XÁC SUẤT THỐNG KÊ

ĐỀ THI KẾT THÚC HỌC PHẦN
HỌC KỲ II NĂM HỌC 2022-2023 (CNTT hệ CLC)
Thời gian làm bài: 90 phút
Đề số: 1

Câu 1 (2 điểm). Một mạng máy tính gồm 20 máy tính bị tấn công bởi một con vi rút máy tính. Con vi rút này thâm nhập được vào từng máy tính với xác suất 0,6 và độc lập với các máy tính khác.

a) Tìm xác suất để con vi rút này thâm nhập được vào ít nhất 16 máy tính.

b) Nếu mạng máy tính gồm 2000 máy tính thì xác suất để con vi rút này thâm nhập được vào ít nhất 1200 máy tính. (*HD:* sử dụng xấp xỉ chuẩn để tính.)

✍ **Lời giải.**

a) Gọi X là số máy tính bị vi rút thâm nhập trong 20 máy.
Ta có $X \sim B(20; 0{,}6)$.
Xác suất để con vi rút này thâm nhập được vào ít nhất 16 máy tính là

$$\mathbb{P}(X \geq 16) = \sum_{x=16}^{20} \mathrm{C}_{20}^{x} \cdot 0{,}6^{x} \cdot (1 - 0{,}6)^{20-x} = 0{,}05095.$$

b) Gọi Y là số máy tính bị vi rút thâm nhập trong 2000 máy.
Ta có $Y \sim B(2000; 0{,}6)$.
Xác suất để con vi rút này thâm nhập được vào ít nhất 1200 máy tính là
$\mathbb{P}(Y \geq 1200) = \sum_{x=1200}^{2000} \mathrm{C}_{2000}^{x} \cdot 0{,}6^{x} \cdot (1 - 0{,}6)^{20-x}$ (không bấm máy tính được)
Tiến hành xấp xỉ phân phối Nhị thức bằng phân phối Chuẩn

$$\begin{aligned}
\mathbb{P}(Y \geq 1200) &= \mathbb{P}(Y \geq 1200 - 0{,}5) \\
&\approx \mathbb{P}\left(Z \geq \frac{1200 - 0{,}5 - np}{\sqrt{np(1-p)}}\right) \\
&= \mathbb{P}\left(Z \geq \frac{1200 - 0{,}5 - 2000 \cdot 0{,}6}{\sqrt{2000 \cdot 0{,}6 \cdot (1 - 0{,}6)}}\right) \\
&\approx \mathbb{P}(Z \geq -0{,}02282) \\
&\approx 0{,}50910.
\end{aligned}$$

□

Câu 2 (2 điểm). Thời gian để cài đặt một phần cứng nào đó là một biến ngẫu nhiên có phân phối chuẩn với độ lệch chuẩn là 5 phút. Một kỹ thuật viên máy tính cài đặt phần cứng này trên 64 máy tính khác nhau, với thời gian cài đặt trung bình là 42 phút.

a) Tìm khoảng tin cậy 96% cho thời gian cài đặt trung bình của tổng thể.

b) Có ý kiến cho rằng "thời gian cài đặt trung bình của tổng thể là trên 40 phút". Hãy kiểm định ý kiến này với mức ý nghĩa 2%.

✍ **Lời giải.**
Gọi X (phút) là thời gian để cài đặt một phần cứng trên một máy tính.
Theo đề bài, ta có $X \sim N(\mu; \sigma^2)$ với $\sigma = 5$: đã biết.
$n = 64$; $\bar{x} = 42$.

a) Tìm khoảng tin cậy 96% cho thời gian cài đặt trung bình của tổng thể.

- **Độ tin cậy:** $96\% \Rightarrow \alpha = 0{,}04$.
  $\Rightarrow z_{1-\alpha/2} = z_{0{,}98} \approx 2{,}05$.
- **Sai số:** $\epsilon = z_{1-\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}$
  $= 2{,}05 \cdot \dfrac{5}{\sqrt{64}}$
  $= \dfrac{41}{32}$
  $= 1{,}28125$.
- **KTC** 96% **cho trung bình** μ **là:**

$$\begin{aligned} \bar{x} - \epsilon \leq \ & \mu \ \leq \bar{x} - \epsilon \\ \Leftrightarrow 42 - 1{,}28125 \leq \ & \mu \ \leq 42 + 1{,}28125 \\ \Leftrightarrow 40{,}71875 \leq \ & \mu \ \leq 43{,}28125 \end{aligned}$$

  $\Rightarrow \mu \in [40{,}71875; 43{,}28125]$.

b) Có ý kiến cho rằng "thời gian cài đặt trung bình của tổng thể là trên 40 phút". Hãy kiểm định ý kiến này với mức ý nghĩa 2%.
$\sigma = 5$: đã biết.
Ta có: $n = 64$; $\bar{x} = 42$.

- **Giả Thuyết KĐ:** $\begin{cases} H_0 : \mu \leq 40 \\ H_1 : \mu > 40 \end{cases}$: KĐ 1 phía, $\mu_0 = 40$.
- **Mức ý nghĩa:** $\alpha = 0{,}02$.
- **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{42 - 40}{\frac{5}{\sqrt{64}}} = 3{,}2.$$

- **Miền bác bỏ:** Bác bỏ H_0 nếu $z_0 > z_{1-\alpha}$.
  Ta có $\alpha = 0{,}02$
  $\Rightarrow z_{1-\alpha} = z_{0{,}98} \approx 2{,}05$.
- **So sánh**
  Ta có: $z_0 > z_{1-\alpha}$
  $\Leftrightarrow 3{,}2 > 2{,}05$ (đúng)
  $\Rightarrow$ bác bỏ $H_0 : \mu \leq 40$.
  Kết luận: Với mức ý nghĩa 2% thì thời gian cài đặt trung bình của tổng thể là trên 40 phút.

