Hướng dẫn giải Bài tập nộp

Note: Danh sách bài tập gồm 28 bài.

- - - - - * * * * * - - - - -

Bài 1. Thời gian cho đến khi cần sạc lại pin cho một máy tính xách tay trong điều kiện bình thường là phân phối chuẩn với trung bình 260 phút và độ lệch chuẩn là 50 phút.

(a) Xác suất pin sử dụng kéo dài hơn bốn giờ là bao nhiêu?

(b) Xác định thời gian sử dụng pin tại những *giá trị phân vị*, là giá trị z sao cho xác suất $\mathbb{P}(Z \le z)$, đạt 25% và 75%. (Hint: tức là tìm a sao cho $\mathbb{P}(Z \le a) = 25\%$ và tìm b sao cho $\mathbb{P}(Z \le b) = 75\%$)

(c) Xác định thời gian sử dụng pin tương ứng với xác suất ít nhất 0,95.

Hướng dẫn: Ta quan tâm đến thời gian cho đến khi cần sạc lại pin cho một máy tính xách tay.

Gọi X là b.n.n. thể hiện thời gian (đơn vị: phút) cho đến khi cần sạc lại pin cho một máy tính xách tay. Theo đề bài, X có phân phối chuẩn, $X \sim \mathcal{N}(\mu;\sigma^2)$ với trung bình $\mu = 260$ phút và độ lệch chuẩn $\sigma = 50$ phút.

■ (a) Xác suất pin sử dụng kéo dài hơn 4 giờ = 240 phút là

$$\mathbb{P}\big(X > 240\big) = 1 - \mathbb{P}\big(X \le 240\big) = 1 - \mathbb{P}\left(\frac{X - \mathbb{E}(X)}{\sqrt{\mathrm{Var}(X)}} \le \frac{240 - \mathbb{E}(X)}{\sqrt{\mathrm{Var}(X)}}\right) = 1 - \mathbb{P}\left(\frac{X-\mu}{\sigma} \le \frac{240-\mu}{\sigma}\right)$$

$$\big(\text{với } Z \sim \mathcal{N}(0;1)\big) \qquad = 1 - \mathbb{P}\Big(Z \le \frac{240-\mu}{\sigma}\Big) = 1 - \Phi\Big(\frac{240-\mu}{\sigma}\Big)$$

$$= 1 - \Phi\Big(\frac{240-260}{50}\Big) = 1 - \Phi\big(-0,4\big) = \Phi\big(0,4\big) = 0,6554;$$

■ (b) Giá trị w_γ là phân vị mức γ của b.n.n. X nếu thoả $\mathbb{P}(X \le w_\gamma) = \gamma$.

▹ Do đó, gọi a là thời gian sử dụng pin tại phân vị 25%, nghĩa là

$$\mathbb{P}\big(X \le a\big) = 25\% = 0,25 \iff \mathbb{P}\left(\frac{X-\mu}{\sigma} \le \frac{a-\mu}{\sigma}\right) = 0,25$$

$$\iff \Phi\Big(\frac{a-\mu}{\sigma}\Big) = 0,25 \iff \frac{a-\mu}{\sigma} = \Phi^{-1}\big(0,25\big) = -\Phi^{-1}\big(1-0,25\big) = -\Phi^{-1}\big(0,75\big)$$

suy ra $a = \mu + \sigma \times \Big(-\Phi^{-1}(0,75)\Big) \approx 260 + 50 \times \big(-0,675\big) = 226,25$ phút ;

▹ Tương tự, gọi b là thời gian sử dụng pin tại phân vị 75%, nghĩa là

$$\mathbb{P}\big(X \le b\big) = 75\% = 0,75 \iff \Phi\Big(\frac{b-\mu}{\sigma}\Big) = 0,75 \iff \frac{b-\mu}{\sigma} = \Phi^{-1}\big(0,75\big)$$

suy ra $b = \mu + \sigma \times \Phi^{-1}(0,75) \approx 260 + 50 \times 0,675 = 293,75$ phút ;

■ (c) Gọi c là thời gian sử dụng pin tương ứng với xác suất ít nhất $0,95$, ta có

$$\mathbb{P}\big(X \le c\big) \ge 0,95 \iff \Phi\Big(\frac{c-\mu}{\sigma}\Big) \ge 0,95 \iff \frac{c-\mu}{\sigma} \ge \Phi^{-1}\big(0,95\big) \iff c \ge \mu + \sigma \times \Phi^{-1}\big(0,95\big)$$

$$\approx 260 + 50 \times 1,645 = 342,25.$$

Vậy, thời gian sử dụng pin tương ứng với xác suất ít nhất $0,95$ là $342,25$ phút. □

Nhắc lại: với $Z \sim \mathcal{N}(0;1)$ thì hàm số $\Phi(a) := \mathbb{P}(Z \le a) = \int_{-\infty}^{a} \frac{1}{2\pi}.e^{-\frac{t^2}{2}}dt.$

và $\mathbb{P}(a < Z < b) = \mathbb{P}(a \le Z < b) = \mathbb{P}(a < Z \le b) = \mathbb{P}(a \le Z \le b) = \Phi(b) - \Phi(a)$.

Bài 2. Đường kính một sợi bán dẫn giả sử có phân phối chuẩn với trung bình $0,5\mu$m và độ lệch chuẩn $0,05$ μm.

(a) Tính xác suất để đường kính của một sợi lớn hơn $0,62$ μm.

(b) Tính xác suất để đường kính của một sợi nằm giữa $0,47$ và $0,63$ μm.

(c) Đường kính sợi của 90% mẫu nhỏ hơn giá trị nào? (*Tức là, tìm x sao cho* $\mathbb{P}(X < x) = 0,9$)

Hướng dẫn: Ta quan tâm đến đường kính sợi bán dẫn.

Gọi X là b.n.n. thể hiện đường kính sợi bán dẫn (đơn vị: μm).

Theo đề bài, X có phân phối chuẩn, $X \sim \mathcal{N}(\mu;\sigma^2)$ với trung bình $\mu = 0,5$ μm và độ lệch chuẩn $\sigma = 0,05$ μm.

■ (a) Xác suất để đường kính của một sợi lớn hơn $0,62$ μm là

$$\mathbb{P}(X > 0,62) = 1 - \mathbb{P}(X \le 0,62) = 1 - \mathbb{P}\left(\frac{X - \mathbb{E}(X)}{\sqrt{\mathrm{Var}(X)}} \le \frac{0,62 - \mathbb{E}(X)}{\sqrt{\mathrm{Var}(X)}}\right) = 1 - \mathbb{P}\left(\frac{X-\mu}{\sigma} \le \frac{0,62-\mu}{\sigma}\right)$$

$$(\text{với } Z \sim \mathcal{N}(0;1)) \quad = 1 - \mathbb{P}\left(Z \le \frac{0,62-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{0,62-\mu}{\sigma}\right)$$

$$= 1 - \Phi\left(\frac{0,62-0,5}{0,05}\right) = 1 - \Phi(2,4) = 1 - 0,9918 = 0,9982;$$

■ (b) Xác suất để đường kính của một sợi nằm giữa $0,47$ và $0,63$ μm là

$$\mathbb{P}(0,47 \le X \le 0,63) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \le \frac{0,63-\mu}{\sigma}\right) - \mathbb{P}\left(\frac{X-\mu}{\sigma} \le \frac{0,47-\mu}{\sigma}\right)$$

$$(\text{với } Z \sim \mathcal{N}(0;1)) \quad = \mathbb{P}\left(Z \le \frac{0,63-\mu}{\sigma}\right) - \mathbb{P}\left(Z \le \frac{0,47-\mu}{\sigma}\right) = \Phi\left(\frac{0,63-\mu}{\sigma}\right) - \Phi\left(\frac{0,47-\mu}{\sigma}\right)$$

$$= \Phi\left(\frac{0,63-0,5}{0,05}\right) - \Phi\left(\frac{0,47-0,5}{0,05}\right)$$

$$= \Phi(2,6) - \Phi(-0,6) = \Phi(2,6) - \left[1 - \Phi(0,6)\right] = 0,9953 - \left[1 - 0,7257\right] = 0,721;$$

■ (c) Đường kính sợi của 90% mẫu nhỏ hơn giá trị c, ta có

$$\mathbb{P}(X < c) = 90\% = 0,9 \Leftrightarrow \Phi\left(\frac{c-\mu}{\sigma}\right) = 0,9 \Leftrightarrow \frac{c-\mu}{\sigma} = \Phi^{-1}(0,9) \Leftrightarrow c = \mu + \sigma \times \Phi^{-1}(0,9)$$

$$\approx 0,5 + 0,05 \times 1,282 = 0,5641.$$

Vậy, 90% mẫu có đường kính nhỏ hơn $0,5641$ μm.

□

Bài 3. Thời gian xử lý của một thuật toán trên máy tính cụ thể có phân phối chuẩn với trung bình là 250 ms và độ lệch chuẩn 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 $\mathbb{P}(248 < \overline{X}_n < 252) = 0,95$.

(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.

Hướng dẫn: Gọi X là thời gian xử lý của thuật toán ở một lần chạy,

theo đề bài, X tuân theo p.p. chuẩn $\mathcal{N}(\mu;\sigma^2)$ với trung bình $\mu = 250$ và độ lệch chuẩn $\sigma = 15$.

■ (a) Gọi X_1 là thời gian xử lý của thuật toán ở lần chạy thứ 1,

X_2 là thời gian xử lý của thuật toán ở lần chạy thứ 2,

....

X_n là thời gian xử lý của thuật toán ở lần chạy thứ n,

□ Ta có $X_1, X_2, ..., X_n$ độc lập và có cùng phân phối xác suất $\mathcal{N}(\mu;\sigma^2)$.

□ Trung bình thời gian xử lý của thuật toán trong n lần chạy là $\overline{X}_n = \dfrac{X_1 + X_2 + ... + X_n}{n}$,

thì $\overline{X}_n$ tuân theo phân phối chuẩn $\mathcal{N}(\mu_{\overline{X}};\sigma^2_{\overline{X}})$, với

- trung bình: $\mathbb{E}(\overline{X}_n) = \mathbb{E}\left(\dfrac{X_1 + ... + X_n}{n}\right) = \dfrac{\mathbb{E}(X_1) + ... + \mathbb{E}(X_n)}{n} = \mu = 250$ ms;
- phương sai: $\mathrm{Var}(\overline{X}_n) = \mathrm{Var}\left(\dfrac{X_1 + ... + X_n}{n}\right) = \dfrac{\mathrm{Var}(X_1) + ... + \mathrm{Var}(X_n)}{n^2} = \dfrac{\sigma^2}{n} = \dfrac{15^2}{n}$;

Ta có

$$\begin{aligned}
\mathbb{P}\big(248 < \overline{X}_n < 252\big) &= \mathbb{P}\left(\frac{248 - \mathbb{E}(\overline{X}_n)}{\sqrt{\mathrm{Var}(\overline{X}_n)}} < \frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{\mathrm{Var}(\overline{X}_n)}} < \frac{252 - \mathbb{E}(\overline{X}_n)}{\sqrt{\mathrm{Var}(\overline{X}_n)}}\right) \\
&= \mathbb{P}\left(\frac{248 - \mathbb{E}(\overline{X}_n)}{\sqrt{\mathrm{Var}(\overline{X}_n)}} \le Z \le \frac{252 - \mathbb{E}(\overline{X}_n)}{\sqrt{\mathrm{Var}(\overline{X}_n)}}\right) \quad (\text{với } Z \sim \mathcal{N}(0;1)) \\
&= \Phi\left(\frac{252 - \mathbb{E}(\overline{X}_n)}{\sqrt{\mathrm{Var}(\overline{X}_n)}}\right) - \Phi\left(\frac{248 - \mathbb{E}(\overline{X}_n)}{\sqrt{\mathrm{Var}(\overline{X}_n)}}\right) = \Phi\left(\frac{252-250}{15/\sqrt{n}}\right) - \Phi\left(\frac{248-250}{15/\sqrt{n}}\right) \\
&= \Phi\left(\frac{2\sqrt{n}}{15}\right) - \Phi\left(-\frac{2\sqrt{n}}{15}\right) = 2\Phi\left(\frac{2\sqrt{n}}{15}\right) - 1;
\end{aligned}$$

Do đó, $\mathbb{P}\big(248 < \overline{X}_n < 252\big) = 0,95 \Leftrightarrow 2\Phi\left(\dfrac{2\sqrt{n}}{15}\right) - 1 = 0,95 \Leftrightarrow \Phi\left(\dfrac{2\sqrt{n}}{15}\right) = \dfrac{1,95}{2} = 0,975 = \Phi(1,96)$,

suy ra $\dfrac{2\sqrt{n}}{15} = 1,96$ nên $n = \left(\dfrac{15 \times 1,96}{2}\right)^2 = 216,09$. Vậy ta tìm được $n = 216$.

■ (b) Gọi Y là số lần thuật toán có thời gian xử lý dài hơn 260ms trong tổng số $m = 300$ lần chạy này,

thì Y có p.p. nhị thức $B(m;p)$

với $p =$ xác suất để thuật toán có thời gian xử lý dài hơn 260ms ở 1 lần chạy $= \mathbb{P}(X > 260)$.

□ Ta tính xác suất để thuật toán có thời gian xử lý dài hơn 260ms ở 1 lần chạy

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

Ta thấy $m = 300$ là một số khá lớn nên có thể dùng phân phối chuẩn để tính xấp xỉ.

Điều kiện để dùng phân phối chuẩn xấp xỉ: $m \times p = 300 \times 0,2514 \ge 5$ và $m \times (1-p) = 300 \times (1 - 0,2514) \ge 5$.

