TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA TOÁN–TIN HỌC

Khoa Toán - Tin học
Fac. of Math. & Computer Science

Bài tập nộp - Lý thuyết thống kê

Họ và tên: Nguyễn Thanh Phong MSSV: 22110154

Ngày 13 tháng 4 năm 2025

Bài tập 1. *Một mẫu ngẫu nhiên cỡ* $n = 3, (X_1; X_2; X_3)$ *được chọn từ tổng thể có trung bình* μ *và phương sai* σ^2*. Với* $a \in \mathbb{R}$*, đặt*

$$T_a = \frac{1}{4}\left(X_1 + aX_2 + X_3\right).$$

(Do $(X_1; X_2; \ldots; X_n)$ *được chọn từ tổng thể có trung bình* μ *và phương sai* σ^2*, nghĩa là* $X_1, X_2, \ldots, X_n$ *là các biến ngẫu nhiên độc lập và có cùng phân phối xác suất*

$$\Longrightarrow \quad \mathbb{E}(X_1) = \mathbb{E}(X_2) = \ldots = \mathbb{E}(X_n) = \mu \quad \text{và} \quad \operatorname{Var}(X_1) = \operatorname{Var}(X_2) = \ldots = \operatorname{Var}(X_n) = \sigma^2)$$

a. Tìm a ***sao cho*** T_a ***là một ước lượng không chệch của*** μ ***nghĩa là*** $\mathbb{E}(T_a) = \mu(*)$***.***

$$\begin{aligned}
\mathbb{E}(T_a) &= \mathbb{E}\left(\frac{1}{4}(X_1 + aX_2 + X_3)\right) \\
&= \mathbb{E}\left(\frac{1}{4}X_1\right) + \mathbb{E}\left(\frac{1}{4}aX_2\right) + \mathbb{E}\left(\frac{1}{4}X_3\right) \\
&= \frac{1}{4}\mu + \frac{a}{4}\mu + \frac{1}{4}\mu = \frac{\mu}{4}(2 + a)
\end{aligned}$$

Từ $(*)$ *ta giải:* $\frac{\mu}{4}(2 + a) = \mu \iff a = 2$

b. Tính phương sai của T_a***.***

$$\begin{aligned}
\operatorname{Var}(T_a) &= \operatorname{Var}\left(\frac{1}{4}(X_1 + aX_2 + X_3)\right) \\
&= \operatorname{Var}\left(\frac{1}{4}X_1\right) + \operatorname{Var}\left(\frac{1}{4}aX_2\right) + \operatorname{Var}\left(\frac{1}{4}X_3\right) \\
&= \frac{1}{16}\operatorname{Var}(X_1) + \frac{a^2}{16}\operatorname{Var}(X_2) + \frac{1}{16}\operatorname{Var}(X_3) \\
&= \frac{\sigma^2}{16}\left(2 + a^2\right) \\
&= \frac{3}{8}\sigma^2 \quad (\textit{thay } a = 2).
\end{aligned}$$

c. Xét trung bình bình phương sai số MSE$(T_a, \mu) = \mathbb{E}\left[(T_a - \mu)^2\right]$***. Tìm giá trị của*** a ***sao cho*** T_a ***là ước lượng tốt nhất theo nghĩa cực tiểu hóa trung bình bình phương sai số.***
Trung bình bình phương sai số (MSE) của T_a *được định nghĩa là:*

$$\operatorname{MSE}(T_a, \mu) = \mathbb{E}\left[(T_a - \mu)^2\right]$$

Ta có:

$$\operatorname{MSE}(T_a, \mu) = \operatorname{Var}(T_a) + \left(\mathbb{E}(T_a) - \mu\right)^2$$

Từ câu (a)*, ta có:*

$$\mathbb{E}(T_a) = \frac{\mu}{4}(2 + a)$$

Do đó:

$$\begin{aligned}(\mathbb{E}(T_a)-\mu)^2 &= \left(\frac{\mu}{4}(2+a)-\mu\right)^2 = \left(\mu\left(\frac{2+a}{4}-1\right)\right)^2 \\ &= \left(\mu\left(\frac{2+a-4}{4}\right)\right)^2 = \left(\mu\left(\frac{a-2}{4}\right)\right)^2 \\ &= \mu^2\left(\frac{a-2}{4}\right)^2\end{aligned}$$

Phương sai của T_a:

$$\operatorname{Var}(T_a) = \frac{\sigma^2}{16}\left(2+a^2\right)$$

Do đó:

$$\operatorname{MSE}(T_a,\mu) = \frac{\sigma^2}{16}\left(2+a^2\right) + \mu^2\left(\frac{a-2}{4}\right)^2$$

Để T_a là ước lượng tốt nhất, ta cần giá trị a cực tiểu hóa MSE. Ta có:

$$\operatorname{MSE}(T_a,\mu) = \frac{\sigma^2}{16}\left(2+a^2\right) + \frac{\mu^2(a-2)^2}{16}(*)$$

Ta tìm min của $(*)$:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}a}\left(\frac{\sigma^2}{16}\left(2+a^2\right)+\frac{\mu^2(a-2)^2}{16}\right)=0 &\iff \frac{\sigma^2}{8}(2a+4)+\frac{\mu^2(a-2)}{8}=0 \\ &\iff \sigma^2(2a+4)+\mu^2(a-2)=0 \\ &\iff \sigma^2(2a+4)+\mu^2a-2\mu^2=0 \\ &\iff a\left(\sigma^2+\mu^2\right)=2\mu^2-4\sigma^2 \\ &\iff a=\frac{2\mu^2-4\sigma^2}{\sigma^2+\mu^2}\end{aligned}$$

Và $\dfrac{\mathrm{d}^2}{\mathrm{d}a^2}\left(\dfrac{\sigma^2}{16}\left(2+a^2\right)+\dfrac{\mu^2(a-2)^2}{16}\right)=\dfrac{\sigma^2}{8}+\dfrac{\mu^2}{8}>0$

Vậy, $a=\dfrac{2\mu^2-4\sigma^2}{\sigma^2+\mu^2}$ *là giá trị cực tiểu hóa trung bình bình phương sai số.*

Bài tập 2. *Cho $\widehat{\Theta}_1$ và $\widehat{\Theta}_2$ là hai ước lượng điểm cho tham số θ. Biết $\mathbb{E}\left(\widehat{\Theta}_1\right)=\mathbb{E}\left(\widehat{\Theta}_2\right)=\theta$, $\operatorname{Var}\left(\widehat{\Theta}_1\right)=\sigma_1^2$ và $\operatorname{Var}\left(\widehat{\Theta}_2\right)=\sigma_2^2$, với $\sigma_1>0,\sigma_2>0$. Đặt $\widehat{\Theta}_3=\alpha\widehat{\Theta}_1+(1-\alpha)\widehat{\Theta}_2$, với $0<\alpha<1$.*

a. $\widehat{\Theta}_3$ là ước lượng không chệch nghĩa là $\mathbb{E}(\widehat{\Theta}_3)=\theta$.

Ta có:

$$\begin{aligned}\mathbb{E}(\widehat{\Theta}_3) &= \mathbb{E}\left(\alpha\widehat{\Theta}_1+(1-\alpha)\widehat{\Theta}_2\right) \\ &= \alpha\mathbb{E}(\widehat{\Theta}_1)+(1-\alpha)\mathbb{E}(\widehat{\Theta}_2) \\ &= \alpha\theta+(1-\alpha)\theta \\ &= \theta\end{aligned}$$

Suy ra: $\widehat{\Theta}_3$ *là ước lượng không chệch cho tham số* θ.
b. Giả sử $\widehat{\Theta}_1$ ***và*** $\widehat{\Theta}_2$ ***độc lập với nhau. Chọn hằng số*** α ***bằng bao nhiêu để cực tiểu hoá phương sai của*** $\widehat{\Theta}_3$***.***

$$\begin{aligned}\operatorname{Var}\widehat{\Theta}_3 &= \operatorname{Var}\left(\alpha\widehat{\Theta}_1 + (1-\alpha)\widehat{\Theta}_2\right)\\ &= \operatorname{Var}(\alpha\widehat{\Theta}_1) + \operatorname{Var}((1-\alpha)\widehat{\Theta}_2)\\ &= \alpha^2 + \sigma_1^2 + (1-\alpha)^2\sigma_2^2 \quad (*)\end{aligned}$$

Ta tìm min của $(*)$.

$$\frac{\mathrm{d}\operatorname{Var}\widehat{\Theta}_3}{\mathrm{d}\alpha} = 2\alpha\sigma_1^2 - 2(1-\alpha)\sigma_2^2 = 2\alpha\sigma_1^2 - 2\sigma_2^2 + 2\alpha\sigma_2^2.$$

Cho đạo hàm cấp 1 bằng 0, ta được:

$$2\alpha\sigma_1^2 - 2\sigma_2^2 + 2\alpha\sigma_2^2 = 0 \iff \alpha(\sigma_1^2 + \sigma_2^2) = \sigma_2^2 \iff \alpha = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \in (0, 1).$$

Và ta tính đạo hàm cấp 2: $\dfrac{\mathrm{d}^2\operatorname{Var}\widehat{\Theta}_3}{\mathrm{d}\alpha^2} = 2\sigma_1^2 + 2\sigma_2^2 > 0$
Suy ra: $\alpha = \dfrac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$ *thì* $\operatorname{Var}\widehat{\Theta}_3$ *cực tiểu hoá.*

Bài tập 3. *Cho* $X_1, X_2, \dots X_n$ *là mẫu ngẫu nhiên lấy từ tổng thể có hàm mật độ xác suất* $f(x;\theta) = \frac{1}{\theta}e^{-x/\theta}$ *với* $0 < x < \infty, 0 < \theta < \infty$, *và bằng 0 trong các trường hợp khác.*
a. Tìm ước lượng hợp lý cực đại $\widehat{\Theta}_{MLE}$ ***cho*** θ***.***
Với $\Theta = (0, +\infty)$, *hàm hợp lý:*

$$\begin{aligned}L(\theta; x_1, ..., x_n) &= f(x_1, \theta) \times f(x_2, \theta) \times ...f(x_n, \theta)\\ &= \frac{1}{\theta}e^{\frac{-x_1}{\theta}} \times \frac{1}{\theta}e^{\frac{-x_2}{\theta}} \times ...\frac{1}{\theta}e^{\frac{-x_n}{\theta}}\\ &= \frac{1}{\theta}e^{-\frac{x_1+x_2+...x_n}{\theta}} = \left(\frac{1}{\theta}\right)^n .e^{\frac{-S_n}{\theta}}, \text{ với } S_n = \sum_{i=1}^{n} x_i\end{aligned}$$

Ước lượng hợp lý cực đại $\widehat{\Theta}_{MLE}$ *là giá trị thoả:*

$$L(\widehat{\Theta}_{MLE}; x_1, ..., x_n) = \max_{\theta\in\Theta} L(\theta; x_1, ..., x_n)$$

Do hàm $\ln$ *là hàm đồng biến, ta có:*

$$\ln(L(\theta; x_1, ..., x_n)) = -n \times \ln(\theta) - \frac{S_n}{\theta}$$

Lấy đạo hàm cấp 1 theo θ, *ta được:*

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\ln(L(\theta; x_1, ..., x_n)) = -\frac{n}{\theta} + \frac{S_n}{\theta^2}$$

Xét cho đạo hàm cấp 1 bằng 0, ta có:

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\ln(L(\theta;x_1,...,x_n))=0 \iff -\frac{n}{\theta}+\frac{S_n}{\theta^2}=0 \iff \theta=\frac{S_n}{n}.$$