□

Câu 3 (2 điểm). Giả sử ta phải chấp nhận hoặc từ chối một lô hàng lớn các linh kiện máy tính. Với mục đích kiểm soát chất lượng, ta thu thập một mẫu 200 linh kiện và tìm thấy 24 linh kiện bị lỗi trong số đó.

a) Tìm khoảng tin cậy 96% cho tỷ lệ linh kiện bị lỗi trong toàn bộ lô hàng.

b) Nhà sản xuất tuyên bố rằng cứ 10 linh kiện trong lô hàng thì có không quá 1 linh kiện bị lỗi. Hỏi rằng ta có đủ bằng chứng để phản bác tuyên bố này không với mức ý nghĩa 2%?

✍ **Lời giải.**

a) Tìm khoảng tin cậy 96% cho tỷ lệ linh kiện bị lỗi trong toàn bộ lô hàng.

Gọi Y là linh kiện bị lỗi trong toàn bộ lô hàng.

Ta có: $n = 200$, $y = 24 \Rightarrow$ tỷ lệ mẫu $\hat{p} = \dfrac{y}{n} = \dfrac{24}{200} = 0{,}12$.

• **Độ tin cậy:** $96\% \Rightarrow \alpha = 0{,}04$.

$\Rightarrow z_{1-\alpha/2} = z_{0,98} \approx 2{,}05$.

• **Sai số:** $\epsilon = z_{1-\alpha/2} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

$$= 2{,}05 \cdot \sqrt{\frac{0{,}12\,(1-0{,}12)}{200}}$$

$$\approx 0{,}04711.$$

• **KTC** 96% **cho tỷ lệ** p **là**

$$\begin{aligned} \hat{p} - \epsilon \leq \ & p \ \leq \hat{p} + \epsilon \\ \Leftrightarrow 0{,}12 - 0{,}04711 \leq \ & p \ \leq 0{,}12 + 0{,}04711 \\ \Leftrightarrow 0{,}07289 \leq \ & p \ \leq 0{,}16711 \end{aligned}$$

$\Rightarrow p \in [0{,}07289; 0{,}16711]$.

b) Nhà sản xuất tuyên bố rằng cứ 10 linh kiện trong lô hàng thì có không quá 1 linh kiện bị lỗi. Hỏi rằng ta có đủ bằng chứng để phản bác tuyên bố này không với mức ý nghĩa 2%?

• **GTKĐ:** $\begin{cases} H_0 : p \leq 0{,}1 \\ H_1 : p > 0{,}1 \end{cases}$: KĐ 1 phía, $p_0 = 0{,}1$.

• **Mức ý nghĩa:** $\alpha = 2\% = 0{,}02$.

• **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{0{,}12 - 0{,}1}{\sqrt{\dfrac{0{,}1(1-0{,}1)}{200}}} \approx 0{,}94281.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $z_0 > z_{1-\alpha}$.

Ta có $\alpha = 0{,}02$

$\Rightarrow z_{1-\alpha} = z_{0,98} \approx 2{,}05$.

• **So sánh và kết luận:**

Ta có: $\quad z_0 > z_{1-\alpha}$

$\Leftrightarrow 0{,}94281 > 2{,}05 \quad$ (sai)

$\Rightarrow$ Chưa đủ cơ sở để bác bỏ $H_0 : p \leq 0{,}1$.

Kết luận: Với mức ý nghĩa 2% thì tỷ lệ linh kiện bị lỗi của lô hàng là không quá 0,1 hay tuyên bố của nhà sản xuất là đúng.

□

Câu 4 (2 điểm). Dữ liệu sau lần lượt là về tuổi thọ pin (đv: giờ) của các laptop có trang bị ổ ghi CD (mẫu X) và không có ổ ghi CD (mẫu Y): $n = 14$; $\bar{x} = 4{,}8$; $s_X = 1{,}6$ và $m = 20$; $\bar{y} = 5{,}3$; $s_Y = 1{,}4$. Giả sử phương sai 2 tổng thể bằng nhau. Hãy kiểm định xem tuổi thọ pin của các laptop có trang bị ổ ghi CD có thấp hơn so với loại không có ổ ghi CD hay không với mức ý nghĩa 2%? Tính $p-$giá trị.

✍ **Lời giải.**

Gọi X_1 (giờ) là tuổi thọ của pin laptop có trang bị ổ ghi CD

X_2 (giờ) là tuổi thọ của pin laptop không có trang bị ổ ghi CD.

Theo đề bài $X_1 \sim N(\mu_1; \sigma_1^2)$; $X_2 \sim N(\mu_2; \sigma_2^2)$

$\sigma_1 = \sigma_2$ chưa biết.
Mẫu 1: $n_1 = 14$; $\bar{x}_1 = 4{,}8$ và $s_1 = 1{,}6$.
Mẫu 2: $n_2 = 20$; $\bar{x}_2 = 5{,}3$ và $s_2 = 1{,}4$.

- **Giả thuyết kiểm định** $\begin{cases} H_0 : \mu_1 \geq \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases}$; $\quad \Rightarrow \mu_1 - \mu_2 = 0$; $\alpha = 0{,}02$.

- **Giá trị thống kê kiểm định**

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(14-1)\cdot 1{,}6^2 + (20-1)\cdot 1{,}4^2}{14+20-2} = \frac{1763}{800}$$

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}} = \frac{4{,}8 - 5{,}3 - 0}{\sqrt{\dfrac{\frac{1763}{800}}{14} + \dfrac{\frac{1763}{800}}{20}}} \approx -0{,}96656.$$

- **Miền bác bỏ:** Bác bỏ H_0 khi $t_0 < -t_{\alpha;n_1+n_2-2}$.
  Ta có $\alpha = 0{,}02$
  $\Rightarrow -t_{\alpha;n_1+n_2-2} = -t_{0{,}02;32} \approx z_{0{,}98} \approx -2{,}05$.

