4%, 3%, 2% và 1%, tương ứng từ ca 1 đến ca 4. Giả sử số lượng sản phẩm của bốn ca có tỷ lệ 3:3:2:2, tương ứng từ ca 1 đến ca 4. Một sản phẩm được chọn ngẫu nhiên để kiểm tra.

(a) Tính xác suất sản phẩm được chọn là phế phẩm.

(b) Tính xác suất sản phẩm được chọn là phế phẩm hoặc được sản xuất bởi ca 3.

(c) Nếu sản phẩm được chọn là phế phẩm, thì xác suất nó là sản phẩm của ca 3 là bao nhiêu?

Lời giải.

Phép thử: Chọn ngẫu nhiên một sản phẩm để kiểm tra

Đặt các biến cố:

A_i: Sản phẩm được chọn được sản xuất bởi ca i $(i = 1,2,3,4)$

B: Sản phẩm được chọn là phế phẩm

Theo đề bài:

$P(A_1) = 0{,}3$	$P(A_2) = 0{,}3$	$P(A_3) = 0{,}2$	$P(A_4) = 0{,}2$
$P(B \mid A_1) = 0{,}04$	$P(B \mid A_2) = 0{,}03$	$P(B \mid A_3) = 0{,}02$	$P(B \mid A_4) = 0{,}01$

(a) Vì $\{A_1, A_2, A_3, A_4\}$ là một **hệ đầy đủ** nên xác suất để sản phẩm được chọn là phế phẩm:

$$\begin{aligned} P(B) &= P(B \mid A_1)\cdot P(A_1) + P(B \mid A_2)\cdot P(A_2) + P(B \mid A_3)\cdot P(A_3) + P(B \mid A_4)\cdot P(A_4) \\ &= 0{,}04\cdot 0{,}3 + 0{,}03\cdot 0{,}3 + 0{,}02\cdot 0{,}2 + 0{,}01\cdot 0{,}2 \\ &= 0{,}027 \end{aligned}$$

(b) Ta có:

$$P(BA_3) = P(B \mid A_3)\cdot P(A_3) = 0{,}02\cdot 0{,}2 = 0{,}004$$

Xác suất sản phẩm được chọn là phế phẩm **hoặc** được sản xuất bởi ca 3:

$$P(B + A_3) = P(B) + P(A_3) - P(BA_3) = 0{,}027 + 0{,}2 - 0{,}004 = 0{,}223$$

(c) Xác suất sản phẩm được chọn được sản xuất bởi ca 3 **nếu** nó là phế phẩm:

$$P(A_3 \mid B) = \frac{P(A_3B)}{P(B)} = \frac{0{,}004}{0{,}027} \approx 0{,}1481$$

■

Câu 2 (3 điểm)

Tổng bức xạ mặt trời hằng ngày (X) ở một địa điểm D trong tháng Mười có hàm mật độ xác suất là:

$$f(x) = \begin{cases} C(x-2)(6-x), & 2 \le x \le 6 \\ 0, & \text{nơi khác} \end{cases}$$

(a) Tìm hằng số C.

(b) Tính giá trị của hàm phân phối xác suất tại 5.

(c) Cho $Y = 3X^2 + 4$. Tính kỳ vọng của Y.

Lời giải.

(a) Ta có:

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_2^6 C(x-2)(6-x)\,dx = \frac{32}{3}C$$

Để $f(x)$ là hàm mật độ xác suất thì:

(i) $\displaystyle\int_{-\infty}^{+\infty} f(x)\,dx = 1 \quad\Leftrightarrow\quad \frac{32}{3}C = 1 \quad\Leftrightarrow\quad C = \frac{3}{32}$ (1)

(ii) $f(x) \ge 0,\quad \forall x \in \mathbb{R} \quad\Leftrightarrow\quad C(x-2)(6-x) \ge 0,\quad \forall x \in [2,6] \quad\Leftrightarrow\quad C \ge 0$ (2)

Từ (1) và (2), suy ra $C = \frac{3}{32}$.

(b) Giá trị của hàm phân phối xác suất tại 5:

$$F_X(5) = \int_{-\infty}^{5} f(x)\,dx = \int_2^5 \frac{3}{32}(x-2)(6-x)\,dx = 0{,}84375$$

(c) Ta có:

$$\mathbb{E}(Y) = \mathbb{E}(3X^2+4) = \int_{-\infty}^{+\infty} (3x^2+4)f(x)\,dx = \int_2^6 \frac{3}{32}(3x^2+4)(x-2)(6-x)\,dx = 54{,}4$$

■

Câu 3 (3 điểm)

Một công ty điện gia dụng ghi nhận rằng thời gian giữa hai lượt bảo hành liên tiếp cho các sản phẩm của họ là một biến ngẫu nhiên có phân phối mũ với trung bình 1,5 ngày (1 ngày = 24 giờ).

(a) Tính xác suất để không có lượt bảo hành nào được tiếp nhận bởi công ty này trong khoảng thời gian 5 giờ.

(b) Tính xác suất để công ty này đã tiếp nhận ít nhất 2 lượt bảo hành trong khoảng thời gian 5 giờ.

(c) Một ngày ngẫu nhiên được chọn để khảo sát, biết rằng công ty này đã tiếp nhận nhiều hơn 1 lượt bảo hành trong ngày này. Tính xác suất để công ty này đã tiếp nhận không quá 3 lượt bảo hành trong ngày này.

Lời giải.

(a) Gọi T (giờ) là thời gian giữa hai lượt bảo hành liên tiếp cho các sản phẩm của công ty.
Theo đề: $T \sim Exp(\lambda_T)$
Ta có:

$$\mathbb{E}(X) = \frac{1}{\lambda_T} \Rightarrow \lambda_T = \frac{1}{\mathbb{E}(X)} = \frac{1}{1{,}5 \cdot 24} = \frac{1}{36}$$

Xác suất để không có lượt bảo hành nào được tiếp nhận bởi công ty trong 5 giờ:

$$P(T > 5) = 1 - P(T \leq 5) = 1 - \left(1 - e^{-\frac{1}{36} \cdot 5}\right) \approx 0{,}8703$$

(b) Gọi X là số lượt bảo hành mà công ty tiếp nhận trong khoảng thời gian 5 giờ.

Ta có: $X \sim \mathcal{P}(\lambda_X)$ với $\lambda_X = \dfrac{5}{36}$

Xác suất để công ty đã tiếp nhận ít nhất 2 lượt bảo hành trong khoảng thời gian 5 giờ:

$$P(X \geq 2) = 1 - P(X < 2) = 1 - \sum_{k=0}^{1} e^{-\frac{5}{36}} \cdot \frac{\left(\frac{5}{36}\right)^k}{k!} \approx 1 - 0{,}9912 = 0{,}0088$$

(c) Gọi Y là số lượt bảo hành mà công ty tiếp nhận trong 1 ngày.

Ta có: $Y \sim \mathcal{P}(\lambda_Y)$ với $\lambda_Y = \dfrac{24}{36} = \dfrac{2}{3}$

Ta có:

$$P(Y \leq 3 \cap Y > 1) = P(2 \leq Y \leq 3) = \sum_{k=2}^{3} e^{-\frac{2}{3}} \cdot \frac{\left(\frac{2}{3}\right)^k}{k!} \approx 0{,}1394$$

$$P(Y > 1) = 1 - P(Y \leq 1) = \sum_{k=0}^{1} e^{-\frac{2}{3}} \cdot \frac{\left(\frac{2}{3}\right)^k}{k!} \approx 0{,}1443$$

Xác suất để công ty đã tiếp nhận không quá 3 lượt bảo hành **biết rằng** công ty đã tiếp nhận nhiều hơn 1 lượt bảo hành trong ngày được khảo sát:

$$P(Y \leq 3 \mid Y > 1) = \frac{P(Y \leq 3 \cap Y > 1)}{P(Y > 1)} = \frac{0{,}1394}{0{,}1443} \approx 0{,}9660$$

Câu 4 (1 điểm)

Một nhà máy sản xuất một loại điện trở với giá trị trung bình 100 ohm và độ lệch chuẩn 10 ohm. Tính xác suất giá trị trung bình của 50 điện trở được chọn ngẫu nhiên nhỏ hơn 95 ohm.

Lời giải.

Gọi X là giá trị của một điện trở được chọn ngẫu nhiên.

Theo đề: $X \sim \mathcal{N}\left(100\ ;\ 10^2\right)$

Gọi Y là giá trị trung bình của 50 điện trở được chọn ngẫu nhiên

Ta có: $Y = \dfrac{1}{n}(X_1 + X_2 + \ldots + X_n)$, trong đó $X_i \sim \mathcal{N}\left(\mu\ ;\ \sigma^2\right)$, $\quad n = 50, \mu = 100, \sigma = 10$

Theo định lý giới hạn trung tâm, ta có $Z = \dfrac{Y - \mu_Y}{\sigma_Y} \sim \mathcal{N}(0\ ;\ 1)$ với:

- $\mu_Y = \dfrac{1}{n} \cdot n\mu = \mu = 100$
- $\sigma_Y = \sqrt{\dfrac{1}{n^2} \cdot n\sigma^2} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{10}{\sqrt{50}} \approx 1{,}4142$

Xác suất giá trị trung bình của các điện trở này nhỏ hơn 95 ohm:

$$\begin{aligned} P(Y < 95) &= P\left(\frac{Y - 100}{1{,}4142} < \frac{95 - 100}{1{,}4142}\right) \\ &\approx P(Z < -3{,}54) \\ &= 1 - \Phi(3{,}54) \approx 1 - 1 = 0 \end{aligned}$$

■

B TÓM TẮT LÝ THUYẾT PHẦN THỐNG KÊ

1 Thống kê mô tả

Một số độ đo thường dùng:

	Tổng thể (kích thước: N)	**Mẫu** (kích thước: n)
Trung bình	$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$	$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Phương sai	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$
Độ lệch chuẩn	$\sigma = \sqrt{\sigma^2}$	$s = \sqrt{s^2}$

2 Khoảng tin cậy

2.1 Khoảng tin cậy cho trung bình

- Khoảng tin cậy: $\mu \in \left[\overline{x} - \varepsilon \,;\, \overline{x} + \varepsilon\right]$
- Sai số ước lượng:

	$n > 30$	$n \leq 30$ (mẫu có pp chuẩn)
σ^2 đã biết	$\varepsilon = z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$	
σ^2 chưa biết	$\varepsilon = z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$	$\varepsilon = t^{n-1}_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$

- Để $\varepsilon \leq \varepsilon^*$ thì $n \geq \left(z_{1-\frac{\alpha}{2}} \frac{\sigma}{\varepsilon^*}\right)^2$ (nếu σ chưa biết thì thay bởi s)

2.2 Khoảng tin cậy cho tỷ lệ

- Khoảng tin cậy: $p \in \left[\widehat{p} - \varepsilon \,;\, \widehat{p} + \varepsilon\right]$
- Sai số ước lượng: $\varepsilon = z_{1-\frac{\alpha}{2}} \sqrt{\frac{\widehat{p}\,(1-\widehat{p}\,)}{n}}$

- Để $\varepsilon \le \varepsilon^*$ thì
  - Trường hợp đã biết $\widehat{p}$: $n \ge \left(\dfrac{z_{1-\frac{\alpha}{2}}}{\varepsilon^*}\right)^2 \widehat{p}\,(1-\widehat{p}\,)$
  - Trường hợp chưa biết $\widehat{p}$: $n \ge \left(\dfrac{z_{1-\frac{\alpha}{2}}}{\varepsilon^*}\right)^2 \cdot 0{,}25$

3 Kiểm định giả thuyết - một mẫu

3.1 So sánh kỳ vọng với một số

- Giả thuyết kiểm định:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \ne \mu_0 \end{cases} \quad ; \quad \begin{cases} H_0 : \mu \le \mu_0 \\ H_1 : \mu > \mu_0 \end{cases} \quad ; \quad \begin{cases} H_0 : \mu \ge \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$

- Thống kê kiểm định:

	$n > 30$	$n \le 30$ (mẫu có pp chuẩn)
σ^2 đã biết	$Z_0 = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0\,;\,1)$	
σ^2 chưa biết	$Z_0 = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim \mathcal{N}(0\,;\,1)$	$T_0 = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim T(n-1)$

- Miền bác bỏ:

Đối thuyết	**Thống kê** Z	**Thống kê** T
$=$	$\lvert z_0 \rvert > z_{1-\frac{\alpha}{2}}$	$\lvert t_0 \rvert > t^{n-1}_{1-\frac{\alpha}{2}}$
$>$	$z_0 > z_{1-\alpha}$	$t_0 > t^{n-1}_{1-\alpha}$
$<$	$z_0 < -z_{1-\alpha}$	$t_0 < -t^{n-1}_{1-\alpha}$

- *p-giá trị:*

Đối thuyết	**Thống kê** Z	**Thống kê** T
$=$	$2\left[1 - \Phi(\lvert z_0 \rvert)\right]$	$2\,P(t^{n-1} \ge \lvert t_0 \rvert)$
$>$	$1 - \Phi(z_0)$	$P(t^{n-1} \ge t_0)$
$<$	$\Phi(z_0)$	$P(t^{n-1} \le t_0)$

3.2 So sánh tỷ lệ với một số

- Giả thuyết kiểm định:

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases} \quad ; \quad \begin{cases} H_0 : p \leq p_0 \\ H_1 : p > p_0 \end{cases} \quad ; \quad \begin{cases} H_0 : p \geq p_0 \\ H_1 : p < p_0 \end{cases}$$

- Thống kê kiểm định:

$$Z_0 = \frac{\widehat{P} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} \sim \mathcal{N}(0\,;\,1)$$

- Miền bác bỏ, *p-giá trị*: xem mục **3.1**

4 Kiểm định giả thuyết - hai mẫu độc lập

4.1 So sánh hai kỳ vọng

- Giả thuyết kiểm định:

$$\begin{cases} H_0 : \mu_1 - \mu_2 = D_0 \\ H_1 : \mu_1 - \mu_2 \neq D_0 \end{cases} \quad ; \quad \begin{cases} H_0 : \mu_1 - \mu_2 \leq D_0 \\ H_1 : \mu_1 - \mu_2 > D_0 \end{cases} \quad ; \quad \begin{cases} H_0 : \mu_1 - \mu_2 \geq D_0 \\ H_1 : \mu_1 - \mu_2 < D_0 \end{cases}$$

- Thống kê kiểm định:
  - **Trường hợp 1:** σ^2 đã biết

$$Z_0 = \frac{\overline{X_1} - \overline{X_2} - D_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0\,;\,1)$$

  - **Trường hợp 2:** σ^2 chưa biết, mẫu lớn

$$Z_0 = \frac{\overline{X_1} - \overline{X_2} - D_0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \sim \mathcal{N}(0\,;\,1)$$

  - **Trường hợp 3:** σ^2 chưa biết, mẫu nhỏ, $\sigma_1^2 = \sigma_2^2$

$$T_0 = \frac{\overline{X_1} - \overline{X_2} - D_0}{\sqrt{S_P^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \sim T(n_1 + n_2 - 2) \quad \text{với} \quad S_P^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}$$

  - **Trường hợp 4:** σ^2 chưa biết, mẫu nhỏ, $\sigma_1^2 \neq \sigma_2^2$

$$T_0 = \frac{\overline{X_1} - \overline{X_2} - D_0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \sim T(df) \quad \text{với} \quad df = \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{\dfrac{(S_1^2/n_1)^2}{n_1 - 1} + \dfrac{(S_2^2/n_2)^2}{n_2 - 1}}$$

- Miền bác bỏ, *p-giá trị*: xem mục **3.1**

4.2 So sánh hai tỷ lệ

- Giả thuyết kiểm định:

$$\begin{cases} H_0 : p_1 - p_2 = D_0 \\ H_1 : p_1 - p_2 \neq D_0 \end{cases} ; \quad \begin{cases} H_0 : p_1 - p_2 \leq D_0 \\ H_1 : p_1 - p_2 > D_0 \end{cases} ; \quad \begin{cases} H_0 : p_1 - p_2 \geq D_0 \\ H_1 : p_1 - p_2 < D_0 \end{cases}$$

- Thống kê kiểm định:

$$Z_0 = \frac{\widehat{P}_1 - \widehat{P}_2 - D_0}{\sqrt{\widehat{P}\,(1 - \widehat{P}\,)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim \mathcal{N}\,(0\,;\,1)$$