Vậy ước lượng hợp lý cực đại θ là: $\widehat{\Theta}_{MLE}=\dfrac{S_n}{n}=\dfrac{X_1+X_2+...+X_n}{n}$.

b. $\widehat{\Theta}_{MLE}$ ***là ước lượng không chệch cho*** θ ***nghĩa là*** $\mathbb{E}(\widehat{\Theta}_{MLE})=\theta$

Ta có:

$$\begin{aligned}\mathbb{E}(\widehat{\Theta}_{MLE})&=\mathbb{E}\left(\frac{X_1+X_2+...+X_n}{n}\right)\\&=\frac{\mathbb{E}(X_1)+\mathbb{E}(X_2)+...+\mathbb{E}(X_n)}{n}\\&=\frac{\theta\times n}{n}=\theta.\end{aligned}$$

Lại có: $X_1,...,X_n$ *là mẫu ngẫu nhiên nên* $X_1,...,X_n$ *là các biến ngẫu nhiên độc lập cùng phân phối.*
Do đó:

$$\begin{aligned}\mathbb{E}(X_1)=\mathbb{E}(X_2)=...=\mathbb{E}(X_n)&=\int_{-\infty}^{+\infty}x.f(x,\theta)dx=\int_0^{+\infty}x.\frac{1}{\theta}.e^{\frac{-x}{\theta}}dx\\&=-xe^{\frac{-x}{\theta}}\Big|_{x=0}^{x\to+\infty}-\left[\theta e^{\frac{-x}{\theta}}\Big|_{x=0}^{x\to+\infty}\right]\\&=0+\theta=\theta.\end{aligned}$$

c. Chứng minh rằng miền bác bỏ tối ưu để kiểm định $H_0:\theta=\theta'=1$ ***so với*** $H_1:\theta=\theta''=2$ ***là***

$$C=\left\{(X_1,X_2,\ldots,X_n):\widehat{\Theta}_{MLE}\geq k\right\},$$

với $k=k(\alpha)$ ***là một số phụ thuộc vào mức ý nghĩa*** α***.***

Phép kiểm định $H_0:\theta=\theta'=1$ *với* $H_1:\theta=\theta''=2$.
Ta sẽ áp dụng bổ đề Neyman–Pearson:
Giả sử: Tồn tại $\epsilon>0$ *và miền* W *sao cho:*

(i) $\dfrac{L(\theta'|x)}{L(\theta''|x)}\leq\epsilon,\forall x\in W.$

(ii) $\dfrac{L(\theta'|x)}{L(\theta''|x)}\geq\epsilon,\forall x\in W^c.$

Ta có:

$L(\theta'|x)=\dfrac{1}{\theta'}e^{\frac{-S_n}{\theta'}}=e^{-S_n}(H_0:\theta=\theta'=1)$

$L(\theta''|x)=\dfrac{1}{\theta''}e^{\frac{-S_n}{\theta''}}=\left(\dfrac{1}{2}\right)^n e^{\frac{-S_n}{2}}(H_1:\theta=\theta''=2)$

nên $\dfrac{L(\theta'|x)}{L(\theta''|x)}=e^{-S_n}\times 2^n\times e^{\frac{S_n}{2}}=2^n\times e^{\frac{-S_n}{2}}.$

Với $x \in W$, ta có:

$$\frac{L(\theta'|x)}{L(\theta''|x)} \leq \epsilon \iff 2^n \times e^{\frac{-S_n}{2}} \leq \epsilon \iff \frac{2^n}{\epsilon} \leq e^{\frac{S_n}{2}}$$
$$\iff \ln\left(\frac{2^n}{\epsilon}\right) \leq \frac{S_n}{2}$$
$$\iff 2 \times n \ln\left(\frac{2}{e^n}\right) \leq S_n \iff \frac{S_n}{n} \geq 2\ln\left(\frac{2}{e^{\frac{1}{n}}}\right)$$

Theo bổ đề Neyman–Pearson, ta có:

$$W = \left\{(X_1, ..., X_n) : \frac{S_n}{n} \geq 2 \times \ln\left(\frac{2}{\epsilon^{\frac{1}{n}}}\right) = k\right\}$$

là miền bác bỏ tối ưu cho phép kiểm định.

d. Trong câu (c), xét TH $n = 1$. Tìm k để mức ý nghĩa của kiếm định là $\alpha = 0,05$.

Xét $n = 1$, theo bổ đề Neyman–Pearson:

$$\alpha = \mathbb{P}\left((X_1, ..., X_n) \in W | H_0 \textit{ đúng }\right) = \mathbb{P}\left(\frac{S_n}{n} \geq k | \theta = \theta' = 1\right) = \mathbb{P}\left(X_1 \geq k | \theta' = 1\right) \ (\textit{với } n = 1)$$
$$= \int_k^{+\infty} f(x, \theta) dx \ (\textit{với } \theta = 1) = \int_k^{+\infty} \frac{1}{\theta} e^{\frac{-x}{\theta}} dx = \int_k^{+\infty} e^{-x} dx = e^{-k}$$

Tìm k để $\alpha = 0,05$

Với $n = 1$, ta có:

$$\alpha = e^{-k} \textit{ nên } \alpha = 0,05 \iff e^{-k} = 0,05 \Longrightarrow k = -\ln(0,05) = \ln(20)$$

e.

$$C = \left\{(X_1, ..., X_n) : \widehat{\Theta}_{MLE} \geq k\right\}$$

có phải miền bác bỏ tối ưu đều cho phép kiểm định $H_0 : \theta = 1$ và $H_1 : \theta > 1$ không?

Điều này tương đương với C là miền bác bỏ tối ưu cho mọi phép kiểm định.

$$\begin{cases} H_0 : & \theta = 1 \\ H_1 : & \theta = \theta_1 \textit{ với } \theta_1 \geq 1. \end{cases} \textit{Với } \theta_1 \textit{ bất kì.}$$

Xét phép kiểm định $H_0 : \theta = 1$ với $H_1 : \theta = \theta_1$

Ta có:

$$\frac{L(1|x)}{L(\theta_1|x)} = e^{-S_n} \times (\theta_1)^n \times e^{\frac{-S_n}{\theta_1}}$$

Tồn tại k: $\dfrac{L(1|x)}{L(\theta_1|x)} \leq k, \forall x \in C$

Ta có:

$$\frac{L(1|x)}{L(\theta_1|x)} \leq k \iff \theta_1^n \times e^{\frac{S_n}{\theta_1} - S_n} \leq k$$
$$\iff \frac{\theta_1^n}{k} \leq e^{S_n - \frac{S_n}{\theta_1}}$$
$$\iff \ln\left(\frac{\theta_1^n}{k}\right) \leq S_n \times \left(1 - \frac{1}{\theta_1}\right)$$
$$\iff \frac{S_n}{n} \geq \frac{\theta_1}{\theta_1 - 1} \times \ln\left(\frac{\theta_1}{k^{\frac{1}{n}}}\right) (\textit{với } k > 0)$$

Bài tập 4. *Cho $X_1, X_2, \ldots, X_n$ là mẫu ngẫu nhiên lấy từ tổng thể có hàm mật độ xác suất $f(x;\theta) = \theta x^{\theta-1}, 0 < x < 1, 0 < \theta < \infty$ và bằng 0 trong các trường hợp khác.*

a. Chứng minh rằng ước lượng hợp lý cực đại cho θ là

$$\widehat{\Theta} \equiv \widehat{\Theta}\left(X_1, \ldots, X_n\right) = \frac{-n}{\ln\left(X_1 X_2 \ldots X_n\right)}$$

Với $\Theta = (0, +\infty)$, hàm hợp lý:

$$\begin{aligned} L(\theta; x_1, ..., x_n) &= f(x_1, \theta) \times f(x_2, \theta) \times ... f(x_n, \theta) \\ &= \theta x_1^{\theta-1} \times \theta x_2^{\theta-1} \times ... \theta \times x_n^{\theta-1} \\ &= \theta^n \times \prod_{i=1}^{n} x_i^{\theta-1} \end{aligned}$$

Ước lượng hợp lý cực đại $\widehat{\Theta}$ là giá trị thoả:

$$L(\widehat{\Theta}; x_1, ..., x_n) = \max_{\theta \in \Theta} L(\theta; x_1, ..., x_n)$$

Do hàm $\ln$ là hàm đồng biến, ta có:

$$\ln(L(\theta; x_1, ..., x_n)) = n.\ln(\theta) + (\theta - 1)\ln(x_1) + ... + (\theta - 1)\ln(x_n)$$

Lấy đạo hàm cấp 1 theo θ, ta được:

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \ln(L(\theta; x_1, ..., x_n)) = \frac{n}{\theta} + \ln\left(\prod_{i=1}^{n} x_i\right)$$

Xét cho đạo hàm cấp 1 bằng 0, ta có:

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \ln(L(\theta; x_1, ..., x_n)) = 0 \iff \frac{n}{\theta} + \ln\left(\prod_{i=1}^{n} x_i\right) = 0 \iff \theta = \frac{-n}{\ln\left(\prod_{i=1}^{n} x_i\right)}$$

Vậy ước lượng hợp lý cực đại θ là: $\widehat{\Theta} \equiv \widehat{\Theta}\left(X_1, \ldots, X_n\right) = \dfrac{-n}{\ln\left(\prod_{i=1}^{n} X_i\right)} = \dfrac{-n}{\ln\left(X_1 X_2 ... X_n\right)}$.

b. Chứng minh rằng miền bác bỏ tối ưu để kiểm định $H_0 : \theta = 1$ so với $H_1 : \theta = 2$ là

$$C = \left\{\left(X_1, \ldots, X_n\right) : \widehat{\Theta}\left(X_1, \ldots, X_n\right) \geq k\right\}$$

Phép kiểm định $H_0 : \theta = \theta' = 1$ với $H_1 : \theta = \theta'' = 2$.
Ta sẽ áp dụng bổ đề Neyman–Pearson:
Giả sử: Tồn tại $\epsilon > 0$ và miền W sao cho:
(i) $\dfrac{L(\theta'|x)}{L(\theta''|x)} \leq \epsilon, \forall x \in W.$

(ii) $\dfrac{L(\theta'|x)}{L(\theta''|x)} \geq \epsilon, \forall x \in W^c.$

Ta có:
$L(\theta'|x) = \theta^n \times \prod_{i=1}^n x_i^{\theta-1} = 1(H_0 : \theta = \theta' = 1)$ *và* $L(\theta''|x) = \theta^n \times \prod_{i=1}^n x_i^{\theta-1} = 2^n \times \prod_{i=1}^n x_i(H_1 : \theta = \theta'' = 2)$
nên $\dfrac{L(\theta'|x)}{L(\theta''|x)} = \dfrac{1}{2^n \times \prod_{i=1}^n x_i}$
Với $x \in W$*, ta có:*

$$\begin{aligned}
\frac{L(\theta'|x)}{L(\theta''|x)} \leq \epsilon &\iff \frac{1}{2^n \times \prod_{i=1}^n x_i} \leq \epsilon \iff \prod_{i=1}^n x_i \geq \frac{1}{2^n \epsilon} \\
&\iff \ln\left(\prod_{i=1}^n x_i\right) \geq \ln\left(\frac{1}{2^n\epsilon}\right) \\
&\iff \frac{1}{\ln\left(\prod_{i=1}^n x_i\right)} \leq \frac{1}{\ln\left(\dfrac{1}{2^n\epsilon}\right)} \\
&\iff -\frac{1}{\ln\left(\prod_{i=1}^n x_i\right)} \geq -\frac{1}{\ln\left(\dfrac{1}{2^n\epsilon}\right)} \\
&\iff \frac{-n}{\ln\left(x_1x_2...x_n\right)} \geq \frac{-n}{\ln\left(\dfrac{1}{2^n\epsilon}\right)}
\end{aligned}$$