Dùng phân phối chuẩn tính xấp xỉ, xác suất có ít nhất 70 lần thuật toán có thời gian xử lý dài hơn 260ms là

$$
\begin{aligned}
\mathbb{P}(Y \geq 70) = 1 - \mathbb{P}(Y < 70) = 1 - \mathbb{P}(Y \leq 69) &= 1 - \mathbb{P}\left(\frac{Y - \mathbb{E}(Y)}{\sqrt{\mathrm{Var}(Y)}} \leq \frac{69 + 0,5 - \mathbb{E}(Y)}{\sqrt{\mathrm{Var}(Y)}}\right) \\
&= 1 - \mathbb{P}\left(\frac{Y - m \times p}{\sqrt{m \times p \times (1-p)}} \leq \frac{69,5 - m \times p}{\sqrt{m \times p \times (1-p)}}\right) \\
&\approx 1 - \Phi\left(\frac{69,5 - m \times p}{\sqrt{m \times p \times (1-p)}}\right) \\
&= 1 - \Phi\left(\frac{69,5 - 300 \times 0,2514}{\sqrt{300 \times 0,2514 \times (1 - 0,2514)}}\right) \\
&\approx 1 - \Phi(-0,79) = \Phi(0,79) = 0,7852 .
\end{aligned}
$$

□

Bài 4. Một nhà máy sản xuất ra bu lông với đường kính tuân theo phân phối chuẩn $\mathcal{N}(4; 0,09)$ (đv: mm). Các bu lông được đo đạc một cách chính xác và những cái nào có đường kính nhỏ hơn 3,5mm hoặc lớn hơn 4,4mm sẽ bị loại bỏ. Chọn ngẫu nhiên một bu lông được máy sản xuất.

(a) Tính xác suất bu lông này được chấp nhận.

(b) Một lô gồm 900 bu lông được xản suất. Khoảng bao nhiêu bu lông trong số này được chấp nhận?

Hướng dẫn: Ta quan tâm đến đường kính của bu lông.

Gọi Y là biến ngẫu nhiên thể hiện đường kính của bu lông. Theo đề bài, Y có phân phối chuẩn, $Y \sim \mathcal{N}(\mu; \sigma^2)$ với trung bình $\mu = 4$ và phương sai $\sigma^2 = 0,09 \quad \Rightarrow \quad \sigma = 0,3$.

■ (a) Theo đề bài, một bu lông sẽ bị loại bỏ nếu đường kính nhỏ hơn 3,5mm hoặc lớn hơn 4,4mm, tức là nếu $Y < 3,5$ hoặc $Y > 4,4$.

$\Longrightarrow$ Ngược lại, bu lông được chấp nhận nếu đường kính "$\geq$ 3,5" và "$\leq$ 4,4", tức là $3,5 \leq Y \leq 4,4$.

Do đó, xác suất để một bu lông được chấp nhận là

$$
\begin{aligned}
\mathbb{P}(3,5 \leq Y \leq 4,4) &= \mathbb{P}\left(\frac{3,5-\mu}{\sigma} \leq \frac{Y-\mu}{\sigma} \leq \frac{4,4-\mu}{\sigma}\right) = \mathbb{P}\left(\frac{3,5-\mu}{\sigma} \leq Z \leq \frac{4,4-\mu}{\sigma}\right) \quad (\text{với } Z \sim \mathcal{N}(0;1)) \\
&= \Phi\left(\frac{4,4-\mu}{\sigma}\right) - \Phi\left(\frac{3,5-\mu}{\sigma}\right) \\
&= \Phi\left(\frac{4,4-4}{0,3}\right) - \Phi\left(\frac{3,5-4}{0,3}\right) \\
&= \Phi\left(\frac{4}{3}\right) - \Phi\left(\frac{-5}{3}\right) \approx 0,909 - 0,048 = 0,861 .
\end{aligned}
$$

■ (b) Một lô gồm $m = 900$ bu lông được sản xuất. Từ câu (a) ta đã tính được

$$p = \text{" xác suất để 1 bu lông được chấp nhận "} = 0,861 .$$

Cho nên trong $m = 900$ bu lông được sản xuất, số bu lông được chấp nhận sẽ khoảng

$$m \times p = 900 \times 0,861 = 774,9 .$$

Vậy, trong tổng số $m = 900$ bu lông được sản xuất sẽ có khoảng 775 bu lông được chấp nhận.

Bài 5. Trọng lượng một túi rau khi thu hoạch tại một nông trại là b.n.n. có phân phối chuẩn với trung bình 550 g và độ lệch chuẩn 125 g. Biết rằng trong một ngày có 4000 túi rau được thu hoạch. Ước lượng số túi rau có trọng lượng lớn hơn 325 g được thu hoạch trong một ngày.
(*Hint: tính $p =$ xác suất để một túi rau có trọng lượng lớn hơn 325 g,*
sau đó lấy xác suất p này nhân với tổng số gói rau thì sẽ ra số gói rau có trọng lượng lớn hơn 325)

Hướng dẫn: Ta quan tâm đến trọng lượng một túi rau.

Gọi X là biến ngẫu nhiên thể hiện trọng lượng của một túi rau.

Theo đề bài, X có phân phối chuẩn, $X \sim \mathcal{N}(\mu; \sigma^2)$ với trung bình $\mu = 550$ và độ lệch chuẩn $\sigma = 125$.

Trong tổng số $m = 4000$ túi rau được thu hoạch, ta muốn ước lượng số túi rau có trọng lượng lớn hơn 325g.

• Trước tiên, ta tính xác suất để một túi rau có trọng lượng lớn hơn 325g là

$$p = \mathbb{P}(X > 325) = 1 - \mathbb{P}(X \leq 325) = 1 - \mathbb{P}\Big(\frac{X-\mu}{\sigma} < \frac{325-\mu}{\sigma}\Big) = 1 - \mathbb{P}\Big(Z < \frac{325-\mu}{\sigma}\Big)$$

$$= 1 - \Phi\Big(\frac{325-\mu}{\sigma}\Big) \qquad (\text{với } Z \sim \mathcal{N}(0;1))$$

$$(\text{ nhắc lại, ta có } \Phi(-a) = 1 - \Phi(a)) \qquad = 1 - \Phi\Big(\frac{325-550}{125}\Big) = 1 - \Phi(-1.8) = \Phi(1.8) \approx 0,964 \ .$$

Cho nên trong $m = 4000$ túi rau được thu hoạch, số túi rau có trọng lượng lớn hơn 325g là

$$m \times p = 4000 \times 0,964 = 3856 \ .$$

Vậy, trong tổng số $m = 4000$ túi rau được thu hoạch sẽ có khoảng 3856 túi có trọng lượng lớn hơn 325g.

□

Bài 6. Xác suất để một bu lông được sản xuất đáp ứng yêu cầu kỹ thuật là 0, 86.

(a) Chọn ngẫu nhiên 50 bu lông được sản xuất. Tính xác suất có ít nhất 48 bu lông đáp ứng yêu cầu kỹ thuật.

(b) Một lô gồm 500 bu lông được sản xuất ra. Dùng một xấp xỉ phù hợp để tính xác suất có ít nhất 250 bu lông đáp ứng yêu cầu kỹ thuật.

Hướng dẫn: ■ (a) Gọi Y là b.n.n. thể hiện số bu lông đáp ứng kỹ thuật trong tổng số 50 bu lông.
Thì Y có phân phối nhị thức $B(n; p)$ với $n = 50$
và $p =$ "xác suất để 1 bu lông đáp ứng kỹ thuật" $= 0,86$ (theo dữ kiện đề bài).
Xác suất để trong số 50 bu lông, có ít nhất 48 bu lông đáp ứng yêu cầu kỹ thuật, là

$$\mathbb{P}\big(Y \geq 48\big) = \mathbb{P}\big(Y = 48\big) + \mathbb{P}\big(Y = 49\big) + \mathbb{P}\big(Y = 50\big)$$
$$= C_n^{48}\, p^{48}(1-p)^{n-48} + C_n^{49}\, p^{49}(1-p)^{n-49} + C_n^{50}\, p^{50}(1-p)^{n-50} \approx 0,0221.$$

■ (c) Gọi W là b.n.n. thể hiện số bu lông đáp ứng kỹ thuật trong tổng số 500 bu lông.
Thì W có phân phối nhị thức $B(m; p)$ với $m = 500$ và $p =$ "xác suất để 1 bu lông đáp ứng kỹ thuật" $= 0,86$.

Ta thấy $0,1 < p < 0,9$, hơn nữa $m \times p = 500 \times 0,86 \geq 5$ và $m \times (1-p) = 500 \times (1-0,86) \geq 5$, do đó, có thể dùng phân phối chuẩn $\mathcal{N}\big(m.p\ ;\ m.p.(1-p)\big)$ để tính xấp xỉ cho phân phối nhị thức $B(m; p)$ của W.

Cụ thể, ta có, xác suất để có ít nhất 250 bu lông đáp ứng yêu cầu kỹ thuật là

$$\mathbb{P}(W \geq 250) = 1 - \mathbb{P}(W < 250) = 1 - \mathbb{P}(W \leq 249) = 1 - \mathbb{P}(W < 249 + 0,5) \quad (\textit{hiệu chỉnh sự liên tục})$$

$$(W \text{ có phân phối nhị thức, } W \sim B(m;p)) \quad = 1 - \mathbb{P}\left(\frac{W - \mathbb{E}(W)}{\sqrt{\text{Var}(W)}} < \frac{249 + 0,5 - \mathbb{E}(W)}{\sqrt{\text{Var}(W)}} \right)$$

$$(\text{dùng } Z \text{ có phân phối chuẩn để xấp xỉ, } Z \sim \mathcal{N}(0;1)) \quad \approx 1 - \mathbb{P}\left(Z < \frac{249 + 0,5 - \mathbb{E}(W)}{\sqrt{\text{Var}(W)}} \right)$$

$$= 1 - \Phi\left(\frac{249 + 0,5 - \mathbb{E}(W)}{\sqrt{\text{Var}(W)}} \right)$$

$$= 1 - \Phi\left(\frac{249 + 0,5 - m.p}{\sqrt{m.p.(1-p)}} \right)$$

$$= 1 - \Phi\left(\frac{249 + 0,5 - 500 \times 0,86}{\sqrt{500 \times 0,86.(1 - 0,86)}} \right)$$

$$\approx 1 - \Phi(-23,26371) = \Phi(23,26371) \approx 1 .$$

□

Bài 7. Chiều dài (đv: cm) của một loài sâu tuân theo phân phối chuẩn $\mathcal{N}(\mu; \sigma^2)$ với $\sigma > 0$.

(a) Bắt ngẫu nhiên một con sâu. Nếu $\mu = 15$ cm và $\sigma = 3,5$ cm, hãy tính xác suất để chiều dài sâu

(i) ngắn hơn $18,5$ cm; (ii) ít nhất $16,75$ cm; (iii) từ $11,5$ cm đến $18,5$ cm.

(b) (Không bắt buộc) Biết rằng 30% số sâu dài ít nhất 16cm, và 15% số sâu ngắn hơn 10 cm. Tìm trung bình μ và độ lệch chuẩn σ của chiều dài của các con sâu.

Hướng dẫn: Gọi Y là biến ngẫu nhiên thể hiện chiều dài của một loài sâu đang xét. Theo đề bài, Y có phân phối chuẩn, $Y \sim \mathcal{N}(\mu; \sigma^2)$ với trung bình mu và độ lệch chuẩn σ.

■ (a) Với $\mu = 15$ cm và $\sigma = 3,5$ cm.

- (i) Xác suất để chiều dài sâu ngắn hơn $18,5$ cm là

$$\mathbb{P}(Y < 18,5) = \mathbb{P}\left(\frac{Y-\mu}{\sigma} < \frac{18,5-\mu}{\sigma}\right) = \Phi\left(\frac{18,5-\mu}{\sigma}\right) = \Phi\left(\frac{18,5-15}{3,5}\right) = \Phi(1) \approx 0,84134 .$$

- (ii) Xác suất để chiều dài sâu ít nhất $16,75$ cm là

$$\mathbb{P}(Y \geq 16,75) = 1 - \mathbb{P}(Y < 16,75) = 1 - \mathbb{P}\left(\frac{Y-\mu}{\sigma} < \frac{16,75-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{16,75-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{16,75-15}{3,5}\right)$$

$$= 1 - \Phi(0.5) \approx 1 - 0,6915 = 0,3085 .$$

- (iii) Xác suất để chiều dài sâu từ $11,5$ cm đến $18,5$ cm là

$$\mathbb{P}(11,5 \leq Y \leq 18,5) = \mathbb{P}\left(\frac{11,5-\mu}{\sigma} \leq \frac{Y-\mu}{\sigma} \leq \frac{18,5-\mu}{\sigma}\right)$$

$$= \Phi\left(\frac{18,5-\mu}{\sigma}\right) - \Phi\left(\frac{11,5-\mu}{\sigma}\right) = \Phi\left(\frac{18,5-15}{3,5}\right) - \Phi\left(\frac{11,5-15}{3,5}\right)$$

$$= \Phi(1) - \Phi(-1) = \Phi(1) - \left[1 - \Phi(1)\right] = 2 \times \Phi(1) - 1 \approx 2 \times 0,84134 - 1 = 0,68268 .$$

■ (b) Giả thiết 30% số sâu dài ít nhất 16 cm, tức là ta có $\mathbb{P}(X \geq 16) = 30\% = 0,3.$

Tương tự, 15% số sâu dài ngắn hơn 10 cm, tức là ta có $\mathbb{P}(X < 10) = 15\% = 0,15.$

Do đó, ta có hệ phương trình sau

$$\begin{cases} \mathbb{P}(X \geq 16) & = 0,3 \\ \mathbb{P}(X < 10) & = 0,15 \end{cases} \Leftrightarrow \begin{cases} \mathbb{P}\Big(\dfrac{X-\mu}{\sigma} < \dfrac{16-\mu}{\sigma}\Big) & = 1 - 0,3 = 0,7 \\ 1 - \mathbb{P}\Big(\dfrac{X-\mu}{\sigma} < \dfrac{10-\mu}{\sigma}\Big) & = 0,85 \end{cases} \Leftrightarrow \begin{cases} \Phi\Big(\dfrac{16-\mu}{\sigma}\Big) & = 0,7 \approx \Phi(0,525) \\ \Phi\Big(-\dfrac{10-\mu}{\sigma}\Big) & = 0,85 \approx \Phi(1,035) \end{cases}$$

vì Φ là hàm đồng biến (hàm tăng) nên ta suy ra

$$\begin{cases} \dfrac{16-\mu}{\sigma} & = 0,525 \\ -\dfrac{10-\mu}{\sigma} & = 1,035 \end{cases} \Leftrightarrow \begin{cases} \mu + \sigma \times 0,525 & = 16 \\ \mu - \sigma \times 1,035 & = 10 \end{cases} \Rightarrow \begin{cases} \sigma & \approx 3,8462\ ; \\ \mu & \approx 13,9808\ . \end{cases}$$

□

Bài 8. Một sản phẩm được đóng gói thành từng bao có khối lượng là một biến ngẫu nhiên có phân phối chuẩn với trung bình $1,42\ kg$ và độ lệch chuẩn $0,025\ kg$.