- **So sánh:**
  Ta có: $\qquad t_0 < -t_{\alpha;n_1+n_2-2}$
  $\Leftrightarrow -0{,}96656 < -2{,}05 \quad$ (sai)
  $\Rightarrow$ chưa đủ cơ sở để bác bỏ $H_0 : \mu_1 \geq \mu_2$.
  Kết luận: Với mức ý nghĩa 2% thì tuổi thọ pin của các laptop có trang bị ổ ghi CD **không** thấp hơn so với loại không có ổ ghi CD.

- Tính p - giá trị:
  Vì bậc tự do $v = n_1 + n_2 - 2 = 14 + 20 - 2 = 32 \geq 30$ nên
  $p-$giá trị $= \mathbb{P}\left(T_{12+15-2} \leq t_0\right) \approx \mathbb{P}\left(Z \leq -0{,}96656\right) = \Phi(-0{,}96656) \approx 0{,}1669$.
  Vì $p-$giá trị $\approx 0{,}1669 > 0{,}02 = \alpha$ nên chưa đủ cơ sở để bác bỏ $H_0 : \mu_1 \geq \mu_2$.

□

Câu 5 (2 điểm). Các máy chủ email hiện đại và các bộ lọc chống email rác cố gắng xác định các email rác và chuyển chúng vào thư mục rác. Có nhiều cách khác nhau để phát hiện email rác, và việc nghiên cứu vẫn đang tiếp tục. Về vấn đề này, một nhân viên về bảo mật thông tin cho rằng khả năng một email là email rác có phụ thuộc vào việc nó có chứa hình ảnh hay không. Dữ liệu sau đây được thu thập trên $n = 1000$ email ngẫu nhiên

	Có hình ảnh	Không hình ảnh
Thư rác	160	240
Không phải thư rác	140	460

Với mức ý nghĩa 2% hãy cho biết ý kiến của nhân viên đó có tin được không?

XÁC SUẤT THỐNG KÊ

ĐỀ THI KẾT THÚC HỌC PHẦN
HỌC KỲ I NĂM HỌC 2022-2023
Thời gian làm bài: 90 phút
Đề số: 1

Câu 1 (2 điểm). 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 hàng 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.

✍ **Lời giải.**

a) Gọi X là số bu lông đáp ứng yêu cầu kỹ thuật trong 50 bu lông.
Khi đó $X \sim B(50; 0{,}86)$.
Xác suất có ít nhất 48 bu lông đáp ứng yêu cầu kỹ thuật là

$$\mathbb{P}(X \geq 48) = \sum_{x=48}^{50} \mathrm{C}_{50}^{x} \cdot 0{,}86^{x} \cdot (1 - 0{,}86)^{50-x} \approx 0{,}0221.$$

b) Gọi Y là số bu lông đáp ứng yêu cầu kỹ thuật trong 500 bu lông.
Khi đó $Y \sim B(500; 0{,}86)$.
Tiến hành xấp xỉ phân phối nhị thức bằng phân phối chuẩn.
Xác suất có ít nhất 250 bu lông đáp ứng yêu cầu kỹ thuật là

$$\begin{aligned}
\mathbb{P}(Y \geq 250) &= \mathbb{P}(Y \geq 250 - 0{,}5) \\
&\approx \mathbb{P}\left(Z \geq \frac{250 - 0{,}5 - np}{\sqrt{np(1-p)}}\right) \\
&= \mathbb{P}\left(Z \geq \frac{250 - 0{,}5 - 500 \cdot 0{,}86}{\sqrt{500 \cdot 0{,}86 \cdot (1 - 0{,}86)}}\right) \\
&= \mathbb{P}(Z \geq -23{,}2637) \\
&\approx 1.
\end{aligned}$$

□

Câu 2 (2 điểm). Một hệ thống tên lửa phản lực sử dụng động cơ đẩy nhiên liệu rắn. Tốc độ cháy của nhiên liệu rắn là một đặc trưng quan trọng của động cơ. Các kỹ sư biết rằng độ lệch chuẩn của tốc độ cháy là 2 cm/s. Các kỹ sư chọn cỡ mẫu là $n = 25$ thu được trung bình mẫu tốc độ cháy là $\bar{x} = 51{,}3$ cm/s. Biết rằng tốc độ cháy tuân theo phân phối chuẩn.

a) Tìm khoảng tin cậy 95% cho tốc độ cháy trung bình của nhiên nhiên liệu.

b) Thông số kỹ thuật yêu cầu tốc độ cháy trung bình của thanh nhiên liệu là 50 cm/s. Dựa trên dữ liệu đã thu thập được, hãy kiểm định xem thông số kỹ thuật này có được đáp ứng hay không với mức ý nghĩa 1%.

✍ **Lời giải.**
Gọi X (cm/s) là tốc độ cháy của thanh nhiên liệu rắn.
Theo đề bài, ta có $X \sim N(\mu; \sigma^2)$ với $\sigma = 2$: đã biết.
$n = 25$; $\bar{x} = 51{,}3$.

a) Tìm khoảng tin cậy 95% cho tốc độ cháy trung bình của nhiên nhiên liệu.

- **Độ tin cậy:** 95% $\Rightarrow \alpha = 0{,}05$.
$\Rightarrow z_{1-\alpha/2} = z_{0{,}975} \approx 1{,}96$.
- **Sai số:** $\epsilon = z_{1-\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}$
$= 1{,}96 \cdot \dfrac{2}{\sqrt{25}}$
$= 0{,}784$.