Theo bổ đề Neyman–Pearson, ta có:

$$W = \left\{ (X_1, ..., X_n) : \frac{-n}{\ln\left(X_1X_2...X_n\right)} \geq \frac{-n}{\ln\left(\dfrac{1}{2^n\epsilon}\right)} = k \right\}$$

là miền bác bỏ tối ưu cho phép kiểm định.
c. Trong câu (b) xét trường hợp $n = 1$***. Tìm*** k ***để mức ý nghĩa của kiểm định là*** $\alpha = 0,05$
Xét $n = 1$*, theo bổ đề Neyman–Pearson:*

$$\begin{aligned}
\alpha = \mathbb{P}\left((X_1, ..., X_n) \in W | H_0 \text{ đúng }\right) &= \mathbb{P}\left(\frac{-n}{\ln\left(X_1X_2...X_n\right)} \geq k \middle| \theta = \theta' = 1\right) \\
&= \mathbb{P}\left(\frac{-1}{\ln(X_1)} \geq k \middle| \theta' = 1\right) (n = 1) \\
&= \mathbb{P}\left(\ln(X_1) \geq \frac{-1}{k} \middle| \theta = 1\right) \\
&= \mathbb{P}(X_1 \geq e^{\frac{-1}{k}} \Big| \theta = 1) \\
&= \int_{e^{\frac{-1}{k}}}^{+\infty} f(x, \theta)dx \ (\text{với } \theta = 1) = \int_{e^{\frac{-1}{k}}}^{1} \theta.x^{\theta-1}dx \\
&= \int_{e^{\frac{-1}{k}}}^{1} 1dx = 1 - e^{\frac{-1}{k}} \ (\text{với } \theta = 1, 0 < x < 1)
\end{aligned}$$

Tìm k *để* $\alpha = 0,05$
Với $n = 1$*, ta có:*

$$\alpha = e^{-k} \text{ nên } \alpha = 0,05 \iff 1-e^{\frac{-1}{k}} = 0,05 \iff e^{\frac{-1}{k}} = 0,95 \iff \frac{-1}{k} = \ln(0,95) \iff k = \frac{-1}{\ln(0,95)}.$$

Vậy $k = \dfrac{-1}{\ln(0,95)}$ *thì* $\alpha = 0,05$ *với* $n = 1$.

d. C có phải là miền bác bỏ tối ưu đều để kiểm định $H_0 : \theta = 1$ so với $H_1 : \theta > 1$ không?

Bài tập 5. *Đo đường kính (đv:* mm*) của một số chi tiết máy do một máy tiện tự động sản xuất, ta ghi nhận được số liệu như sau*

Đường kính	*12,00*	*12,05*	*12,10*	*12,15*	*12,20*	*12,25*	*12,30*	*12,35*	*12,40*
Số lượng	*2*	*3*	*7*	*9*	*10*	*8*	*6*	*5*	*3*

a. Tính trung bình mẫu, phương sai mẫu và độ lệch chuẩn mẫu.

Cỡ mẫu: $n = \sum_{i=1}^{9} n_i = 53$

Trung bình mẫu: $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{9} n_i x_i = \dfrac{646,95}{53} \approx 12,2066$

Phương sai mẫu: $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{9} n_i(x_i - \overline{x})^2 \approx 0,0106.$

Độ lệch chuẩn mẫu: $S = \sqrt{S^2} \approx 0,1030.$

b. Tìm khoảng tin cậy (KTC) 98% cho đường kính trung bình của các chi tiết máy.

Gọi $\mu =$ *đường kính trung bình của các chi tiết máy.*
Lấy mẫu: cỡ mẫu $n = 53$*, trung bình mẫu* $\overline{x} \approx 12,2066$*, phương sai mẫu* $S^2 \approx 0,0106$*. Phương sai tổng thể* **không biết** và $n > 30$.
$\Longrightarrow$ *TH2: Tính theo phân phối Gauss* $\mathcal{N}(0,1)$*, độ tin cậy:* $1 - \alpha = 98\% = 0,98 \Longrightarrow \alpha = 0,02$
Dung sai: $\epsilon = z_{1-\frac{\alpha}{2}}.\dfrac{S}{\sqrt{n}} = z_{0,99}.\dfrac{S}{\sqrt{n}} = 2,325.\dfrac{0,1030}{\sqrt{53}} \approx 0,033.$

Vậy khoảng tin cậy 98% *cho đường kính trung bình của các chi tiết máy là:*

$$\overline{x} - \epsilon \leq \mu \leq \overline{x} + \epsilon \Longrightarrow \mu \in [12,1736; 12,2396].$$

c. Nếu muốn sai số ước lượng KTC cho trung bình không quá $E = 0,02$ mm với độ tin cậy 98% thì phải lấy cỡ mẫu tối thiểu bao nhiêu?

Ta có sai số ước lượng: $\epsilon = z_{0,99}.\dfrac{S}{\sqrt{n}}$
Để sai số ước lượng

$$\epsilon \leq 0,02 \iff z_{0,99}.\frac{S}{\sqrt{n}} \leq 0,02 \iff n \geq (z_{0,99})^2.\left(\frac{S}{0,02}\right)^2 = 2,325^2.\left(\frac{0,103}{0,02}\right)^2 \approx 143,24.$$

Vậy phải lấy tối thiểu cỡ mẫu là $n = 144$.

d. Những chi tiết máy có đường kính nhỏ hơn 12,1 và lớn hơn 12,30 được xem là không đạt tiêu chuẩn kĩ thuật. Tìm khoảng tin cậy 96% cho tỷ lệ chi tiết máy không đạt tiêu chuẩn.

Gọi $p =$ *tỷ lệ chi tiết máy không đạt tiêu chuẩn.*
Lấy mẫu: $n = 53, Y = 13.$
$\Longrightarrow$ *Tỷ lệ mẫu:* $\widehat{p} = \dfrac{Y}{n} = \dfrac{13}{53}$
Kiểm tra điều kiện: $n\widehat{p} = 13 \geq 5, n(1-\widehat{p}) = 40 \geq 5$
Mức ý nghĩa: $1-\alpha = 96\% = 0,96 \Longrightarrow \alpha = 0,04.$
Dung sai (Sai số KTC):

$$\epsilon = z_{1-\frac{\alpha}{2}}.\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = z_{0,98}.\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 2,055.\sqrt{\frac{13}{53}.\frac{40}{53.53}} \approx 0,1215.$$

Vậy khoảng tin cậy 96% *cho tỷ lệ chi tiết máy không đạt tiêu chuẩn là:*

$$\widehat{p} - \epsilon \leq p \leq \widehat{p} + \epsilon \iff \frac{13}{53} - 0,1215 \leq p \leq \frac{13}{53} + 0,1215 \iff p \in \left[\frac{13}{53} - 0,1215; \frac{13}{53} + 0,1215\right]$$

Bài tập 6. *Một khảo sát về chiều cao* $X($ cm$)$ *của một giống cây trồng, người ta quan sát một mẫu và có kết quả sau*

Chiều cao (cm)	*100*	*110*	*120*	*130*	*140*	*150*	*160*
Số cây	*10*	*10*	*15*	*30*	*10*	*10*	*15*

Giả sử chiều cao X *có phân phối chuẩn.*

Giải

Gọi $\mu =$ *là chiều cao trung bình của giống cây trồng.*
Lấy mẫu: cỡ mẫu $n = 10 + 10 + 15 + 30 + 10 + 10 + 15 = 100$

Trung bình mẫu: $\overline{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} n_i x_i = 131$

Phương sai mẫu: $s^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n} n_i(x_i - \overline{x})^2 \approx 332,3232$

$\Longrightarrow s = \sqrt{s^2} \approx 18,2297$

a. Ước lượng chiều cao trung bình của giống cây trồng trên với độ tin cậy 95%.

Phương sai tổng thể σ^2 *không biết, cỡ mẫu* $n = 100 > 30$
$\Longrightarrow$ *TH2: Tính theo phân phối Gauss* $\mathcal{N}(0,1)$*. Mức ý nghĩa:* $1-\alpha = 95\% = 0,95 \iff \alpha = 0,05$

Dung sai: $\epsilon = z_{1-\frac{\alpha}{2}} \times \dfrac{s}{\sqrt{n_x}} = z_{0,975} \times \dfrac{18,2297}{\sqrt{100}} = 1,96 \times \dfrac{18,2297}{\sqrt{100}} \approx 3,573$
Vậy khoảng tin cậy 95% *cho chiều cao trung bình của giống cây trồng trên là:*

$$\overline{x} - \varepsilon \leq \mu \leq \overline{x} + \epsilon \iff 127,427 \leq \mu \leq 134,573 \iff \mu \in [127,427; 134,573].$$

b. Những cây trồng có chiều cao từ 135 cm trở lên được gọi là những cây "cao". Hãy ước lượng tỷ lệ những cây cao với độ tin cậy 96%.

Gọi k = số lượng những cây trồng có chiều cao từ 135cm trở lên.
Lấy mẫu: $n = 100, k = 35$
Gọi p là tỷ lệ những cây cao trong số những giống cây trồng này.
Tỉ lệ mẫu: $\widehat{p} = \frac{k}{n} = \frac{35}{100} = 0,35.$
Kiểm tra điều kiện: $n\widehat{p} = 35 \geq 5$ *và* $n(1-\widehat{p}) = 65 \geq 5.$
Mức ý nghĩa: $1 - \alpha = 0,96 \iff \alpha = 0,04.$
Dung sai:

$$\epsilon = z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = z_{0,98} \times \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 2,055 \times \sqrt{\frac{0,35.0,65}{100}} \approx 0,098$$

Vậy khoảng tin cậy 96% cho tỷ lệ những cây cao trong số những giống cây trồng trên là:

$$\widehat{p} - \epsilon \leq p \leq \widehat{p} + \epsilon \iff 0,252 \leq p \leq 0,448$$

c. Một chuyên gia lâm nghiệp cho rằng chiều cao trung bình của giống cây này là 133 cm. Hãy kiểm định ý kiến trên với mức ý nghĩa 3%.

Kiểm định giả thuyết:

$$\begin{cases} H_0 : \mu = 133 \\ H_1 : \mu \neq 133 \end{cases}$$

Mức ý nghĩa: $\alpha = 3\% = 0,03$
Phương sai tổng thể σ^2 *chưa biết và cỡ mẫu* $n = 100 > 30$
Thống kê kiểm định:

$$z_0 = \frac{\overline{X} - \mu}{s_x/\sqrt{n_x}} = \frac{131 - 133}{18,2297/\sqrt{100}} \approx -1,0971$$

Miền bác bỏ: Ta sẽ bác bỏ H_0 *nếu* $|z_0| > z_{1-\frac{\alpha}{2}} = z_{0,985} \approx 2,17.$

Vì $|z_0| = 1,0971 < z_{0,985} \approx 2,17$ *nên chưa đủ cơ sở để bác bỏ* H_0.

Vậy với 97% độ tin cậy rằng chiều cao trung bình của giống cây là 133cm.

d. Người ta áp dụng phương pháp mới trong việc trồng và chăm sóc cây. Sau một thời gian, khảo sát 100 cây đã trồng theo phương pháp mới được số liệu sau

Chiều cao (cm)	*100*	*110*	*120*	*130*	*140*	*150*	*160*
Số cây	*6*	*10*	*20*	*34*	*12*	*7*	*11*

Với mức ý nghĩa 2%, kiểm định xem phương pháp mới có làm tăng chiều cao trung bình của cây không.