(a) Tìm xác suất để khối lượng một bao, được chọn ngẫu nhiên, nằm giữa $1,37kg$ và $1,45kg$.

(b) Tìm số bao trung bình, trong số 5000 bao được đóng gói, có khối lượng dưới $1,35\ kg$.

Hướng dẫn: Gọi X là khối lượng của một bao sản phẩm.

Theo đề bài, X có p.p. chuẩn $\mathcal{N}(\mu;\sigma^2)$ với trung bình $\mu = 1,42$ và độ lệch chuẩn $\sigma = 0,025$.

■ (a) Xác suất để khối lượng của một bao sản phẩm nằm giữa $1,37kg$ và $1,45kg$ là

$$\begin{aligned} \mathbb{P}(1,37 \leq X \leq 1,45) &= \mathbb{P}\Big(\frac{1,37-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{1,45-\mu}{\sigma}\Big) = \mathbb{P}\Big(\frac{1,37-\mu}{\sigma} \leq Z < \frac{1,45-\mu}{\sigma}\Big) \\ &= \Phi\Big(\frac{1,45-\mu}{\sigma}\Big) - \Phi\Big(\frac{1,37-\mu}{\sigma}\Big) \qquad (\text{với } Z \sim \mathcal{N}(0;1)) \\ &= \Phi\Big(\frac{1,45-1,42}{0,025}\Big) - \Phi\Big(\frac{1,37-1,42}{0,025}\Big) \\ &= \Phi(1,2) - \Phi(-2) = \Phi(1,2) - 1 + \Phi(2) \approx 0,8849 - 1 + 0,9772 = 0,8621\ . \end{aligned}$$

■ (b) Trong tổng số $m = 5000$ bao sản phẩm, ta muốn ước lượng số bao có khối lượng dưới $1,35\ kg$.

• Trước tiên, ta tính xác suất để một bao sản phẩm có khối lượng dưới $1,35kg$ là

$$\begin{aligned} p = \mathbb{P}(X < 1,35) &= \mathbb{P}\Big(\frac{X-\mu}{\sigma} < \frac{1,35-\mu}{\sigma}\Big) = \mathbb{P}\Big(Z < \frac{1,35-\mu}{\sigma}\Big) \qquad (\text{với } Z \sim \mathcal{N}(0;1)) \\ &= \Phi\Big(\frac{1,35-\mu}{\sigma}\Big) \\ &= \Phi\Big(\frac{1,35-1,42}{0,025}\Big) = \Phi(-2.8) = 1 - \Phi(2.8) \approx 1 - 0,9974 = 0,0026\ . \end{aligned}$$

Cho nên trong $m = 5000$ bao, số bao sản phẩm có khối lượng dưới $1,35kg$ là

$$m \times p = 5000 \times 0,0026 = 13\ .$$

Vậy, trong tổng số $m = 5000$ bao sản phẩm sẽ có khoảng 13 bao có trọng lượng dưới $1,35kg$.

□

Bài 9. Trong một dây chuyền đóng gói sản phẩm của một nhà máy, giả sử thời gian X (phút) mà một nhân viên hoàn thành đóng gói mỗi sản phẩm là biến ngẫu nhiên tuân theo phân phối chuẩn với trung bình $\mu = 7$ phút và độ lệch chuẩn $\sigma > 0$ phút. Biết rằng 10% số nhân viên của nhà máy hoàn thành đóng gói sản phẩm trong thời gian ít hơn 6 phút.

(a) Tìm giá trị của σ.

(b) Chọn ngẫu nhiên 10 nhân viên, tính xác suất để có ít hơn 2 nhân viên hoàn thành đóng gói sản phẩm trong thời gian ít hơn 6 phút.

(c) Nếu chọn ngẫu nhiên n nhân viên ($n > 50$). Bằng cách sử dụng xấp xỉ chuẩn có hiệu chỉnh liên tục, ta tính được xác suất để có ít hơn 5 trong số n nhân viên này sẽ hoàn thành đóng gói sản phẩm trong thời gian ít hơn 6 phút là $0,3446$. Tìm giá trị của n.

Hướng dẫn: ■ (a) 10% số nhân viên hoàn thành đóng gói trong thời gian ít hơn 6 phút nên ta có

$$\mathbb{P}\big(X < 6\big) = 10\% = 0,1; \quad \text{mà} \quad \mathbb{P}\big(X < 6\big) = \mathbb{P}\left(\frac{X-\mu}{\sigma} < \frac{6-\mu}{\sigma}\right) = \Phi\left(\frac{6-\mu}{\sigma}\right)$$

nên

$$\mathbb{P}\big(X < 6\big) = 0,1 \Leftrightarrow \Phi\left(\frac{6-\mu}{\sigma}\right) = 0,1 \Rightarrow \frac{6-\mu}{\sigma} = \Phi^{-1}\big(0,1\big) \approx -1,282 \Rightarrow \sigma = \frac{6-\mu}{-1,282} = \frac{6-7}{-1,282} \approx 0,78;$$

■ (b) Gọi Y là b.n.n. thể hiện số nhân viên hoàn thành trong thời gian ít hơn 6 phút trong tổng số 10 nhân viên, thì Y có phân phối nhị thức $B(m;p)$ với $m = 10$

và $p =$ "xác suất để 1 nhân viên hoàn thành trong ít hơn 6 phút" $= 0,1$.

Xác suất để có ít hơn 2 nhân viên hoàn thành đóng gói sản phẩm trong thời gian ít hơn 6 phút là

$$\begin{aligned}\mathbb{P}(Y < 2) = \mathbb{P}(Y = 0) + \mathbb{P}(Y = 1) &= C_m^0\, p^0\big(1-p\big)^{m-0} + C_m^1\, p^1\big(1-p\big)^{m-1}\\ &= C_{10}^0\, p^0\big(1-p\big)^{10-0} + C_{10}^1\, p^1\big(1-p\big)^{10-1} \approx 0,7361 \ ;\end{aligned}$$

■ (c) Gọi W là b.n.n. thể hiện số nhân viên hoàn thành trong thời gian ít hơn 6 phút, trong tổng số n nhân viên. Dùng phân phối Gauss $\mathcal{N}(0;1)$ tính xấp xỉ cho xác suất để có ít hơn 5 nhân viên hoàn thành trong thời gian ít hơn 6 phút là

$$\begin{aligned}\mathbb{P}(W < 5) = \mathbb{P}(W \le 4) = \mathbb{P}(W \le 4 + 0,5) &= \mathbb{P}\left(\frac{W - n\times p}{\sqrt{n\times p\times(1-p)}} \le \frac{4,5 - n\times p}{\sqrt{n\times p\times(1-p)}}\right)\\ &\approx \Phi\left(\frac{4,5 - n\times p}{\sqrt{n\times p\times(1-p)}}\right) = 0,3446\end{aligned}$$

suy ra

$$\begin{aligned}\frac{4,5 - n\times p}{\sqrt{n\times p\times(1-p)}} \approx \Phi^{-1}(0.3446) = -0,4 &\Leftrightarrow \big(4,5 - n\times p\big)^2 = n\times p\times(1-p)\times 0,4^2\\ &\Leftrightarrow p^2.n^2 - \Big[2.p\times 4,5 + p.(1-p).0,16\Big].n + (4,5)^2 = 0\end{aligned}$$

giải pt bậc hai ở trên với điều kiện $n > 50$ ta được $n \approx 53,8$ suy ra $n = 54$.

(Quan trọng là lập luận và các hướng giải, đáp số gần 54 vẫn được chấp nhận)

□ (Không bắt buộc) Có thể kiểm tra lại giá trị $n = 54$ bằng cách thay $n = 54$ vào tính

$$\mathbb{P}(W < 5) \approx \Phi\left(\frac{4,5 - n\times p}{\sqrt{n\times p\times(1-p)}}\right) = \Phi\left(\frac{4,5 - 54\times 0,1}{\sqrt{54\times 0,1\times(1-0,1)}}\right) \approx 0,3415.$$

Ta thấy với $n = 54$ thì tính được $\mathbb{P}(W < 5)$ có sai số khoảng 10^{-3} so với giá trị $0,3446 \Rightarrow$ chấp nhận được.

□

Bài 10. 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 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. Xác định phân phối xác suất của Y, và tính xác suất cho việc có từ 10 đến 20 hành khách bị phạt, tức là $\mathbb{P}(10 \leq Y \leq 20)$.

Hướng dẫn: Gọi X là trọng lượng hành lý ký gửi của một hành khách.

Theo đề bài, X tuân theo phân phối chuẩn $\mathcal{N}(\mu;\sigma^2)$ với trung bình $\mu = 20,41$ và độ lệch chuẩn $\sigma = 1,45$.

■ (a) Tỷ lệ hành khác đi máy bay phải chịu khoản phí hành lý quá tải là

$$\mathbb{P}(X > 22,67) = 1 - \mathbb{P}\left(\frac{X-\mu}{\sigma} \leq \frac{22,67-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{22,67-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{22,67-20,41}{1,45}\right)$$
$$\approx 1 - \Phi(1,56) = 1 - 0,9406 = 0,0594.$$

Hoặc có thể xấp xỉ $1 - \Phi\left(\dfrac{22,67-20,41}{1,45}\right) = 1 - \Phi(1,558621) \approx 1 - 0,9402 = 0,0598.$

■ (b) Ta có Y là số hành khách bị phạt vì có hành lý quá tải trong tổng số 200 hành khách, thì Y có phân phối nhị thức $B(n;p)$ với $n = 200$
và $p =$ xác suất để 1 hành khách bị phạt vì hành lý quá tải $= 0,0594$ (tính ở câu (a)).

Xác suất có từ 10 đến 20 hành khách bị phạt là

$$\begin{aligned}\mathbb{P}(10 \leq Y \leq 20) &= \mathbb{P}(Y=10) + \mathbb{P}(Y=11) + \dots + \mathbb{P}(Y=20)\\ &= C_n^{10}p^{10}(1-p)^{n-10} + C_n^{11}p^{11}(1-p)^{n-11} + \dots + C_n^{20}p^{20}(1-p)^{n-20} \approx 0,7465.\end{aligned}$$

□ Cách khác: ta thấy $n = 200$ là một số khá lớn, kiểm tra điều kiện $n \times p \geq 5$ và $n \times (1-p) \geq 5$ thỏa, nên ta có thể dùng phân phối Gauss để tính xấp xỉ cho phân phối nhị thức của Y như sau

$$\begin{aligned}\mathbb{P}(10 \leq Y \leq 20) &= \mathbb{P}(Y \leq 20) - \mathbb{P}(Y < 10) = \mathbb{P}(Y \leq 20) - \mathbb{P}(Y \leq 9)\\ &= \mathbb{P}(Y \leq 20 + 0,5) - \mathbb{P}(Y \leq 9 + 0,5) \quad (\text{hiệu chỉnh sự liên tục})\\ &= \mathbb{P}\left(\frac{Y-\mathbb{E}(Y)}{\sqrt{\mathrm{Var}(Y)}} \leq \frac{20,5-\mathbb{E}(Y)}{\sqrt{\mathrm{Var}(Y)}}\right) - \mathbb{P}\left(\frac{Y-\mathbb{E}(Y)}{\sqrt{\mathrm{Var}(Y)}} \leq \frac{9,5-\mathbb{E}(Y)}{\sqrt{\mathrm{Var}(Y)}}\right)\\ &\approx \mathbb{P}\left(Z \leq \frac{20,5-n\times p}{\sqrt{n\times p\times(1-p)}}\right) - \mathbb{P}\left(Z \leq \frac{9,5-n\times p}{\sqrt{n\times p\times(1-p)}}\right) \quad (Z \text{ có p.p. Gauss } \mathcal{N}(0;1))\\ &= \Phi\left(\frac{20,5-n\times p}{\sqrt{n\times p\times(1-p)}}\right) - \Phi\left(\frac{9,5-n\times p}{\sqrt{n\times p\times(1-p)}}\right)\\ &= \Phi\left(\frac{20,5-200\times 0,0594}{\sqrt{200\times 0,0594\times(1-0,0594)}}\right) - \Phi\left(\frac{9,5-200\times 0,0594}{\sqrt{200\times 0,0594\times(1-0,0594)}}\right)\\ &\approx \Phi(2,58) - \Phi(-0,71)\\ &= \Phi(2,58) - \Big[1-\Phi(0,71)\Big] = 0,9951 - \Big[1-0,7611\Big] = 0,7562.\end{aligned}$$

□

Bài 11. Giả sử tuổi thọ bóng đèn do một công ty sản xuất xấp xỉ phân phối chuẩn với độ lệch chuẩn là 40 giờ. Một mẫu 30 bóng đèn cho thấy tuổi thọ trung bình là 780 giờ.

a) Hãy tìm khoảng tin cậy 96% cho tuổi thọ trung bình của tất cả các bóng đèn do công ty sản xuất.

b) Nếu muốn sai số ước lượng không quá 10 giờ, thì phải quan sát ít nhất bao nhiêu bóng đèn.

Hướng dẫn: Ta quan tâm tuổi thọ trung bình bóng đèn. Đặt $\mu =$ tuổi thọ trung bình của bóng đèn.

□ Theo đề bài, " tuổi thọ bóng đèn xấp xỉ phân phối chuẩn với độ lệch chuẩn là 40 giờ "

$\Longrightarrow$ độ lệch chuẩn $\sigma = 40$ $\Longrightarrow$ phương sai $\sigma^2 = 40^2$ đã biết.

□ Lấy mẫu: cỡ mẫu $n = 30$ bóng đèn và trung bình mẫu $\overline{x} = 780$ giờ.