- **KTC** 95% **cho trung bình** μ **là:**

$$\begin{aligned} \bar{x} - \epsilon \leq & \quad \mu \quad \leq \bar{x} - \epsilon \\ \Leftrightarrow 51{,}3 - 0{,}784 \leq & \quad \mu \quad \leq 51{,}3 + 0{,}784 \\ \Leftrightarrow 50{,}516 \leq & \quad \mu \quad \leq 52{,}084 \end{aligned}$$

$\Rightarrow \mu \in [50{,}516; 52{,}084]$.

b) Thông số kỹ thuật yêu cầu tốc độ cháy trung bình của thanh nhiên liệu là 50 cm/s. Dựa trên dữ liệu đã thu thập được, hãy kiểm định xem thông số kỹ thuật này có được đáp ứng hay không với mức ý nghĩa 1%.
$\sigma = 2$: đã biết.
Ta có: $n = 25$; $\bar{x} = 51{,}3$.

- **Giả Thuyết KĐ:** $\begin{cases} H_0 : \mu = 50 \\ H_1 : \mu \neq 50 \end{cases}$: KĐ 2 phía, $\mu_0 = 50$.
- **Mức ý nghĩa:** $\alpha = 0{,}01$.
- **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{51{,}3 - 50}{\frac{2}{\sqrt{25}}} = 3{,}25.$$

- **Miền bác bỏ:** Bác bỏ H_0 nếu $|z_0| > z_{1-\alpha/2}$.
Ta có $\alpha = 0{,}01$
$\Rightarrow z_{1-\alpha/2} = z_{0{,}995} \approx 2{,}58$.
- **So sánh**
Ta có: $\quad |z_0| > z_{1-\alpha/2}$
$\Leftrightarrow |3{,}25| > z_{1-\alpha/2}$
$\Leftrightarrow 3{,}25 > 2{,}58 \quad$ (sai)
$\Rightarrow$ Chưa đủ cơ sở để bác bỏ $H_0 : \mu = 50$.
Kết luận: Với mức ý nghĩa 1% thì tốc độ cháy trung bình của thanh nhiên liệu là 50 cm/s hay thông số kỹ thuật này được đáp ứng.

□

Câu 3 (2 điểm). Một nhà sản xuất chất bán dẫn sản xuất bộ điều khiển sử dụng trong công nghệ động cơ ô tô. Nhà sản xuất bán dẫn lấy mẫu ngẫu nhiên gồm 200 sản phẩm và thấy rằng có bốn sản phẩm bị lỗi.

a) Tìm khoảng tin cậy 95% cho tỷ lệ lỗi của sản phẩm.

b) Khách hàng yêu cầu tỷ lệ lỗi của sản phẩm phải dưới 0,05. Hỏi nhà sản xuất có thể chứng minh khả năng đáp ứng ở mức chất lượng này cho khách hàng không? Sử dụng mức ý nghĩa 1%.

✍ **Lời giải.**

a) Tìm khoảng tin cậy 95% cho tỷ lệ lỗi của sản phẩm.
Gọi Y là số sản phẩm bị lỗi.
Ta có: $n = 200$, $y = 4 \Rightarrow$ tỷ lệ mẫu $\hat{p} = \frac{y}{n} = \frac{4}{200} = 0{,}02$.
- **Độ tin cậy:** 95% $\Rightarrow \alpha = 0{,}05$.

$\Rightarrow z_{1-\alpha/2} = z_{0{,}975} \approx 1{,}96$.

- **Sai số:** $\epsilon = z_{1-\alpha/2} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

  $= 1{,}96 \cdot \sqrt{\dfrac{0{,}02\,(1-0{,}02)}{200}}$

  $\approx 0{,}0194.$
- **KTC** 95% **cho tỷ lệ** p **là**

$$\begin{aligned} \hat{p}-\epsilon \leq &\ p \ \leq \hat{p}+\epsilon \\ \Leftrightarrow 0{,}02-0{,}0194 \leq &\ p \ \leq 0{,}02+0{,}0194 \\ \Leftrightarrow 0{,}0006 \leq &\ p \ \leq 0{,}0394 \end{aligned}$$

$\Rightarrow p \in [0{,}0006; 0{,}0394].$

b) Khách hàng yêu cầu tỷ lệ lỗi của sản phẩm phải dưới 0,05. Hỏi nhà sản xuất có thể chứng minh khả năng đáp ứng ở mức chất lượng này cho khách hàng không? Sử dụng mức ý nghĩa 1%.

- **GTKĐ:** $\begin{cases} H_0 : p \geq 0{,}05 \\ H_1 : p < 0{,}05 \end{cases}$: KĐ 1 phía, $p_0 = 0{,}05$.
- **Mức ý nghĩa:** $\alpha = 1\% = 0{,}01$.
- **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\hat{p}-p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{0{,}02-0{,}05}{\sqrt{\dfrac{0{,}05(1-0{,}05)}{200}}} \approx -1{,}9467.$$

- **Miền bác bỏ:** Bác bỏ H_0 nếu $z_0 < -z_{1-\alpha}$.

  Ta có $\alpha = 0{,}01$

  $\Rightarrow -z_{1-\alpha} = -z_{0{,}99} \approx -2{,}33.$
- **So sánh và kết luận:**

  Ta có: $z_0 < -z_{1-\alpha}$

  $\Leftrightarrow -1{,}9467 < -2{,}33$ (sai)

  $\Rightarrow$ Chưa đủ cơ sở để bác bỏ $H_0 : p \geq 0{,}05$.