Gọi Y là chiều cao của một giống cây được trồng bằng phương pháp mới.
Ta có : $Y \sim \mathcal{N}(0,1).$
$n_y = 6 + 10 + 20 + 34 + 12 + 7 + 11 = 100$

$\Longrightarrow \overline{Y} = \dfrac{1}{n_y}\sum_{i=1}^{n} n_i y_i = \dfrac{100 \times 6 + ... + 160 \times 11}{100} = 130,1$

$\Longrightarrow s_y^2 = \dfrac{1}{n_y - 1}\sum_{i=1}^{n} n_i (y_i - \overline{y})^2 = \dfrac{1}{100-1}[6 \times (100-131)^2 + ... + 11 \times (160-131)^2] \approx 255,5455$

$\Longrightarrow s_y = \sqrt{s_y^2} \approx 15,9858$

Phát biểu giả thuyết và đối thuyết:

$$\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 < \mu_2 \end{cases}$$

Mức ý nghĩa: $\alpha = 2\% = 0,02$

Tính giá trị thống kê kiểm định:

Vì σ_1^2 *và* σ_2^2 *chưa biết và cỡ mẫu* $n_x = n_y = 100 > 30$ *(lớn) nên*

$$z_0 = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}}} = \frac{131 - 130,1 - 0}{\sqrt{\dfrac{332,3232}{100} + \dfrac{255,5455}{100}}} \approx 0,3712$$

Bác bỏ H_0 *khi* $z_0 < -z_{1-\alpha}$

Với $\alpha = 0,02 \Longrightarrow z_{1-\alpha} = z_{0,98} \approx 2,0537$

Do đó: $z_0 = 0,3712 > -z_{1-\alpha} = -2,0537$

Suy ra: Không đủ cơ sở để bác bỏ H_0.

Vậy với 98% độ tin cậy rằng phương pháp mới không làm tăng chiều cao trung bình của cây.

e. Có ý kiến cho rằng phương pháp mới làm tăng tỷ lệ cây "cao". Với mức ý nghĩa 1% hãy kiểm tra ý kiến này.

Gọi Z *và* T *lần lượt là số cây cao được trồng bằng phương pháp cũ và phương pháp mới.*

Ta có: $Z \sim B(n = 100, p_1)$ *và* $T \sim B(m = 100, p_2)$, $z = 35$, $t = 30$

Phát biểu giả thuyết và đối thuyết:

$$\begin{cases} H_0 : p_1 - p_2 = 0 \\ H_1 : p_1 - p_2 < 0 \end{cases}$$

Do đó ta có: $D_0 = 0$

Mức ý nghĩa: $\alpha = 1\% = 0,01$

Tính giá trị thống kê kiểm định, dưới giả định H_0 *đúng:*

$$z_0 = \frac{\widehat{P_1} - \widehat{P_2} - D_0}{\sqrt{\widehat{P}(1-\widehat{P})\left(\frac{1}{n}+\frac{1}{m}\right)}} \sim \mathcal{N}(0,1)$$

với

$$\widehat{P_1} = \frac{z}{n} = \frac{35}{100} = 0,35; \quad \widehat{P_2} = \frac{t}{m} = \frac{30}{100} = 0,3; \quad \widehat{P} = \frac{z+t}{n+m} = \frac{65}{200} = 0,325$$

Ta có

$$z_0 = \frac{0,35 - 0,3 - 0}{\sqrt{0,325 \times (1-0,325) \times \left(\frac{1}{100}+\frac{1}{100}\right)}} \approx 0,7549$$

Bác bỏ H_0 *khi* $z_0 < -z_{1-\alpha}$

Với $\alpha = 0,01 \Rightarrow z_{1-\alpha} = z_{0,99} \approx 2,3263$

Do đó: $z_0 = 0,7549 > -z_{1-\alpha} = -2,3263$

Suy ra: Không đủ cơ sở để bác bỏ H_0.

Vậy với 99% độ tin cậy rằng phương pháp mới không làm tăng tỷ lệ cây "cao".

Bài tập 7. *Tuổi thọ một loại pin (đv: giờ) do một nhà máy sản xuất có phân phối chuẩn với độ lệch chuẩn là 1,25 giờ. Khảo sát mẫu ngẫu nhiên 10 viên pin thấy trung bình tuổi thọ là 40,5 giờ.*

a. Có đủ bằng chứng để kết luận rằng tuổi thọ trung bình loại pin này lớn hơn 40 giờ hay không? Sử dụng $\alpha = 0,05$. Tính p-giá trị.

Gọi μ = *tuổi thọ trung bình của loại pin này.*

Giả thuyết kiểm định:

$$\begin{cases} H_0 : \mu \leq \mu_0 = 40; & \text{Tuổi thọ trung bình của loại pin này không lớn hơn 40 giờ} \\ H_1 : \mu \geq \mu_0 = 40; & \text{Tuổi thọ trung bình của loại pin này lớn hơn 40 giờ} \end{cases}.$$

Mức ý nghĩa: $\alpha = 0,05$.
Theo đề bài: Độ lệch chuẩn tổng thể $\sigma = 1,25$, *phương sai tổng thể* σ^2 ***đã biết***.

$\Longrightarrow$ *Tính theo phân phối Gauss* $\mathcal{N}(0,1)$.

Lấy mẫu, cỡ mẫu $n = 10$, *trung bình mẫu:* $\overline{x} = 40,5$.
Thống kê kiểm định:

$$z_0 = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{40,5 - 40}{1,25/\sqrt{10}} \approx 1,2649.$$

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

Kết luận: Vì $z_0 = 1,2649 < 1,96 = z_{0,975}$ nên chưa đủ cơ sở để bác bỏ H_0.

Vậy với mức ý nghĩa 5% tuổi thọ trung bình của loại pin này không lớn hơn 40 giờ.
Tính $p_{giá\ trị} = 1 - \Phi(z_0) = 1 - \Phi(1,2649) = 1 - 0,897 = 0,103.$

b. Nếu tuổi thọ loại pin thực sự là 42 giờ, tính xác suất sai lầm loại II.

$$\begin{aligned}\beta = \mathbb{P}(\textit{không bác bỏ } H_0 \,|\, H_1) &= \mathbb{P}(z_0 \leq z_{1-\alpha} \,|\, \mu = 42)\\ &= \mathbb{P}\left(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \leq 1,645 \,\middle|\, \mu = 42\right)\\ &= \mathbb{P}\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} + \frac{\mu - \mu_0}{\sigma/\sqrt{n}} \leq 1,645 \,\middle|\, \mu = 42\right) (*)\end{aligned}$$

Ta có: $\dfrac{\overline{X} - \mathbb{E}(\overline{X})}{\sqrt{Var(\overline{X})}} = \dfrac{X - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1).$

Do đó:

$$\begin{aligned}(*) \iff \beta &= \mathbb{P}\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq 1,645 - \frac{\mu - \mu_0}{\sigma/\sqrt{n}} \,\middle|\, \mu = 42\right)\\ &= \mathbb{P}\left(z \leq 1,645 - \frac{\mu - \mu_0}{\sigma/\sqrt{n}} \,\middle|\, \mu = 42\right) \ \textit{Với } Z \sim \mathcal{N}(0,1).\\ &= \Phi\left(1,645 - \frac{42 - 40}{1,25/\sqrt{10}}\right)\\ &= \Phi(-3,4146) = 1 - \Phi(3,4146) \approx 1 - 0,9997 = 0,0003.\end{aligned}$$

Bài tập 8. *Người ta đo hàm lượng natri (đv:* mg*) của hai mươi hộp bắp hữu cơ* 300 g *của một nhà sản xuất A và thu được dữ liệu như sau*

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.*

Giải:

Gọi $\mu =$ *là hàm lượng natri trung bình trong các hộp bắp hữu cơ của nhà sản xuất A*
Bước 1: *Giả thuyết kiểm định*

$$\begin{cases} H_0 : \mu \leq \mu_0 = 130 \textit{ các hộp bắp hữu cơ của nhà sản xuất A đảm bảo tiêu chuẩn}\\ H_1 : \mu \geq \mu_0 = 130 \textit{ các hộp bắp hữu cơ của nhà sản xuất A không đảm bảo tiêu chuẩn}\end{cases}$$

Bước 2: *Mức ý nghĩa:* $\alpha = 0,02$

Bước 3: *Lấy mẫu: cỡ mẫu* $n = 20$*, trung bình mẫu* $\overline{x} = \dfrac{\sum_{i=1}^{n} x_i}{n} = 129,747$

Phương sai mẫu: $s^2 = \sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} = 0,7681$

Phương sai tổng thể σ^2 ***không biết****, cỡ mẫu* $n = 20$

$\Longrightarrow$ *TH3: Tính theo phân phối t–Student*

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

Bước 4: *Miền bác bỏ:*

Ta sẽ bác bỏ H_0 *nếu* $t_0 > t^{n-1}_{1-\alpha} = t^{19}_{0,98} \approx 2,242$

Bước 5: *Kết luận:*

Vì $t_0 = -1,291 < 2,242$ *nên chưa đủ cơ sở để bác bỏ* H_0

Vậy với mức ý nghĩa 2%, các hộp bắp hữu cơ của nhà sản xuất A đảm bảo tiêu chuẩn.

- $p_{giá\ trị} = \mathbb{P}(T_{n-1} \geq t_0 = \mathbb{P}(T_{19} \geq -1,291) = 1 - \mathbb{P}(T_{19} \leq -1,291) = \mathbb{P}(T_{19} \leq 1,291) \approx 0,8914$

Vậy $p_{giá\ trị} \approx 0,8914$.

Bài tập 9. *Do cholesterol (đơn vị mg%) cho một nhóm người, ta ghi nhận lại được*

Chol.	150 − 160	160 − 170	170 − 180	180 − 190	190 − 200	200 − 210
Số người	*3*	*9*	*11*	*3*	*2*	*1*

Cho rằng độ cholesterol tuân theo phân phối chuẩn.

a. Tính trung bình mẫu $\overline{x}$ ***và phương sai mẫu*** s^2***.***

Cỡ mẫu: $n = 29$

Trung bình mẫu: $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{6} n_i x_i \approx 173,2759$

Phương sai mẫu: $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{6} n_i(x_i - \overline{x})^2 \approx 143,3498$

b. Tìm khoảng ước lượng cho trung bình cholesterol trong dân số ở độ tin cậy 0,95. Nếu ta muốn độ tin cậy tăng lên thì khoảng ước lượng này sẽ rộng ra hay thu hẹp lại? (Giải thích ngắn gọn mà không cần thực hiện các tính toán).

Gọi μ *là trung bình cholesterol trong dân số.*

Lấy mẫu, cỡ mẫu $n = 29$*; trung bình mẫu* $\overline{x} = 173,2659$*; phương sai mẫu* $s^2 = 143,3498$.

Phương sai tổng thể σ^2 ***không biết****, cỡ mẫu* $n = 29 < 30 \Longrightarrow$ *TH3: Tính theo phân phối t-Student.*

Độ tin cậy: $1 - \alpha = 0,95 \Rightarrow \alpha = 0,05$

Dung sai: $\epsilon = t^{n-1}_{1-\frac{\alpha}{2}}\dfrac{s}{\sqrt{n}} = t^{28}_{0,975}\sqrt{\dfrac{s^2}{n}} = 2,0484 \times \sqrt{\dfrac{143,3498}{29}} \approx 4,5542$

Vậy khoảng tin cậy 95% cho trung bình cholesterol trong dân số là:

$$\overline{x} - \epsilon \leq \mu \leq \overline{x} + \epsilon \quad \Longleftrightarrow \quad \mu \in [168,7217; 177,8301]$$

Muốn độ tin cậy tăng lên thì độ lệch của trung bình cholesterol sẽ tăng hơn so với tế. Do đó, khoảng ước lượng này sẽ rộng ra.

c. Có tài liệu cho biết lượng cholesterol trung bình là $\mu_0 = 175\text{mg}\%$. Giá trị này có phù hợp với mẫu quan sát không? (Mức ý nghĩa 4%)

Gọi μ là trung bình cholesterol trong dân số.