■ a) Độ tin cậy $1 - \alpha = 96\% = 0.96 \Longrightarrow \alpha = 0.04$.

• Dung sai ε (trong trường hợp phương sai đã biết) được xác định bởi

$$\varepsilon = z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = z_{0.98} \cdot \frac{\sigma}{\sqrt{n}} = z_{0.98} \cdot \frac{40}{\sqrt{30}} \approx 2,055 \times \frac{40}{\sqrt{30}} \approx 14,998.$$

• Vậy khoảng tin cậy 96% cho tuổi thọ trung bình của các bóng đèn do công ty sản xuất là

$$\overline{x} - \varepsilon < \mu < \overline{x} + \varepsilon \iff 780 - 14,998 < \mu < 780 + 14,998 \iff 765,002 < \mu < 794,998 .$$

■ b) Sai số ước lượng không quá 10 giờ, nghĩa là $\varepsilon \leq 10$, mà ta có $\varepsilon = z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$, nên

$$\varepsilon \leq 10 \iff z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq 10 \iff n \geq \left(z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{10}\right)^2$$

$$\iff n \geq \left(2,055 \times \frac{40}{10}\right)^2 \iff n \geq 67,5684 .$$

Vậy nếu muốn sai số ước lượng không quá 10 giờ ở độ tin cậy 96%, thì phải quan sát ít nhất 68 bóng đèn.

□

Bài 12. Đem cân một số trái cây vừa thu hoạch được bảng số liệu

X (g)	200-210	210-220	220-230	230-240	240-250
Số trái	12	17	20	18	15

a) Tìm khoảng ước lượng trọng lượng trung bình của trái cây với độ tin cậy 0, 95 và 0, 99.

b) Muốn sai số ước lượng không quá $E = 2$g ở độ tin cậy 99% phải quan sát ít nhất bao nhiêu trái cây ?

Hướng dẫn: Ta quan tâm trọng lượng trung bình trái cây. Đặt $\mu =$ trọng lượng trung bình của trái cây.

□ Lấy mẫu: cỡ mẫu $n = 12 + 17 + 20 + 18 + 15 = 82$ và trung bình mẫu được tính như sau

$$\overline{x} = \frac{\sum \overline{x}_i \cdot n_i}{\sum n_i} = \frac{205 \times 12 + 215 \times 17 + 225 \times 20 + 235 \times 18 + 245 \times 15}{n} = 225,8537 ,$$

(xét khoảng thứ nhất với trọng lượng 200-210g nên ta lấy giá trị đại diện là 205g.)

■ a) Trong trường hợp này cỡ mẫu $n = 82 \geq 30$ là mẫu lớn và phương sai σ^2 KHÔNG biết.

Ta cần tính phương sai mẫu như sau:

$$s^2 = \frac{\sum(\overline{x}_i - \overline{x})^2 \cdot n_i}{n-1} = \frac{(205-\overline{x})^2 \times 12 + (215-\overline{x})^2 \times 17 + (225-\overline{x})^2 \times 20 + (235-\overline{x})^2 \times 18 + (245-\overline{x})^2 \times 15}{n-1}$$

$$= 175,8055.$$

□ Độ tin cậy $1-\alpha = 0.95 \implies \alpha = 0.05.$

• Dung sai ε (trong trường hợp mẫu lớn , phương sai KHÔNG biết) được xác định bởi

$$\varepsilon = z_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = z_{0,975} \cdot \frac{s}{\sqrt{n}} = z_{0.975} \cdot \frac{\sqrt{175,8055}}{\sqrt{82}} \approx 1,96 \times \frac{\sqrt{175,8055}}{\sqrt{82}} \approx 2,8698 \ .$$

• Vậy khoảng tin cậy 95% cho trọng lượng trung bình của trái cây là

$$\overline{x} - \varepsilon < \mu < \overline{x} + \varepsilon \iff 222,9838 < \mu < 228,7236 \ .$$

□ Độ tin cậy $1-\alpha = 0.99 \implies \alpha = 0.01.$

• Dung sai ε (trong trường hợp mẫu lớn , phương sai KHÔNG biết) được xác định bởi

$$\varepsilon = z_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = z_{0,995} \cdot \frac{s}{\sqrt{n}} = z_{0.995} \cdot \frac{\sqrt{175,8055}}{\sqrt{82}} \approx 2,575 \times \frac{\sqrt{175,8055}}{\sqrt{82}} \approx 3,7704 \ .$$

• Vậy khoảng tin cậy 95% cho trọng lượng trung bình của trái cây là

$$\overline{x} - \varepsilon < \mu < \overline{x} + \varepsilon \iff 222,0833 < \mu < 229,6241 \ .$$

■ b) Sai số ước lượng không quá 2 g, nghĩa là $\varepsilon \leq 2$, mà ta có $\varepsilon = z_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$,
nên ở độ tin cậy $1-\alpha = 99\% = 0,99$, thì

$$\varepsilon \leq 2 \iff z_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \leq 2 \iff n \geq \left(z_{1-\frac{\alpha}{2}} \cdot \frac{s}{2}\right)^2$$

$$\iff n \geq \left(2,575 \times \frac{\sqrt{175,8055}}{2}\right)^2 \iff n \geq 291,4251 \ .$$

Vậy nếu muốn sai số ước lượng không quá 2 g ở độ tin cậy 99%, thì phải quan sát ít nhất 292 trái cây. □

Bài 13. Người ta đo ion Na^+ trên một số người và ghi nhận lại được kết quả như sau

$$129, \ 132, \ 140, \ 141, \ 138, \ 143, \ 133, \ 137, \ 140, \ 143, \ 138, \ 140 \ .$$

a) Tính trung bình mẫu và phương sai mẫu.

b) Ước lượng trung bình tổng thể ở độ tin cậy 0, 95.

Hint: trung bình mẫu $= \overline{x}$ *và phương sai mẫu* $= S^2$ *=> tính từ số liệu, dữ liệu thực nghiệm đề cho.*

trung bình tổng thể $= \mu$ *(không biết!!) và phương sai tổng thể* $= \sigma^2$ *(ở bài này,* σ^2 *cũng không biết!)*

Hướng dẫn: Ta quan tâm chỉ số trung bình ion Na^+ của một người.
Đặt $\mu =$ chỉ số trung bình ion Na^+ của một người.

■ a) Lấy mẫu: cỡ mẫu $n = 12$ (vì ta thấy dữ liệu là chỉ số đo được của 12 người)
và trung bình mẫu $\overline{x}$ được tính như sau

$$\overline{x} = \frac{\sum x_i}{n} = \frac{129 + 132 + 140 + \ \dots \ + 143 + 138 + 140}{12} \approx 137,833 \ ,$$

Trong trường hợp này cỡ mẫu $n = 12 < 30$ là mẫu nhỏ và phương sai σ^2 KHÔNG biết.

Ta cần tính phương sai mẫu s^2 như sau:

$$s^2 = \frac{\sum(x_i - \overline{x})^2}{n-1} = \frac{(129-\overline{x})^2 + (132-\overline{x})^2 + (140-\overline{x})^2 + \dots + (143-\overline{x})^2 + (138-\overline{x})^2 + (140-\overline{x})^2}{n-1} \approx 19,424.$$

■ b) Độ tin cậy $1-\alpha = 0.95 \implies \alpha = 0.05$.

• Dung sai ε (trong trường hợp mẫu nhỏ , phương sai KHÔNG biết) được xác định bởi

$$\varepsilon = t^{n-1}_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = t^{12-1}_{0,975} \cdot \frac{s}{\sqrt{n}} = t^{11}_{0,975} \cdot \frac{\sqrt{19,424}}{\sqrt{12}} \approx 2,201 \times \frac{\sqrt{19,424}}{\sqrt{12}} \approx 2,8 \ .$$

• Vậy khoảng tin cậy 95% cho chỉ số trung bình ion Na^+ là

$$\overline{x} - \varepsilon < \mu < \overline{x} + \varepsilon \iff 135,033 < \mu < 140,633 \ .$$

□

Bài 14. 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 $0,06$.

Hướng dẫn: Ta quan tâm tỷ lệ lỗi trong các module phần mềm của dự án này.

□ Theo đề bài, tỷ lệ mẫu $\widehat{p} = 0,25$.

■ (a) KTC 95%: Độ tin cậy $1-\alpha = 95\% = 0,95 \Rightarrow \alpha = 0,05$.

Gọi n là cỡ mẫu, KTC 95% có dung sai được tính bởi

$$\varepsilon = z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\widehat{p}\times(1-\widehat{p})}{n}} = z_{0,975} \times \sqrt{\frac{\widehat{p}\times(1-\widehat{p})}{n}} = 1,96 \times \sqrt{\frac{\widehat{p}\times(1-\widehat{p})}{n}} = 0,06 \ ,$$

suy ra

$$\sqrt{\frac{\widehat{p}\times(1-\widehat{p})}{n}} = \frac{0,06}{1,96} = \frac{3}{98};$$

KTC 99%: Độ tin cậy $1-\alpha = 99\% = 0,99 \Rightarrow \alpha = 0,01$.

Dung sai của KTC 99% là

$$\varepsilon = z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\widehat{p}\times(1-\widehat{p})}{n}} = z_{0,995} \times \sqrt{\frac{\widehat{p}\times(1-\widehat{p})}{n}} \approx 2,575 \times \frac{0,06}{1,96} \approx 0,0788 \ ,$$

• Vậy khoảng tin cậy 99% cho tỷ lệ lỗi trong các module phần mề của dự án này là

$$\widehat{p} - \varepsilon < p < \widehat{p} + \varepsilon \Leftrightarrow 0,25 - 0,0788 < p < 0,25 + 0,0788 \Leftrightarrow 0,1712 < p < 0,3288 \ .$$

■ (b) Với độ tin cậy 99%, tìm n để sai số KTC thỏa $\varepsilon \le 0,06$. Ta có

$$\varepsilon = z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\widehat{p}\times(1-\widehat{p})}{n}} = z_{0,995} \times \sqrt{\frac{\widehat{p}\times(1-\widehat{p})}{n}} \le 0,06 \Leftrightarrow \left(\frac{z_{0,995}}{0,06} \times \sqrt{\widehat{p}\times(1-\widehat{p})}\right)^2 \le n$$

$$\Leftrightarrow \left(\frac{2,575}{0,06} \times \sqrt{0,25\times(1-0,25)}\right)^2 \le n \Leftrightarrow 345,3451 \le n.$$

Vậy để sai số KTC 99% không vượt quá $0,06$ thì cỡ mẫu phải thỏa $n \ge 346$.

□

Bài 15. Một mẫu ngẫu nhiên gồm 50 mũ bảo hiểm được sử dụng bởi những người đi xe máy đã được thử nghiệm va chạm, và 18 trong số những chiếc mũ bảo hiểm đã bị thiệt hại.

a) Tìm khoảng tin cậy 95% cho tỷ lệ mũ bảo hiểm loại này sẽ cho thấy thiệt hại từ thử nghiệm này.

b) (Không bắt buộc) Bao nhiêu mũ bảo hiểm phải được kiểm tra để sai số khi ước lượng giá trị thực của p nhỏ hơn $0,02$ với độ tin cậy 95%?

Hướng dẫn: Ta quan tâm tỷ lệ mũ bảo hiểm bị thiệt hại. Đặt $p =$ tỷ lệ mũ bảo hiểm bị thiệt hại.

□ Theo đề bài: " mẫu ngẫu nhiên gồm 50 mũ được thử nghiệm và thấy có 18 cái bị thiệt hại "

$\Longrightarrow$ cỡ mẫu $n = 50$ mũ bảo hiểm và tỷ lệ mẫu $\widehat{p} = \dfrac{18}{n} = \dfrac{18}{50}$.

■ a) Độ tin cậy $1-\alpha = 95\% = 0.95 \quad \Longrightarrow \quad \alpha = 0.05$.

• Dung sai :

$$\varepsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\widehat{p}.(1-\widehat{p})}{n}} = z_{0.975} \cdot \sqrt{\frac{\widehat{p}.(1-\widehat{p})}{n}} = 1,96 \times \sqrt{\frac{\frac{18}{50}.\left(1-\frac{18}{50}\right)}{50}} \approx 0,133.$$

• Vậy khoảng tin cậy 96% cho tỷ lệ mũ bảo hiểm bị thiệt hại là

$$\widehat{p}-\varepsilon < p < \widehat{p}+\varepsilon \iff \frac{18}{50} - 0,133 < p < \frac{18}{50} + 0,133 \iff 0,227 < p < 0,493 .$$

■ b) Sai số ước lượng nhỏ hơn 2%, nghĩa là $\varepsilon < 2\%$, mà ta có $\varepsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\dfrac{\widehat{p}.(1-\widehat{p})}{n}}$, nên

$$\varepsilon < 2\% \iff z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\widehat{p}.(1-\widehat{p})}{n}} < 0,02 \iff n > \left(z_{1-\frac{\alpha}{2}} \cdot \frac{\sqrt{\widehat{p}.(1-\widehat{p})}}{0,02}\right)^2$$

$$\iff n > \left(1,96 \times \frac{\sqrt{\frac{18}{50}.\left(1-\frac{18}{50}\right)}}{0,02}\right)^2 \iff n > 2212,762 .$$

Vậy muốn sai số ước lượng nhỏ hơn 2% ở độ tin cậy 95%, thì phải quan sát ít nhất 2213 mũ bảo hiểm. □

Bài 16. Nhiệt độ nước trung bình tại một trạm bơm nước cung cấp cho một khu dân cư không được vượt quá 100^oF. Những khảo sát trong quá khứ đã chỉ ra rằng độ lệch chuẩn của nhiệt độ là 2^oF. Đo nhiệt độ nước tại trạm bơm trong 9 ngày được chọn ngẫu nhiên và tính được nhiệt độ trung bình là 98^oF. Biết rằng nhiệt độ nước tuân theo phân phối chuẩn.
(a) Có bằng chứng cho thấy nhiệt độ nước có thể chấp nhận được hay không, mức ý nghĩa 0.05?
(b) Tính mức ý nghĩa nhỏ nhất khi bạn đưa ra kết luận bác bỏ giả thuyết H_0.
(Hint: p-giá trị = mức ý nghĩa nhỏ nhất để bác bỏ H_0)

Hướng dẫn: Ta quan tâm nhiệt độ nước trung bình. Đặt $\mu =$ nhiệt độ nước trung bình.