- Miền bác bỏ, *p-giá trị*: xem mục **3.1**

5 Hồi quy tuyến tính

- Từ bảng dữ liệu, ta có: $n \; ; \; \sum_{i=1}^{n} x_i \; ; \; \sum_{i=1}^{n} x_i^2 \; ; \; \sum_{i=1}^{n} y_i \; ; \; \sum_{i=1}^{n} y_i^2 \; ; \; \sum_{i=1}^{n} x_i y_i$
- Phương trình đường thẳng hồi quy: $\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1\, x$
- Ta có:
  - $S_{xy} = \sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i = \sum x_i y_i - n\,\overline{x}\,\overline{y}$
  - $S_{xx} = \sum x_i^2 - \frac{1}{n} \left(\sum x_i\right)^2 = \sum x_i^2 - n\,\overline{x}^{\,2}$
  - $S_{yy} = \sum y_i^2 - \frac{1}{n} \left(\sum y_i\right)^2 = \sum y_i^2 - n\,\overline{y}^{\,2}$
- Hệ số:
  - $\widehat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$
  - $\widehat{\beta}_0 = \frac{1}{n} \sum y_i - \widehat{\beta}_1 \frac{1}{n} \sum x_i = \overline{y} - \widehat{\beta}_1\, \overline{x}$
- Độ đo sự biến thiên của dữ liệu:
  - $SSR = \widehat{\beta}_1\, S_{xy}$
  - $SST = S_{yy}$
  - $SSE = SST - SSR$
- Hệ số xác định: $R^2 = \frac{SSR}{SST} = r_{XY}^2 \quad \in [0;1]$
- Hệ số tương quan: $r_{XY} = \frac{S_{xy}}{\sqrt{S_{xx}\, S_{yy}}} \quad \in [-1;1]$ (cùng dấu với $\widehat{\beta}_1$)

B1 CUỐI KỲ HÈ 2024-2025

Câu 1 (6 điểm)

Quan sát cân nặng của bé trai và bé gái lúc sơ sinh, ta có kết quả như sau:

Trọng lượng (gam)	3000–3200	3200–3400	3400–3600	3600–3800	3800–4000
Số bé trai	7	14	20	15	9
Số bé gái	8	20	20	10	7

(a) Có ý kiến cho rằng cân nặng trung bình của bé trai cao hơn cân nặng trung bình của bé gái. Với mức ý nghĩa 3%, hãy cho nhận xét về ý kiến đó. Tính *p-giá trị*.

(b) Với mức ý nghĩa 3%, hãy so sánh tỷ lệ bé trai có cân nặng dưới 3400 gam với tỷ lệ bé gái có cân nặng dưới 3400 gam.

(c) Gộp hai mẫu lại thành một. Tính trung bình và độ lệch chuẩn của mẫu gộp. Sau đó, ước lượng cân nặng trung bình của trẻ sơ sinh với độ tin cậy 98%.

(d) Ước lượng tỷ lệ trẻ sơ sinh có cân nặng hơn 3600 gam với độ tin cậy 99%. Nếu muốn sai số ước lượng không quá 5% thì cần quan sát thêm ít nhất bao nhiêu trẻ?

Lời giải.

(a) Gọi X_1 (gam) và X_2 (gam) lần lượt là cân nặng của bé trai và bé gái lúc sơ sinh.
Gọi μ_1, σ_1, μ_2, σ_2 lần lượt là trung bình và độ lệch chuẩn của cân nặng bé trai và bé gái.
Ta có: σ_1, σ_2 chưa biết, $n_1 = n_2 = 65 > 30$

- Giả thuyết kiểm định: $\begin{cases} H_0 : \mu_1 - \mu_2 \leq 0 \\ H_1 : \mu_1 - \mu_2 > 0 \end{cases}$
- Dưới giả định H_0 đúng, thống kê kiểm định:

$$Z_0 = \frac{\overline{X_1} - \overline{X_2} - D_0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \sim \mathcal{N}(0\,;\,1)$$

- Ta có: $\bar{x}_1 \approx 3515{,}3846$ $\quad s_1 \approx 240{,}5922$ $\quad n_1 = 65$
  $\bar{x}_2 \approx 3463{,}0769$ $\quad s_2 \approx 234{,}2336$ $\quad n_2 = 65$
  $D_0 = 0$

  suy ra:

$$z_0 = \frac{3515{,}3846 - 3463{,}0769 - 0}{\sqrt{\dfrac{240{,}5922^2}{65} + \dfrac{234{,}2336^2}{65}}} \approx 1{,}2559$$

- Miền bác bỏ: $W_{0,03} = \{ z_0 : z_0 > z_{1-\alpha} = z_{0,97} = 1{,}88 \}$
- Ta có: $z_0 = 1{,}2559 < 1{,}88 \quad \Rightarrow \quad z_0 \notin W_{0,03}$

Vậy với mức ý nghĩa 3%, chưa đủ cơ sở bác bỏ H_0, nghĩa là cân nặng trung bình của bé trai **không** cao hơn cân nặng trung bình của bé gái.

$$p\text{-giá trị} = 1 - \Phi(z_0) \approx 1 - \Phi(1{,}26) = 1 - 0{,}8962 = 0{,}1038$$

(b) Gọi Y_1 và Y_2 lần lượt là số bé trai và số bé gái có cân nặng dưới 3400 gam.

Ta có: $Y_1 \sim B(n_1\,;\,p_1)$, $\quad Y_2 \sim B(n_2\,;\,p_2)$

- Giả thuyết kiểm định: $\begin{cases} H_0 : p_1 - p_2 = 0 \\ H_1 : p_1 - p_2 \neq 0 \end{cases}$
- Dưới giả định H_0 đúng, thống kê kiểm định:

$$Z_0 = \frac{\widehat{P}_1 - \widehat{P}_2 - D_0}{\sqrt{\widehat{P}\,(1-\widehat{P}\,)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \sim \mathcal{N}(0\,;\,1)$$

- Ta có: $\widehat{p}_1 = \dfrac{21}{65}, \quad \widehat{p}_2 = \dfrac{28}{65}, \quad \widehat{p} = \dfrac{21+28}{65+65} = \dfrac{49}{130}, \quad n_1 = n_2 = 65, \quad D_0 = 0$

  suy ra:

$$z_0 = \frac{\dfrac{21}{65} - \dfrac{28}{65} - 0}{\sqrt{\dfrac{49}{130}\left(1 - \dfrac{49}{130}\right)\left(\dfrac{1}{65} + \dfrac{1}{65}\right)}} \approx -1{,}2669$$

- Miền bác bỏ: $W_{0,03} = \{ z_0 : |z_0| > z_{1-\frac{\alpha}{2}} = z_{0,985} = 2{,}17 \}$
- Ta có: $|z_0| = 1{,}2669 < 2{,}17 \quad \Rightarrow \quad z_0 \notin W_{0,03}$

Vậy với mức ý nghĩa 3%, chưa đủ cơ sở bác bỏ H_0, nghĩa là tỷ lệ bé trai và tỷ lệ bé gái có cân nặng dưới 3400 gam là như nhau.

(c) Gọi X (gam) là cân nặng của trẻ sơ sinh.

Gọi μ, σ lần lượt là trung bình và độ lệch chuẩn của cân nặng trẻ sơ sinh.

Ta có: σ chưa biết, $\quad n = 130 > 30$

- Phân phối mẫu: $Z = \dfrac{\overline{X} - \mu}{S/\sqrt{n}} \sim \mathcal{N}(0\,;\,1)$
- Ta có: $\overline{x} = 3489{,}2308, \quad s = 237{,}9649, \quad n = 130$

  Độ tin cậy 98% $\quad \Rightarrow \quad \alpha = 0{,}02 \quad \Rightarrow \quad z_{1-\frac{\alpha}{2}} = z_{0,99} = 2{,}33$
- Sai số ước lượng:

$$\varepsilon = z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} = 2{,}33 \cdot \frac{237{,}9649}{\sqrt{130}} \approx 48{,}6292$$

Vậy khoảng tin cậy 98% cho cân nặng trung bình của trẻ sơ sinh là:

$$\mu \in \left[\mu - \varepsilon\ ;\ \mu + \varepsilon\right] = \left[3440{,}6016\ ;\ 3537{,}86\right] \quad \text{(gam)}$$

(d) Gọi Y là số trẻ sơ sinh có cân nặng hơn 3600 gam trong mẫu. Ta có: $Y \sim B(n\ ;\ p)$

- Phân phối mẫu: $Z = \dfrac{\widehat{P} - p}{\sqrt{\dfrac{\widehat{P}(1-\widehat{P})}{n}}} \sim \mathcal{N}(0\ ;\ 1)$
- Ta có: $\widehat{p} = \dfrac{41}{130} \approx 0{,}3154, \quad n = 130$

  Độ tin cậy 99% $\quad \Rightarrow \quad \alpha = 0{,}01 \quad \Rightarrow \quad z_{1-\frac{\alpha}{2}} = z_{0{,}995} = 2{,}58$
- Sai số ước lượng:

$$\varepsilon = z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 2{,}58 \cdot \sqrt{\frac{\frac{41}{130}\left(1 - \frac{41}{130}\right)}{130}} \approx 0{,}1051$$

Vậy khoảng tin cậy 99% cho tỷ lệ trẻ sơ sinh có cân nặng trên 3600 gam là:

$$p \in \left[\widehat{p} - \varepsilon\ ;\ \widehat{p} + \varepsilon\right] = \left[0{,}2103\ ;\ 0{,}4205\right]$$

Để sai số ước lượng không quá $\varepsilon^* = 0{,}05$ thì:

$$\varepsilon = z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \le \varepsilon^* \quad \Rightarrow \quad n \ge \left(\frac{z_{1-\frac{\alpha}{2}}}{\varepsilon^*}\right)^2 \widehat{p}(1-\widehat{p}) = \left(\frac{2{,}58}{0{,}05}\right)^2 \frac{41}{130}\left(1 - \frac{41}{130}\right) \approx 575$$

Vậy để sai số ước lượng không quá 0,05 (với cùng độ tin cậy) thì cần quan sát **thêm** ít nhất:

$$575 - 130 = 445 \quad \text{(trẻ sơ sinh)}$$

■

Câu 2 (1,5 điểm)

Các khiếm khuyết xảy ra trong nội thất bằng chất dẻo được sử dụng cho xe ô tô tuân theo phân phối Poisson với trung bình 0,02 lỗ hổng trên mỗi panel.

(a) Nếu 50 panel được kiểm tra, tính xác suất không có khiếm khuyết nào?

(b) Kiểm tra 200 panel, xác suất có không quá 50 panel có ít nhất một khiếm khuyết là bao nhiêu?

Lời giải.

(a) Gọi X là số khiếm khuyết trên một tấm panel.

Theo đề: $X \sim \mathcal{P}(\lambda = 0{,}02)$

Gọi X_{50} là số khiếm khuyết trên 50 tấm panel.

Suy ra: $X_{50} \sim \mathcal{P}(\lambda_{50} = 50 \cdot 0{,}02 = 1)$

Xác suất để 50 tấm panel được kiểm tra không có khiếm khuyết nào:

$$P(X_{50} = 0) = \frac{e^{-1} \cdot 1^0}{0!} \approx 0{,}3679$$

(b) Xác suất một tấm panel có ít nhất một khiếm khuyết:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{e^{-0{,}02} \cdot 0{,}02^0}{0!} \approx 0{,}0198$$

Gọi Y là số panel có ít nhất một khiếm khuyết trong 200 tấm được kiểm tra.

Ta có: $Y \sim B\,(200\,;\,0{,}0198)$

Xấp xỉ phân phối nhị thức bởi phân phối chuẩn, ta được $Z = \dfrac{Y - \mu_Y}{\sigma_Y} \sim \mathcal{N}\,(0\,;\,1)$ với:

- $\mu_Y = 200 \cdot 0{,}0198 = 3{,}96$
- $\sigma_Y = \sqrt{200 \cdot 0{,}0198 \cdot (1 - 0{,}0198)} \approx 1{,}9702$

Xác suất có không quá 50 panel có ít nhất một khiếm khuyết:

$$P(Y \leq 50) \approx \Phi\left(\frac{50 - 3{,}96 + 0{,}5}{1{,}9702}\right) \approx \Phi(23{,}62) \approx 1$$

■

Câu 3 (2,5 điểm)

Cho bảng dữ liệu sau:

Tốc độ lắng đọng SO_2	14	18	40	43	45	112
Mức hao hụt khối lượng của thép	280	350	470	500	560	1200

Chọn biến độc lập là tốc độ lắng đọng SO_2 (mg/m^2/ngày), và biến phụ thuộc là mức hao hụt khối lượng của thép (g/m^2).

(a) Vẽ biểu đồ phân tán (scatter plot). Mô hình hồi quy tuyến tính đơn giản có hợp lý trong tình huống này không?

(b) Xác định phương trình của đường hồi quy tuyến tính đơn y theo x.

(c) Nếu tốc độ lắng đọng SO_2 tăng 2 (mg/m^2/ngày) thì mức hao hụt khối lượng của thép thay đổi như thế nào?

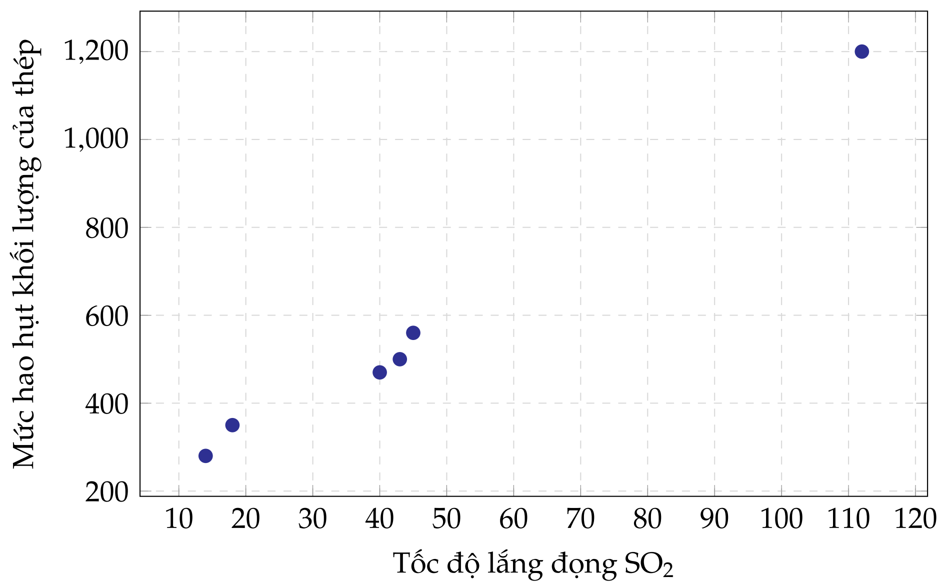
(d) Tính hệ số xác định R^2 và giải thích ý nghĩa của kết quả tìm được.

Lời giải.

(a) Biến độc lập (x): Tốc độ lắng đọng SO_2

Biến phụ thuộc (y): Mức hao hụt khối lượng của thép

Từ biểu đồ phân tán (bên dưới), ta thấy mô hình hồi quy tuyến tính đơn giản là hợp lý trong tình huống này.