Bước 1: *Giả thuyết kiểm định:*

$$\begin{cases} H_0 : \mu = \mu_0 = 175mg\% : \textit{Giá trị phù hợp với mẫu quan sát.} \\ H_1 : \mu \neq \mu_0 = 175mg\% : \textit{Giá trị không phù hợp với mẫu quan sát} \end{cases}$$

Bước 2: *Mức ý nghĩa: $\alpha = 4\% = 0,04$*

Bước 3: *Phương sai tổng thể **không biết**, cỡ mẫu $n = 29 < 30 \Longrightarrow$ Tính theo phân phối Student.*

Thống kê kiểm định:

$$t_0 = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{173,2759 - 175}{\sqrt{143,3448/29}} \approx -0,7755$$

Bước 4: *Miền bác bỏ: Ta sẽ bác bỏ H_0 nếu $|t_0| > t^{n-1}_{1-\frac{\alpha}{2}} = t^{28}_{0,98} \approx 2,188$*

Bước 5: *Kết luận: Vì $|t_0| = 0,7755 < 2,188 = t^{28}_{0,98}$ nên chưa đủ cơ sở để bác bỏ H_0.*

Vậy với mức ý nghĩa 4% giá trị này phù hợp với mẫu quan sát.

d. Khi đo cholesterol trên một nhóm 40 người ở địa phương khác thì được $\bar{x}_2 = 176\text{mg}\%, s_2^2 = 145\ (mg\%)^2$. Hỏi lượng cholesterol trung bình ở 2 địa phương này có khác nhau hay không với mức ý nghĩa 3% ? Giả sử phương sai hai tổng thể bằng nhau.

Gọi μ_1 là biến ngẫu nhiên thể hiện trung bình cholesterol của dân số ở địa phương ban đầu.

Gọi μ_2 là biến ngẫu nhiền thể hiện trung bình cholesterol của dân số ở địa phương khác.

Bước 1: *Giả thuyết kiểm định:*

$$\begin{cases} H_0 : \mu_1 - \mu_2 = 0 = D_0 : \textit{Lượng cholesterol trung bình ở 2 địa phương này không khác nhau.} \\ H_1 : \mu_1 - \mu_2 \neq 0 = D_0 : \textit{Lượng cholesterol trung bình ở 2 địa phương này khác nhau.} \end{cases}$$

Bước 2: *Mức ý nghĩa: $\alpha = 3\% = 0,03$*

Bước 3: *Mẫu 1, cỡ mẫu $n_1 = 29$; trung bình mẫu: $\overline{x} = 173,2759$; phương sai mẫu: $s_1 = 143,3498$. Mẫu 2, cỡ mẫu $n_2 = 40$; trung bình mẫu: $\overline{x} = 176$; phương sai mẫu: $s_2 = 145$. Theo đề bài: Phương sai mẫu của tổng thể $\sigma_1^2 = \sigma_2^2$ **không biết**, $n = 29 < 30 \Longrightarrow$ TH 3.1*

Phương sai mẫu chung:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{28.143,3498 + 39.145}{28 + 39} \approx 144,3104$$

Tổng thể kiểm định:

$$t_0 = \frac{\overline{x_1} - \overline{x_2} - D_0}{\sqrt{\frac{s_p}{n_1} + \frac{s_p}{n_2}}} = \frac{173,2759 - 176}{\sqrt{\frac{144,3104}{29} + \frac{144,3104}{40}}} \approx -0,9298 \text{ và } t_0 \sim T(67)$$

Bước 4: *Miền bác bỏ: Ta sẽ bác bỏ H_0 nếu $|t_0| > t_{1-\frac{\alpha}{2}}^{n_!+n_2-2} = t_{0,985}^{67} \approx 2,2569$*

Bước 5: *Kết luận: Vì $|t_0| = 0,9298 < 2,2569 = t_{0,985}^{67}$ nên chưa đủ cơ sở để bác bỏ H_0.*
Vậy với mức ý nghĩa 3% lượng cholesterol trung bình ở hai địa phương này không khác nhau.

Bài tập 10. *Đo chỉ số chất béo X (đv: %) trong sữa bò (của 125 con bò thuộc một giống bò sữa lai mới của Hà Lan), ta được bảng số liệu sau*

X	3,5	3,8	4,5	5,2	5,6	6,4	6,8
n_i	2	8	35	40	20	15	5

Giả thiết rằng X có phân phối chuẩn.

(a) Tìm khoảng tin cậy 99% cho trung bình chỉ số chất béo trong sữa giống bò lai trên.

Đối tượng khảo sát có cờ mẫu lớn (125 > 30) và chưa biết phương sai tổng thể. Ta có trung bình mẫu:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{7} n_i x_i = \frac{2.3,5 + 8.3,8 + 35.4,5 + 40.5,2 + 20.5,6 + 15.6,4 + 5.6,8}{125} = 5,1592$$

và phương sai mẫu:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{7} n_i\left(x_i - \bar{x}\right)^2 \approx 0,6129$$

Độ tin cậy $1-\alpha = 99\% \Longrightarrow \alpha = 0,01$ nên ta có phân vị $z_{1-\frac{\alpha}{2}} = z_{0,995} = 2,58$. Dung sai:

$$\varepsilon = z_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}} = 2,58\sqrt{\frac{0,6129}{125}} \approx 0,1806$$

Vậy khoảng tin cậy 99% cho trung bình chỉ số chất béo trong sữa giống bò lai trên là:

$$\mu \in [\bar{x} - \varepsilon, \bar{x} + \varepsilon] = [5,1592 - 0,1806; 5,1592 + 0,1806] = [4,9786; 5,3398]$$

(b) Biết trung bình chỉ số chất béo trong sữa giống bò thuần chủng (giống bò cũ) là 4,65. Việc lai tạo có cho trung bình chỉ số chất béo của sữa bò tăng lên hay không, với $\alpha = 1\%$?

Gọi μ là trung bình chỉ số chất béo trong sữa của giống bò mới.
"Trung bình chỉ số chất béo trong sửa giống bò thuần chủng (giống bò cū) là 4,65 " $\Rightarrow \mu_0 = 4,65$.
"Việc lai tạo có cho trung bình chỉ số chất béo của sữa bò tăng lên hay không" $\Rightarrow H_1 : \mu > \mu_0 = 4,65$.

Ta tiến hành kiểm định giả thuyết $\begin{cases} H_0 : \mu = 4,65 \\ H_1 : \mu > 4,65 \end{cases}$

Thống kê kiểm định:

$$z_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{5,1592 - 4,65}{\sqrt{\dfrac{0,6129}{125}}} = 7,272$$

Ta bác bỏ H_0 *nếu* $z_0 > z_{1-\alpha} = z_{0,99} = 2,33$. *Ta có* $z_0 = 7,272 > 2,33 \Longrightarrow$ *Bác bỏ* H_0.

Vậy với mức ý nghĩa 1%*, việc lai tạo có cho trung bình chỉ số chất béo của sữa bò tăng lên.*

(c) Sữa bò được đánh giá là loại 1 nếu chỉ số chất béo nằm trong khoảng từ 4,0 đến 6,0. Có ý kiến cho rằng ít nhất 70% ***lượng sữa bò của giống bò lai mới này thuộc loại 1. Hãy kiểm định ý kiến trên với mức ý nghĩa*** 5%***. Tính*** p***-giá trị.***

Đặt Y *là số con bò mà sữa của chúng thuộc loại 1. Ta tính:*

$$\widehat{P} = \frac{Y}{n} = \frac{35 + 40 + 20}{125} = 0,76$$

Đặt p là tỷ lệ lượng sữa của giống bò mới mà thuộc loại 1 . "Có ý kiến cho rằng ít nhất 70% *lượng sữa bò của giống bò lai mới này thuộc loại 1"* $\Longrightarrow \begin{cases} p_0 = 0,7 \\ H_0 : p \geq p_0 = 0,7 \end{cases}$

Ta kiểm tra: $n \cdot p_0 = 125 \cdot 0,7 = 87,5 \geq 5$ *và* $n(1 - p_0) = 37,5 \geq 5$.
Ta tiến hành kiểm định giả thuyết

$$\begin{cases} H_0 : p \geq 0,7 \\ H_1 : p < 0,7 \end{cases}$$

Thống kê kiểm định:

$$z_0 = \frac{\widehat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0,76 - 0,7}{\sqrt{\frac{0,7.0,3}{125}}} \approx 1,464$$

Ta bác bỏ H_0 *nếu* $z_0 < -z_{1-\alpha} = -z_{0,95} = -1,64$. *Ta có* $z_0 = 1,464 > -1,64 \Rightarrow$ *Không bác bỏ* H_0.

Vậy với mức ý nghĩa 5%*, ý kiến trên là đúng. Ta tính p-giá trị:* $p = \Phi(z_0) = \Phi(1,464) = 0,9284$.

Bài tập 11. *Một máy sản xuất ống kính interocular có một máy nghiền mới sẽ được đánh giá là đủ điều kiện vận hành 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%*. Khảo sát một mẫu ngẫu nhiên gồm 260 ống kính thì thấy có 09 ống kính bị lỗi. Xây dựng và kiểm định giả thuyết phù hợp để xác định xem máy nghiền mới có đủ điều kiện vận hành hay không với mức ý nghĩa 0,05? Tính p-giá trị.*

Giải:

Bước 1: *Đặt* p = *tỷ lệ % của các ống kính được đánh bóng có chứa khuyết tật bề mặt.*
Theo đề $p_0 = 2\% = 0,02$
Giả thuyết kiểm định:

$$\begin{cases} H_0 : p \geq p_0 = 0,02 \\ H_1 : p < p_0 = 0,02 \end{cases}$$

Bước 2: *Mức ý nghĩa:* $\alpha = 0,05$

Bước 3: *Lấy mẫu: cỡ mẫu n = 260, tỷ lệ mẫu* $\widehat{p} = \dfrac{Y}{n} = \dfrac{9}{260}$

Tính giá trị thống kê kiểm định:

$$z_0 = \frac{\widehat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{\dfrac{9}{260} - 0,02}{\sqrt{\dfrac{0,02 \times 0,98}{260}}} \approx 1,6833$$

Bước 4: *Miền bác bỏ: Ta sẽ bác bỏ* H_0 *nếu* $z_0 < -z_{1-\alpha} = -z_{0,95} = -1,645$

Bước 5: *Kết luận:*
Vì $z_0 \approx 1,6833 > -1,645 = -z_{0,95}$ *nên chưa đủ cơ sở để bác bỏ* H_0. *Vậy với mức ý nghĩa* $\alpha = -0,05$, *máy mới này không đủ điều kiện.*

- $p_{\text{giá trị}} = \Phi(z_0) = \Phi(1,6833) \approx 0,9538$

Bài tập 12. *Trong một nhà máy sản xuất đồ uống, hai máy đóng chai tự động được sử dụng để đóng những chai nước có thể tích thực là 16,0 ounces. Giả sử thể tích nước trong các chai được đóng bởi hai máy trên tuân theo phân phối chuẩn với độ lệch chuẩn lần lượt là* $\sigma_1 = 0,02$ *và* $\sigma_2 = 0,025$ *ounces. Một kỹ sư quản lý chất lượng cho rằng thể tích thực của các chai nước do hai nhà máy đóng chai thực hiện là như nhau. Một mẫu ngẫu nhiên gồm 10 chai nước từ mỗi máy, cho biết*

Máy 1		*Máy 2*	
16,03	16,01	16,02	16,03
16,04	15,96	15,97	16,04
16,05	15,98	15,96	16,02
16,05	16,02	16,01	16,01
16,02	15,99	15,99	16

Với mức ý nghĩa 5%, khẳng định của người kỹ sư có đúng không? Tính p-giá trị.