□ Theo đề bài, "Những khảo sát trong quá khứ đã chỉ ra rằng độ lệch chuẩn của nhiệt độ là 2°F"

$\Rightarrow$ độ lệch chuẩn $\sigma = 2 \quad \Rightarrow$ phương sai $\sigma^2 = 4$ đã biết.

□ Lấy mẫu:

" Đo nhiệt độ nước trong 9 ngày được chọn ngẫu nhiên và tính được nhiệt độ trung bình là 98^oF "

$\Rightarrow$ cỡ mẫu $n = 9$ và trung bình mẫu $\overline{x} = 98$.

" Nhiệt độ nước trung bình hạ lưu từ ống tháp xả giải nhiệt của nhà máy không được vượt quá 100°F.
Có bằng chứng gì cho ta thấy nhiệt độ nước có thể chấp nhận được hay không "

$\Longrightarrow$ $H_0: \ \mu \leq 100$ $\Longrightarrow$ $H_1: \ \mu > 100.$ (*Dấu " = " luôn nằm ở giả thuyết* H_0)

• Giả thuyết cần kiểm định $\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu > \mu_0 \end{cases} \iff \begin{cases} H_0: & \mu = \mu_0 = 100 \\ H_1: & \mu > \mu_0 \end{cases}$;

$\left(\text{cũng có thể viết} \ \begin{cases} H_0: & \mu \leq \mu_0 = 100 \\ H_1: & \mu > \mu_0 \end{cases} \right)$

• Mức ý nghĩa $\alpha = 5\% = 0,05$.

• Tính thống kê kiểm định: Lấy mẫu: cỡ mẫu $n = 9$ có trung bình mẫu $\overline{x} = 98$;
Phương sai $\sigma^2 = 4$ đã biết $\Rightarrow$ dùng phân phối chuẩn.

Ta có thống kê kiểm định:

$$z_0 = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{98 - 100}{2/\sqrt{9}} = -3 \ ;$$

• Miền bác bỏ: ta sẽ bác bỏ H_0 nếu $z_0 > z_{1-\alpha} = z_{0,95} = 1,645$.

• Kết luận: do $z_0 = -3 < 1,645 = z_{1-\alpha} \Longrightarrow$ KHÔNG bác bỏ H_0 tức là $\mu = \mu_0 = 100$;
(hoặc $\mu \leq \mu_0 = 100$ nếu lúc đầu đặt $H_0: \mu \leq \mu_0 = 100$)

Vậy nhiệt độ nước trung bình KHÔNG vượt quá 100^oF, với mức ý nghĩa $\alpha = 5\%$.
Do đó nhiệt độ nước là chấp nhận được.

■ Tính p-giá trị: trong trường hợp này ta có p-giá trị $= 1 - \Phi(z_0) = 1 - \Phi(-3) = \Phi(3) \approx 0,9987$.
Ta thấy p-giá trị $= 0,9987 > 0,05 = \alpha \Longrightarrow$ KHÔNG bác bỏ H_0
$\Longrightarrow$ có cùng kết luận khi sử dụng miền bác bỏ.

□

Bài 17. Đo hàm lượng natri (đv: mg) của hai mươi hộp bắp hữu cơ 300g của nhà sản xuất A, thu được dữ liệu

131, 15	130, 69	130, 91	129, 54	129, 64	128, 77	130, 72	128, 33	128, 24	129, 78
129, 65	130, 14	129, 29	128, 71	129	129, 39	130, 42	129, 53	130, 12	130, 92

Giả sử theo tiêu chuẩn, hàm lượng natri trung bình không được phép vượt quá 130mg.
Sử dụng mức ý nghĩa 0, 02, dữ liệu trên có cho thấy các hộp bắp hữu cơ của nhà sản xuất A đảm bảo tiêu chuẩn hay không? Tính p-giá trị.

Hướng dẫn: Ta quan tâm hàm lượng natri trung bình của hộp bắp hữu cơ.
Đặt $\mu =$ hàm lượng natri trung bình của hộp bắp hữu cơ của nhà sản xuất A.

□ Lấy mẫu: cỡ mẫu $n = 20$ và

• Trung bình mẫu:

$$\overline{x} = \frac{131,15 + 129,65 + 130,69 + 130,14 + ... + 129,78 + 130,92}{n} = 129,747;$$

- Phương sai mẫu: $s^2 = \dfrac{(131,15 - \overline{x})^2 + (129,65 - \overline{x})^2 + ... + (129,78 - \overline{x})^2 + (130,92 - \overline{x})^2}{n-1} \approx 0,7681;$

- Độ lệch chuẩn mẫu: $s = \sqrt{\text{phương sai mẫu}} = \sqrt{s^2} \approx 0,8764;$

" Giả sử theo tiêu chuẩn, hàm lượng natri trung bình không được phép vượt quá 130mg. Sử dụng mức ý nghĩa 0, 02, dữ liệu trên có cho thấy các hộp bắp hữu cơ của nhà sản xuất A đảm bảo tiêu chuẩn hay không?

$\Longrightarrow \quad H_0 : \ \mu \leq 130 \quad \Longrightarrow \quad H_1 : \ \mu > 130.$ (*Dấu " = " luôn nằm ở giả thuyết* H_0)

- Giả thuyết cần kiểm định $\begin{cases} H_0 : \ \mu \ \leq \ \mu_0 = 130 & : \text{đảm bảo tiêu chuẩn} \\ H_1 : \ \mu \ > \ \mu_0 & : \text{KHÔNG đảm bảo tiêu chuẩn} \end{cases}$;

- Mức ý nghĩa $\alpha = 0,02$.
- Tính thống kê kiểm định: Phương sai σ^2 KHÔNG biết và cỡ mẫu $n = 20 < 30$ (mẫu nhỏ)

$\Rightarrow$ dùng phân phối t-Student với bậc tự do $(n-1)$.

Ta có thống kê kiểm định: $t_0 = \dfrac{\overline{x} - \mu_0}{s/\sqrt{n}} = \dfrac{129,747 - 130}{0,7681/\sqrt{20}} = -1,291$;

- Miền bác bỏ: ta sẽ bác bỏ H_0 nếu $t_0 > t_{1-\alpha}^{n-1} = t_{0,98}^{19} \approx 2,1$.
- Kết luận: do $t_0 = -1,291 < 2,1 = t_{1-\alpha}^{n-1} \Longrightarrow$ KHÔNG bác bỏ H_0 tức là $\mu \leq \mu_0 = 130$;

Vậy hàm lượng natri trung bình của hộp ngũ cốc ≤ 130 mg, với mức ý nghĩa $\alpha = 0,02$.

■ Tính p-giá trị: gọi T_{19} là biến ngẫu nhiên có phân phối t-Student với bậc tự do $(n-1) = 19$, ta có

$$p\text{-giá trị} = \mathbb{P}(T_{19} \geq t_0) = \mathbb{P}(T_{19} \geq -1,291) = 1 - \mathbb{P}(T_{19} < -1,291)$$
$$= \mathbb{P}(T_{19} < 1,291) \approx 0,85.$$

Ta thấy p-giá trị $= 0,85 > 0,02 = \alpha \quad \Longrightarrow$ KHÔNG bác bỏ H_0

$\Longrightarrow$ có cùng kết luận khi sử dụng miền bác bỏ.

□

Bài 18. Biết trọng lượng X (g/quả) của mỗi quả trứng có phân phối chuẩn. Đem cân 100 quả trứng có kết quả sau:

x_i	155	160	165	170	175	180	185
n_i	5	12	14	25	24	14	6

Cho biết trứng có trọng lượng **lớn hơn** 170g là trứng loại một.

(a) Tìm khoảng tin cậy 97% cho trọng lượng trứng trung bình. Nếu ta muốn sai số ước lượng không quá 0, 1g thì cần khảo sát thêm bao nhiêu trứng?

(b) Tìm khoảng tin cậy 98% cho tỷ lệ trứng loại một.

(c) Có ý kiến rằng trọng lượng trứng trung bình lớn hơn 170g/quả. Kiểm định ý kiến trên với mức ý nghĩa 1%.

(d) Có ý kiến cho là **ít nhất** 65% số trứng thuộc loại một. Kiểm định ý kiến trên với mức ý nghĩa 1%.

Hướng dẫn: Ta quan tâm trọng lượng trứng trung bình. Đặt $\mu =$ trọng lượng trứng trung bình.

☐ Lấy mẫu: cỡ mẫu $n = 5 + 12 + 14 + 25 + 24 + 14 + 6 = 100$;

- Trung bình mẫu:

$$\overline{x} = \frac{155 \times 5 + 160 \times 12 + ... + 180 \times 14 + 185 \times 6}{n} = 170,85;$$

- Phương sai mẫu:

$$s^2 = \frac{(155 - \overline{x})^2 \times 5 + (160 - \overline{x})^2 \times 12 + ... + (180 - \overline{x})^2 \times 14 + (185 - \overline{x})^2 \times 6}{n-1} \approx 60,1288;$$

- Độ lệch chuẩn mẫu: $s = \sqrt{\text{phương sai mẫu}} = \sqrt{s^2} \approx 7,7543;$

■ (a) Độ tin cậy $1 - \alpha = 97\% = 0,97$ suy ra $\alpha = 0,03$.

- Phương sai σ^2 không biết, cỡ mẫu $n = 100 \geq 30$ (mẫu lớn) $\Rightarrow$ TH2: sử dụng phân phối Gauss.
- Sai số KTC (dung sai):

$$\varepsilon = z_{1-\frac{\alpha}{2}} \times \frac{s}{\sqrt{n}} = z_{0,985} \times \frac{s}{\sqrt{n}} \approx 2,17 \times \frac{7,7543}{\sqrt{100}} \approx 1,6827 \ ;$$

- Vậy KTC 97% cho doanh thu trung bình một ngày của cửa hàng là

$$\overline{x} - \varepsilon \leq \mu \leq \overline{x} + \varepsilon \Leftrightarrow 170,85 - 1,6827 \leq \mu \leq 170,85 + 1,6827 \Leftrightarrow 169,1673 \leq \mu \leq 172,5327 \ .$$

■ Để sai số ước lượng không quá 0, 1g thì

$$\varepsilon \leq 0,1 \Leftrightarrow z_{1-\frac{\alpha}{2}} \times \frac{s}{\sqrt{n}} \leq 0,1 \Leftrightarrow n \geq \left(z_{1-\frac{\alpha}{2}} \times \frac{s}{0,1}\right)^2 = 2,17^2 \times \frac{60,1288}{0,1^2} = 28314,05.$$

■ (b) Gọi $p =$ tỷ lệ trứng loại 1. Trứng là loại 1 $\Leftrightarrow$ có trọng lượng **trên** 170g.

Ta có: cỡ mẫu $n = 100$, số trứng gà loại 1 là $Y = 24 + 14 + 6 = 44$, nên tỷ lệ mẫu là $\widehat{p} = \frac{Y}{n} = \frac{44}{100}$;

- Độ tin cậy $1 - \alpha = 98\% = 0,98$ suy ra $\alpha = 0,02$.
- Sai số KTC (dung sai):

$$\widetilde{\varepsilon} = z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\widehat{p} \times (1 - \widehat{p})}{n}} = z_{0,99} \times \sqrt{\frac{\widehat{p} \times (1 - \widehat{p})}{n}} = 2,325 \times \sqrt{\frac{\frac{44}{100} \times \left(1 - \frac{44}{100}\right)}{100}} \approx 0,1154;$$

- Vậy KTC 98% cho tỷ lệ trứng loại 1 là

$$\widehat{p} - \widetilde{\varepsilon} \leq p \leq \widehat{p} + \widetilde{\varepsilon} \Leftrightarrow \frac{44}{100} - 0,1154 \leq p \leq \frac{44}{100} + 0,1154 \Leftrightarrow 0,3246 \leq p \leq 0,5554;$$

■ (c) " Có ý kiến cho rằng trọng lượng trứng trung bình lớn hơn 170g/quả.

$\Longrightarrow H_1 : \mu > 170 \Longrightarrow H_0 : \mu \leq 170$. (*Dấu " = " luôn nằm ở giả thuyết* H_0)

- Giả thuyết cần kiểm định $\begin{cases} H_0 : \mu \leq \mu_0 = 170 \\ H_1 : \mu > \mu_0 \end{cases}$;

- Mức ý nghĩa $\alpha = 0,01$.

• Tính thống kê kiểm định: Phương sai σ^2 KHÔNG biết và cỡ mẫu $n = 100 \geq 30$ (mẫu lớn)

$\Rightarrow$ dùng phân phối Gauss.

Ta có thống kê kiểm định: $z_0 = \dfrac{\overline{x} - \mu_0}{s/\sqrt{n}} = \dfrac{170,85 - 170}{7,7543/\sqrt{100}} = 1,0962$;

• Miền bác bỏ: ta sẽ bác bỏ H_0 nếu $z_0 > z_{1-\alpha} = z_{0,99} \approx 2,325$.

• Kết luận: do $z_0 = 1,0962 < 2,325 = z_{1-\alpha} \Rightarrow$ KHÔNG bác bỏ H_0 tức là $\mu \leq \mu_0 = 170$;

Vậy trọng lượng trung bình của trứng ≤ 170 g/quả, với mức ý nghĩa $\alpha = 0,01$.

■ (d) Có ý kiến là "50% số trứng thuộc loại một "

$\Longrightarrow H_0: \ p = 0,5 \quad \Longrightarrow H_1: \ p \neq 0,5$ (*Dấu " = " luôn nằm ở giả thuyết H_0*)

• Giả thuyết cần kiểm định $\begin{cases} H_0: & p = p_0 = 0,5 \\ H_1: & p \neq p_0 \end{cases}$;

• Mức ý nghĩa $\alpha = 0,01$.