(b) Từ bảng dữ liệu, ta có:

$$n = 6; \quad \sum_{i=1}^{n} x_i = 272; \quad \sum_{i=1}^{n} x_i^2 = 18538; \quad \sum_{i=1}^{n} y_i = 3360; \quad \sum_{i=1}^{n} y_i^2 = 2425400; \quad \sum_{i=1}^{n} x_i y_i = 210120.$$

Ta có:

- $$S_{xy} = \sum x_i y_i - \frac{1}{n}\sum x_i \sum y_i = 210120 - \frac{1}{6}\cdot 272 \cdot 3360 = 57800$$
- $$S_{xx} = \sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2 = 18538 - \frac{1}{6}\cdot 272^2 \approx 6207{,}3333$$

Suy ra:

- $$\widehat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{57800}{6207{,}3333} \approx 9{,}3116$$
- $$\widehat{\beta}_0 = \frac{1}{n}\sum y_i - \widehat{\beta}_1 \frac{1}{n}\sum x_i = \frac{1}{6}\cdot 3360 - 9{,}3116 \cdot \frac{1}{6}\cdot 272 \approx 137{,}8741$$

Vậy phương trình đường hồi quy tuyến tính:

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x = 137{,}8741 + 9{,}3116\, x$$

(c) Nếu tốc độ lắng đọng SO_2 tăng 2 (mg/m^2/ngày) thì mức hao hụt khối lượng của thép cũng tăng một lượng:

$$2\widehat{\beta}_1 = 2 \cdot 9{,}3116 = 18{,}6232 \quad (g/m^2)$$

(d) Ta có:

- $SSR = \widehat{\beta}_1 S_{xy} = 9{,}3116 \cdot 57800 = 538210{,}48$
- $SST = \sum y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2 = 2425400 - \frac{1}{6} \cdot 3360^2 = 543800$

Hệ số xác định:

$$R^2 = \frac{SSR}{SST} = \frac{538210{,}48}{543800} \approx 0{,}9897$$

Ý nghĩa: 98,97% sự thay đổi trong mức hao hụt khối lượng của thép được giải thích bởi tốc độ lắng đọng SO_2, 1,03% còn lại được giải thích bởi các yếu tố khác.

■

B2 CUỐI KỲ II 2024-2025

Câu 1 (2 điểm)

Trong một tế bào, có một phân tử enzyme đang di chuyển dọc theo một sợi protein dài (coi như đường thẳng một chiều). Enzyme này đang tìm một chỗ để thực hiện phản ứng. Ban đầu, enzyme đứng ở vị trí giữa sợi (gọi là vị trí số 0). Mỗi giây, enzyme có thể di chuyển một bước sang phải với xác suất p, hoặc di chuyển một bước sang trái với xác suất $1-p$. Giả sử enzyme đã di chuyển trong n giây (n bước).

(a) Sau n bước, xác suất enzyme đang ở vị trí k trên sợi là bao nhiêu?

(b) Giả sử $n = 300$ và $p = 1/2$. Hỏi xác suất enzyme đã đi cách xa vị trí ban đầu ít nhất 40 bước là bao nhiêu? (Sử dụng xấp xỉ chuẩn có hiệu chỉnh).

Lời giải.

(a) Gọi X là một bước đi của enzyme, trong đó $X = 1$ nếu sang phải, $X = -1$ nếu sang trái.

Vị trí của enzyme sau n bước: $S_n = \sum_{i=1}^{n} X_i$

Gọi r là số bước sang phải sau n bước. Vì enzyme đang ở vị trí k nên khi đó ta có:

$$r - (n - r) = k \quad \Rightarrow \quad r = \frac{n+k}{2}$$

Xác suất để enzyme ở vị trí k trên sợi sau n bước:

$$P(S_n = k) = \begin{cases} C_n^r \cdot p^r \cdot (1-p)^{n-r} & \text{nếu } (n+k) \,\vdots\, 2 \\ 0 & \text{nếu khác} \end{cases}, \quad \text{với } r = \frac{n+k}{2}$$

(b) Bảng phân phối xác suất của X:

X	1	-1
p_X	p	$1-p$

Ta có:

- $\mathbb{E}(X) = 1 \cdot p + (-1) \cdot (1-p) = 2p - 1$
- $\mathbb{E}(X^2) = 1^2 \cdot p + (-1)^2 \cdot (1-p) = 1$
- $\text{Var}(X) = \mathbb{E}(X^2) - \left[\mathbb{E}(X)\right]^2 = 1 - (2p-1)^2 = 4p(1-p)$

Theo định lý giới hạn trung tâm, ta có $Z_n = \dfrac{S_n - \mu_n}{\sigma_n} \sim \mathcal{N}(0\,;\,1)$ với:

- $\mu_n = n\mu_X = n(2p-1) = 0$
- $\sigma_n = \sqrt{n\text{Var}(X)} = \sqrt{4np(1-p)} = 10\sqrt{3}$

Xác suất enzyme đã đi cách xa vị trí ban đầu ít nhất 40 bước:

$$\begin{aligned} P(|S_n| \geq 40) \;=\; 2P(S_n \geq 40) \;&=\; 2\left[1 - P(S_n < 40)\right] \\ &\approx\; 2\left[1 - \Phi\left(\frac{40 - 0 - 0{,}5}{10\sqrt{3}}\right)\right] \\ &\approx\; 2\left[1 - \Phi(2{,}28)\right] \\ &=\; 2(1 - 0{,}9887) \;=\; 0{,}0226 \end{aligned}$$

■

Câu 2 (2 điểm)

Trong số 100 người được phát hiện ngẫu nhiên là mắc ung thư phổi, có 67 người đã tử vong trong vòng 5 năm kể từ khi phát hiện bệnh.

(a) Tìm khoảng tin cậy 95% cho tỷ lệ người mắc ung thư phổi sẽ tử vong trong vòng 5 năm.

(b) Cần lấy mẫu thêm ít nhất bao nhiêu người nữa để dung sai của khoảng tin cậy ở câu **(a)** nhỏ hơn 0,02?

Lời giải.

(a) Gọi Y là số người mắc ung thư phổi tử vong trong vòng 5 năm trong mẫu. Ta có: $Y \sim B(n\,;\,p)$

- Phân phối mẫu: $Z = \dfrac{\widehat{P} - p}{\sqrt{\dfrac{\widehat{P}(1-\widehat{P})}{n}}} \sim \mathcal{N}(0\,;\,1)$
- Ta có: $\widehat{p} = \dfrac{67}{100} = 0{,}67, \quad n = 100$

  Độ tin cậy 95% $\Rightarrow$ $\alpha = 0{,}05$ $\Rightarrow$ $z_{1-\frac{\alpha}{2}} = z_{0{,}975} = 1{,}96$
- Sai số ước lượng:

$$\varepsilon = z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 1{,}96 \cdot \sqrt{\frac{0{,}67\,(1-0{,}67)}{100}} \approx 0{,}0922$$

Vậy khoảng tin cậy 95% cho tỷ lệ người mắc ung thư phổi tử vong trong vòng 5 năm:

$$p \in \left[\widehat{p} - \varepsilon\,;\, \widehat{p} + \varepsilon\right] = \left[0{,}5778\,;\,0{,}7622\right]$$

(b) Để sai số ước lượng nhỏ hơn $\varepsilon^* = 0{,}02$ thì:

$$\varepsilon = z_{1-\frac{\alpha}{2}} \sqrt{\frac{\widehat{p}\,(1-\widehat{p}\,)}{n}} < \varepsilon^* \quad \Rightarrow \quad n > \left(\frac{z_{1-\frac{\alpha}{2}}}{\varepsilon^*}\right)^2 \widehat{p}\,(1-\widehat{p}\,) = \left(\frac{1{,}96}{0{,}02}\right)^2 0{,}67(1-0{,}67) \approx 2124$$

Vậy để sai số ước lượng không quá 0,02 (với cùng độ tin cậy) thì cần quan sát **thêm** ít nhất:

$$2124 - 100 = 2024 \quad \text{(người)}$$

■

Câu 3 (3,5 điểm)

Cân nặng của trẻ sơ sinh tại hai quận liền kề ở một tỉnh được thu thập ngẫu nhiên, với các dữ liệu sau:

Thống kê	Quận A	Quận B
Cỡ mẫu (n)	53	44
Trung bình mẫu ($\overline{x}$)	3,08 kg	3,27 kg
Phương sai mẫu (s^2)	1,07 kg^2	1,01 kg^2

(a) Tìm khoảng tin cậy 98% cho cân nặng trung bình của trẻ sơ sinh tại quận A.

(b) Với mức ý nghĩa 1%, hãy kiểm định ý kiến cho rằng cân nặng trung bình của trẻ sơ sinh tại quận A lớn hơn 3kg.

(c) Với mức ý nghĩa 1%, hãy kiểm định ý kiến cho rằng cân nặng trung bình của trẻ sơ sinh là như nhau ở cả hai quận. Tính *p-giá trị* tương ứng với kiểm định này.

Lời giải.

(a) Gọi X (kg) là cân nặng của trẻ sơ sinh quận A.
Gọi μ, σ lần lượt là trung bình và độ lệch chuẩn của cân nặng trẻ sơ sinh quận A.
Ta có: σ chưa biết, $n = 53 > 30$

- Phân phối mẫu: $Z = \dfrac{\overline{X} - \mu}{S/\sqrt{n}} \sim \mathcal{N}(0\,;\,1)$
- Ta có: $\overline{x} = 3{,}08$, $s^2 = 1{,}07$, $n = 53$
  Độ tin cậy 98% $\Rightarrow$ $\alpha = 0{,}02$ $\Rightarrow$ $z_{1-\frac{\alpha}{2}} = z_{0{,}99} = 2{,}33$
- Sai số ước lượng:

$$\varepsilon = z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} = 2{,}33 \cdot \frac{\sqrt{1{,}07}}{\sqrt{53}} \approx 0{,}3311$$

Vậy khoảng tin cậy 98% cho cân nặng trung bình của trẻ sơ sinh quận A là:

$$\mu \in \left[\mu - \varepsilon\ ;\ \mu + \varepsilon\right] = \left[2{,}7489\ ;\ 3{,}4111\right] \quad (\text{kg})$$

(b) Gọi X, μ, σ tương tự câu (**a**).

Ta có: σ chưa biết, $n = 53 > 30$

- Giả thuyết kiểm định: $\begin{cases} H_0 : \mu \leq 3 \\ H_1 : \mu > 3 \end{cases}$
- Dưới giả định H_0 đúng, thống kê kiểm định:

$$Z_0 = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim \mathcal{N}(0\ ;\ 1)$$

- Ta có: $\overline{x} = 3{,}08$, $\mu_0 = 3$, $s^2 = 1{,}07$, $n = 53$

  suy ra:

$$z_0 = \frac{3{,}08 - 3}{\dfrac{\sqrt{1{,}07}}{\sqrt{53}}} \approx 0{,}5630$$

- Miền bác bỏ: $W_{0,01} = \{\, z_0 : z_0 > z_{1-\alpha} = z_{0,99} = 2{,}33 \,\}$
- Ta có: $z_0 = 0{,}5630 < 2{,}33 \quad \Rightarrow \quad z_0 \notin W_{0,01}$

Vậy với mức ý nghĩa 1%, chưa đủ cơ sở bác bỏ H_0, nghĩa là cân nặng trung bình của trẻ sơ sinh quận A **không** lớn hơn 3kg.

(c) Gọi X_1 (kg) và X_2 (kg) lần lượt là cân nặng của trẻ sơ sinh quận A và quận B.

Gọi μ_1, σ_1, μ_2, σ_2 lần lượt là trung bình và độ lệch chuẩn của cân nặng trẻ quận A và quận B.

Ta có: σ_1, σ_2 chưa biết, $n_1 = 53 > 30$, $n_2 = 44 > 30$

- Giả thuyết kiểm định: $\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \end{cases}$
- Dưới giả định H_0 đúng, thống kê kiểm định:

$$Z_0 = \frac{\overline{X_1} - \overline{X_2} - D_0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \sim \mathcal{N}(0\ ;\ 1)$$

- Ta có: $\overline{x}_1 = 3{,}08 \qquad s_1^2 = 1{,}07 \qquad n_1 = 53$

  $\overline{x}_2 = 3{,}27 \qquad s_2^2 = 1{,}01 \qquad n_2 = 44$

  $D_0 = 0$

  suy ra:

$$z_0 = \frac{3{,}08 - 3{,}27 - 0}{\sqrt{\dfrac{1{,}07}{53} + \dfrac{1{,}01}{44}}} \approx -0{,}9147$$

- Miền bác bỏ: $W_{0,01} = \{ z_0 : |z_0| > z_{1-\frac{\alpha}{2}} = z_{0,995} = 2{,}58 \}$
- Ta có: $|z_0| = 0{,}9147 < 2{,}58 \quad \Rightarrow \quad z_0 \notin W_{0,01}$

Vậy với mức ý nghĩa 1%, chưa đủ cơ sở bác bỏ H_0, nghĩa là cân nặng trung bình trẻ sơ sinh ở cả hai quận là như nhau.

$$p\text{-giá trị} = 2\left[1 - \Phi(|z_0|)\right] \approx 2\left[1 - \Phi(0{,}91)\right] = 2\,(1 - 0{,}8186) = 0{,}3628$$

■

Câu 4 (2,5 điểm)

Dữ liệu dưới đây cho biết cân nặng (tính bằng kg) và huyết áp tâm thu (tính bằng mmHg) của 10 nam giới được chọn ngẫu nhiên trong độ tuổi từ 25 đến 30:

Đối tượng	1	2	3	4	5	6	7	8	9	10
Cân nặng (kg)	74,8	75,7	81,6	70,3	96,2	79,4	86,2	95,3	90,7	67,6
Huyết áp (mmHg)	130	133	150	128	151	146	150	140	148	125

(a) Chọn biến cân nặng là biến độc lập, biến huyết áp là biến phụ thuộc. Hãy tìm phương trình hồi quy tuyến tính đơn huyết áp theo cân nặng. Giải thích ý nghĩa của hệ số góc $\widehat{\beta}_1$ nhận được.

(b) Hãy dự đoán huyết áp của một nam giới có cân nặng 80 kg.

(c) Tìm hệ số tương quan mẫu và giải thích ý nghĩa của kết quả tìm được.

(d) Tính hệ số xác định R^2 và giải thích ý nghĩa của kết quả tìm được.

Lời giải.

(a) Biến độc lập (x): Cân nặng

Biến phụ thuộc (y): Huyết áp

Từ bảng dữ liệu, ta có:

$$n = 10;\ \sum_{i=1}^{n} x_i = 817{,}8;\ \sum_{i=1}^{n} x_i^2 = 67793{,}76;\ \sum_{i=1}^{n} y_i = 1401;\ \sum_{i=1}^{n} y_i^2 = 197219;\ \sum_{i=1}^{n} x_i y_i = 115294{,}7.$$

Ta có:

- $S_{xy} = \sum x_i y_i - \frac{1}{n}\sum x_i \sum y_i = 115294{,}7 - \frac{1}{10}\cdot 817{,}8 \cdot 1401 = 720{,}92$
- $S_{xx} = \sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2 = 67793{,}76 - \frac{1}{10}\cdot 817{,}8^2 = 914{,}076$

Suy ra:

- $\widehat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}} = \dfrac{720{,}92}{914{,}076} \approx 0{,}7887$
- $\widehat{\beta}_0 = \dfrac{1}{n}\sum y_i - \widehat{\beta}_1 \dfrac{1}{n}\sum x_i = \dfrac{1}{10} \cdot 1401 - 0{,}7887 \cdot \dfrac{1}{10} \cdot 817{,}8 \approx 75{,}6001$

Vậy phương trình đường hồi quy tuyến tính:

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x = 75{,}6001 + 0{,}7887\, x$$

Ý nghĩa của $\widehat{\beta}_1$: Nếu cân nặng tăng 1 kg thì huyết áp tăng 0,7887 mmHg (về mặt trung bình).

(b) Khi $x = 80$ thì:

$$\widehat{y} = 75{,}6001 + 0{,}7887 \cdot 80 = 138{,}6961$$

Vậy một nam giới có cân nặng 80 kg thì có huyết áp dự đoán là 138,6961 mmHg.

(c) Ta có:

$$S_{yy} = \sum y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2 = 197219 - \frac{1}{10} \cdot 1401^2 = 938{,}9$$

Hệ số tương quan:

$$r_{XY} = \frac{S_{xy}}{\sqrt{S_{xx}\, S_{yy}}} = \frac{720{,}92}{\sqrt{914{,}076 \cdot 938{,}9}} \approx 0{,}7782$$

Ý nghĩa: Cân nặng và huyết áp có tương quan dương (vì $r_{XY} > 0$) và có mối liên hệ tuyến tính khá mạnh (vì r_{XY} có giá trị gần với 1).

(d) Hệ số xác định:

$$R^2 = r_{XY}^2 = 0{,}7782^2 \approx 0{,}6056$$

Ý nghĩa: 60,56% sự thay đổi trong huyết áp được giải thích bởi cân nặng, 39,44% còn lại được giải thích bởi các yếu tố khác.