  Kết luận: Với mức ý nghĩa 1% thì tỷ lệ lỗi của sản phẩm không dưới 0,05 hay nhà sản xuất không thể chứng minh khả năng đáp ứng ở mức chất lượng này cho khách hàng.

□

Câu 4 (2 điểm). Một nhà sản xuất công bố rằng độ bền kéo trung bình của sợi A cao hơn độ bền kéo trung bình của sợi B. Để kiểm định công bố này, 50 mẫu của mỗi loại sợi được kiểm tra dưới các điều kiện tương tự. Loại sợi A có độ bền kéo trung bình là 86,7 kg với độ lệch chuẩn mẫu 6,28 kg, trong khi loại sợi B có độ bền kéo trung bình là 85,4 kg với độ lệch chuẩn mẫu là 5,61 kg. Hãy kiểm định công bố của nhà sản xuất bằng cách sử dụng mức ý nghĩa 1%.

✍ **Lời giải.**

Gọi X_1 (kg) là độ bền kéo của loại sợi A,

X_2 (kg) là độ bền kéo của loại sợi B.

Giả sử: $X_1 \sim N(\mu_1; \sigma_1^2)$, $X_2 \sim N(\mu_2; \sigma_2^2)$

σ_1^2; σ_2^2: **chưa có thông tin**.

Mẫu 1: $n_1 = 50$; $\bar{x}_1 = 86{,}7$; $s_1 \approx 6{,}28$.

Mẫu 2: $n_2 = 50$; $\bar{x}_2 = 85{,}4$; $s_2 \approx 5{,}61$.

⁎ **Kiểm định phương sai**

- Giả thuyết kiểm định $\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2. \end{cases}$

- Mức ý nghĩa $\alpha = 0{,}01$.
- Giá trị thống kê kiểm định

$$f_0 = \frac{s_1^2}{s_2^2} = \frac{6{,}28^2}{5{,}61^2} \approx 1{,}2531.$$

- Miền bác bỏ: bác bỏ H_0 khi $f_0 > f_{\alpha/2,n_1-1,n_2-1}$ hoặc $f_0 < \dfrac{1}{f_{\alpha/2,n_2-1,n_1-1}}$
  Ta có $\alpha = 0{,}01 \Rightarrow f_{\alpha/2,n_1-1,n_2-1} = f_{\alpha/2,n_2-1,n_1-1} = f_{0{,}005,49,49} = 2{,}11$.

- So sánh:
  Ta có $f_0 > f_{\alpha/2,n_1-1,n_2-1}$ $\Leftrightarrow 1{,}2531 > 2{,}11$ (sai) hoặc $f_0 < \dfrac{1}{f_{\alpha/2,n_2-1,n_1-1}}$ $\Leftrightarrow 1{,}2531 < \dfrac{1}{2{,}11}$ $\Leftrightarrow 1{,}2531 < 0{,}4739$ (sai)
  $\Rightarrow$ chưa đủ cơ sở để bác bỏ $H_0 : \sigma_1^2 = \sigma_2^2$
  $\Rightarrow \sigma_1^2 = \sigma_2^2$.

⁎ Kiểm định hai trung bình khi $\sigma_1^2 = \sigma_2^2$

1) Giả thuyết kiểm định: $\begin{cases} H_0 : \mu_1 \le \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases}$ kiểm định 1 phía, $\Delta_0 = \mu_1 - \mu_2 = 0$.

2) Mức ý nghĩa: $\alpha = 0{,}01$.

3) Giá trị thống kê kiểm định

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(50-1)\cdot 6{,}28^2 + (50-1)\cdot 5{,}61^2}{50+50-2} = 35{,}45525$$

$$\Rightarrow t_0 = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}} = \frac{86{,}7 - 85{,}4 - 0}{\sqrt{\dfrac{35{,}45525}{50} + \dfrac{35{,}45525}{50}}} \approx 1{,}0916.$$

4) Miền bác bỏ: bác bỏ H_0 khi $t_0 > t_{\alpha,n_1+n_2-2}$.
   Ta có $\alpha = 0{,}01 \Rightarrow t_{\alpha,n_1+n_2-2} = t_{0{,}01,98} \approx z_{0{,}99} \approx 2{,}33$.

5) So sánh
   Ta có $t_0 > t_{\alpha,n_1+n_2-2}$
   $\Leftrightarrow 1{,}0916 > 2{,}33$ (sai)
   $\Rightarrow$ chưa đủ cở sở để bác bỏ $H_0 : \mu_1 \le \mu_2$.
   Kết luận: Với mức ý nghĩa 1%, độ bền kéo trung bình của sợi A không cao hơn so với độ bền kéo trung bình của sợi B hay công bố của nhà sản xuất không đúng.

□

Câu 5 (2 điểm). 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 của 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 y theo x. Giải thích ý nghĩa của $\hat{\beta}_1$ nhận được.

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

✍ Lời giải.
Giả sử X, Y thỏa mãn mô hình hồi quy tuyến tính đơn biến.

a) Tìm phương trình hồi quy tuyến tính y theo x.
Từ bảng dữ liệu, dùng máy cầm tay, ta ước lượng được: $\hat{\beta}_0 \approx 27{,}8571$; $\hat{\beta}_1 \approx 0{,}2079$.
Phương trình hồi quy tuyến tính đơn y theo x là $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 27{,}8571 + 0{,}2079x$.
⋇ Giải thích ý nghĩa của $\hat{\beta}_1$ nhận được:
Vì $\hat{\beta}_1 = 0{,}2079 > 0$ nên khi vật có khối lượng tăng 1 g thì chiều dài của sợi dây tăng 0,2079 mm.

b) Bạn dự đoán chiều dài sợi dây là bao nhiêu nếu một vật có trọng lượng 375 g được treo lên?
Vật có trọng lượng 375 g $\Rightarrow x_0 = 375$

$$\Rightarrow \hat{y}_0 = 27{,}8571 + 0{,}2079 \cdot 375 = 105{,}8196.$$

Vậy nếu một vật có trọng lượng 375 g được treo lên thì chiều dài sợi dây là 105,8196 mm.