Giải:

Ta quan tâm:
μ_1: *Thể tích thực trung bình của các chai nước do nhà máy 1 thực hiện.*

μ_2: *Thể tích thực trung bình của các chai nước do nhà máy 2 thực hiện.*

Bước 1: *Giả thuyết kiểm định*

$$\begin{cases} H_0: \mu_1 - \mu_2 = D_0 = 0: & \text{Thể tích thực của các chai nước do 2 nhà máy thực hiện là như nhau} \\ H_1: \mu_1 - \mu_2 \neq D_0 = 0: & \text{Thể tích thực của các chai nước do 2 nhà máy thực hiện là khác nhau} \end{cases}$$

Bước 2: *Mức ý nghĩa* $\alpha = 5\% = 0,05$

Bước 3: *Lấy mẫu:*

Mẫu 1: cỡ mẫu $n = 10$, *trung bình mẫu* $\overline{x} = 16,015$ *(ounces)*

Mẫu 2: cỡ mẫu $m = 10$, *trung bình mẫu* $\overline{y} = 16,005$ *(ounces)*

Theo đề bài ta có độ lệch chuẩn tổng thể $\sigma_1 = 0,02, \sigma_2 = 0,025$

$\Longrightarrow$ *TH1: Tính theo phân phối Gauss* $\mathcal{N}(0,1)$

• *Tính thống kê kiểm định*

$$z_0 = \frac{\overline{x} - \overline{y} - D_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} = \frac{16,015 - 16,005 - 0}{\frac{0,02^2}{10} + \frac{0,025^2}{10}} \approx 0,9877$$

Bước 4: *Miền bác bỏ*
Ta sẽ bác bỏ H_0 *nếu* $|z_0| > z_{1-\frac{\alpha}{2}} = z_{0,975} = 1,96$

Bước 5: *Kết luận*
Vì $|z_0| = 0,9877 < 1,96 = z_{0,975}$

$\Longrightarrow$ *Chưa đủ cơ sở để bác bỏ* H_0 *với mức ý nghĩa* $\alpha = 0,05$

Vậy với mức ý nghĩa $\alpha = 0,05$ *khẳng định của kỹ sư là đúng.*
Tính $p_{value} = 2.[1 - \Phi(|z_0|)] = 2.[1 - \Phi(0,9877)] \approx 0,3232.$

Bài tập 13. *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)*

Giải:

Ta quan tâm:
μ_1: *Hiệu suất trung bình của phản ứng hóa học dùng chất xúc tác 1.*

μ_2: *Hiệu suất trung bình của phản ứng hóa học dùng chất xúc tác 2.*

Bước 1: *Giả thuyết kiểm định*

$$\begin{cases} H_0 : \mu_1 - \mu_2 \geq D_0 = 0 : \ \textit{Chất xúc tác 2 không tạo ra hiệu suất trung bình hơn chất xúc tác 1.} \\ \\ H_1 : \mu_1 - \mu_2 < D_0 = 0 : \ \textit{Chất xúc tác 2 tạo ra hiệu suất trung bình hơn chất xúc tác 1.} \end{cases}$$

Bước 2: *Mức ý nghĩa* $\alpha = 0,01$

Bước 3: *Lấy mẫu :*
Mẫu 1: $n = 12$, *trung bình mẫu* $\overline{x} = 86(\%), s_1 = 3(\%)$

Mẫu 2: $m = 15$, *trung bình mẫu* $\overline{x} = 89(\%), s_1 = 2(\%)$

Theo đề bài, phương sai tổng thể σ_1^2 *và* σ_2^2 ***không biết***, $n < 30, m < 30, \sigma_1^2 = \sigma_2^2 \Longrightarrow$ *TH 3.1*

• *Tính thống kê kiểm định*

Phương sai mẫu chung: $s_p^2 = \dfrac{(n-1)s_1^2 + (m-1)s_2^2}{n-1+m-1} = \dfrac{11.3^2 + 14.2^2}{11+14} = 6,2$

$$t_0 = \frac{\overline{x} - \overline{y} - D_0}{\sqrt{\frac{s_p^2}{n} + \frac{s_p^2}{m}}} = \frac{86 - 89 - 0}{\sqrt{\frac{6,2}{12} + \frac{6,2}{15}}} = \frac{-10\sqrt{93}}{31} \approx -3,1109 \text{ và } t_0 \sim T(25)$$

Bước 4: *Miền bác bỏ:*

Ta sẽ bác bỏ H_0 *nếu* $t_0 < t^{25}_{1-\alpha} = -t^{25}_{0,99} \approx -2,4851$

Bước 5: *Vì* $t_0 = -3,1109 < -2,4851 = -t^{25}_{0,99}$

$\Rightarrow$ *Với độ tin cậy 99%, ta sẽ bác bỏ* H_0*, chấp nhận* H_1.

- $p_{\text{giá trị}} = \mathbb{P}(T_{25} \leq t_0) = \mathbb{P}(T_{25} \leq -3,1109) = 1 - \mathbb{P}(T_{25} \leq 3,1109) \approx 1 - 0,9966 = 0,0034$

Vì $p_{\text{giá trị}} = 0,0034 < 0,01 = \alpha \Longrightarrow$ *bác bỏ* H_0

Vậy với độ tin cậy 99%, chất xúc tác 2 tạo ra hiệu suất trung bình cao hơn chất xúc tác 1.

Bài tập 14. *Trong một nghiên cứu để ước tính tỷ lệ cư dân trong một thành phố nào đó và cư dân vùng ngoại ô có ủng hộ việc xây dựng nhà máy năng lượng, người ta thấy rằng có 65 trong 100 cư dân thành thị ủng hộ việc xây dựng nhà máy, trong khi chỉ 58 trong 125 cư dân ngoại ô ủng hộ. Có sự khác biệt có ý nghĩa nào giữa tỷ lệ cư dân thành thị và cư dân ngoại ô trong việc ủng hộ xây dựng nhà máy năng lượng hay không, với mức ý nghĩa 2% ? Tính p-giá trị.*

Giải

Ta quan tâm:

$$\begin{cases} p_1: & \text{Tỷ lệ dân thành thị ủng hộ việc xây dựng nhà máy năng lượng.} \\ p_2: & \text{Tỷ lệ dân ngoại ô ủng hộ việc xây dựng nhà máy năng lượng.} \end{cases}$$

Giả thuyết kiểm định:

$$\begin{cases} H_0: p_1 - p_2 = 0 = D_0: & \text{Không có sự khác biệt trong việc ủng hộ xây dựng nhà máy năng lượng} \\ H_1: p_1 - p_2 \neq 0 = D_0; & \text{Có sự khác biệt trong việc ủng hộ xây dựng nhà máy năng lượng} \end{cases}.$$

Mức ý nghĩa: $\alpha = 0,02$.

Lấy mẫu:

$$\begin{cases} \text{Mẫu 1:} \quad \text{Cỡ mẫu } n = 100, \ \text{số người ủng hộ } X = 65 \Longrightarrow \text{Tỷ lệ mẫu } \widehat{p_1} = \dfrac{X}{n} = 0,65 \\ \\ \text{Mẫu 2:} \quad \text{Cỡ mẫu } m = 125, \ \text{số người ủng hộ } Y = 58 \Longrightarrow \text{Tỷ lệ mẫu } \widehat{p_2} = \dfrac{Y}{m} = 0,464 \end{cases}$$

Tỷ lệ mẫu gộp: $\widehat{p} = \dfrac{X+Y}{n+m} = \dfrac{65+58}{100+125} \approx 0,5468$.

Thống kê kiểm định:

$$z_0 = \frac{\widehat{p_1} - \widehat{p_2} - D_0}{\sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{n} + \frac{1}{m}\right)}} = \frac{0,65 - 0,464 - 0}{\sqrt{0,5468 \times 0,4583\left(\frac{1}{100} + \frac{1}{125}\right)}} \approx 2,7849.$$

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

Kết luận: Vì $|z_0| = 2,7849 > 2,325 = z_{0,99}$ *nên là bác bỏ* H_0*, chấp nhận* H_1.

Vậy với độ tin cậy 98%, có sự khác biệt giữa tỉ lệ cư dân thành thị và cư dân ngoại ô trong việc ủng hộ xây dựng nhà máy năng lượng.

Tính $p_{\text{giá trị}} = 2.[1 - \Phi(|z_0|)] = 2.[1 - \Phi(2,7849)] \approx 2.(1 - 0,9973) = 0,0054.$

Bài tập 15. *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 $\alpha = 5\%$.

Giải:

Ta quan tâm:
μ_D = *độ sai khác trung bình lượng đường trong máu của 1 bệnh nhân trước và sau khi áp dụng phương pháp.*
Phương pháp có hiệu quả $\Longleftrightarrow$ *lượng đường trong máu trước* $>$ *lượng đường trong máu sau khi áp dụng phương pháp.*
$\Longleftrightarrow \mu_D > 0 \Rightarrow H_1 : \mu_D > 0 \Rightarrow H_0 \leq 0$
Bước 1: *Thống kê kiểm định:*

$$\begin{cases} H_0 : \mu_D \leq D_0 = 0 \\ H_1 : \mu > D_0 = 0 \end{cases}$$

Bước 2: *Mức ý nghĩa:* $\alpha = 5\% = 0,05$
Bước 3: *Lấy mẫu: cỡ mẫu n = 10*
Độ sai khác theo cặp trong mẫu

$$\begin{array}{c|c} d_1 = 268 - 106 = 162 & d_6 = 228 - 101 = 127 \\ d_2 = 225 - 186 = 39 & d_7 = 246 - 211 = 35 \\ d_3 = 252 - 223 = 29 & d_8 = 289 - 176 = 122 \\ d_4 = 192 - 110 = 82 & d_9 = 231 - 194 = 37 \\ d_5 = 307 - 203 = 105 & d_{10} = 185 - 203 = -18 \end{array}$$

Trung bình mẫu cho độ sai khác:

$$\overline{d} = \frac{d_1 + ... + d_n}{n} = 71,9$$

Phương sai mẫu của độ sai khác:

$$s_d^2 = \frac{(d_1 - \overline{d})^2 + (d_2 - \overline{d})^2 + ... + (d_n - \overline{d})^2}{n-1} = 3160,6778.$$

$\Rightarrow$ *độ lệch chuẩn mẫu:* $s_d = \sqrt{s_d^2} = 56,2199.$

- *Thống kê kiểm định:* $t_0 = \dfrac{\overline{d} - D_0}{s_d/\sqrt{n}} = \dfrac{71,9 - 0}{56,2199/\sqrt{10}} \approx 4,0443.$

Miền bác bỏ: *Ta sẽ bác bỏ* H_0 *nếu* $t_0 > t^{n-1}_{1-\alpha} = t^{10-1}_{1-0,05} = t^{9}_{0,95} = 1,8331$

Bước 5: *Kết luận:*

Vì $t_0 = 4,0443 > 1,8331 = t^{n-1}_{1-\alpha}$

$\Longrightarrow$ *Bác bỏ* H_0 $\Longrightarrow$ *Chấp nhận* $H_1 : \mu_D > 0$

Vậy với mức ý nghĩa 5%, phương pháp có hiệu quả giúp giảm lượng đường trong máu của bệnh nhân.

Bài tập 16. *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.