■

B3 CUỐI KỲ II 2024-2025 (CNTT-23CLC)

Câu 1 (2 điểm)

Một người đăng ký một gói cước data 4G có tối đa 30 GB lưu lượng trong 1 tháng (30 ngày), hết lưu lượng hệ thống dừng truy cập. Biết rằng nhu cầu sử dụng lưu lượng mỗi ngày của người này độc lập nhau và có cùng phân phối chuẩn $\mathcal{N}\left(1{,}1\ ;\ 0{,}2^2\right)$ (đơn vị: GB).

(a) Tính xác suất để trong 1 tháng (30 ngày) người này không sử dụng hết lưu lượng.

(b) Tính xác suất để người này bị dừng truy cập trong vòng 4 tuần (28 ngày).

(c) Nếu nhà mạng thêm ràng buộc mỗi ngày được sử dụng tối đa 1 GB, hết lưu lượng hệ thống dừng truy cập, thì trong 1 tháng có trung bình bao nhiêu ngày người này bị dừng truy cập?

Lời giải.

(a) Gọi X (GB) là lưu lượng 4G mà người này sử dụng mỗi ngày.

Theo đề: $X \sim \mathcal{N}\left(1{,}1\ ;\ 0{,}2^2\right)$

Gọi S_n là (GB) là lưu lượng 4G mà người này sử dụng trong n ngày.

Suy ra: $S_{30} \sim \mathcal{N}\left(\mu_{30}\ ;\ \sigma_{30}^2\right)$ với:

- $\mu_{30} = 30 \cdot 1{,}1 = 33$
- $\sigma_{30} = \sqrt{30 \cdot 0{,}2^2} \approx 1{,}0954$

Xác suất để người này không sử dụng hết lưu lượng trong 1 tháng:

$$P(S_{30} < 30) = P\left(\frac{S_{30} - 33}{1{,}0954} < \frac{30 - 33}{1{,}0954}\right) \approx P(Z_{30} < -2{,}74) = 1 - \Phi(2{,}74) = 1 - 0{,}9969 = 0{,}0031$$

(b) Ta có: $S_{30} \sim \mathcal{N}\left(\mu_{28}\ ;\ \sigma_{28}^2\right)$ với:

- $\mu_{28} = 28 \cdot 1{,}1 = 30{,}8$
- $\sigma_{28} = \sqrt{28 \cdot 0{,}2^2} \approx 1{,}0583$

Xác suất để người này bị dừng truy cập trong vòng 4 tuần:

$$\begin{aligned} P(S_{28} > 30) &= 1 - P(S_{28} \leq 30) = 1 - P\left(\frac{S_{28} - 30{,}8}{1{,}0583} \leq \frac{30 - 30{,}8}{1{,}0583}\right) \\ &\approx 1 - P(Z_{28} \leq -0{,}76) \\ &= 1 - \left[1 - \Phi(0{,}76)\right] = \Phi(0{,}76) = 0{,}7764 \end{aligned}$$

(c) Xác suất người này bị dừng truy cập trong ngày:

$$\begin{aligned} P(X>1) = 1-P(X\le 1) &= 1-P\left(\frac{X-1{,}1}{0{,}2}\le\frac{1-1{,}1}{0{,}2}\right)\\ &= 1-P(Z\le -0{,}5)\\ &= 1-\left[1-\Phi(0{,}5)\right] = \Phi(0{,}5)=0{,}6915 \end{aligned}$$

Gọi Y là số ngày người này bị dừng truy cập trong 1 tháng.

Ta có: $Y\sim B(30\,;\,0{,}6915)$

Trung bình số ngày người này bị dừng truy cập trong 1 tháng:

$$\mathbb{E}(Y)=np=30\cdot 0{,}6915=20{,}745\quad(\text{ngày})$$

■

Câu 2 (3 điểm)

Thời gian phản hồi của hệ thống máy chủ khi xử lý yêu cầu của người dùng là một biến ngẫu nhiên có phân phối chuẩn với độ lệch chuẩn là 5 ms (mili giây). Một kỹ sư hệ thống ghi nhận thời gian phản hồi trung bình của hệ thống trên 25 yêu cầu là 42 ms.

(a) Tính khoảng tin cậy 96% cho thời gian phản hồi trung bình của hệ thống.

(b) Có một nhận định rằng: "Thời gian phản hồi trung bình của hệ thống là trên 40 ms". Hãy kiểm định nhận định này với mức ý nghĩa 2%.

(c) So với một hệ thống khác, thời gian phản hồi trung bình là 45 ms trên 16 yêu cầu. Kiểm tra xem có sự khác biệt đáng kể giữa thời gian phản hồi của hai hệ thống này không, với mức ý nghĩa 2%. Giả sử hai phương sai tổng thể là bằng nhau.

Lời giải.

(a) Gọi X (ms) là thời gian phản hồi của hệ thống.

Ta có: $X\sim\mathcal{N}(\mu\,;\,\sigma^2)$ với $\sigma=5$

- Phân phối mẫu: $Z=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim\mathcal{N}(0\,;\,1)$
- Ta có: $\overline{x}=42,\quad \sigma=5,\quad n=25$

  Độ tin cậy 96% $\Rightarrow$ $\alpha=0{,}04$ $\Rightarrow$ $z_{1-\frac{\alpha}{2}}=z_{0,98}=2{,}06$
- Sai số ước lượng:

$$\varepsilon=z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}=2{,}06\cdot\frac{5}{\sqrt{25}}=2{,}06$$

Vậy khoảng tin cậy 96% cho thời gian phản hồi trung bình của hệ thống là:

$$\mu \in \left[\mu - \varepsilon\,;\ \mu + \varepsilon\right] = \left[39{,}94\,;\ 44{,}06\right] \quad \text{(ms)}$$

(b) Gọi X tương tự câu (**a**).

- Giả thuyết kiểm định: $\begin{cases} H_0 : \mu \le 40 \\ H_1 : \mu > 40 \end{cases}$
- Dưới giả định H_0 đúng, thống kê kiểm định:

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0\,;\ 1)$$

- Ta có: $\overline{x} = 42,\quad \mu_0 = 40,\quad \sigma = 5,\quad n = 25$
  suy ra:

$$z_0 = \frac{42 - 40}{\dfrac{5}{\sqrt{25}}} = 2$$

- Miền bác bỏ: $W_{0{,}02} = \{\, z_0 : z_0 > z_{1-\alpha} = z_{0{,}98} = 2{,}06 \,\}$
- Ta có: $z_0 = 2 < 2{,}06 \quad \Rightarrow \quad z_0 \notin W_{0{,}02}$

Vậy với mức ý nghĩa 2%, chưa đủ cơ sở bác bỏ H_0, nghĩa là thời gian phản hồi trung bình của hệ thống **không** lớn hơn 40 ms.

(c) Gọi X_1 (ms) và X_2 (ms) lần lượt là thời gian phản hồi của hệ thống hiện tại và hệ thống khác.
Ta có: $X_1 \sim \mathcal{N}\left(\mu_1\,;\ \sigma_1^2\right),\quad X_2 \sim \mathcal{N}\left(\mu_2\,;\ \sigma_2^2\right)$ với $\sigma_1 = \sigma_2 = 5$

- Giả thuyết kiểm định: $\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \ne 0 \end{cases}$
- Dưới giả định H_0 đúng, thống kê kiểm định:

$$Z_0 = \frac{\overline{X_1} - \overline{X_2} - D_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0\,;\ 1)$$

- Ta có: $\overline{x}_1 = 42 \quad \sigma_1 = 5 \quad n_1 = 25$
  $\overline{x}_2 = 45 \quad \sigma_2 = 5 \quad n_2 = 16$
  $D_0 = 0$
  suy ra:

$$z_0 = \frac{42 - 45 - 0}{\sqrt{\dfrac{5^2}{25} + \dfrac{5^2}{16}}} \approx -1{,}8741$$

- Miền bác bỏ: $W_{0{,}02} = \{\, z_0 : |z_0| > z_{1-\frac{\alpha}{2}} = z_{0{,}99} = 2{,}33 \,\}$
- Ta có: $|z_0| = 1{,}8741 < 2{,}33 \quad \Rightarrow \quad z_0 \notin W_{0{,}02}$

Vậy với mức ý nghĩa 2%, chưa đủ cơ sở bác bỏ H_0, nghĩa là không có sự khác biệt đáng kể giữa thời gian phản hồi trung bình của hai hệ thống.

■

Câu 3 (3 điểm)

Công ty phần mềm kiểm tra chất lượng hai phiên bản phần mềm:
Phiên bản 1: 150 người dùng thử nghiệm, 18 người gặp lỗi.
Phiên bản 2: 200 người dùng thử nghiệm, 30 người gặp lỗi.

(a) Tính khoảng tin cậy 97% cho tỷ lệ lỗi của phiên bản 1.

(b) Hãy kiểm định tuyên bố "tỷ lệ lỗi của phần mềm là 10%" của phiên bản 1, với mức ý nghĩa 2%. Tìm *p-giá trị*.

(c) Hỏi có sự khác biệt giữa tỷ lệ lỗi của hai phiên bản phần mềm hay không với mức ý nghĩa 2%. Tìm *p-giá trị*.

Lời giải.

(a) Gọi Y là số người gặp lỗi khi dùng phần mềm phiên bản 1. Ta có: $Y \sim B(n\,;\,p)$

- Phân phối mẫu: $Z = \dfrac{\widehat{P} - p}{\sqrt{\dfrac{\widehat{P}\left(1 - \widehat{P}\right)}{n}}} \sim \mathcal{N}(0\,;\,1)$

- Ta có: $\widehat{p} = \dfrac{18}{150} = 0{,}12, \quad n = 150$

  Độ tin cậy 97% $\Rightarrow$ $\alpha = 0{,}03$ $\Rightarrow$ $z_{1-\frac{\alpha}{2}} = z_{0{,}985} = 2{,}17$

- Sai số ước lượng:

$$\varepsilon = z_{1-\frac{\alpha}{2}} \sqrt{\frac{\widehat{p}\left(1 - \widehat{p}\right)}{n}} = 2{,}17 \cdot \sqrt{\frac{0{,}12\,(1 - 0{,}12)}{150}} \approx 0{,}0576$$

Vậy khoảng tin cậy 97% cho tỷ lệ lỗi của phiên bản 1 là:

$$p \in \left[\widehat{p} - \varepsilon\,;\, \widehat{p} + \varepsilon\right] = \left[0{,}0624\,;\, 0{,}1776\right]$$

(b) Gọi Y tương tự câu **(a)**.

- Giả thuyết kiểm định: $\begin{cases} H_0 : p = 0{,}1 \\ H_1 : p \neq 0{,}1 \end{cases}$

- Dưới giả định H_0 đúng, thống kê kiểm định:

$$Z_0 = \frac{\widehat{P} - p_0}{\sqrt{\dfrac{p_0\left(1 - p_0\right)}{n}}} \sim \mathcal{N}\left(0\ ;\ 1\right)$$

- Ta có: $\widehat{p} = 0{,}12, \quad \widehat{p}_0 = 0{,}1 \quad n = 150$
  suy ra:

$$z_0 = \frac{0{,}12 - 0{,}1}{\sqrt{\dfrac{0{,}1\left(1 - 0{,}1\right)}{150}}} \approx 2{,}5820$$

- Miền bác bỏ: $W_{0,02} = \{\, z_0 : |z_0| > z_{1-\frac{\alpha}{2}} = z_{0,99} = 2{,}33 \,\}$
- Ta có: $|z_0| = 2{,}5820 > 2{,}33 \quad \Rightarrow \quad z_0 \in W_{0,03}$

Vậy với mức ý nghĩa 3%, bác bỏ H_0 và chấp nhận H_1, nghĩa là tỷ lệ lỗi của phần mềm **khác** 10%.

$$p\text{-giá trị} = 2\left[1 - \Phi(|z_0|)\right] \approx 2\left[1 - \Phi(2{,}58)\right] = 2\left(1 - 0{,}9951\right) = 0{,}0098$$

(c) Gọi Y_1 và Y_2 lần lượt là số người dùng gặp lỗi ở phiên bản 1 và phiên bản 2.
Ta có: $Y_1 \sim \mathcal{N}\left(n_1\ ;\ p_1\right), \quad Y_2 \sim \mathcal{N}\left(n_2\ ;\ p_2\right)$

- Giả thuyết kiểm định: $\begin{cases} H_0 : p_1 - p_2 = 0 \\ H_1 : p_1 - p_2 \neq 0 \end{cases}$
- Dưới giả định H_0 đúng, thống kê kiểm định:

$$Z_0 = \frac{\widehat{P}_1 - \widehat{P}_2 - D_0}{\sqrt{\widehat{P}\left(1 - \widehat{P}\right)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \sim \mathcal{N}\left(0\ ;\ 1\right)$$

- Ta có: $\widehat{p}_1 = \dfrac{18}{150} = 0{,}12 \qquad \widehat{p}_2 = \dfrac{30}{200} = 0{,}15 \qquad \widehat{p} = \dfrac{18 + 30}{150 + 200} = \dfrac{24}{175}$

  $n_1 = 150 \qquad n_2 = 200 \qquad D_0 = 0$

  suy ra:

$$z_0 = \frac{0{,}12 - 0{,}15 - 0}{\sqrt{\dfrac{24}{175}\left(1 - \dfrac{24}{175}\right)\left(\dfrac{1}{150} + \dfrac{1}{200}\right)}} \approx -0{,}8074$$

- Miền bác bỏ: $W_{0,02} = \{\, z_0 : |z_0| > z_{1-\frac{\alpha}{2}} = z_{0,99} = 2{,}33 \,\}$
- Ta có: $|z_0| = 0{,}8074 < 2{,}33 \quad \Rightarrow \quad z_0 \notin W_{0,02}$

Vậy với mức ý nghĩa 2%, chưa đủ cơ sở bác bỏ H_0, nghĩa là không có sự khác biệt giữa tỷ lệ lỗi của hai phiên bản phần mềm.

$$p\text{-giá trị} = 2\left[1 - \Phi(|z_0|)\right] \approx 2\left[1 - \Phi(0{,}81)\right] = 2\left(1 - 0{,}7910\right) = 0{,}418$$

■

Câu 4 (2 điểm)

Một trường đại học nghiên cứu mối quan hệ giữa số giờ ôn thi (biến độc lập X) và điểm thi (biến phụ thuộc Y) của sinh viên. Dữ liệu thu thập từ 10 sinh viên được tóm tắt như sau:

$$\sum_{i=1}^{10} x_i = 65; \quad \sum_{i=1}^{10} x_i^2 = 505; \quad \sum_{i=1}^{10} y_i = 72{,}5; \quad \sum_{i=1}^{10} y_i^2 = 546{,}25; \quad \sum_{i=1}^{10} x_i y_i = 512{,}5.$$

(a) Lập phương trình đường thẳng hồi quy tuyến tính đơn giữa số giờ ôn thi và điểm thi.

(b) Dự báo điểm thi của sinh viên nếu họ ôn thi 6 giờ.

(c) Xác định hệ số xác định R^2 để đánh giá mức độ phù hợp của mô hình hồi quy.

Lời giải.

(a) Ta có:

- $S_{xy} = \sum x_i y_i - \frac{1}{n}\sum x_i \sum y_i = 512{,}5 - \frac{1}{10} \cdot 65 \cdot 72{,}5 = 41{,}25$
- $S_{xx} = \sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2 = 505 - \frac{1}{10} \cdot 65^2 = 82{,}5$

Suy ra:

- $\widehat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{41{,}25}{82{,}5} = 0{,}5$
- $\widehat{\beta}_0 = \frac{1}{n}\sum y_i - \widehat{\beta}_1 \frac{1}{n}\sum x_i = \frac{1}{10} \cdot 72{,}5 - 0{,}5 \cdot \frac{1}{10} \cdot 65 = 4$

Vậy phương trình đường hồi quy tuyến tính:

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x = 4 + 0{,}5x$$

(b) Khi $x = 6$ thì:

$$\widehat{y} = 4 + 0{,}5 \cdot 6 = 7$$

Vậy khi một sinh viên ôn thi 6 (giờ) thì có điểm dự đoán là 7 (điểm).

(c) Ta có:

- $SSR = \widehat{\beta}_1 S_{xy} = 0{,}5 \cdot 41{,}25 = 20{,}625$
- $SST = \sum y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2 = 546{,}25 - \frac{1}{10} \cdot 72{,}5^2 = 20{,}625$

Hệ số xác định:

$$R^2 = \frac{SSR}{SST} = \frac{20{,}625}{20{,}625} = 1$$

Đánh giá: Mô hình hồi quy trên hoàn toàn phù hợp để dự đoán điểm thi của sinh viên dựa trên số giờ ôn thi của họ.