• Tính thống kê kiểm định: $z_0 = \dfrac{\widehat{p} - p_0}{\sqrt{\dfrac{p_0.(1-p_0)}{n}}} = \dfrac{\dfrac{44}{100} - 0,5}{\sqrt{\dfrac{0,5.(1-0,5)}{100}}} = = -1,2$;

• Miền bác bỏ: ta sẽ bác bỏ H_0 nếu $|z_0| > z_{1-\frac{\alpha}{2}} = z_{0,995} \approx 2,575$.

• Kết luận: do $|z_0| = 1,2 < 2,575 = z_{1-\frac{\alpha}{2}} \Longrightarrow$ KHÔNG bác bỏ H_0 tức là $p = p_0 = 0,5$;

Vậy tỷ lệ trứng loại 1 là $p = 0,5$, với mức ý nghĩa $\alpha = 1\%$. □

Bài 19. Một nhà máy sản xuất ống kính interocular có một máy nghiền mới. Máy nghiền mới này được coi là đủ điều kiện nếu có bằng chứng cho thấy tỷ lệ phần trăm của các thấu kính được đánh bóng có chứa khuyết tật bề mặt là dưới 2%. Người ta khảo sát một mẫu ngẫu nhiên gồm 300 thấu kính thì thấy 8 ống kính bị lỗi.

a) Xây dựng và kiểm tra giả thuyết phù hợp để xác định xem máy mới này có đủ điều kiện hay không, với mức ý nghĩa 0.05, đồng thời tìm p-giá trị.

b) (không bắt buộc) Giải thích thêm cho kết luận trên bằng cách sử dụng khoảng tin cậy cho p.

Hướng dẫn: Ta quan tâm tỷ lệ thấu kính được đánh bóng có chứa khuyết tật (bị lỗi).

Đặt $p =$ tỷ lệ thấu kính được đánh bóng có chứa khuyết tật (bị lỗi).

□ Theo đề bài, " khảo sát một mẫu ngẫu nhiên gồm 300 thấu kính thì thấy 8 ống kính bị lỗi ".

$\Longrightarrow$ cỡ mẫu $n = 300$ và tỷ lệ mẫu $\widehat{p} = \dfrac{8}{n} = \dfrac{8}{300}$.

" Máy nghiền được coi là đủ điều kiện nếu tỷ lệ phần trăm của các thấu kính được đánh bóng có chứa khuyết tật bề mặt là dưới 2%

kiểm tra giả thuyết phù hợp để xác định xem máy mới này có đủ điều kiện hay không "

$\Longrightarrow H_1: \ p < 2\% = 0,02 \quad \Longrightarrow H_0: \ p \geq 0,02$. (*Dấu " = " luôn nằm ở giả thuyết H_0*)

• Giả thuyết cần kiểm định $\begin{cases} H_0: & p = p_0 = 0,02 & \text{: KHÔNG đủ điều kiện} \\ H_1: & p < p_0 & \text{: đủ điều kiện} \end{cases}$;

(*Dấu " = " luôn nằm ở giả thuyết H_0*) $\left(\text{cũng có thể viết } \begin{cases} H_0: & p \geq p_0 = 0,02 \\ H_1: & p < p_0 \end{cases} \right)$

• Mức ý nghĩa $\alpha = 0,05$.

• Tính thống kê kiểm định: với cỡ mẫu $n = 300$ và tỷ lệ mẫu $\widehat{p} = \dfrac{8}{300}$, ta có

$$z_0 = \frac{\widehat{p} - p_0}{\sqrt{\dfrac{p_0.\left(1 - p_0\right)}{n}}} = \frac{\dfrac{8}{300} - 0,02}{\sqrt{\dfrac{0,02.\left(1 - 0,02\right)}{300}}} \approx 0,825 \ ;$$

• Miền bác bỏ: ta sẽ bác bỏ H_0 nếu $z_0 < -z_{1-\alpha} = -z_{0,95} = -1,645$.

• Kết luận: do $z_0 = 0,825 > -1,645 = -z_{1-\alpha} \implies$ KHÔNG bác bỏ H_0 tức là $p = p_0 = 0,02$.
(hoặc $p \geq p_0 = 0,02$ nếu lúc đầu đặt $H_0 : p \geq p_0 = 0,02$)

Vậy tỷ lệ thấu kính bị lỗi KHÔNG dưới 2%, nên máy mới KHÔNG đủ điều kiện, mức ý nghĩa $\alpha = 5\%$.

■ Tính p-giá trị: trong trường hợp này ta có p-giá trị $= \Phi(z_0) = \Phi(0,825) \approx 0,7953$.

Ta thấy p-giá trị $= 0,7953 > 0,05 = \alpha \implies$ KHÔNG bác bỏ H_0 . □

Bài 20. Hai loại nhựa phù hợp để sử dụng cho một nhà sản xuất linh kiện điện tử. Sức mạnh chịu sự phá huỷ của loại nhựa này là quan trọng. Biết $\sigma_1 = \sigma_2 = 1$ psi. Từ hai mẫu ngẫu nhiên có kích thước $n_1 = 10$ và $n_2 = 12$, ta có được $\overline{x}_1 = 162,5$ và $\overline{x}_2 = 155$. Công ty sẽ không áp dụng nhựa loại 1 trừ khi sức chịu phá vỡ trung bình của nó vượt quá nhựa loại 2 một mức hơn 10 psi.
Trên cơ sở thông tin đó, ta có nên sử dụng nhựa loại 1? Sử dụng $\alpha = 0,05$ và tìm p-giá trị.
(*Hint: " Công ty sẽ không áp dụng nhựa loại 1 trừ khi sức chịu phá vỡ trung bình của nó vượt quá nhựa loại 2 một mức hơn 10 psi "* $\Rightarrow H_1 : \mu_1 - \mu_2 > D_0 = 10$)

Hướng dẫn: Ta muốn so sánh sức chịu phá vỡ trung bình của nhựa loại 1 so với loại 2.

Đặt $\mu_1 =$ sức chịu phá vỡ trung bình của loại 1 và $\mu_2 =$ sức chịu phá vỡ trung bình của loại 2.

" Công ty sẽ không áp dụng nhựa loại 1 trừ khi sức chịu phá vỡ trung bình của nó vượt quá nhựa loại 2 một mức hơn 10 psi "

$\implies H_1 : \mu_1 - \mu_2 > 10 \implies H_0 : \mu_1 - \mu_2 \leq 10$ (*Dấu " = " luôn nằm ở giả thuyết H_0*)

• Giả thuyết cần kiểm định $\begin{cases} H_0 : \mu_1 - \mu_2 = D_0 = 10 & : \text{KHÔNG nên sử dụng nhựa loại 1} \\ H_1 : \mu_1 - \mu_2 > D_0 & : \text{nên sử dụng nhựa loại 1} \end{cases}$;

$\left(\text{cũng có thể viết} \begin{cases} H_0 : \mu_1 - \mu_2 \leq D_0 = 10 \\ H_1 : \mu_1 - \mu_2 > D_0 \end{cases} \right)$

• Mức ý nghĩa $\alpha = 0,05$.

• Tính thống kê kiểm định:

Lấy mẫu: nhựa loại 1 : cỡ mẫu $n = 10$ có trung bình mẫu $\overline{x} = 162,5$;

nhựa loại 2 : cỡ mẫu $m = 12$ có trung bình mẫu $\overline{y} = 155$,

Phương sai $\sigma_1^2 = \sigma_2^2 = 1^2$ đã biết $\implies$ dùng phân phối chuẩn.

Ta có thống kê kiểm định: $$z_0 = \frac{\overline{x} - \overline{y} - D_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} = \frac{162,5 - 155 - 10}{\sqrt{\frac{1^2}{10} + \frac{1^2}{12}}} \approx -5,839 \;;$$

• Miền bác bỏ: ta sẽ bác bỏ H_0 nếu $z_0 > z_{1-\alpha} = z_{0,95} = 1,645$.

• Kết luận: do $z_0 = -5,839 < 1,645 = z_{1-\alpha}$ $\Longrightarrow$ KHÔNG bác bỏ H_0, tức là $\mu_1 - \mu_2 = D_0 = 10$

(hoặc $\mu_1 - \mu_2 \leq 10$ nếu lúc đầu đặt $H_0 : \mu_1 - \mu_2 \leq 10$)

Vậy sức chịu phá vỡ trung bình của nhựa loại 1 KHÔNG cao hơn độ sức chịu phá vỡ trung bình của nhựa loại 2 một mức hơn 10 psi, với mức ý nghĩa $\alpha = 1\%$.

Do đó KHÔNG nên sử dụng nhựa loại 1, với mức ý nghĩa $\alpha = 1\%$.

■ Tính p-giá trị: trong trường hợp này ta có p-giá trị $= 1 - \Phi(z_0) = 1 - \Phi(-5,839) = \Phi(5,839) \approx 1$.

Ta thấy p-giá trị $\approx 1 > 0,05 = \alpha$ $\Longrightarrow$ KHÔNG bác bỏ H_0 □

Bài 21. Hai chất xúc tác có thể được sử dụng trong một phản ứng hoá học. Mười hai phản ứng được cho sử dụng chất xúc tác 1, dẫn đến hiệu suất trung bình là 86 (đv: %) và độ lệch chuẩn mẫu là 3. Mười lăm phản ứng được cho sử dụng chất xúc tác 2, và kết quả là hiệu suất trung bình là 89 với độ lệch chuẩn mẫu là 2. Giả sử hiệu suất các phản ứng xấp xỉ phân phối chuẩn với cùng độ lệch chuẩn. Có bằng chứng để khẳng định rằng chất xúc tác 2 tạo ra hiệu suất trung bình cao hơn chất xúc tác 1 hay không? Sử dụng $\alpha = 0,01$. (*Yêu cầu dùng cả 2 phương pháp: miền bác bỏ và p-giá trị*)

Hướng dẫn: Ta muốn so sánh hiệu suất trung bình của chất 1 so với hiệu suất của chất 2.

Đặt μ_1 = hiệu suất của chất xúc tác 1 và μ_2 = hiệu suất của chất xúc tác 2.

" Có bằng chứng chất xúc tác 2 tạo ra hiệu suất trung bình cao hơn chất xúc tác 1 hay không? "

$\Longrightarrow$ $H_1 : \mu_1 < \mu_2$ hay $\mu_1 - \mu_2 < 0$ $\Longrightarrow$ $H_0 : \mu_1 - \mu_2 \geq 0$.

• Giả thuyết cần kiểm định $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases} \iff \begin{cases} H_0 : \mu_1 - \mu_2 = D_0 = 0 \\ H_1 : \mu_1 - \mu_2 < D_0 \end{cases}$;

(*Dấu "=" luôn nằm ở giả thuyết H_0*) $\left(\text{cũng có thể viết } \begin{cases} H_0 : \mu_1 - \mu_2 \geq D_0 = 0 \\ H_1 : \mu_1 - \mu_2 < D_0 \end{cases}\right)$

■ Mức ý nghĩa $\alpha = 0,01$.

■ Tính thống kê kiểm định:

• Lấy mẫu: chất A : cỡ mẫu $n = 12$ có trung bình mẫu $\overline{x} = 86$, độ lệch chuẩn mẫu $s_1 = 3$;

chất B : cỡ mẫu $m = 15$ có trung bình mẫu $\overline{y} = 89$, độ lệch chuẩn mẫu $s_1 = 2$;

• Phương sai σ_1^2 và σ_2^2 KHÔNG biết với cỡ mẫu $n = 12 < 30$ và $m = 15 < 30$ là cỡ mẫu nhỏ

$\Longrightarrow$ dùng phân phối t-Student với bậc tự do $df = n - 1 + m - 1$.

• Dữ kiện đề bài: " Giả sử hiệu suất các phản ứng xấp xỉ phân phối chuẩn với cùng độ lệch chuẩn "

$\Longrightarrow$ $\sigma_1 = \sigma_2$ hay phương sai bằng nhau $\sigma_1^2 = \sigma_2^2$

$$\Longrightarrow \text{tính } s_p^2 = \frac{(n-1).s_1^2 + (m-1).s_2^2}{n-1+m-1} = \frac{(12-1).3^2 + (15-1).2^2}{12-1+15-1} = 6,2.$$

• Ta có thống kê kiểm định: $$t_0 = \frac{\overline{x} - \overline{y} - D_0}{\sqrt{\frac{s_p^2}{n} + \frac{s_p^2}{m}}} = \frac{86 - 89 - 0}{\sqrt{\frac{6,2}{12} + \frac{6,2}{15}}} \approx -3,111 \;;$$

• Miền bác bỏ: ta sẽ bác bỏ H_0 nếu $t_0 < -t_{1-\alpha}^{n-1+m-1} = -t_{0,99}^{12-1+15-1} = -t_{0,99}^{25} = -2,4851$ (tra bảng phân phối Student để tìm giá trị cho $t_{0,99}^{25}$)

• Kết luận: do $t_0 = -3,111 < -2,4851 = -t_{1-\alpha}^{n-1+m-1}$ $\Rightarrow$ bác bỏ H_0 $\Rightarrow$ chấp nhận H_1 tức là $\mu_1 - \mu_2 < D_0 = 0$ hay $\mu_1 < \mu_2$.

Vậy chất xúc tác 2 tạo ra hiệu suất trung bình cao hơn hiệu suất của chất xúc tác 1, với mức ý nghĩa $\alpha = 1\%$.

■ Tính p-giá trị: với biến ngẫu nhiên $T_{n-1+m-1}$ có phân phối Student$(n-1+m-1)$ ta có

$$p\text{-giá trị} = \mathbb{P}\big(T_{n-1+m-1} < t_0\big) = \mathbb{P}\big(T_{25} < -3,111\big) = 1 - \mathbb{P}\big(T_{25} < 3,111\big)$$

tra bảng phân phối Student với bậc tự do $df = n-1+m-1 = 25$ ta thấy

$$\mathbb{P}\big(T_{25} \le 2,7874\big) = 0,995 \text{ và } \mathbb{P}\big(T_{25} \le 3,7251\big) = 0,9995 \Longrightarrow \mathbb{P}\big(T_{25} < 3,111\big) \approx \frac{0.995 + 0.9995}{2} = 0.9975;$$

$$\Longrightarrow \quad p\text{-giá trị} = 1 - \mathbb{P}\big(T_{25} < 3,111\big) \approx 1 - 0,9975 = 0,0025 .$$

□

Bài 22. Để tìm ra liệu một loại huyết thanh mới có kiềm hãm được bệnh bạch cầu hay không, 9 con chuột, tất cả các con đều trong giai đoạn tiến triển của bệnh, được chọn. Năm con chuột nhận được trị liệu và 4 con không. Thời gian sống, theo năm, từ thời điểm thí nghiệm bắt đầu là như sau

Trị liệu	2,1	5,3	1,4	4,6	0,9
Không trị liệu	1,9	0,5	2,8	3,1	

Tại mức ý nghĩa 0,05, huyết thanh có thể được nói là có hiệu quả hay không?
Giả sử hai tổng thể có phân phối chuẩn với các phương sai bằng nhau.