Giải:

Ta quan tâm: μ_1 *là thời gian sống trung bình của các con chuột nhận được trị liệu tính từ lúc thí nghiệm bắt đầu.*
μ_2 *là thời gian sống trung bình của ác con chuột không được trị liệu tính từ lúc thí nghiệm bắt đầu.*
Giả thuyết kiểm đinh:

$$\begin{cases} H_0 : \mu_1 - \mu_2 \leq 0 = D_0 : \textit{Huyết thanh không có hiệu quả.} \\ H_1 : \mu_1 - \mu_2 > 0 = D_0 : \textit{Huyết thanh có hiệu quả.} \end{cases}$$

Mức ý nghĩa: $\alpha = 0,05$
Lấy mẫu:
Mẫu 1: cỡ mẫu $n = 5$*, trung bình mẫu* $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i = 2,86$*;*

phương sai mẫu $s_1^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = 3,883$

Mẫu 2: cỡ mẫu $m = 4$*, trung bình mẫu* $\overline{y} = \dfrac{1}{m}\sum_{i=1}^{m} y_i = 2,075$*;*

phương sai mẫu $s_2^2 = \dfrac{1}{m-1}\sum_{i=1}^{m}(y_i - \overline{y})^2 = 1,3625$

Theo đề bài: Phương sai tổng thể $\sigma_1^2 = \sigma_2^2$ ***không biết****;* $n < 30$*;* $m < 30 \Rightarrow$ *Tính theo TH 3.1*
Phương sai mẫu chung:

$$s_p^2 = \frac{(n-1)s_1^2 + (m-1)s_2^2}{n-1+m-1} = \frac{4.3,883 + 3,1,3625}{4+3} \approx 2,8028$$

Thống kê kiểm định:

$$t_0 = \frac{\overline{x} - \overline{y} - D_0}{\sqrt{\frac{s_p^2}{n} + \frac{s_p^2}{m}}} = \frac{2,86 - 2,075 - 0}{\sqrt{\frac{2,8028}{5} + \frac{2,0828}{4}}} \approx 0,699 \quad \text{và } t_0 \sim T(7)$$

Miền bác bỏ: Ta sẽ bác bỏ H_0 *nếu* $t_0 > t_{1-\alpha}^{n+m-2} = t_{0,95}^{7} = 1,8946$
Kết luận: Vì $t_0 = 0,699 < 1,8946 = t_{0,95}^{7}$ *nên chưa đủ cơ sở để bác bỏ* H_0.

Vậy với mức ý nghĩa 5%, huyết thanh không có hiệu quả.

Bài tập 17. *Tạp chí Y học New England báo cáo một thử nghiệm để đánh giá hiệu quả của phẫu thuật trên những người được chẩn đoán mắc bệnh* J. *Một nửa số mẫu ngẫu nhiên của 695 (là 347) người trong nghiên cứu đã phẫu thuật và 18 người trong số họ cuối cùng đã thiệt mạng vì bệnh* J *này, so với 31 trong số 348 người không phẫu thuật. Có bằng chứng nào cho thấy rằng phẫu thuật giảm tỷ lệ những người thiệt mạng vì bệnh* J *hay không? Sử dụng* $\alpha = 0,05$. *Tính p-giá trị.*

Giải

Ta quan tâm:
p_1: *Tỷ lệ người thiệt mạng vì bệnh J trong những người đã phẫu thuật.*
p_2: *Tỷ lệ người thiệt mạng vì bệnh J trong những người không phẫu thuật.*
Giả thuyết kiểm định:

$$\begin{cases} H_0 : p_1 - p_2 \geq 0 = D_0 : \text{Phẫu thuật không giúp giảm tỷ lệ người thiệt mạng vì bệnh } J \\ H_1 : p_1 - p_2 < 0 = D_0 : \text{Phẫu thuật giúp giảm tỷ lệ người thiệt mạng vì bệnh } J \end{cases}$$

Mức ý nghĩa: $\alpha = 0,05$
Lấy mẫu:
Mẫu 1: cỡ mẫu $n = 347$, *số người thiệt mạng* $X = 18 \Longrightarrow$ *Tỷ lệ mẫu* $\widehat{p_1} = \frac{X}{n} = \frac{18}{347} \approx 0,0519$
Mẫu 2: cỡ mẫu $n = 348$, *số người thiệt mạng* $Y = 31 \Longrightarrow$ *Tỷ lệ mẫu* $\widehat{p_2} = \frac{Y}{n} = \frac{31}{348} \approx 0,0891$
Tỷ lệ mẫu gộp:

$$\widehat{p} = \frac{X+Y}{n+m} = \frac{18+31}{347+348} \approx 0,0705$$

Thống kê kiểm định:

$$z_0 \frac{\widehat{p_1} - \widehat{p_2} - D_0}{\sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{n} + \frac{1}{m}\right)}} = \frac{0,0519 - 0,0891 - 0}{\sqrt{0,0705 \times 0,9295.\left(\frac{1}{347} + \frac{1}{348}\right)}} \approx -1,9155$$

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

Kết luận: Vì $z_0 = -1,9155 < -1,645 = -z_{0,95}$ nên ta bác bỏ H_0, chấp nhận H_1.

Vậy với độ tin cậy 95%, phẫu thuật giúp giảm tỷ lệ người thiệt mạng do bệnh J.

Tính $p_{giá\ trị}$:

$$p\_giá\ trị = \Phi(z_0) = \Phi(-1,645) = 1 - \Phi(1,645) = 1 - 0,95 = 0,05$$

Bài tập 18. *Một loài hoa có 3 giống* A, B, C. *Mỗi giống hoa có thể cho hoa đỏ hoặc hoa trắng. Số liệu thống kê được cho trong bảng sau*

Màu Loài	A	B	C
Hoa đỏ	58	102	65
Hoa trắng	102	118	75

Với mức ý nghĩa 0,05 , hỏi đặc tính màu hoa và giống hoa có độc lập nhau hay không?

Giải:

Kiểm định giả thuyết:

$$\begin{cases} H_0 : \text{Màu hoa và giống hoa độc lập nhau.} \\ H_1 : \text{Màu hoa và giống hoa không độc lập nhau.} \end{cases}$$

Từ bảng số liệu trên ta có được:
$s = 3, \quad r = 2$

$n_1 = 58 + 102 + 65 = 225, \quad n_2 = 102 + 118 + 75 = 295$

$m_1 = 58 + 102 = 160, \quad m_2 = 102 + 118 = 220, \quad m_3 = 65 + 75 = 140$

Do đó: $n = n_1 + n_2 = m_1 + m_2 + m_3 = 520$

$$e_{11} = \frac{n_1 m_1}{n} \approx 69,2308 \qquad e_{12} = \frac{n_1 m_2}{n} \approx 95,1923 \qquad e_{13} = \frac{n_1 m_3}{n} \approx 60,5769$$
$$e_{21} = \frac{n_2 m_1}{n} \approx 90,7692 \qquad e_{22} = \frac{n_2 m_2}{n} \approx 124,8077 \qquad e_{23} = \frac{n_2 m_3}{n} \approx 79,4231$$

Tính giá trị thống kê Q^2

$$Q^2 = \sum_{i=1}^{2} \sum_{j=1}^{3} \frac{n_{ij}^2}{e_{ij}} - n = \left(\frac{58^2}{69,2308} + ... + \frac{75^2}{79,4231} \right) - 520 \approx 4,6389$$

Bác bỏ H_0 *khi* $Q^2 > \chi^2_{(r-1)(s-1)}(\alpha)$

Ta có: $\chi^2_{(r-1)(s-1)}(\alpha) = \chi^2_2(0,05) \approx 5,991$

Suy ra: $Q^2 = 4,6389 < \chi^2_2(0,05) = 5,991$

Kết luận: Không đủ cơ sở để bác bỏ H_0. *Vậy đặc tính màu hoa và giống hoa độc lập với nhau.*

Bài tập 19. *Chọn ngẫu nhiên 500 người ở một khu vực dân cư, dữ liệu về nhóm máu của những người này được tổng hợp trong bảng dưới đây*

A	B	AB	O
75	*150*	*15*	*260*

Theo từ điển y khoa thì tỷ lệ nhóm máu A, B, AB và O trong dân số lần lượt là $0,18; 0,28; 0,05$ *và* $0,49$.

Với mức ý nghĩa 1%, dữ liệu về nhóm máu của 500 người được thu thập có phù hợp với tỷ lệ theo lý thuyết trong từ điển y khoa hay không?

Giải

Gọi X là biến ngẫu nhiên thể hiện nhóm máu của 1 người trong 500 người này.

$$X(\Omega) \in \{A, B, AB, O\}$$

$$\begin{array}{c|c|c} P_i & E_i = n \times P_i & O_i \\ \mathbb{P}(X = A) = 0,18 & E_1 = 500.0,18 = 90 & O_1 = 75 \\ \mathbb{P}(X = B) = 0,28 & E_2 = 500.0,28 = 140 & O_2 = 150 \\ \mathbb{P}(X = AB) = 0,05 & E_3 = 500.0,05 = 25 & O_3 = 15 \\ \mathbb{P}(X = O) = 0,49 & E_4 = 500.0,46 = 245 & O_4 = 260 \end{array}$$

Giả thuyết cần kiểm định:

$$\begin{cases} H_0 : X \text{ tuân theo từ điển y khoa.} \\ H_1 : X \text{ không tuân theo từ điển y khoa.} \end{cases}$$

Mức ý nghĩa: $\alpha = 0,01$
Thống kê kiểm định:

$$Q^2 = \sum_{j=1}^{4} \frac{(O_j - E_j)^2}{E_j} = \frac{(75-90)^2}{90} + \frac{(150-140)^2}{140} + \frac{(15-25)^2}{25} + \frac{(260-245)^2}{245} \approx 8,13$$

Miền bác bỏ: Ta sẽ bác bỏ H_0 nếu $Q^2 > \chi^2_{\alpha;k-r-1} = \chi^2_{0,01;(4-0-1)} = \chi^2_{0,01;3} = 11,345$

Kết luận: Vì $Q^2 = 8,13 < 11,345 = \chi^2_{0,01;3}$ *nên chưa đủ cơ sở để bác bỏ H_0.*

Vậy với mức ý nghĩa 1%, mẫu nhóm máu thu được tuân theo từ điển y khoa.

Bài tập 20. *Điểm thi của 200 sinh viên trong một lớp học được cho bởi bảng sau*

Điểm thi	$(50; 60]$	$(60; 70]$	$(70; 80]$	$(80; 90]$	$(90; 100]$
Số sinh viên	*12*	*36*	*90*	*44*	*18*

Có ý kiến cho rằng điểm thi của sinh viên tuân theo phân phối chuẩn với điểm trung bình bằng 75 và độ lệch chuẩn bằng 8. Với $\alpha = 5\%$, hãy kiểm định ý kiến trên.

Giải

Gọi X là biến ngẫu nhiên thể hiện số điểm thi của 1 sinh viên trong lớp.
Chia miền giá trị thành 5 khoảng.