■

B4 CUỐI KỲ I 2024-2025 (CA 1)

Câu 1 (2 điểm)

Thời gian xử lý của một thuật toán trên một máy tính cụ thể có phân phối chuẩn với trung bình là 250 ms và độ lệch chuẩn là 15 ms. Gọi $\overline{X}_n$ là trung bình thời gian xử lý của thuật toán trong một mẫu ngẫu nhiên gồm n lần chạy.

(a) Tìm n sao cho $P(248 < \overline{X}_n < 252) = 0{,}95$.

(b) Chạy thuật toán 300 lần. Sử dụng xấp xỉ chuẩn có hiệu chỉnh để tìm xác suất có ít nhất 70 lần trong số đó có thời gian xử lý dài hơn 260 ms.

Lời giải.

(a) Gọi X (ms) là thời gian xử lý của thuật toán trong một lần chạy.

Theo đề: $X \sim \mathcal{N}\left(250\,;\,15^2\right)$

Ta có: $\overline{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$

Theo định lý giới hạn trung tâm, ta có $Z_n = \dfrac{\overline{X}_n - \mu}{\sigma} \sim \mathcal{N}\left(0\,;\,1\right)$ với:

- $\mu = \dfrac{1}{n} \cdot n\sigma_X = \sigma_X = 250$
- $\sigma = \sqrt{\dfrac{1}{n^2} \cdot n\sigma_X^2} = \dfrac{\sigma_X}{\sqrt{n}} = \dfrac{15}{\sqrt{n}}$

Ta có:

$$\begin{aligned}
& P(248 < \overline{X}_n < 252) &&= 0{,}95 \\
\Leftrightarrow\ & P\left(\frac{248-250}{15/\sqrt{n}} < \frac{\overline{X}_n - 250}{15/\sqrt{n}} < \frac{252-250}{15/\sqrt{n}}\right) &&= 0{,}95 \\
\Leftrightarrow\ & \Phi\left(\frac{2\sqrt{n}}{15}\right) - \Phi\left(-\frac{2\sqrt{n}}{15}\right) &&= 0{,}95 \\
\Leftrightarrow\ & 2\,\Phi\left(\frac{2\sqrt{n}}{15}\right) - 1 &&= 0{,}95 \\
\Leftrightarrow\ & \Phi\left(\frac{2\sqrt{n}}{15}\right) &&= 0{,}975 \\
\Leftrightarrow\ & \frac{2\sqrt{n}}{15} &&= 1{,}96 \\
\Leftrightarrow\ & n &&\approx 217
\end{aligned}$$

(b) Ta có:

$$\begin{aligned} P(X > 260) &= 1 - P(X \le 260) = 1 - P\left(\frac{X-250}{15} \le \frac{260-250}{15}\right) \\ &\approx 1 - P(Z < 0{,}67) \\ &= 1 - \Phi(0{,}67) = 1 - 0{,}7486 = 0{,}2514 \end{aligned}$$

Gọi Y là số lần thuật toán có thời gian xử lý dài hơn 260 ms trong 300 lần chạy.

Ta có: $Y \sim B\,(300\,;\,0{,}2514)$

Xấp xỉ phân phối nhị thức bởi phân phối chuẩn, ta có : $Z = \dfrac{Y - \mu_Y}{\sigma_Y} \sim \mathcal{N}\,(0\,;\,1)$ với:

- $\mu_Y = 300 \cdot 0{,}2514 = 75{,}42$
- $\sigma_Y = \sqrt{300 \cdot 0{,}2514 \cdot (1 - 0{,}2514)} \approx 7{,}5139$

Xác suất có ít nhất 70 lần thuật toán có thời gian xử lý dài hơn 260 ms:

$$P(Y \ge 70) = 1 - P(Y < 70) \approx 1 - \Phi\left(\frac{70 - 75{,}42 - 0{,}5}{7{,}5139}\right) \approx 1 - \Phi(-0{,}79) = \Phi(0{,}79) = 0{,}7852$$

■

Câu 2 (2 điểm)

Một công ty phát triển phần mềm muốn ước lượng tỷ lệ lỗi p (bugs) trong các module phần mềm của một dự án lớn. Dựa trên một mẫu kiểm thử lớn, tỷ lệ mẫu $\widehat{p}$ là 0,25 và dung sai (sai số) của khoảng tin cậy 95% cho p là 0,06.

(a) Tìm khoảng tin cậy 99% cho p.

(b) Tìm cỡ mẫu tối thiểu để dung sai của khoảng tin cậy trong câu **(a)** không quá 0,06.

Lời giải.

(a) Gọi Y là số lỗi trong module phần mềm. Ta có: $Y \sim B\,(n\,;\,p)$

- Phân phối mẫu: $Z = \dfrac{\widehat{P} - p}{\sqrt{\dfrac{\widehat{P}\,(1 - \widehat{P})}{n}}} \sim \mathcal{N}\,(0\,;\,1)$
- Ta có: $\widehat{p} = 0{,}25$

  Độ tin cậy 95% $\Rightarrow$ $\alpha_1 = 0{,}05$ $\Rightarrow$ $z_{1-\frac{\alpha_1}{2}} = z_{0{,}975} = 1{,}96$

  Độ tin cậy 99% $\Rightarrow$ $\alpha_2 = 0{,}01$ $\Rightarrow$ $z_{1-\frac{\alpha_1}{2}} = z_{0{,}995} = 2{,}58$

- Sai số ước lượng:

$$\varepsilon_1 = z_{1-\frac{\alpha_1}{2}} \sqrt{\frac{\widehat{p}\,(1-\widehat{p}\,)}{n}} \quad (1) \quad ; \quad \varepsilon_2 = z_{1-\frac{\alpha_2}{2}} \sqrt{\frac{\widehat{p}\,(1-\widehat{p}\,)}{n}} \quad (2)$$

Từ (1) và (2), suy ra:

$$\varepsilon_2 = \frac{z_{1-\frac{\alpha_2}{2}}\ \varepsilon_1}{z_{1-\frac{\alpha_2}{2}}} = \frac{2{,}58 \cdot 0{,}06}{1{,}96} \approx 0{,}0790$$

Vậy khoảng tin cậy 99% cho p là:

$$p \in \left[\widehat{p} - \varepsilon\ ;\ \widehat{p} + \varepsilon\right] = \left[0{,}171\ ;\ 0{,}329\right]$$

(b) Để sai số ước lượng không quá $\varepsilon^* = 0{,}06$ thì:

$$\varepsilon = z_{1-\frac{\alpha_2}{2}} \sqrt{\frac{\widehat{p}\,(1-\widehat{p}\,)}{n}} \leq \varepsilon^* \quad \Rightarrow \quad n \geq \left(\frac{z_{1-\frac{\alpha_2}{2}}}{\varepsilon^*}\right)^2 \widehat{p}\,(1-\widehat{p}\,) = \left(\frac{2{,}58}{0{,}06}\right)^2 0{,}25\ (1-0{,}25) = 346{,}6875$$

Vậy để sai số ước lượng không quá 0,06 (với độ tin cậy 99%) thì cỡ mẫu tối thiểu là 347.

■

Câu 3 (4 điểm)

Mục tiêu của một nghiên cứu là so sánh hiệu suất của hai hệ thống lưu trữ dữ liệu trong môi trường đám mây, phục vụ cho các ứng dụng doanh nghiệp. Hệ thống A và hệ thống B đều được triển khai để lưu trữ và truy xuất dữ liệu từ các ứng dụng quản lý cơ sở dữ liệu. Việc so sánh hiệu suất của hai hệ thống này là rất quan trọng để lựa chọn hệ thống tối ưu cho doanh nghiệp, đảm bảo tốc độ truy xuất và độ bền vững của dữ liệu. Dữ liệu thu thập được từ hai hệ thống như sau:

- Hệ thống A: số thử nghiệm $m = 20$, trung bình thời gian truy xuất dữ liệu $\overline{x} = 150$ ms, độ lệch chuẩn $s_1 = 12$ ms.
- Hệ thống B: số thử nghiệm $n = 20$, trung bình thời gian truy xuất dữ liệu $\overline{y} = 160$ ms, độ lệch chuẩn $s_2 = 14$ ms.

Biết rằng thời gian truy xuất dữ liệu ở hai hệ thống có phân phối chuẩn.

(a) Có giả thuyết cho rằng trung bình thời gian truy xuất dữ liệu của hệ thống A ít hơn 155 ms. Hãy kiểm định giả thuyết này với mức ý nghĩa 5%.

(b) Liệu trung bình thời gian truy xuất của hệ thống B có dài hơn hệ thống A không? Giả sử phương sai của hai hệ thống là bằng nhau, và sử dụng mức ý nghĩa 5%.

Lời giải.

(a) Gọi X (ms) là thời gian truy xuất dữ liệu của hệ thống A.

Ta có: $X \sim \mathcal{N}\left(\mu\,;\,\sigma^2\right)$, với σ chưa biết, $n = 20 < 30$

- Giả thuyết kiểm định: $\begin{cases} H_0 : \mu \geq 155 \\ H_1 : \mu < 155 \end{cases}$
- Dưới giả định H_0 đúng, thống kê kiểm định:

$$T_0 = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim T(19)$$

- Ta có: $\overline{x} = 150$, $\mu_0 = 155$, $s = 12$, $n = 20$

  suy ra:

$$t_0 = \frac{150 - 155}{\dfrac{12}{\sqrt{20}}} \approx -1{,}8634$$

- Miền bác bỏ: $W_{0,05} = \{\, t_0 : t_0 < -\,t_{1-\alpha}^{n-1} = -\,t_{0,95}^{19} = -1{,}7291 \,\}$
- Ta có: $t_0 = -1{,}8634 < -1{,}7291 \quad \Rightarrow \quad t_0 \in W_{0,05}$

Vậy với mức ý nghĩa 5%, bác bỏ H_0 và chấp nhận H_1, nghĩa là trung bình thời gian truy xuất của hệ thống ít hơn 155 ms.

(b) Gọi X_1 (ms) và X_2 (ms) lần lượt là thời truy xuất dữ liệu của hệ thống A và hệ thống B.

Ta có: $X_1 \sim \mathcal{N}\left(\mu_1\,;\,\sigma_1^2\right)$, σ_1 chưa biết, $n_1 = 20 < 30$

$X_2 \sim \mathcal{N}\left(\mu_2\,;\,\sigma_2^2\right)$, σ_2 chưa biết, $n_2 = 20 < 30$

$\sigma_1^2 = \sigma_2^2$

- Giả thuyết kiểm định: $\begin{cases} H_0 : \mu_1 - \mu_2 \geq 0 \\ H_1 : \mu_1 - \mu_2 < 0 \end{cases}$
- Dưới giả định H_0 đúng, thống kê kiểm định:

$$T_0 = \frac{\overline{X_1} - \overline{X_2} - D_0}{\sqrt{S_P^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \sim T(38)$$

- Ta có: $\overline{x}_1 = 150$ $s_1 = 12$ $n_1 = 20$

  $\overline{x}_2 = 160$ $s_2 = 14$ $n_2 = 20$

  $D_0 = 0$

  suy ra:

$$s_P^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{19 \cdot 12^2 + 19 \cdot 14^2}{20 + 20 - 2} = 170$$

$$t_0 = \frac{150 - 160 - 0}{\sqrt{170\left(\dfrac{1}{20} + \dfrac{1}{20}\right)}} \approx -2{,}4254$$

- Miền bác bỏ: $W_{0,05} = \{ t_0 : t_0 < -t_{1-\alpha}^{n_1+n_2-2} = -t_{0,95}^{38} \approx -z_{0,95} = -1,65 \}$
- Ta có: $t_0 = -2,4254 < -1,65 \quad \Rightarrow \quad t_0 \in W_{0,05}$

Vậy với mức ý nghĩa 5%, bác bỏ H_0 và chấp nhận H_1, nghĩa là trung bình thời gian truy xuất của hệ thống B dài hơn hệ thống A.

Note: Khi n lớn ($n > 30$) thì $t_{1-\alpha}^{n} \approx z_{1-\alpha}$.

■

Câu 4 (2 điểm)

Bảng dưới đây trình bày thời gian phản hồi trung bình (y, đv: s) của một hệ thống máy chủ đối với các kích thước cơ sở dữ liệu khác nhau (x, đv: GB).

x (GB)	0,25	0,75	1,25	1,75	2,25	2,75	3,25	3,75	4,00	4,30	4,55
y (s)	1,087	1,228	1,583	1,798	1,939	2,138	2,172	2,315	2,455	2,735	2,954

(a) Tìm phương trình hồi quy tuyến tính đơn y theo x. Giải thích ý nghĩa của $\widehat{\beta}_1$ nhận được.

(b) Hãy dự đoán thời gian phản hồi của hệ thống khi kích thước cơ sở dữ liệu là 1,55 GB.

Lời giải.

(a) Từ bảng dữ liệu, ta có:

$$n = 11;\ \sum_{i=1}^{n} x_i = 28,85;\ \sum_{i=1}^{n} x_i^2 = 97,6925;\ \sum_{i=1}^{n} y_i = 22,404\ \sum_{i=1}^{n} y_i^2 \approx 49,0692;\ \sum_{i=1}^{n} x_i y_i = 67,3217.$$

Ta có:

- $S_{xy} = \sum x_i y_i - \dfrac{1}{n}\sum x_i \sum y_i = 67,3217 - \dfrac{1}{11}\cdot 28,85 \cdot 22,404 \approx 8,5621$
- $S_{xx} = \sum x_i^2 - \dfrac{1}{n}\left(\sum x_i\right)^2 = 97,6925 - \dfrac{1}{11}\cdot 28,85^2 = 22,0268$

Suy ra:

- $\widehat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}} = \dfrac{8,5621}{22,0268} \approx 0,3887$
- $\widehat{\beta}_0 = \dfrac{1}{n}\sum y_i - \widehat{\beta}_1 \dfrac{1}{n}\sum x_i = \dfrac{1}{11}\cdot 22,404 - 0,3887 \cdot \dfrac{1}{11}\cdot 28,85 \approx 1,0173$

Vậy phương trình đường hồi quy tuyến tính:

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x = 1,0173 + 0,3887\,x$$

Ý nghĩa của $\widehat{\beta}_1$: Nếu kích thước cơ sở dữ liệu tăng 1 GB thì thời gian phản hồi của hệ thống tăng 0,3887 s (về mặt trung bình).

(b) Khi $x = 1,55$ thì:

$$\widehat{y} = 1,0173 + 0,3887 \cdot 1,55 \approx 1,6198$$

Vậy khi kích thước cơ sở dữ liệu là 1,55 GB thì hệ thống có thời gian phản hồi dự đoán là 1,6198 s.

■

B5 CUỐI KỲ I 2024-2025 (CA 2)

Câu 1 (2 điểm)

Điện áp đầu ra của một bộ nguồn có phân phối chuẩn với trung bình là 12,0V và độ lệch chuẩn là 0,3V. Gọi $\overline{X}_n$ là trung bình điện áp đầu ra của một mẫu ngẫu nhiên gồm n bộ nguồn.

(a) Với $n = 30$, tìm xác suất $P(11{,}8 < \overline{X}_{30} < 12{,}2)$.

(b) Kiểm tra ngẫu nhiên 200 bộ nguồn. Sử dụng xấp xỉ chuẩn có hiệu chỉnh để tìm xác suất có ít nhất 40 bộ nguồn có điện áp thấp hơn 11,7V.

Lời giải.

(a) Gọi X (V) là điện áp đầu ra của một bộ nguồn.