Hướng dẫn: Ta muốn so sánh tuổi thọ trung bình chuột được trị liệu so với KHÔNG trị liệu.

Đặt μ_1 = tuổi thọ trung bình nhóm được trị liệu và μ_2 = tuổi thọ trung bình KHÔNG trị liệu.

" huyết thanh có thể được nói là có hiệu quả hay không? "

$\Longrightarrow$ $H_1 : \mu_1 > \mu_2$ hay $\mu_1 - \mu_2 > 0$ $\Longrightarrow$ $H_0 : \mu_1 - \mu_2 \le 0$.

• Giả thuyết cần kiểm định $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases} \Leftrightarrow \begin{cases} H_0 : \mu_1 - \mu_2 = D_0 = 0 : \text{KHÔNG hiệu quả} \\ H_1 : \mu_1 - \mu_2 > D_0 \quad : \text{có hiệu quả} \end{cases}$;

(*Dấu "=" luôn nằm ở giả thuyết H_0*) $\left(\text{cũng có thể viết } \begin{cases} H_0 : \mu_1 - \mu_2 \le D_0 = 0 \\ H_1 : \mu_1 - \mu_2 > D_0 \end{cases}\right)$

■ Mức ý nghĩa $\alpha = 0,05$.

■ Tính thống kê kiểm định:

• Lấy mẫu: có trị liệu : cỡ mẫu $n = 5$ có trung bình mẫu $\overline{x} = 2,86$,
độ lệch chuẩn mẫu $s_1 = \sqrt{3,883} \approx 1,9705$;

KHÔNG trị liệu : cỡ mẫu $m = 4$ có trung bình mẫu $\overline{y} = 2,075$,
độ lệch chuẩn mẫu $s_2 = \sqrt{1,3625} \approx 1,1673$;

• Phương sai σ_1^2 và σ_2^2 KHÔNG biết với cỡ mẫu $n = 5 < 30$ và $m = 4 < 30$ là cỡ mẫu nhỏ

$\Longrightarrow$ dùng phân phối t-Student với bậc tự do $df = n - 1 + m - 1$.

• Dữ kiện đề bài: " Giả sử hai tổng thể có phân phối chuẩn với các phương sai bằng nhau "

$\Longrightarrow$ phương sai bằng nhau $\sigma_1^2 = \sigma_2^2$

$$\Longrightarrow \quad \text{tính} \quad s_p^2 = \frac{(n-1).s_1^2 + (m-1).s_2^2}{n-1+m-1} = \frac{(5-1)\times 3,883 + (4-1)\times 1,3625}{5-1+4-1} \approx 2,8028.$$

• Ta có thống kê kiểm định:

$$t_0 = \frac{\overline{x} - \overline{y} - D_0}{\sqrt{\dfrac{s_p^2}{n} + \dfrac{s_p^2}{m}}} = \frac{2,86 - 2,075 - 0}{\sqrt{\dfrac{2,8028}{5} + \dfrac{2,8028}{4}}} \approx 0,699 \;;$$

• Miền bác bỏ: ta sẽ bác bỏ H_0 nếu $t_0 > t_{1-\alpha}^{n-1+m-1} = t_{0,95}^{5-1+4-1} = t_{0,95}^{7} = 1,8946$

(tra bảng phân phối Student để tìm giá trị cho $t_{0,95}^7$)

• Kết luận: do $t_0 = 0,699 < 1,8946 = t_{1-\alpha}^{n-1+m-1}$ $\Rightarrow$ KHÔNG bác bỏ H_0 tức là $\mu_1 - \mu_2 \le 0$.

Vậy tuổi thọ trung bình nhóm được trị liệu KHÔNG cao hơn nhóm KHÔNG được trị liệu,
tức là huyết thanh KHÔNG hiệu quả, với mức ý nghĩa $\alpha = 5\%$. □

Bài 23. Một mẫu ngẫu nhiên 500 cư dân trưởng thành của quận Maricopa chỉ ra rằng 385 người ủng hộ việc tăng giới hạn tốc độ đường cao tốc lên 75 dặm một giờ, và một mẫu khác gồm 400 cư dân của quận Pima đã chỉ ra rằng 267 người ủng hộ giới hạn tăng lên. Những dữ liệu này có cho thấy có sự khác biệt trong việc hỗ trợ tăng giới hạn tốc độ cho cư dân của hai quận này hay không? Sử dụng $\alpha = 0,02$, và tính p-giá trị cho kiểm định này là bao nhiêu?

Hướng dẫn: Ta muốn so sánh giữa tỷ lệ dân cư quận Maricopa so với tỷ lệ dân cư quận Pima trong việc ủng hộ tăng giới hạn tốc độ đường cao tốc.

Đặt p_1 = tỷ lệ dân quận Maricopa ủng hộ và p_2 = tỷ lệ dân quận Pima ủng hộ.

" có sự khác biệt trong việc hỗ trợ tăng giới hạn tốc độ cho cư dân của hai quận này hay không? "

$\Longrightarrow$ ta muốn so sánh tỷ lệ ủng hộ của cư dân ở 2 quận:

là bằng nhau (" = ") hay khác nhau (" $\neq$ ").

$\Longrightarrow$ $H_1 :\ p_1 \neq p_2$ hay $p_1 - p_2 \neq 0$ $\Longrightarrow$ $H_0 :\ p_1 = p_2$.

• Giả thuyết cần kiểm định
$$\begin{cases} H_0 : & p_1 = p_2 \\ H_1 : & p_1 \neq p_2 \end{cases} \iff \begin{cases} H_0 : & p_1 - p_2 = D_0 = 0 \\ H_1 : & \mu_1 - \mu_2 \neq D_0 \end{cases} ;$$

(*Dấu " = " luôn nằm ở giả thuyết* H_0)

■ Mức ý nghĩa: $\alpha = 0,02$. ■ Tính thống kê kiểm định:

• Lấy mẫu: cư dân quận Maricopa : cỡ mẫu $n = 500$ có tỷ lệ mẫu $\widehat{p}_1 = \dfrac{385}{n} = \dfrac{385}{500}$;

cư dân quận Pima : cỡ mẫu $m = 400$ có tỷ lệ mẫu $\widehat{p}_2 = \dfrac{267}{m} = \dfrac{267}{400}$;

• Ta có $\widehat{p} = \dfrac{385+267}{n+m} = \dfrac{385+267}{500+400} = \dfrac{652}{900}$,

$$z_0 = \frac{\widehat{p}_1 - \widehat{p}_2 - D_0}{\sqrt{\widehat{p}.(1-\widehat{p}).\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}} = \frac{\dfrac{385}{500} - \dfrac{267}{400} - 0}{\sqrt{\dfrac{652}{900}.\left(1 - \dfrac{652}{900}\right).\left(\dfrac{1}{500} + \dfrac{1}{400}\right)}} \approx 3,42 \;;$$

• Miền bác bỏ: ta sẽ bác bỏ H_0 nếu $|z_0| > z_{1-\frac{\alpha}{2}} = z_{0,975} = 1,96$.

• Kết luận: do $|z_0| = |3,42| = 3,42 > 1,96 = z_{1-\frac{\alpha}{2}} \Longrightarrow$ bác bỏ H_0,

$\Longrightarrow$ chấp nhận H_1, tức là $\mu_1 - \mu_2 \neq D_0 = 0$;

Vậy tỷ lệ ủng hộ của cư dân ở 2 quận là KHÁC NHAU , với mức ý nghĩa $\alpha = 5\%$.

■ Tính p-giá trị: p-giá trị $= 2.\left[1 - \Phi(|z_0|)\right] = 2.\left[1 - \Phi(3,43)\right] \approx 2.\left(1 - 0,9997\right) = 0,0006$.

ta thấy p-giá trị $= 0,0006 < 0,05 = \alpha \Longrightarrow$ bác bỏ H_0. □

Bài 24. Đo chỉ số chất béo trong sữa bò X (tính theo % của 130 con bò lai F_1 Hà-Ấn), được bảng số liệu

X	3,3	3,9	4,5	5,1	5,7	6,3	6,9
n_i	2	8	35	43	22	15	5

Giả thiết rằng X có phân phối chuẩn.

(a) Hãy ước lượng trung bình chỉ số chất béo trong sữa giống bò lai trên ở độ tin cậy 99%.

(b) Nếu muốn ước lượng trung bình chỉ số chất béo trong sữa của bò lai F_1-Hà-Ấn với sai số ước lượng không quá 0,15% và độ tin cậy 99% thì phải lấy mẫu tối thiểu bao nhiêu?

(c) Biết rằng trung bình chỉ số chất béo trong sữa của giống bò Hà thuần chủng là 4,95. Hỏi việc lai tạo trên có cho trung bình chỉ số chất béo trong sữa bò tăng lên không với mức ý nghĩa 1%?

Hướng dẫn: Ta quan tâm chỉ số chất béo trung bình. Đặt $\mu =$ số chất béo trung bình trong sữa bò.
Phương sai σ^2 KHÔNG biết.

□ Lấy mẫu: cỡ mẫu $n = 2+8+35+43+22+15+5 = 130$;

trung bình mẫu $\overline{x} = \dfrac{3,3\times 2 + 3,9\times 8 + 4,5\times 35 + ... + 6,9\times 5}{n} \approx 5,1462$;

và phương sai mẫu $s^2 = \dfrac{(3,3-\overline{x})^2\times 2 + (3,9-\overline{x})^2\times 8 + ... + (6,9-\overline{x})^2\times 5}{n} \approx 0,5895$.

■ (a) Độ tin cậy $1-\alpha = 99\% = 0.99 \Longrightarrow \alpha = 0.01$.

• Dung sai ε (trường hợp phương sai KHÔNG biết và cỡ mẫu lớn $n = 130 \geq 30$) được xác định bởi

$$\varepsilon = z_{1-\frac{\alpha}{2}}\cdot\frac{s}{\sqrt{n}} = z_{0.995}\cdot\frac{s}{\sqrt{n}} = z_{0.995}\cdot\frac{\sqrt{0,5895}}{\sqrt{130}} \approx 2,575\times\frac{\sqrt{0,5895}}{\sqrt{130}} \approx 0,1734 .$$

• Vậy khoảng tin cậy 99% cho chỉ số chất béo trung bình trong sữa bò là

$$\overline{x}-\varepsilon < \mu < \overline{x}+\varepsilon \Leftrightarrow 5,1462 - 0,1734 < \mu < 5,1462 + 0,1734 \Leftrightarrow 4,9728 < \mu < 5,3196 .$$

■ (b) Độ tin cậy $1-\alpha = 99\% = 0.99 \Longrightarrow \alpha = 0.01$.

□ Sai số ước lượng không quá 0,15 (đv: %),

Lưu ý: μ, $\overline{x}$, σ và ε có cùng đơn vị, và trong bài này có cùng đơn vị là %.

Do đó, sai số ước lượng không quá 0, 15 (đv: %), nghĩa là $\varepsilon \le 0,15$, mà ta có $\varepsilon = z_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$, nên

$$\varepsilon \le 0,15 \Leftrightarrow z_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \le 0,15 \Leftrightarrow n \ge \left(z_{1-\frac{\alpha}{2}} \cdot \frac{s}{0,15}\right)^2 \Leftrightarrow n \ge \left(z_{0,995} \cdot \frac{s}{0,15}\right)^2$$

$$\Leftrightarrow n \ge \left(2,575 \times \frac{\sqrt{0,5895}}{0,15}\right)^2$$

$$\Leftrightarrow n \ge 173,7224 .$$

Vậy muốn sai số ước lượng không quá 0, 15 (%) ở độ tin cậy 99%, thì phải quan sát ít nhất 174 con bò.

■ (c) " Biết rằng trung bình chỉ số chất béo trong sữa của giống bò Hà thuần chủng là 4, 95. Hỏi việc lai tạo trên có cho trung bình chỉ số chất béo trong sữa bò tăng lên không "

$\Longrightarrow$ $H_1: \ \mu > 4,95$ $\Longrightarrow$ $H_0: \ \mu \le 4,95$ (*Dấu " = " luôn nằm ở giả thuyết* H_0)

• Giả thuyết cần kiểm định $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases} \Leftrightarrow \begin{cases} H_0: \mu = \mu_0 = 4,95 & : \text{KHÔNG tăng} \\ H_1: \mu > \mu_0 & : \text{có tăng} \end{cases}$;

$\left(\text{cũng có thể viết } \begin{cases} H_0: \mu \le \mu_0 = 4,95 \\ H_1: \mu > \mu_0 \end{cases}\right)$

• Mức ý nghĩa $\alpha = 1\% = 0,01$.

• Tính thống kê kiểm định:

Lấy mẫu: cỡ mẫu $n = 130$ có trung bình mẫu $\overline{x} = 5,1462$ và độ lệch chuẩn mẫu $s = \sqrt{0,5895}$;

Phương sai σ^2 KHÔNG biết và cỡ mẫu $n = 130 \ge 30$ là mẫu lớn $\Rightarrow$ dùng phân phối chuẩn.

Ta có thống kê kiểm định: $z_0 = \dfrac{\overline{x} - \mu_0}{s/\sqrt{n}} = \dfrac{5,1462 - 4,95}{\sqrt{0,5895}/\sqrt{130}} = 2,9136$;

• Miền bác bỏ: ta sẽ bác bỏ H_0 nếu $z_0 > z_{1-\alpha} = z_{0,99} = 2,325$.