□

XÁC SUẤT THỐNG KÊ	ĐỀ THI KẾT THÚC HỌC PHẦN HỌC KỲ I NĂM HỌC 2022-2023 *Thời gian làm bài: 90 phút* **Đề số: 2**

Câu 1 (2 điểm). Một máy sản xuất ra bu lông với đường kính tuân theo phân phối chuẩn $N(4; 0.09)$ (đv: mm). Các bu lông được đo đạc một cách chính xác và những cái nào có đường kính nhỏ hơn 3.5 mm hoặc lớn hơn 4.4 mm 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 500 bu lông được sản xuất ra. Gọi X là số bu lông được chấp nhận của lô này. Tính $\mathbb{E}(X)$ và $Var(X)$.

✍ Lời giải.
Gọi Y (mm) là đường kính của bu lông.
Theo đề bài, ta có $Y \sim N(4; 0.09) \Rightarrow \mu_Y = 4$; $\sigma_Y^2 = 0.09 \Rightarrow \sigma_Y = 0.3$.

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

$$\mathbb{P}(3.5 \le Y \le 4.4) = \mathbb{P}\left(\frac{3.5-4}{0.3} \le Z \le \frac{4.4-4}{0.3}\right) = \mathbb{P}\left(-\frac{5}{3} \le Z \le \frac{4}{3}\right) = \int_{-\frac{5}{3}}^{\frac{4}{3}} f(z)\,\mathrm{d}z \approx 0.8610.$$

b) Một lô gồm 500 bu lông được sản xuất ra. Gọi X là số bu lông được chấp nhận của lô này. Tính $\mathbb{E}(X)$ và $Var(X)$.
Với X là số bu lông được chấp nhận của lô gồm 500 bu lông. Khi đó $X \sim B(500; 0.8610)$.
Do đó

- $\mathbb{E}(X) = np = 500 \cdot 0.8610 = 430.5$.
- $Var(X) = np(1-p) = 500 \cdot 0.8610 \cdot (1 - 0.8610) = 59.8395$.

□

Câu 2 (2 điểm). Nhiệt độ nước trung bình hạ lưu từ ống tháp xả giải nhiệt của nhà máy điện không được lớn hơn 100°F. Kinh nghiệm quá khứ đã chỉ ra rằng độ lệch chuẩn của nhiệt độ là 2°F. Nhiệt độ nước được đo trên chín ngày được lựa chọn ngẫu nhiên, và nhiệt độ trung bình được tìm thấy là 98°F. Biết rằng nhiệt độ nước tuân theo phân phối chuẩn.

a) Tìm khoảng tin cậy 96% cho nhiệt độ nước trung bình.

b) Có bằng chứng gì cho ta thấy nhiệt độ nước có thể chấp nhận được hay không với mức ý nghĩa 2%?

✍ **Lời giải.**

Gọi X (°F) là nhiệt độ hạ lưu từ ống tháp xả giải nhiệt của nhà máy điện.
Theo đề bài, ta có $X \sim N(\mu; \sigma^2)$ với $\sigma = 2$: đã biết.
$n = 9$; $\bar{x} = 98$.

a) Tìm khoảng tin cậy 96% cho nhiệt độ nước trung bình.
• **Độ tin cậy:** 96% $\Rightarrow \alpha = 0.04$.
$\Rightarrow z_{1-\alpha/2} = z_{0.98} \approx 2.05$.
• **Sai số:** $\epsilon = z_{1-\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}$
$= 2.05 \cdot \dfrac{2}{\sqrt{9}}$
$= \dfrac{41}{30}$
≈ 1.3667.
• **KTC 96% cho trung bình μ là:**

$$\begin{aligned} \bar{x} - \epsilon \leq &\ \mu \leq \bar{x} - \epsilon \\ \Leftrightarrow 98 - 1.3667 \leq &\ \mu \leq 98 + 1.3667 \\ \Leftrightarrow 96.6333 \leq &\ \mu \leq 99.3667 \end{aligned}$$

$\Rightarrow \mu \in [96.6333; 99.3667]$.

b) Có bằng chứng gì cho ta thấy nhiệt độ nước có thể chấp nhận được hay không với mức ý nghĩa 2%?
$\sigma = 2$: đã biết.
Ta có: $n = 9$; $\bar{x} = 98$.
• **Giả Thuyết KĐ:** $\begin{cases} H_0 : \mu \leq 100 \\ H_1 : \mu > 100 \end{cases}$: KĐ 1 phía, $\mu_0 = 100$.
• **Mức ý nghĩa:** $\alpha = 0.02$.
• **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{98 - 100}{\frac{2}{\sqrt{9}}} = -3.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $z_0 > z_{1-\alpha}$.
Ta có $\alpha = 0.02$
$\Rightarrow z_{1-\alpha} = z_{0.98} \approx 2.05$.
• **So sánh và kết luận:**
Ta có: $z_0 > z_{1-\alpha}$
$\Leftrightarrow -3 > 2.05$ (sai)

$\Rightarrow$ Chưa đủ cơ sở để bác bỏ $H_0 : \mu \leq 100$.
Kết luận: Với mức ý nghĩa 2% thì nhiệt độ nước không vượt quá 100°F hay nhiệt độ có thể chấp nhận được.