Theo đề: $X \sim \mathcal{N}\left(12\ ;\ 0{,}3^2\right)$

Ta có: $\overline{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$

Theo định lý giới hạn trung tâm, ta có $Z_n = \dfrac{\overline{X}_n - \mu}{\sigma} \sim \mathcal{N}(0\ ;\ 1)$ với:

- $\mu = \dfrac{1}{n} \cdot n\sigma_X = \sigma_X = 12$
- $\sigma = \sqrt{\dfrac{1}{n^2} \cdot n\sigma_X^2} = \dfrac{\sigma_X}{\sqrt{n}} = \dfrac{0{,}3}{\sqrt{30}} \approx 0{,}0548$

Ta có:

$$\begin{aligned}
P(11{,}8 < \overline{X}_{30} < 12{,}2) &= P\left(\frac{11{,}8-12}{0{,}0548} < \frac{\overline{X}_{30}-12}{0{,}0548} < \frac{12{,}2-12}{0{,}0548}\right) \\
&\approx P(-3{,}65 < Z < 3{,}65) \\
&= \Phi(3{,}65) - \Phi(-3{,}65) \\
&= 2\,\Phi(3{,}65) - 1 \ \approx\ 2 \cdot 1 - 1 \ =\ 1
\end{aligned}$$

(b) Ta có:

$$P(X < 11{,}7) = P\left(\frac{X-12}{0{,}3} < \frac{11{,}7-12}{0{,}3}\right) \approx P(Z < -1) = 1 - \Phi(1) = 1 - 0{,}8413 = 0{,}1587$$

Gọi Y là số bộ nguồn có điện áp thấp hơn 11,7V trong 200 bộ nguồn.

Ta có: $Y \sim B\,(200\ ;\ 0{,}1587)$

Xấp xỉ phân phối nhị thức bởi phân phối chuẩn, ta có : $Z = \dfrac{Y - \mu_Y}{\sigma_Y} \sim \mathcal{N}(0\ ;\ 1)$ với:

- $\mu_Y = 200 \cdot 0{,}1587 = 31{,}74$
- $\sigma_Y = \sqrt{200 \cdot 0{,}1587 \cdot (1 - 0{,}1587)} \approx 5{,}1675$

Xác suất có ít nhất 40 bộ nguồn có điện áp thấp hơn 11,7V:

$$P(Y \geq 40) = 1 - P(Y < 40) \approx 1 - \Phi\left(\frac{40 - 31{,}74 - 0{,}5}{5{,}1675}\right)$$
$$\approx 1 - \Phi(1{,}50) = 1 - 0{,}9332 = 0{,}0668$$

■

Câu 2 (2 điểm)

Một công ty sản xuất cảm biến nhiệt độ muốn ước lượng tỷ lệ cảm biến bị lỗi (p) trong một lô hàng lớn. Dựa trên một mẫu lớn các cảm biến, tỷ lệ mẫu $\hat{p}$ là 0,28 và dung sai (sai số) của khoảng tin cậy 95% cho p là 0,055.

(a) Tìm khoảng tin cậy 99% cho p.

(b) Tìm cỡ mẫu tối thiểu để dung sai của khoảng tin cậy trong câu **(a)** không quá 0,055.

Lời giải.

(a) Gọi Y là số cảm biến bị lỗi trong lô hàng. Ta có: $Y \sim B(n\,;\,p)$

- Phân phối mẫu: $Z = \dfrac{\hat{P} - p}{\sqrt{\dfrac{\hat{P}(1-\hat{P})}{n}}} \sim \mathcal{N}(0\,;\,1)$
- Ta có: $\hat{p} = 0{,}28$
  Độ tin cậy 95% $\Rightarrow$ $\alpha_1 = 0{,}05$ $\Rightarrow$ $z_{1-\frac{\alpha_1}{2}} = z_{0,975} = 1{,}96$
  Độ tin cậy 99% $\Rightarrow$ $\alpha_2 = 0{,}01$ $\Rightarrow$ $z_{1-\frac{\alpha_1}{2}} = z_{0,995} = 2{,}58$
- Sai số ước lượng:

$$\varepsilon_1 = z_{1-\frac{\alpha_1}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad (1) \quad ; \quad \varepsilon_2 = z_{1-\frac{\alpha_2}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad (2)$$

  Từ (1) và (2), suy ra:

$$\varepsilon_2 = \frac{z_{1-\frac{\alpha_2}{2}}\ \varepsilon_1}{z_{1-\frac{\alpha_2}{2}}} = \frac{2{,}58 \cdot 0{,}055}{1{,}96} \approx 0{,}0724$$

Vậy khoảng tin cậy 99% cho p là:

$$p \in \left[\hat{p} - \varepsilon\,;\ \hat{p} + \varepsilon\right] = \left[0{,}2076\,;\ 0{,}3524\right]$$

(b) Để sai số ước lượng không quá $\varepsilon^* = 0{,}055$ thì:

$$\varepsilon = z_{1-\frac{\alpha_2}{2}} \sqrt{\frac{\widehat{p}\,(1-\widehat{p}\,)}{n}} \le \varepsilon^* \quad \Rightarrow \quad n \ge \left(\frac{z_{1-\frac{\alpha_2}{2}}}{\varepsilon^*}\right)^2 \widehat{p}\,(1-\widehat{p}\,) = \left(\frac{2{,}58}{0{,}055}\right)^2 0{,}28\;(1-0{,}28) = 443{,}6133$$

Vậy để sai số ước lượng không quá 0,055 (với độ tin cậy 99%) thì cỡ mẫu tối thiểu là 444.

■

Câu 3 (4 điểm)

Mục tiêu của một nghiên cứu là kiểm tra và so sánh độ bền kéo của hai loại vật liệu composite được sử dụng trong sản xuất các bộ phận máy bay. Các vật liệu này cần có khả năng chịu lực tốt để đảm bảo độ an toàn và hiệu suất của máy bay. Việc kiểm tra độ bền kéo của chúng là rất quan trọng trong việc đánh giá khả năng ứng dụng trong ngành hàng không. Dữ liệu thu thập được từ hai loại vật liệu như sau:

- Vật liệu A: cỡ mẫu $n_A = 15$, trung bình độ bền kéo $\overline{x}_A = 1200$ MPa, độ lệch chuẩn mẫu $s_A = 50$ MPa.
- Vật liệu B: cỡ mẫu $n_B = 10$, trung bình độ bền kéo $\overline{x}_B = 1185$ MPa, độ lệch chuẩn mẫu $s_B = 45$ MPa.

Biết rằng bền kéo của hai vật liệu có phân phối chuẩn.

(a) Có giả thuyết cho rằng trung bình độ bền kéo của vật liệu A là cao hơn 1180 MPa. Hãy kiểm định giả thuyết này với mức ý nghĩa 5%.

(b) Liệu trung bình độ bền kéo của vật liệu B có thấp hơn vật liệu A hay không? Giả sử phương sai của hai loại vật liệu là bằng nhau, và sử dụng mức ý nghĩa 5%.

Lời giải.

(a) Gọi X (MPa) là độ bền kéo của vật liệu A.

Ta có: $X \sim \mathcal{N}\left(\mu\,;\,\sigma^2\right)$, với σ chưa biết, $n = 15 < 30$

- Giả thuyết kiểm định: $\begin{cases} H_0 : \mu \le 1180 \\ H_1 : \mu > 1180 \end{cases}$
- Dưới giả định H_0 đúng, thống kê kiểm định:

$$T_0 = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim T(14)$$

- Ta có: $\overline{x} = 1200,\quad \mu_0 = 1180,\quad s = 50,\quad n = 15$
  suy ra:
$$t_0 = \frac{1200 - 1180}{\dfrac{50}{\sqrt{15}}} \approx 1{,}5492$$

- Miền bác bỏ: $W_{0,05} = \{\, t_0 : t_0 > t_{1-\alpha}^{n-1} = t_{0,95}^{14} = 1{,}7613 \,\}$

- Ta có: $t_0 = 1{,}5492 < 1{,}7613 \quad \Rightarrow \quad t_0 \notin W_{0,05}$

Vậy với mức ý nghĩa 5%, chưa đủ cơ sở bác bỏ H_0, nghĩa là trung bình độ bền kéo của vật liệu A **không** cao hơn 1180 MPa.

(b) Gọi X_1 (ms) và X_2 (ms) lần lượt là độ bền kéo của vật liệu A và vật liệu B.

Ta có: $X_1 \sim \mathcal{N}\left(\mu_1\,;\,\sigma_1^2\right)$, σ_1 chưa biết, $n_1 = 15 < 30$

$X_2 \sim \mathcal{N}\left(\mu_2\,;\,\sigma_2^2\right)$, σ_2 chưa biết, $n_2 = 10 < 30$

$\sigma_1^2 = \sigma_2^2$

- Giả thuyết kiểm định: $\begin{cases} H_0 : \mu_1 - \mu_2 \leq 0 \\ H_1 : \mu_1 - \mu_2 > 0 \end{cases}$

- Dưới giả định H_0 đúng, thống kê kiểm định:
$$T_0 = \frac{\overline{X_1} - \overline{X_2} - D_0}{\sqrt{S_P^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \sim T(23)$$

- Ta có: $\overline{x}_1 = 1200 \quad s_1 = 50 \quad n_1 = 15$
  $\overline{x}_2 = 1185 \quad s_2 = 45 \quad n_2 = 10$
  $D_0 = 0$
  suy ra:
$$s_P^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{14\cdot 50^2 + 9\cdot 45^2}{15+10-2} = \frac{53225}{23}$$
$$t_0 = \frac{1200 - 1185 - 0}{\sqrt{\dfrac{53225}{23}\left(\dfrac{1}{15} + \dfrac{1}{10}\right)}} \approx 0{,}7638$$

- Miền bác bỏ: $W_{0,05} = \{\, t_0 : t_0 > t_{1-\alpha}^{n_1+n_2-2} = t_{0,95}^{23} = 1{,}7139 \,\}$

- Ta có: $t_0 = 0{,}7638 < 1{,}7139 \quad \Rightarrow \quad t_0 \notin W_{0,05}$

Vậy với mức ý nghĩa 5%, chưa đủ cơ sở bác bỏ H_0, nghĩa là trung bình độ bền kéo của vật liệu B **không** thấp hơn vật liệu A.

■

Câu 4 (2 điểm)

Bài báo "Đo đạc thực nghiệm sự truyền nhiệt bức xạ trong hệ thống dòng hỗn hợp khí-rắn" (G. Han, K. Tuzla, và J. Chen, Tạp chí AIChe, 2002:1910-1916) thảo luận về việc hiệu chỉnh một thiết bị đo bức xạ. Một số phép đo đã được thực hiện đối với các giá trị sức điện động của thiết bị đo (x, đv: V) và dòng bức xạ (y, đv: kW/m^2). Dữ liệu như trong bảng sau:

Đầu ra tín hiệu (x)	1,08	2,42	4,17	4,46	5,17	6,92	7,43	7,89	8,31	8,69
Dòng bức xạ (y)	15	31	51	55	67	89	93	94	95	97

(a) Tìm phương trình hồi quy tuyến tính đơn y theo x. Giải thích ý nghĩa của $\widehat{\beta_1}$ nhận được.

(b) Hãy dự đoán dòng bức xạ nếu thiết bị đo đọc được 3V.

Lời giải.

(a) Từ bảng dữ liệu, ta có:

$$n = 10;\ \sum_{i=1}^{n} x_i = 56{,}54;\ \sum_{i=1}^{n} x_i^2 = 380{,}9478;\ \sum_{i=1}^{n} y_i = 687\ \sum_{i=1}^{n} y_i^2 = 55141;\ \sum_{i=1}^{n} x_i y_i = 4576{,}49.$$

Ta có:

- $S_{xy} = \sum x_i y_i - \frac{1}{n}\sum x_i \sum y_i = 4576{,}49 - \frac{1}{10}\cdot 56{,}54\cdot 687 = 692{,}192$
- $S_{xx} = \sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2 = 380{,}9478 - \frac{1}{10}\cdot 56{,}54^2 \approx 61{,}2706$

Suy ra:

- $\widehat{\beta_1} = \frac{S_{xy}}{S_{xx}} = \frac{692{,}192}{61{,}2706} \approx 11{,}2973$
- $\widehat{\beta_0} = \frac{1}{n}\sum y_i - \widehat{\beta_1}\frac{1}{n}\sum x_i = \frac{1}{10}\cdot 687 - 11{,}2973\cdot\frac{1}{10}\cdot 56{,}54 \approx 4{,}8251$

Vậy phương trình đường hồi quy tuyến tính:

$$\widehat{y} = \widehat{\beta_0} + \widehat{\beta_1}x = 4{,}8251 + 11{,}2973\,x$$

Ý nghĩa của $\widehat{\beta_1}$: Nếu đầu ra tín hiệu tăng 1 V thì dòng bức xạ tăng 11,2973 kW/m^2 (về mặt trung bình).

(b) Khi $x = 3$ thì:

$$\widehat{y} = 4{,}8251 + 11{,}2973\cdot 3 = 38{,}717$$

Vậy khi thiết bị đo đọc được đầu ra tín hiệu là 3 V thì dòng bức xạ dự đoán là 38,717 kW/m^2.

■

B6 CUỐI KỲ II 2023-2024

Câu 1 (4 điểm)

Các sinh viên tổ chức một cuộc thi ước tính độ dài của một phút mà không cần dựa vào đồng hồ đeo tay hoặc đồng hồ treo tường hay bất kỳ một thiết bị đo thời gian nào. Thời gian ước đoán (đơn vị: giây) của những người chơi được liệt kê bên dưới.

69 81 39 65 42 21 60 63 66 48 64 70 96 91 65 58

Giả sử rằng thời gian ước đoán của sinh viên tuân theo một phân phối chuẩn.

(a) Tìm khoảng tin cậy 95% cho trung bình số giây dự đoán của các sinh viên.

(b) Sử dụng mức ý nghĩa 0,05 để kiểm tra khẳng định rằng thời gian ước đoán của sinh viên là từ một tổng thể có giá trị trung bình bằng 60 giây. Có vẻ như sinh viên khá giỏi trong việc ước tính một phút?

(c) Ban tổ chức dự định nhân rộng cuộc thi và sẽ trao quà cho những ai đoán số giây chính xác trong khoảng ±5 giây (tức là từ 55 tới 65 giây). Dựa trên kết quả thử nghiệm này, hãy tính khoảng tin cậy 95% cho số người sẽ trúng thưởng nếu cuộc thi cũng có 16 người tham gia.

Lời giải.

(a) Gọi X (giây) là số giây dự đoán của sinh viên.

Ta có: $X \sim \mathcal{N}\left(\mu\,;\,\sigma^2\right)$ với σ chưa biết, $n = 16 < 30$

- Phân phối mẫu: $T = \dfrac{\overline{X} - \mu}{S/\sqrt{n}} \sim T(15)$
- Ta có: $\overline{x} = 62{,}375$, $s \approx 18{,}8569$, $n = 16$

  Độ tin cậy 95% $\Rightarrow$ $\alpha = 0{,}05$ $\Rightarrow$ $t^{n-1}_{1-\frac{\alpha}{2}} = t^{15}_{0{,}975} = 2{,}1314$
- Sai số ước lượng:

$$\varepsilon = t^{n-1}_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} = 2{,}1314 \cdot \frac{18{,}8569}{\sqrt{16}} \approx 10{,}0479$$

Vậy khoảng tin cậy 95% cho số giây dự đoán của sinh viên là:

$$\mu \in \left[\mu - \varepsilon\,;\,\mu + \varepsilon\right] = \left[52{,}3271\,;\,72{,}4229\right] \quad \text{(giây)}$$

(b) Gọi X tương tự câu **(a)**.

- Giả thuyết kiểm định: $\begin{cases} H_0 : \mu = 60 \\ H_1 : \mu \neq 60 \end{cases}$

- Dưới giả định H_0 đúng, thống kê kiểm định:

$$T_0 = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim T(15)$$

- Ta có: $\overline{x} = 62{,}375, \quad \mu_0 = 60, \quad s \approx 18{,}8569, \quad n = 16$
  suy ra:

$$t_0 = \frac{62{,}305 - 60}{\dfrac{18{,}8569}{\sqrt{16}}} \approx 0{,}5038$$

- Miền bác bỏ: $W_{0,05} = \{\, t_0 : |t_0| > t^{n-1}_{1-\frac{\alpha}{2}} = t^{15}_{0,975} = 2{,}1314 \,\}$

- Ta có: $|t_0| = 0{,}5038 < 2{,}1314 \quad \Rightarrow \quad t_0 \notin W_{0,05}$

Vậy với mức ý nghĩa 5%, chưa đủ cơ sở bác bỏ H_0, nghĩa là trung bình thời gian ước đoán của sinh viên là 60 giây.