• Kết luận: do $z_0 = 2,9136 > 2,325 = z_{1-\alpha}$ $\Longrightarrow$ bác bỏ H_0

$\Longrightarrow$ chấp nhận H_1, tức là $\mu > \mu_0 = 4,95$;

Vậy việc lai tạo cho trung bình chỉ số chất béo trong sữa tăng lên, với mức ý nghĩa $\alpha = 1\%$. □

Bài 25. Tạp chí Y học New England báo cáo một thử nghiệm để đánh giá hiệu quả của phẫu thuật trên những người đàn ông được chẩn đoán mắc bệnh J. Một nửa số mẫu ngẫu nhiên của 695 (là 347) nam giới trong nghiên cứu đã phẫu thuật và 18 người trong số họ cuối cùng đã qua đời vì bệnh J so với 31 trong số 348 người không phẫu thuật. Có bằng chứng nào cho thấy rằng phẫu thuật giảm tỷ lệ những người qua đời vì bệnh J hay không? Sử dụng $\alpha = 0,05$. Tính p-giá trị.

Hướng dẫn: Ta so sánh tỷ lệ tử vong giữa nhóm được phẫu thuật so với nhóm KHÔNG phẫu thuật.

Đặt p_1 = tỷ lệ tử vong nhóm phẫu thuật và p_2 = tỷ lệ tử vong nhóm KHÔNG phẫu thuật.

" Có bằng chứng nào cho thấy rằng phẫu thuật giảm tỷ lệ những người chết vì bệnh J hay không? "

$\Longrightarrow$ $H_1: \ p_1 < p_2$ hay $p_1 - p_2 < 0$ $\Longrightarrow$ $H_0: \ p_1 \ge p_2$.

• Giả thuyết cần kiểm định $\begin{cases} H_0: p_1 = p_2 \\ H_1: p_1 < p_2 \end{cases} \Longleftrightarrow \begin{cases} H_0: p_1 - p_2 = D_0 = 0 \\ H_1: \mu_1 - \mu_2 < D_0 \end{cases}$;

(*Dấu " = " luôn nằm ở giả thuyết* H_0) $\left(\text{cũng có thể viết } \begin{cases} H_0: p_1 - p_2 \ge D_0 = 0 \\ H_1: p_1 - p_2 < D_0 \end{cases}\right)$

■ Mức ý nghĩa: $\alpha = 0,05$.

■ Tính thống kê kiểm định:

• Lấy mẫu: nhóm phẫu thuật : cỡ mẫu $n = 347$ có tỷ lệ mẫu $\widehat{p}_1 = \frac{18}{n} = \frac{18}{347}$;

nhóm KHÔNG phẫu thuật : cỡ mẫu $m = 348$ có tỷ lệ mẫu $\widehat{p}_2 = \frac{31}{m} = \frac{31}{348}$;

• Ta có $\widehat{p} = \frac{18+31}{n+m} = \frac{18+31}{347+348} = \frac{49}{695}$,

$$z_0 = \frac{\widehat{p}_1 - \widehat{p}_2 - D_0}{\sqrt{\widehat{p}.(1-\widehat{p}).\left(\frac{1}{n}+\frac{1}{m}\right)}} = \frac{\frac{18}{347} - \frac{31}{348} - 0}{\sqrt{\frac{49}{695}.\left(1-\frac{49}{695}\right).\left(\frac{1}{347}+\frac{1}{348}\right)}} \approx -1,916 \ ;$$

• Miền bác bỏ: ta sẽ bác bỏ H_0 nếu $z_0 < -z_{1-\alpha} = -z_{0,95} = -1,645$.

• Kết luận: do $z_0 = -1,916 < -1,645 = -z_{1-\alpha} \implies$ bác bỏ H_0,

$\implies$ chấp nhận H_1, tức là $p_1 - p_2 < 0$;

Vậy phẫu thuật giúp GIẢM tỷ lệ tử vong , với mức ý nghĩa $\alpha = 5\%$.

■ Tính p-giá trị: p-giá trị $= \Phi(z_0) = \Phi(-1,916) = 1 - \Phi(1,916) \approx 1 - 0,9723 = 0,0277$.

ta thấy p-giá trị $= 0,0277 < 0,05 = \alpha \implies$ bác bỏ H_0. □

Bài 26. Một bác sĩ dinh dưỡng nghiên cứu một chế độ ăn kiêng và tập thể dục mới để làm giảm lượng đường trong máu của các bệnh nhân bị bệnh tiểu đường. 10 bệnh nhận bị bệnh tiểu đường được chọn để thử nghiệm chương trình này, bảng kết quả bên dưới cho biết lượng đường trong máu trước và sau khi các bệnh nhân tham gia chương trình

Trước	268	225	252	192	307	228	246	298	231	185
Sau	106	186	223	110	203	101	211	176	194	203

Số liệu được cung cấp có đủ bằng chứng để kết luận rằng chế độ ăn kiêng và tập thể dục có tác dụng làm giảm lượng đường trong máu hay không? Sử dụng mức ý nghĩa $\alpha = 5\%$.

Hướng dẫn: Bài tập tương tự đã sửa trên lớp lý thuyết.

Bài 27. Một người sử dụng một sợi dây thun (hoặc một lò xo) có độ đàn hồi cao để làm một cái cân đơn giản. Anh ta treo các vật nặng lên dây và ghi lại chiều dài sợi dây cho mỗi lần cân. Dưới đây là dữ liệu cho một số lần cân:

Khối lượng g (x)	50	100	150	200	250	300	350	400
Chiều dài mm (y)	37	48	60	71	80	90	102	109

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

(b) Bạn dự đoán chiều dài sợi dây là bao nhiêu nếu một vật có trọng lượng 375g được treo lên?

(c) Tính hệ số xác định R^2. Nhận xét mối liên hệ tuyến tính giữa chiều dài sợi dây và trọng lượng vật nặng.

(d) Tính hệ số tương quan r_{xy}.

Hướng dẫn: ■ (a) Phương trình đường thẳng hồi quy tuyến tính có dạng $y = \widehat{\beta}_0 + \widehat{\beta}_1 \cdot x$.

Công thức tính $\widehat{\beta}_1$ và $\widehat{\beta}_0$:

$$\widehat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\left(\sum_{i=1}^{n} x_i \cdot y_i\right) - n \cdot \overline{x} \cdot \overline{y}}{\left(\sum_{i=1}^{n} x_i^2\right) - n.(\,\overline{x}\,)^2} \qquad \text{và} \qquad \widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \cdot \overline{x} \ .$$

Ta có

- $\sum_{i=1}^{n} x_i \cdot y_i = 50 \times 37 + 100 \times 48 + 150 \times 60 + ... + 400 \times 109 = 156150$;
- $\overline{x} = \dfrac{x_1 + x_2 + ... + x_n}{n} = \dfrac{50 + 100 + 150 + ... + 400}{8} = 225$;
- $\overline{y} = \dfrac{y_1 + y_2 + ... + y_n}{n} = \dfrac{37 + 48 + 60 + ... + 109}{8} \approx 74,625$;
- $\sum_{i=1}^{n} x_i^2 = 50^2 + 100^2 + 150^2 + ... + 400^2 = 510000$;

thay vào công thức ta được $S_{xy} = 21825$ và $S_{xx} = 105000$, do đó

$$\widehat{\beta}_1 \approx 0,2079 \qquad \text{và} \qquad \widehat{\beta}_0 \approx 27,8571 \ .$$

Vậy phương trình đường thẳng hồi quy tuyến tính là $y = \widehat{\beta}_0 + \widehat{\beta}_1 \cdot x = 27,8571 + 0,2079 \cdot x$.

□ Ý nghĩa của hệ số $\widehat{\beta}_1$: hệ số $\widehat{\beta}_1$ là hệ số góc của phương trình đường thẳng hồi quy.

Vì $\widehat{\beta}_1 = 0,2079 > 0 \implies$ nếu x tăng thì y cũng tăng theo.

Chẳng hạn, nếu x tăng thêm (hoặc giảm đi) 1 đơn vị thì y sẽ tăng thêm (hoặc giảm đi) 0,2079 đơn vị.

■ (b) Với $x = 375$, thay vào phương trình đường thẳng hồi quy tìm được ở câu (a) ta được

$$y = \widehat{\beta}_0 + \widehat{\beta}_1.x = 27,8571 + 0,2079 \times 375 = 105,8196 \ .$$

Vậy nếu treo một vật nặng 375g thì chiều dài của sợi dây sẽ khoảng 105,8196 mm .

■ (c) Ta có hệ số xác định $R^2 = \dfrac{SSR}{SST}$ (lưu ý $0 \le R^2 \le 1$) với

$SSR = \widehat{\beta}_1.S_{xy} \approx 0,2079 \times 21825 = 4537,418$, và $SST = \sum_{i=1}^{n} (y_i - \overline{y})^2 = 4547,875$, suy ra $R^2 \approx 0,9975$.

Vì R^2 gần 1, nên giữa y (chiều dài sợi dây) và x (trọng lượng vật nặng) có mối liên hệ tuyến tính mạng.

■ (d) Hệ số tương quan r_{xy} được tính bởi $r_{xy} = \dfrac{S_{xy}}{\sqrt{S_{xx} \times SST}} = \dfrac{21825}{\sqrt{105000 \times 4547,875}} \approx 0,9987$;

Vì $r_{xy} > 0$ nên giữa x và y có mối tương quan dương,

và vì r_{xy} gần 1 nên mối liên hệ tuyến tính thuận giữa x và y mạnh.

□

Bài 28. Một nghiên cứu về mối liên hệ giữa tỷ lệ cây xanh (X - đơn vị: m ^{2} / người) và nhiệt độ trung bình trong mùa hè (Y - đơn vị: $°C$), khảo sát tại 12 thành phố (các thành phố này có cùng kiểu khí hậu địa lý), thu được dữ liệu sau

$$\sum_{j=1}^{12} x_j = 241; \quad \sum_{j=1}^{12} x_j^2 = 7281; \quad \sum_{j=1}^{12} x_j . y_j = 6404; \quad \sum_{j=1}^{12} y_j = 341,5; \quad \sum_{j=1}^{12} y_j^2 = 9813,25;$$

(a) Tìm phương trình đường thẳng hồi quy tuyến tính bằng phương pháp bình phương nhỏ nhất.

(b) Nếu một thành phố (cùng kiểu khí hậu này) có tỷ lệ cây xanh $x_0 = 24$ m^2/người, dựa vào đường thẳng hồi quy tìm được ở câu (a) hãy dự đoán nhiệt độ trung bình trong mùa hè của thành phố đó.

(c) Tính hệ số xác định R^2 và nhận xét về mối liên hệ tuyến tính giữa tỷ lệ cây xanh và nhiệt độ trung bình trong mùa hè.

(d) Tính hệ số tương quan r_{xy}.

Hướng dẫn: ■ (a) Phương trình đường thẳng hồi quy tuyến tính có dạng $y = \widehat{\beta}_0 + \widehat{\beta}_1 . x$.

Ước lượng cho hệ số $\widehat{\beta}_1$ và $\widehat{\beta}_0$ là $\widehat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}} = \dfrac{\left(\sum_{i=1}^{n} x_i . y_i\right) - n . \overline{x} . \overline{y}}{\left(\sum_{i=1}^{n} x_i^2\right) - n.(\overline{x})^2}$ và $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 . \overline{x}$.

Từ dữ kiện đề bài, ta có $n = 12$ ("khảo sát tại 12 thành phố")

- $\sum_{i=1}^{n} x_i . y_i = 6404$ và $\sum_{i=1}^{n} x_i^2 = 7281$;
- $\overline{x} = \dfrac{x_1 + x_2 + ... + x_n}{n} = \dfrac{241}{12} = 20,0833$;
- $\overline{y} = \dfrac{y_1 + y_2 + ... + y_n}{n} = \dfrac{341,5}{12} \approx 28,4583$;

thay vào công thức ta được $S_{xy} = -454,4583$ và $S_{xx} = 2440,9167$, do đó

$$\widehat{\beta}_1 \approx -0,1862 \quad \text{và} \quad \widehat{\beta}_0 \approx 32,1975 .$$

Vậy phương trình đường thẳng hồi quy tuyến tính là $y = \widehat{\beta}_0 + \widehat{\beta}_1 . x = 32,1975 - 0,1862 . x$.

□ Ý nghĩa của hệ số $\widehat{\beta}_1$: hệ số $\widehat{\beta}_1$ là hệ số góc của phương trình đường thẳng hồi quy.

Vì $\widehat{\beta}_1 = -0,1862 < 0 \implies$ nếu x tăng (hoặc giảm) thì y sẽ giảm (hoặc tăng) theo.

Chẳng hạn, nếu x tăng thêm 1 đơn vị thì y sẽ giảm đi $-0,1862$ đơn vị.

■ (b) Với $x_0 = 24$, dự đoán nhiệt độ trung bình trong mùa hè $\widehat{y} = 32,1975 - 0,1862 \times 24 = 27,7287\,°\mathrm{C}$;

■ (c) Ta có hệ số xác định $R^2 = \dfrac{SSR}{SST}$ (lưu ý $0 \leq R^2 \leq 1$)

với $SSR = \widehat{\beta}_1.S_{xy} \approx -0,1862 \times (-454,4583) = 84,6201$,

và $SST = \sum_{i=1}^{n} (y_i - \overline{y})^2 = \left(\sum_{i=1}^{n} y_i^2\right) - n \times (\overline{y})^2 = 94,7292,$ nên $R^2 \approx 0,8933$.

Vì R^2 khá gần 1, nên giữa y (nhiệt độ trung bình trong mùa hè) và x (tỷ lệ cây xanh) có mối liên hệ tuyến tính mạnh.

■ (d) Hệ số tương quan r_{xy} (lưu ý $-1 \leq r_{xy} \leq 1$) được tính bởi

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx} \times SST}} = \frac{-454,4583}{\sqrt{2440,9167 \times 94,7292}} \approx -0,9451;$$

Vì $r_{xy} < 0$ nên giữa x và y có mối tương quan nghịch,

mà r_{xy} rất gần -1 nên mối liên hệ tuyến tính nghịch giữa x và y là mạnh.

□

- - - Hết - - -