$\square$

Câu 3 (2 điểm). Trong một mẫu ngẫu nhiên gồm 85 vòng bi trục khuỷu động cơ ôtô trong đó có 10 vòng bi độ nhám bề mặt hoàn thiện vượt quá các thông số kỹ thuật (gọi tắt là vượt chuẩn).

a) Tìm khoảng tin cậy 96% cho tỷ lệ vòng bi vượt chuẩn.

b) Dữ liệu này có cho thấy rằng tỷ lệ vòng bi vượt chuẩn là cao hơn 0.10 hay không với mức ý nghĩa 2%.

✍ **Lời giải.**

a) Tìm khoảng tin cậy 96% cho tỷ lệ vòng bi vượt chuẩn. Gọi Y là số vòng bi vượt chuẩn.
Ta có: $n = 85$, $y = 10 \Rightarrow$ tỷ lệ mẫu $\hat{p} = \dfrac{y}{n} = \dfrac{10}{85} = \dfrac{2}{17}$.
• **Độ tin cậy:** $96\% \Rightarrow \alpha = 0.04$.
$\Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.98} \approx 2.05$.
• **Sai số:** $\epsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

$$= 2.05 \cdot \sqrt{\frac{\frac{2}{17}\left(1-\frac{2}{17}\right)}{85}}$$

≈ 0.0716.
• **KTC** 96% **cho tỷ lệ** p **là**

$$\begin{aligned} \hat{p} - \epsilon \leq\ & p \leq \hat{p} + \epsilon \\ \Leftrightarrow \frac{2}{17} - 0.0716 \leq\ & p \leq \frac{2}{17} + 0.0716 \\ \Leftrightarrow 0.0460 \leq\ & p \leq 0.1892 \end{aligned}$$

$\Rightarrow p \in [0.0460; 0.1892]$.

b) Dữ liệu này có cho thấy rằng tỷ lệ vòng bi vượt chuẩn là cao hơn 0.10 hay không với mức ý nghĩa 2%.
• **GTKĐ:** $\begin{cases} H_0 : p \leq 0.10 \\ H_1 : p > 0.10 \end{cases}$: KĐ 1 phía, $p_0 = 0.10$.
• **Mức ý nghĩa:** $\alpha = 2\% = 0.02$.
• **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{\dfrac{2}{17} - 0.10}{\sqrt{\dfrac{0.10(1-0.10)}{85}}} \approx 0.5423.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $z_0 > z_{1-\alpha}$.
Ta có $\alpha = 0.02$
$\Rightarrow z_{1-\alpha} = z_{0.98} \approx 2.05$.
• **So sánh và kết luận:**
Ta có: $z_0 > z_{1-\alpha}$
$\Leftrightarrow 0.5423 > 2.05$ (sai)
$\Rightarrow$ Chưa đủ cơ sở để bác bỏ $H_0 : p \leq 0.10$.
Kết luận: Với mức ý nghĩa 2% thì tỷ lệ vòng bi vượt chuẩn không cao hơn 0.10.

$\square$

Câu 4 (2 điểm). Trong một nghiên cứu để ước tính tỷ lệ cư dân trong một thành phố nào đó và các vùng ngoại ô của nó ủng hộ việc xây dựng nhà máy năng lượng hạt nhân, người ta thấy rằng 63 trong 100 cư dân thành thị ủng hộ việc xây dựng, trong khi chỉ 59 trong 125 cư dân ngoại ô là ủng hộ. Hỏi có sự khác biệt giữa tỷ lệ cư dân thành thị và ngoại ô ủng hộ việc xây dựng nhà máy hạt nhân hay không với mức ý nghĩa 2%? Sử dụng $p-$giá trị.

✍ **Lời giải.**

Mẫu 1: $y_1 = 63$; $n_1 = 100 \Rightarrow \hat{p}_1 = \dfrac{63}{100} = 0.63.$

Mẫu 2: $y_2 = 59$; $n_2 = 125 \Rightarrow \hat{p}_2 = \dfrac{59}{125} = 0.427.$

$$\Rightarrow \hat{p} = \frac{y_1 + y_2}{n_1 + n_2} = \frac{63 + 59}{100 + 125} = \frac{122}{225}.$$

- Giả thuyết kiểm định: $\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{cases}; \qquad \Rightarrow \Delta_0 = p_1 - p_2 = 0.$

- Giá trị thống kê kiểm định
$$z_0 = \frac{\hat{p}_1 - \hat{p}_2 - \Delta_0}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} = \frac{0.63 - 0.427 - 0}{\sqrt{\dfrac{122}{225}\left(1 - \dfrac{122}{225}\right)\left(\dfrac{1}{100} + \dfrac{1}{125}\right)}} \approx 3.03699.$$

- $p-$giá trị $= 2\left(1 - \Phi(|3.03699|)\right) = 2\left(1 - 0.99881\right) = 0.00238.$
  Ta có $p-$giá trị $= 0.00238 < 0.02 = \alpha$
  $\Rightarrow$ bác bỏ $H_0 : p_1 = p_2$.
  Kết luận: Với mức ý nghĩa 2% thì có sự khác biệt có ý nghĩa giữa tỷ lệ cư dân thành thị và ngoại ô trong ủng hộ việc xây dựng nhà máy hạt nhân.

$\square$

Câu 5 (2 điểm). Một thí nghiệm đã được thực hiện để nghiên cứu sự thay đổi độ hòa tan của hóa chất Y trong nước. Khối lượng hòa tan (tính bằng kg) trong 1 lít nước ở các nhiệt độ nước khác nhau được cho trong bảng dưới đây

Nhiệt độ °C (x)	15	20	25	30	35	50	70
Khối lượng của Y (y)	2.1	2.6	2.9	3.3	4.0	5.1	7.0

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

b) Bạn dự đoán khối lượng hòa tan là bao nhiêu nếu nhiệt độ nước là 42°C?