(c) Gọi Y là số người trúng thưởng trong cuộc thi. Ta có: $Y \sim B\,(n\,;\,p)$

- Phân phối mẫu: $Z = \dfrac{\widehat{P} - p}{\sqrt{\dfrac{\widehat{P}\,(1 - \widehat{P}\,)}{n}}} \sim \mathcal{N}\,(0\,;\,1)$

- Ta có: $\widehat{p} = \dfrac{6}{16} = 0{,}375, \quad n = 16$

  Độ tin cậy 95% $\quad \Rightarrow \quad \alpha = 0{,}05 \quad \Rightarrow \quad z_{1-\frac{\alpha}{2}} = z_{0,975} = 1{,}96$

- Sai số ước lượng:

$$\varepsilon = z_{1-\frac{\alpha}{2}} \sqrt{\frac{\widehat{p}\,(1 - \widehat{p}\,)}{n}} = 1{,}96 \cdot \sqrt{\frac{0{,}375\,(1 - 0{,}375)}{16}} \approx 0{,}2372$$

Khoảng tin cậy 95% cho tỷ lệ người trúng thưởng là:

$$p \in \left[\,\widehat{p} - \varepsilon\,;\,\widehat{p} + \varepsilon\,\right] = \left[\,0{,}1378\,;\,0{,}6122\,\right]$$

Suy ra khoảng tin cậy 95% cho số người trúng thưởng là:

$$y \in \left[\,16 \cdot 0{,}1378\,;\,16 \cdot 0{,}6122\,\right] = \left[\,2{,}2048\,;\,9{,}7952\,\right] \approx \left[\,3\,;\,9\,\right] \quad \text{(người)}$$

■

Câu 2 (3 điểm)

Một nghiên cứu được thực hiện tại Mỹ trong thập niên 80, về sự gia tăng chất phóng xạ trong răng sữa của trẻ em, cụ thể là chất strontium-90. Dữ liệu hàm lượng strontium-90 (đơn vị: mBq/gr canxi) trong răng của 12 trẻ em ở New York và ở Pennsylvania được thu thập:

Pennsylvania	135	132	149	130	151	163	151	142	156	133	138	161
New York	133	140	142	131	134	129	128	140	140	140	137	143

Giả sử rằng, hàm lượng strontium-90 trong răng của trẻ em ở hai thành phố là tuân theo phân phối chuẩn, với phương sai đồng nhất.

(a) Sử dụng mức ý nghĩa 0,05 để kiểm tra khẳng định rằng lượng strontium-90 trung bình của trẻ em Pennsylvania lớn hơn lượng trung bình của trẻ em New York.

(b) Một nhà nghiên cứu cho rằng, tỷ lệ trẻ em có hàm lượng strontium-90 trong răng từ 140 mBq/gr canxi trở lên là giống nhau cho hai thành phố. Bạn có đồng tình với nhận định này? Hãy kiểm định với mức ý nghĩa 0,05.

Lời giải.

(a) Gọi X_1 (mBq/gr canxi) và X_2 (mBq/gr canxi) lần lượt là hàm lượng strontitum-90 trong răng của trẻ em Pennsylvania và New York.

Ta có: $X_1 \sim \mathcal{N}\left(\mu_1 ; \sigma_1^2\right)$, σ_1 chưa biết, $n_1 = 12 < 30$

$X_2 \sim \mathcal{N}\left(\mu_2 ; \sigma_2^2\right)$, σ_2 chưa biết, $n_2 = 12 < 30$

$\sigma_1^2 = \sigma_2^2$

- Giả thuyết kiểm định: $\begin{cases} H_0 : \mu_1 - \mu_2 \leq 0 \\ H_1 : \mu_1 - \mu_2 > 0 \end{cases}$
- Dưới giả định H_0 đúng, thống kê kiểm định:

$$T_0 = \frac{\overline{X_1} - \overline{X_2} - D_0}{\sqrt{S_P^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \sim T(22)$$

- Ta có: $\overline{x}_1 \approx 145{,}0833$ $\quad s_1 \approx 11{,}6186$ $\quad n_1 = 12$

  $\overline{x}_2 \approx 136{,}4167$ $\quad s_2 \approx 5{,}2129$ $\quad n_2 = 12$

  $D_0 = 0$

  suy ra:

$$s_P^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{11 \cdot 11{,}6186^2 + 11 \cdot 5{,}2129^2}{12+12-2} \approx 81{,}0831$$

$$t_0 = \frac{145{,}0833 - 136{,}4167 - 0}{\sqrt{81{,}0831\left(\frac{1}{12} + \frac{1}{12}\right)}} \approx 2{,}3575$$

- Miền bác bỏ: $W_{0,05} = \{\, t_0 : t_0 > t_{1-\alpha}^{n_1+n_2-2} = t_{0,95}^{22} = 1{,}7171 \,\}$
- Ta có: $t_0 = 2{,}3575 > 1{,}7171 \quad \Rightarrow \quad t_0 \in W_{0,05}$

Vậy với mức ý nghĩa 5%, bác bỏ H_0 và chấp nhận H_1, nghĩa là lượng strontium-90 trung bình của trẻ em Pennsylvania lớn hơn của trẻ em New York.

(b) Gọi Y_1 và Y_2 lần lượt là số trẻ em có lượng strontium-90 từ 140 mBq/gr canxi trở lên ở Pennsylvania và New York.

Ta có: $Y_1 \sim \mathcal{N}(n_1\,;\,p_1)$, $\quad Y_2 \sim \mathcal{N}(n_2\,;\,p_2)$

- Giả thuyết kiểm định: $\begin{cases} H_0 : p_1 - p_2 = 0 \\ H_1 : p_1 - p_2 \neq 0 \end{cases}$
- Dưới giả định H_0 đúng, thống kê kiểm định:

$$Z_0 = \frac{\widehat{P}_1 - \widehat{P}_2 - D_0}{\sqrt{\widehat{P}\,(1-\widehat{P}\,)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim \mathcal{N}(0\,;\,1)$$

- Ta có: $\widehat{p}_1 = \frac{7}{12} \qquad \widehat{p}_2 = \frac{6}{12} \qquad \widehat{p} = \frac{7+6}{12+12} = \frac{13}{24}$

  $n_1 = 12 \qquad n_2 = 12 \qquad D_0 = 0$

  suy ra:

$$z_0 = \frac{\frac{7}{12} - \frac{6}{12} - 0}{\sqrt{\frac{13}{24}\left(1 - \frac{13}{24}\right)\left(\frac{1}{12} + \frac{1}{12}\right)}} \approx 0{,}4097$$

- Miền bác bỏ: $W_{0,05} = \{\, z_0 : |z_0| > z_{1-\frac{\alpha}{2}} = z_{0,975} = 1{,}96 \,\}$
- Ta có: $|z_0| = 0{,}4097 < 1{,}96 \quad \Rightarrow \quad z_0 \notin W_{0,05}$

Vậy với mức ý nghĩa 5%, chưa đủ cơ sở bác bỏ H_0, nghĩa là tỷ lệ trẻ em có lượng strontium-90 từ 140 mBq/gr canxi trở lên là giống nhau cho hai thành phố.

■

Câu 3 (3 điểm)

Một nhóm các nhà nghiên cứu nghĩ rằng chỉ số IQ có thể bị ảnh hưởng bởi thể tích não (đơn vị: cm^3). Để làm sáng tỏ nghi vấn này, họ đã thực hiện một nghiên cứu, trong đó, họ đo chỉ số IQ và thể tích não của 11 người tình nguyện viên. Kết quả như sau:

IQ	87	101	103	96	101	96	93	88	97	114	113
Thể tích não	1027	1281	1051	1079	1173	1079	1067	1104	1029	1100	1204

Dựa vào dữ liệu trên:

(a) Tìm hệ số tương quan tuyến tính giữa chỉ số IQ và thể tích não. Nhận xét về kết quả.

(b) Thiết lập mô hình hồi quy tuyến tính dự đoán chỉ số IQ theo thể tích não. Ước lượng các hệ số trong mô hình. Nhận xét và giải thích kết quả ước lượng của các hệ số.

(c) Nếu một người có thể tích não là 1200 cm^3 thì chỉ số IQ sẽ có thể là bao nhiêu?

(d) Từ dữ liệu này, nhóm nghiên cứu đã tính được hệ số xác định R^2 cho mô hình xây dựng ở câu **(b)** là 0,1947. Hãy giải thích kết quả này? Liệu rằng, mô hình được xây dựng có đủ tốt để dự đoán chỉ số IQ dựa trên thể tích não?

Lời giải.

(a) X: Thể tích não; Y: Chỉ số IQ (**Note:** Nếu không đọc kĩ đề sẽ dễ bị đặt nhầm 2 biến X, Y)

Từ bảng dữ liệu, ta có:

$$n = 11;\ \sum_{i=1}^{n} x_i = 12194;\ \sum_{i=1}^{n} x_i^2 = 13580464;\ \sum_{i=1}^{n} y_i = 1089;\ \sum_{i=1}^{n} y_i^2 = 108579;\ \sum_{i=1}^{n} x_i y_i = 1210272.$$

Ta có:

- $S_{xy} = \sum x_i y_i - \frac{1}{n}\sum x_i \sum y_i = 1210272 - \frac{1}{11}\cdot 1089 \cdot 12194 = 3066$
- $S_{xx} = \sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2 = 13580464 - \frac{1}{11}\cdot 12194^2 \approx 62860{,}7273$
- $S_y = \sum y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2 = 108579 - \frac{1}{11}\cdot 1089^2 = 768$

Hệ số tương quan:

$$r_{XY} = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \frac{3066}{\sqrt{62860{,}7273 \cdot 768}} \approx 0{,}4413$$

Nhận xét: Thể tích não và chỉ số IQ có tương quan dương (vì $r_{XY} > 0$) và có mối liên hệ tuyến tính khá yếu (vì r_{XY} có giá trị gần với 0).

(b) Ta có:

- $\widehat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{3066}{62860{,}7273} \approx 0{,}0488$

- $\widehat{\beta}_0 = \dfrac{1}{n}\sum y_i - \widehat{\beta}_1 \dfrac{1}{n}\sum x_i = \dfrac{1}{11} \cdot 1089 - 0{,}0488 \cdot \dfrac{1}{11} \cdot 12194 \approx 44{,}9030$

Vậy phương trình đường hồi quy tuyến tính:

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x = 44{,}903 + 0{,}0488\, x$$

Ý nghĩa của $\widehat{\beta}_1$: Nếu thể tích não tăng 1 cm^3 thì chỉ số IQ tăng 0,0488 đơn vị (về mặt trung bình).

(c) Khi $x = 1200$ thì:

$$\widehat{y} = 44{,}903 + 0{,}0488 \cdot 1200 = 103{,}463$$

Vậy một người có thể tích não 1200 cm^3 thì có chỉ số IQ dự đoán là 103,463 đơn vị.

(d) Ý nghĩa của R^2 : 19,47% sự thay đổi trong chỉ số IQ được giải thích bởi thể tích não, 80,53% còn lại được giải thích bởi các yếu tố khác.

Nhận xét: Mô hình trên không đủ tốt để dự đoán chỉ số IQ dựa trên thể tích não.

■

B7 CUỐI KỲ I 2023-2024 (CA 2)

Câu 1 (3 điểm)

Trong một kho chứa các sản phẩm do máy A và máy B sản xuất, người ta lấy ngẫu nhiên 200 chi tiết thì thấy có 10 chi tiết từ máy A và 190 chi tiết từ máy B. Đo các chi tiết do máy A sản xuất, người ta thu được chiều dài của chúng như sau (đơn vị cm):

23,4 23,5 23,8 24,1 24,4 25,2 25,7 26,1 24,8 24,9

Đo chiều dài các chi tiết do máy B sản xuất, người ta tính được trung bình mẫu là 25,0 cm và độ lệch chuẩn mẫu là 0,85 cm. Giả sử chiều dài các chi tiết do mỗi máy sản xuất đều có phân phối chuẩn.

(a) Tìm khoảng tin cậy 95% cho tỷ lệ các chi tiết do máy A sản xuất có trong kho.

(b) Tìm khoảng tin cậy 95% cho chiều dài trung bình của các chi tiết do máy A sản xuất. Với cùng độ tin cậy, nếu muốn sai số ước lượng bé hơn 0,2 thì cần khảo sát ít nhất bao nhiêu chi tiết máy?

(c) Nếu chiều dài các chi tiết do mỗi máy sản xuất đều tuân theo phân phối chuẩn với độ lệch chuẩn 0,9 cm, có thể kết luận các chi tiết do hai máy sản suất có chiều dài trung bình khác nhau hay không? Biết rằng mức ý nghĩa bằng 1%. Giả sử phương sai hai tổng thể bằng nhau.

Lời giải.

(a) Gọi Y là số chi tiết do máy A sản xuất trong mẫu. Ta có: $Y \sim B(n\,;\,p)$

- Phân phối mẫu: $Z = \dfrac{\widehat{P} - p}{\sqrt{\dfrac{\widehat{P}(1-\widehat{P})}{n}}} \sim \mathcal{N}(0\,;\,1)$
- Ta có: $\widehat{p} = \dfrac{10}{200} = 0{,}05, \quad n = 200$

  Độ tin cậy 95% $\Rightarrow$ $\alpha = 0{,}05$ $\Rightarrow$ $z_{1-\frac{\alpha}{2}} = z_{0,975} = 1{,}96$
- Sai số ước lượng:

$$\varepsilon = z_{1-\frac{\alpha}{2}} \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 1{,}96 \cdot \sqrt{\frac{0{,}05\,(1-0{,}05)}{200}} \approx 0{,}0302$$

Vậy khoảng tin cậy 95% cho tỷ lệ chi tiết do máy A sản xuất là:

$$p \in \left[\widehat{p} - \varepsilon \ ; \ \widehat{p} + \varepsilon\right] = \left[0{,}0198 \ ; \ 0{,}0802\right]$$

(b) Gọi X (cm) là chiều dài của chi tiết do máy A sản xuất.
Ta có: $X \sim \mathcal{N}\left(\mu \ ; \ \sigma^2\right)$ với σ chưa biết, $n = 10 < 30$

- Phân phối mẫu: $T = \dfrac{\overline{X} - \mu}{S/\sqrt{n}} \sim T(9)$
- Ta có: $\overline{x} = 24{,}59$, $s \approx 0{,}9146$, $n = 10$
  Độ tin cậy 95% $\Rightarrow$ $\alpha = 0{,}05$ $\Rightarrow$ $t^{n-1}_{1-\frac{\alpha}{2}} = t^{9}_{0{,}975} = 2{,}2622$
- Sai số ước lượng:

$$\varepsilon = t^{n-1}_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} = 2{,}2622 \cdot \frac{0{,}9146}{\sqrt{10}} \approx 0{,}6543$$

Vậy khoảng tin cậy 95% cho chiều dài trung bình của chi tiết do máy A sản xuất là:

$$\mu \in \left[\mu - \varepsilon \ ; \ \mu + \varepsilon\right] = \left[23{,}9357 \ ; \ 25{,}2443\right] \quad \text{(cm)}$$

Độ tin cậy 95% $\Rightarrow$ $\alpha = 0{,}05$ $\Rightarrow$ $z_{1-\frac{\alpha}{2}} = z_{0{,}975} = 1{,}96$

Để sai số ước lượng nhỏ hơn $\varepsilon^* = 0{,}2$ thì:

$$\varepsilon = z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} < \varepsilon^* \quad \Rightarrow \quad n > \left(z_{1-\frac{\alpha}{2}} \frac{s}{\varepsilon^*}\right)^2 = \left(1{,}96 \cdot \frac{0{,}9146}{0{,}2}\right)^2 \approx 80{,}3368$$

Vậy để sai số ước lượng nhỏ hơn 0,2 (với cùng độ tin cậy) thì cần khảo sát ít nhất 81 chi tiết máy.