Khoảng chia P P_i	*Tần số lý thuyết* $E_i = n \times P_i$	*Tần số thực nghiệm* O_i
$\mathbb{P}(X \leq 60) = 0,0304$	$E_1 = 200.0,0304 = 6,08$	$O_1 = 12$
$\mathbb{P}(60 < X \leq 70) = 0,2356$	$E_2 = 200.0,2356 = 47,12$	$O_1 = 36$
$\mathbb{P}(70 < X \leq 80) = 0,468$	$E_3 = 200.0,468 = 93,6$	$O_1 = 9$
$\mathbb{P}(80 < X \leq 90) = 0,2356$	$E_4 = 200.0,2356 = 47,12$	$O_1 = 44$
$\mathbb{P}(90 < X) = 0,0304$	$E_5 = 200.0,0304 = 6,08$	$O_1 = 18$

Giả thuyết cần kiểm định:

$$\begin{cases} H_0 : X \sim \mathcal{N}(75, 64) \\ H_1 : X \textit{ không tuân theo } \mathcal{N}(75, 64) \end{cases}$$

Mức ý nghĩa: $\alpha = 0,05$

Thống kê kiểm định:

$$Q^2 = \sum_{j=1}^{5} \frac{(O_j - E_j)^2}{E_j} = \frac{(12-6,08)^2}{6,08} + \frac{(36-47,12)^2}{47,12} + \frac{(90-93,6)^2}{93,6} + \frac{(44-47,12)^2}{47,12} + \frac{(18-6,08)^2}{6,08} \approx 32,103$$

Miền bác bỏ: Ta sẽ bác bỏ H_0 *nếu* $Q^2 > \chi^2_{\alpha;k-r-1} = \chi^2_{0,05;(5-0-1)} = \chi^2_{0,05;4} = 9,488$

Kết luận: Vì $Q^2 = 32,103 > 9,488 = \chi^2_{0,05;4}$ *nên ta bác bỏ* H_0*, chấp nhận* H_1*.*

Vậy với độ tin cậy 95%, điểm thi của sinh viên không tuân theo phân phối chuẩn với điểm trung bình bằng 75 và độ lệch chuẩn bằng 8.

Bài tập 21. *Một công ty điện lực ở thành phố* A *thực hiện khảo sát lượng điện tiêu thụ của 14 ngày trong mùa hè, với mục đích tìm mối liên hệ giữa nhiệt độ trong một ngày mùa hè (X - đơn vị $°F$) với lượng điện tiêu thụ (Y - đơn vị:* mKW). *Cho biết*

$$\sum_{i=1}^{14} x_i = 1196; \quad \sum_{i=1}^{14} x_i^2 = 102674; \quad \sum_{i=1}^{14} x_i y_i = 27365; \quad \sum_{i=1}^{14} y_i = 319,1; \quad \sum_{i=1}^{14} y_i^2 = 7301,29;$$

Giải:

Ta có: $n = 14$

Trung bình mẫu:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{14} x_i = \frac{1}{14} \times 1196 \approx 85,4286$$

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{14} y_i = \frac{1}{14} \times 319,1 \approx 22,7929$$

Phương sai mẫu:

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{14}(x_i - \overline{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{14} x_i^2 - n.\overline{x}^2\right) \approx 38,5662$$

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{14}(y_i - \overline{y})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{14} y_i^2 - n.\overline{y}^2\right) \approx 2,1586$$

$\Longrightarrow s_x = \sqrt{s_y^2} \approx 6,2102$

$\Longrightarrow s_y = \sqrt{s_y^2} \approx 1,4692$

a. Tìm khoảng tin cậy 99% lượng điện tiêu thụ trung bình của thành phố trong một ngày mùa hè. Biết rằng lượng điện tiêu thụ Y tuân theo phân phối chuẩn.

Ta có: $Y \sim \mathcal{N}(0,1)$
Gọi μ *là lượng điện tiêu thụ trung bình của thành phố trong một ngày mùa hè.*
Ta áp dụng trường hợp 3 vì $n = 14 < 30$, Y *có phân phối chuẩn.*
Do đó: $1 - \alpha = 0,99 \Longrightarrow \alpha = 0,01$

$$\Rightarrow t_{1-\frac{\alpha}{2}}^{n-1} = t_{0,995}^{13} \approx 3,0123$$

Dung sai: $\varepsilon = t_{1-\frac{\alpha}{2}}^{n-1} \times \dfrac{s_y}{\sqrt{n}} = 3,0123 \times \dfrac{1,4692}{\sqrt{14}} \approx 1,1828$

Kết luận: Khoảng tin cậy 99% cho lượng điện tiêu thụ trung bình của thành phố trong một ngày hè là:

$$\overline{y} - \varepsilon \leq \mu \leq \overline{y} + \varepsilon \iff 31,6101 \leq \mu \leq 23,9757$$

b. Khảo sát lượng điện tiêu thụ Z trong 16 ngày mùa hè cùng năm ở thành phố B, tính được $\bar{z} = 26,25(\text{mKW})$ và $s_Z = 1,62$; Có thể khẳng định rằng lượng điện tiêu thụ của hai thành phố trong mùa hè là như nhau hay không, giả sử phương sai hai tổng thể bằng nhau? $(\alpha = 0,01)$.

Ta có: $n_y = 14$; $\overline{y} = 22,7929$; $s_y = 1,4692$

$$n_z = 16;\ \overline{z} = 26,25; \quad s_z = 1,62$$

Kiểm định giả thuyết:

$$\begin{cases} H_0 : \mu_y = \mu_z \\ H_1 : \mu_y \neq \mu_z \end{cases}$$

Mức ý nghĩa: $\alpha = 0,01$

Do $\sigma_y^2 = \sigma_z^2$ *và mẫu nhỏ nên*

- *Phương sai mẫu chung:*

$$s_p^2 = \frac{(n_y - 1)s_y^2 + (n_z - 1)s_z^2}{n_y + n_z - 2} = \frac{(1,4692)^2.13 + 15.(1,62)^2}{28} \approx 2,4081$$

- *Thống kê:*

$$t_0 = \frac{\overline{y} - \overline{z} - (\mu_y - \mu_z)}{s_p\sqrt{\dfrac{1}{n_y} + \dfrac{1}{n_z}}} = \frac{22,7929 - 26,25 - 0}{1,5518\sqrt{\dfrac{1}{14} + \dfrac{1}{16}}} \approx -6,0875$$

Đặt $df = n_y + n_z - 2 = 28$

Do đó: $t_{1-\frac{\alpha}{2}}^{df} = t_{0,995}^{28} \approx 2,7633$

Bác bỏ H_0 *khi* $|t_0| > t_{1-\frac{\alpha}{2}}^{df}$

Ta có: $|t_0| = 6,0875 > 2,7633$ *nên bác bỏ* H_0.

Kết luận: Với 99% độ tin cậy lượng điện tiêu thụ của hai thành phố trong mùa hè là khác nhau.

c. Tìm đường thẳng hồi quy ước lượng biểu diễn lượng điện tiêu thụ theo nhiệt độ.

Ta có: $\overline{x} = 85,4286; \quad \overline{y} = 22,7929$

Do đó: $S_{xy} = \sum_{i=1}^{14} x_i y_i - n\overline{x}.\overline{y} = 27365 - 14 \times 85,4286 \times 22,7929 \approx 104,6825$

$$S_{xx} = \sum_{i=1}^{14} x_i^2 - n\overline{x}^2 = 102674 - 14 \times (85,4286)^2 \approx 501,3602$$

Suy ra: $\widehat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}} = \dfrac{104,6825}{501,3602} \approx 0,2088$

$$\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1\overline{x} = 22,7929 - 0,2088 \times 85,4286 \approx 4,9554$$

Vậy đường thẳng hồi quy ước lượng biểu diễn lượng điện tiêu thụ theo nhiệt độ là:

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x = 4,9554 + 0,2088x$$

d. Tính hệ số xác định R^2 và hệ số tương quan mẫu r_{xy}. Nhận xét về mối liên hệ giữa nhiệt độ ngày hè và lượng điện tiêu thụ.

Ta có: $SSR = \sum_{i=1}^{n} (\widehat{y}_i - \overline{y})^2 = \widehat{\beta}_1 \times S_{xy} = 0,2088 \times 104,6825 \approx 21,8577$

$$SST = \sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n} y_i^2 - n.\overline{y}^2 \approx 28,0619$$

Do đó: $R^2 = \dfrac{SSR}{SST} = \dfrac{21,8577}{28,0619} \approx 0,7789$

Vì $0 < R^2 < 1$ *nên nhiệt độ ngày hè và lượng điện tiêu thụ có mối liên hệ tuyến tính yếu. Vậy 77,89% sự biến thiên lượng điện tiêu thụ được giải thích bởi nhiệt độ ngày hè.*

Bài tập 22. *Trong cấu tạo một loại dây thừng, người ta quan tâm đến hàm lượng nylon x (đơn vị: %) ảnh hưởng như thế nào đến lực căng y (đơn vị: psi) (lực kéo tối đa trước khi sợi dây bị đứt). Số liệu bên dưới cho kết quả đo tương ứng (x, y) của 8 sợi dây:*

Hàm lượng nylon	*0*	*10*	*20*	*20*	*30*	*40*	*50*	*50*
Lực căng	*160*	*240*	*320*	*340*	*395*	*450*	*510*	*520*

a. Tìm phương trình hồi quy tuyến tính đơn biểu diễn mối liên hệ của x và y.

Ta có: $n = 8; \sum_{i=1}^{8} x_i = 220; \sum_{i=1}^{8} x_i^2 = 8400; \sum_{i=1}^{8} x_iy_i = 96950; \sum_{i=1}^{8} y_i = 2935; \sum_{i=1}^{8} y_i^2 = 1190225.$

Ta có:

$$S_{xy} = \left(\sum_{i=1}^{8} x_iy_i\right) - n.\overline{x}.\overline{y} = 96950 - 8.\frac{220}{8}.\frac{2935}{8} = 16237,5$$

$$S_{xx} = \left(\sum_{i=1}^{8} x_i^2\right) - n.(\overline{x})^2 = 8400 - 8.\left(\frac{220}{8}\right)^2 = 2350.$$

Suy ra các hệ số:

$$\widehat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{162375,5}{2350} \approx 6,9096.$$

$$\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1.\overline{x} = \frac{2935}{8} - 6,9096.\frac{220}{8} = 176,861.$$

Vậy phương trình hồi quy tuyến tính đơn cần tìm là: $y = \widehat{\beta}_0 + \widehat{\beta}_1 x = 176,861 + 6,9096x.$

b. Dự đoán lực căng của một sợi dây có hàm lượng nylon bằng 45.

Vì $x = 45$ *nên ta thay vào phương trình đường thẳng hồi quy đã tìm được ở câu a, do đó*

$$y = \widehat{\beta}_0 + \widehat{\beta}_1 x = 176,841 + 6,9096.45 = 487,773$$

Vậy lực căng của một sợi dây có hàm lượng nylon bằng 45 là: 487,773 (psi)

c. Tính hệ số xác định R^2 và hệ số tương quan mẫu r_{xy}. Nhận xét về mối liên hệ giữa hàm lượng nylon và lực căng của dây thừng.

Ta có: $SSR = \sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2 = \widehat{\beta}_1 \times S_{xy} = 6,9096 \times 16237,5 \approx 112194,63$

$$SST = \sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n} y_i^2 - n.\overline{y}^2 = 1190225 - 8.(366,875)^2 \approx 113446,875$$

Do đó: $R^2 = \dfrac{SSR}{SST} = \dfrac{112194,63}{113446,875} \approx 0,989$

Ta có: $r_{xy}^2 = R^2 = 0,989$*. Lại có:* r_{xy} *và* S_{xy} *cùng dấu mà* $S_{xy} = 16237,5 > 0$*, do đó* $r_{xy} > 0$.
Vậy hệ số tương quan mẫu $r_{xy} = \sqrt{R^2} = \sqrt{0,989} \approx 0,9945$.
Vì R^2 *gần 1 nên mối liên hệ tuyến tính giữa hàm lượng nylon và lực căng sợi dây là mối liên hệ tuyến tính mạnh.*
Ngoài ra do $r_{xy} > 0$ *nên mối liên hệ giữa hàm lượng nylon và lực căng sợi dây là mối liên hệ thuận.*