(c) Gọi X_1 (cm) và X_2 (cm) lần lượt là chiều dài của chi tiết do máy A và máy B sản xuất.
Ta có: $X_1 \sim \mathcal{N}\left(\mu_1 \ ; \ \sigma_1^2\right)$, $X_2 \sim \mathcal{N}\left(\mu_2 \ ; \ \sigma_2^2\right)$, với $\sigma_1 = \sigma_2 = 0{,}9$

- Giả thuyết kiểm định: $\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \end{cases}$
- Dưới giả định H_0 đúng, thống kê kiểm định:

$$Z_0 = \frac{\overline{X_1} - \overline{X_2} - D_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0 \ ; \ 1)$$

- Ta có: $\overline{x}_1 = 24{,}59$ $\quad \sigma_1 = 0{,}9$ $\quad n_1 = 10$
  $\overline{x}_2 = 25$ $\quad \sigma_2 = 0{,}9$ $\quad n_2 = 190$
  $D_0 = 0$
  suy ra:

$$z_0 = \frac{24{,}59 - 25 - 0}{\sqrt{\dfrac{0{,}9^2}{10} + \dfrac{0{,}9^2}{190}}} \approx -1{,}4041$$

- Miền bác bỏ: $W_{0,01} = \{ z_0 : |z_0| > z_{1-\frac{\alpha}{2}} = z_{0,995} = 2{,}58 \}$
- Ta có: $|z_0| = 1{,}4041 < 2{,}58 \quad \Rightarrow \quad z_0 \notin W_{0,01}$

Vậy với mức ý nghĩa 1%, chưa đủ cơ sở bác bỏ H_0, nghĩa là chiều dài trung bình các chi tiết do hai máy sản xuất là như nhau.

■

Câu 2 (2,5 điểm)

Một công ty sản xuất thuốc diệt côn trùng muốn so sánh hiệu quả của hai loại thuốc. Các kỹ sư hóa chất chọn hai căn phòng có kích thước giống nhau và phun vào tương ứng hai loại thuốc diệt côn trùng 1 và 2. Sau đó ở mỗi phòng, họ thả vào 132 con côn trùng các loại và đếm số côn trùng chết sau 1 giờ. Kết quả cho thấy có 117 côn trùng chết ở phòng 1 và 112 côn trùng chết ở phòng 2.

(a) Với mức ý nghĩa $\alpha = 5\%$, có thể khẳng định rằng loại thuốc diệt côn trùng 1 có hiệu quả diệt trên 90% số lượng côn trùng hay không? Tính *p-giá trị*.

(b) Với dữ liệu thực nghiệm, ta có đủ bằng chứng để kết luận rằng hiệu quả của hai loại thuốc diệt côn trùng là như nhau hay không? $\alpha = 1\%$.

Lời giải.

(a) Gọi Y là số lượng côn trùng chết ở phòng 1. Ta có: $Y \sim B(n\,;\,p)$

- Giả thuyết kiểm định: $\begin{cases} H_0 : p \leq 0{,}9 \\ H_1 : p > 0{,}9 \end{cases}$
- Dưới giả định H_0 đúng, thống kê kiểm định:

$$Z_0 = \frac{\widehat{P} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} \sim \mathcal{N}(0\,;\,1)$$

- Ta có: $\widehat{p} = \dfrac{117}{132} = \dfrac{39}{44}$, $\quad \widehat{p}_0 = 0{,}9 \quad n = 132$

  suy ra:

$$z_0 = \frac{\dfrac{39}{44} - 0{,}9}{\sqrt{\dfrac{0{,}9(1-0{,}9)}{132}}} \approx -0{,}5222$$

- Miền bác bỏ: $W_{0,05} = \{ z_0 : z_0 > z_{1-\alpha} = z_{0,95} = 1{,}65 \}$
- Ta có: $z_0 = -0{,}5222 < 1{,}65 \quad \Rightarrow \quad z_0 \notin W_{0,05}$

Vậy với mức ý nghĩa 5%, chưa đủ cơ sở bác bỏ H_0, nghĩa là thuốc diệt côn trùng loại 1 **không** có hiệu quả diệt trên 90% số lượng côn trùng.

$$p\text{-giá trị} = 1 - \Phi(z_0) \approx 1 - \Phi(-0{,}52) = \Phi(0{,}52) = 0{,}6985$$

(b) Gọi Y_1 và Y_2 lần lượt là số lượng côn trùng chết ở phòng 1 và phòng 2.

Ta có: $Y_1 \sim \mathcal{N}(n_1\,;\,p_1)$, $Y_2 \sim \mathcal{N}(n_2\,;\,p_2)$

- Giả thuyết kiểm định: $\begin{cases} H_0 : p_1 - p_2 = 0 \\ H_1 : p_1 - p_2 \neq 0 \end{cases}$
- Dưới giả định H_0 đúng, thống kê kiểm định:

$$Z_0 = \frac{\widehat{P}_1 - \widehat{P}_2 - D_0}{\sqrt{\widehat{P}(1-\widehat{P})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim \mathcal{N}(0\,;\,1)$$

- Ta có: $\widehat{p}_1 = \frac{117}{132} = \frac{39}{44}$ $\quad \widehat{p}_2 = \frac{112}{132} = \frac{28}{33}$ $\quad \widehat{p} = \frac{117+112}{132+132} = \frac{229}{264}$

  $n_1 = 132$ $\quad n_2 = 132$ $\quad D_0 = 0$

  suy ra:

$$z_0 = \frac{\frac{39}{44} - \frac{28}{33} - 0}{\sqrt{\frac{229}{264}\left(1 - \frac{229}{264}\right)\left(\frac{1}{132}+\frac{1}{132}\right)}} \approx 0{,}9074$$

- Miền bác bỏ: $W_{0,01} = \{ z_0 : |z_0| > z_{1-\frac{\alpha}{2}} = z_{0,995} = 2{,}58 \}$
- Ta có: $|z_0| = 0{,}9074 < 2{,}58 \quad \Rightarrow \quad z_0 \notin W_{0,01}$

Vậy với mức ý nghĩa 1%, chưa đủ cơ sở bác bỏ H_0, nghĩa là hiệu quả của hai loại thuốc diệt côn trùng là như nhau.

■

Câu 3 (3 điểm)

Trong một nghiên cứu về mối liên hệ giữa việc tiếp xúc tiếng ồn và bệnh tăng huyết áp, người ta cho một nhóm $n = 18$ người tiếp xúc với các nguồn âm thanh có cường độ khác nhau và ghi nhận được kết quả thực nghiệm như sau, trong đó x là tiếng ồn (Đơn vị: đê-xi-ben) và y là độ tăng huyết áp (Đơn vị: mi-li-mét thủy ngân):

$$\sum_{i=1}^{n} x_i = 1352; \quad \sum_{i=1}^{n} y_i = 444{,}848; \quad \sum_{i=1}^{n} x_i^2 = 106024; \quad \sum_{i=1}^{n} y_i^2 = 11460{,}641; \quad \sum_{i=1}^{n} x_i y_i = 34853{,}976.$$

(a) Tìm phương trình đường thẳng hồi quy tuyến tính bằng phương pháp bình phương nhỏ nhất. Cho nhận xét về mô hình.

(b) Nếu một người tiếp xúc tiếng ồn có cường độ bằng $x = 68$ (đê-xi-ben) thì độ tăng huyết áp trung bình là bao nhiêu?

(c) Tính hệ số xác định R^2, từ đó suy ra hệ số tương quan mẫu và nhận xét về mối liên hệ tuyến tính giữa tiếng ồn và độ tăng huyết áp.

Lời giải.

(a) Ta có:

- $S_{xy} = \sum x_i y_i - \dfrac{1}{n}\sum x_i \sum y_i = 34853{,}976 - \dfrac{1}{18}\cdot 1352 \cdot 444{,}848 \approx 1440{,}9484$
- $S_{xx} = \sum x_i^2 - \dfrac{1}{n}\left(\sum x_i\right)^2 = 106024 - \dfrac{1}{18}\cdot 1352^2 \approx 4473{,}7778$

Suy ra:

- $\widehat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}} = \dfrac{1440{,}9484}{4473{,}7778} \approx 0{,}3221$
- $\widehat{\beta}_0 = \dfrac{1}{n}\sum y_i - \widehat{\beta}_1 \dfrac{1}{n}\sum x_i = \dfrac{1}{18}\cdot 444{,}848 - 0{,}3221 \cdot \dfrac{1}{18}\cdot 1352 \approx 20{,}2915$

Vậy phương trình đường hồi quy tuyến tính:

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x = 20{,}2915 + 0{,}3221\, x$$

Nhận xét: Tiếng ồn và độ tăng huyết áp có tương quan dương. Khi tiếng ồn tăng 1 đê-xi-ben thì độ tăng huyết áp tăng 0,3221 mi-li-mét thủy ngân.

(b) Khi $x = 68$ thì:

$$\widehat{y} = 20{,}2915 + 0{,}3221 \cdot 68 = 42{,}1943$$

Vậy khi một người tiếp xúc tiếng ồn có cường độ bằng 68 (đê-xi-ben) thì độ tăng huyết áp dự đoán là 42,1943 (mi-li-mét thủy ngân).

(c) Ta có:

- $SSR = \widehat{\beta}_1 S_{xy} = 0{,}3221 \cdot 1440{,}9484 \approx 464{,}1295$
- $SST = \sum y_i^2 - \dfrac{1}{n}\left(\sum y_i\right)^2 = 11460{,}641 - \dfrac{1}{18}\cdot 444{,}848^2 \approx 466{,}7664$

Hệ số xác định:

$$R^2 = \frac{SSR}{SST} = \frac{464{,}1295}{466{,}7664} \approx 0{,}9944$$

Suy ra hệ số tương quan:

$$|r_{XY}| = \sqrt{R^2} = \sqrt{0,9944} \approx 0,9972 \quad \Rightarrow \quad r_{XY} = 0,9972 \quad (\text{vì } \widehat{\beta}_1 > 0)$$

Nhận xét: Tiếng ồn và độ tăng huyết áp có mối liên hệ tuyến tính mạnh (vì r_{XY} có giá trị gần với 1).

■

Câu 4 (1,5 điểm)

Giả sử thời gian X (phút) để một sinh viên được chọn ngẫu nhiên hoàn thành một bài kiểm tra ngắn tuân theo phân phối chuẩn với trung bình $\mu = 14$ (phút) và phương sai σ^2. Biết rằng 10% số sinh viên hoàn thành bài kiểm tra trong thời gian ít hơn 12 phút.

(a) Tìm giá trị của σ.

(b) Nếu giảng viên chọn ngẫu nhiên 15 sinh viên. Tìm xác suất có ít hơn 2 sinh viên hoàn thành bài kiểm tra trong thời gian ít hơn 12 phút.

(c) Nếu giảng viên chọn ngẫu nhiên n sinh viên ($n > 100$). Bằng cách sử dụng xấp xỉ chuẩn có hiệu chỉnh liên tục, ta tính được xác suất để có ít hơn 9 trong số n sinh viên này sẽ hoàn thành bài kiểm tra trong thời gian ít hơn 12 phút là 0,3085. Tìm giá trị của n.

Lời giải.

(a) Ta có: $X \sim \mathcal{N}(14\,;\,\sigma^2)$

$$\begin{aligned}
& P(X < 12) &&= 0,1 \\
\Leftrightarrow \quad & P\left(\frac{X-14}{\sigma} < \frac{12-14}{\sigma}\right) &&= 0,1 \\
\Leftrightarrow \quad & \Phi\left(\frac{-2}{\sigma}\right) &&= 0,1 \\
\Leftrightarrow \quad & 1 - \Phi\left(\frac{2}{\sigma}\right) &&= 0,1 \\
\Leftrightarrow \quad & \Phi\left(\frac{2}{\sigma}\right) &&= 0,9 \\
\Leftrightarrow \quad & \frac{2}{\sigma} &&\approx 1,28 \\
\Leftrightarrow \quad & \sigma &&= 1,5625
\end{aligned}$$

(b) Gọi Y là số sinh viên hoàn thành bài kiểm tra trong thời gian ít hơn 12 phút trong 15 sinh viên.
Ta có: $Y \sim B(15\,;\,0,1)$

Xác suất có ít hơn 2 sinh viên hoàn thành bài kiểm tra trong thời gian ít hơn 12 phút:

$$P(X < 2) = \sum_{k=0}^{1} C_{15}^{k} \cdot 0{,}1^{k} \cdot (1 - 0{,}1)^{15-k} \approx 0{,}5490$$

(c) Gọi S là số sinh viên hoàn thành bài kiểm tra trong thời gian ít hơn 12 phút trong n sinh viên.
Ta có: $S \sim B\left(n\ ;\ 0{,}1\right)$

Xấp xỉ phân phối nhị thức bởi phân phối chuẩn, ta có $Z = \dfrac{S - \mu_S}{\sigma_S} \sim \mathcal{N}\left(0\ ;\ 1\right)$ với:

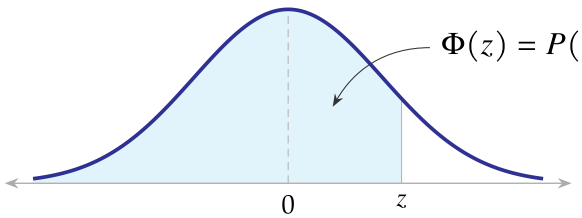
- $\mu_S = n \cdot 0{,}1 = 0{,}1n$
- $\sigma_Y = \sqrt{n \cdot 0{,}1 \cdot (1 - 0{,}1)} = 0{,}3\sqrt{n}$

Ta có:

$$\begin{aligned}
& P(S < 9) &&= 0{,}3085 \\
\Leftrightarrow\quad & \Phi\left(\frac{9 - 0{,}1n - 0{,}5}{0{,}3\sqrt{n}}\right) &&= 0{,}3085 \\
\Leftrightarrow\quad & \Phi\left(\frac{0{,}1n - 8{,}5}{0{,}3\sqrt{n}}\right) &&= 0{,}6915 \quad (= \Phi(0{,}5)\,) \\
\Leftrightarrow\quad & \frac{0{,}1n - 8{,}5}{0{,}3\sqrt{n}} &&= 0{,}5 \\
\Leftrightarrow\quad & n &&= 100
\end{aligned}$$

■

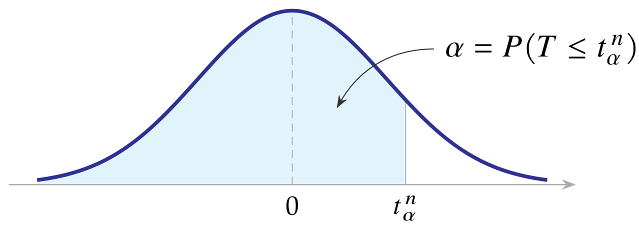
BẢNG TRA PHÂN PHỐI CHUẨN TẮC



$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6699	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9762	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

BẢNG TRA PHÂN PHỐI STUDENT



n \ α	0.60	0.75	0.90	0.95	0.975	0.99	0.995	0.9995
1	0.3249	1.0000	3.0777	6.3138	12.7062	31.8205	63.6567	636.6192
2	0.2887	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248	31.5991
3	0.2767	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409	12.9240
4	0.2707	0.7407	1.5332	2.1318	2.7764	3.7469	4.6041	8.6103
5	0.2672	0.7267	1.4759	2.0150	2.5706	3.3649	4.0321	6.8688
6	0.2648	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074	5.9588
7	0.2632	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995	5.4079
8	0.2619	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554	5.0413
9	0.2610	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498	4.7809
10	0.2602	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693	4.5869
11	0.2596	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058	4.4370
12	0.2590	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545	4.3178
13	0.2586	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123	4.2208
14	0.2582	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768	4.1405
15	0.2579	0.6912	1.3406	1.7531	2.1314	2.6025	2.9467	4.0728
16	0.2576	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208	4.0150
17	0.2573	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982	3.9651
18	0.2571	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784	3.9216
19	0.2569	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609	3.8834
20	0.2567	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453	3.8495
21	0.2566	0.6864	1.3232	1.7207	2.0796	2.5176	2.8314	3.8193
22	0.2564	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188	3.7922
23	0.2563	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073	3.7676
24	0.2562	0.6848	1.3178	1.7109	2.0639	2.4922	2.7969	3.7452
25	0.2561	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874	3.7245
26	0.2560	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787	3.7066
27	0.2559	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707	3.6896
28	0.2558	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633	3.6739
29	0.2557	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564	3.6594
30	0.2556	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500	3.6460
40	0.2550	0.6807	1.3031	1.6839	2.0211	2.4233	2.7045	3.5510
60	0.2545	0.6786	1.2958	1.6706	2.0003	2.3901	2.6603	3.4602
120	0.2539	0.6765	1.2886	1.6577	1.9799	2.3578	2.6174	3.3735