✍ **Lời giải.**

Giả sử X, Y thỏa mãn mô hình hồi quy tuyến tính đơn biến.

a) Tìm phương trình hồi quy tuyến tính y theo x.
Từ bảng dữ liệu, dùng máy cầm tay, ta ước lượng được: $\hat{\beta}_0 \approx 0.7549$; $\hat{\beta}_1 \approx 0.0886$.
Phương trình hồi quy tuyến tính đơn y theo x là $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 0.7549 + 0.0886x$.
⁕ Giải thích ý nghĩa của $\hat{\beta}_1$ nhận được:
Vì $\hat{\beta}_1 = 0.0886 > 0$ nên khi tăng nhiệt độ lên 1°C thì khối lượng hòa tan tăng 0.0886 kg.

b) Bạn dự đoán khối lượng hòa tan là bao nhiêu nếu nhiệt độ nước là 42°C.
Khi nhiệt độ nước là 42°C $\Rightarrow x_0 = 42$
$$\Rightarrow \hat{y}_0 = 0.7549 + 0.0886 \cdot 42 = 4.4761.$$
Vậy nếu nhiệt độ nước là 42°C thì khối lượng hòa tan là 4.4761 kg.

$\square$

XÁC SUẤT THỐNG KÊ

ĐỀ CUỐI KỲ II NĂM HỌC 2021-2022

Thời gian làm bài: 90 phút

Đề số: 1

Câu 1 (4 điểm). Biết trọng lượng X (g/quả) của mỗi quả trứng có phân phối chuẩn. Đem cân 100 quả trứng ta 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** 170 g là trứng loại một.

a) Tìm khoảng tin cậy 97% cho trọng lượng trứng trung bình.

b) Tìm khoảng tin cậy 98% cho tỷ lệ trứng loại một. Nếu ta muốn sai số ước lượng không quá 0.1 g thì cần khảo sát thêm bao nhiêu trứng?

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

d) Có ý kiến cho rằng 50% số trứng thuộc loại một. Hãy kiểm định ý kiến trên với mức ý nghĩa 1%.

✍ Lời giải.

Gọi X (cm/s) là tốc độ cháy của nhiên liệu rắn.
Theo đề bài, ta có $X \sim N(\mu; \sigma^2)$
σ^2: chưa biết.
Từ bảng dữ liệu, ta tính được: $n = 100$; $\bar{x} = 170.85$; $s \approx 7.7543$.

a) Tìm khoảng tin cậy 97% cho trọng lượng trứng trung bình

• **Độ tin cậy:** $97\% \Rightarrow \alpha = 0.03$.
$\Rightarrow t_{\alpha/2;n-1} = t_{0.015;99} \approx z_{0.985} \approx 2.17$.

• **Sai số:** $\epsilon = t_{\alpha/2;n-1} \cdot \dfrac{s}{\sqrt{n}}$
$= 2.17 \cdot \dfrac{7.7543}{\sqrt{100}}$
≈ 1.6827.

• **KTC** 97% **cho trung bình** μ **là:**

$$\begin{aligned} \bar{x} - \epsilon \leq & \quad \mu \quad \leq \bar{x} - \epsilon \\ \Leftrightarrow 170.85 - 1.6827 \leq & \quad \mu \quad \leq 170.85 + 1.6827 \\ \Leftrightarrow 169.1673 \leq & \quad \mu \quad \leq 172.5327 \end{aligned}$$

$\Rightarrow \mu \in [169.1673; 172.5327]$.

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

Gọi Y là số trứng loại một.
Ta có: $n = 100$, $y = 24 + 14 + 6 = 44 \Rightarrow$ tỷ lệ mẫu $\hat{p} = \dfrac{y}{n} = \dfrac{44}{100} = 0.44$.

• **Độ tin cậy:** $98\% \Rightarrow \alpha = 0.02$.
$\Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.99} \approx 2.33$.

- **Sai số:** $\epsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$
$= 2.33 \cdot \sqrt{\dfrac{0.44\,(1-0.44)}{100}}$
$\approx 0.1157.$
- **KTC** 98% **cho tỷ lệ** p **là**

$$\begin{aligned} \hat{p}-\epsilon \leq & \;\; p \;\; \leq \hat{p}+\epsilon \\ \Leftrightarrow 0.44-0.1157 \leq & \;\; p \;\; \leq 0.44+0.1157 \\ \Leftrightarrow 0.3243 \leq & \;\; p \;\; \leq 0.5557 \end{aligned}$$

$\Rightarrow p \in [0.3243; 0.5557]$.
⋇ Nếu ta muốn sai số ước lượng không quá 0.1 g thì cần khảo sát thêm bao nhiêu trứng?
Để $\epsilon \leq 0.1$ thì

$$n \geq \left(\frac{z_{1-\alpha/2}}{\epsilon}\right)^2 \cdot \hat{p}(1-\hat{p}) = \left(\frac{2.33}{0.1}\right)^2 \cdot 0.44(1-0.44) \approx 133.7681.$$

$\Rightarrow n \geq 134.$
Do đó, để $\epsilon \leq 0.1$ thì phải khảo sát ít nhất 134 quả trứng.
Vậy để $\epsilon \leq 0.1$ thì phải khảo sát **thêm** $134 - 100 = 34$ quả trứng nữa.

c) Có ý kiến cho rằng trọng lượng trứng trung bình lớn hơn 170 g/quả. Hãy kiểm định ý kiến trên ứng với mức ý nghĩa 1%.
σ^2: chưa biết.
Ta có: $n = 100$; $\bar{x} = 170.85$; $s \approx 7.7543$.
- **Giả Thuyết KĐ:** $\begin{cases} H_0 : \mu \leq 170 \\ H_1 : \mu > 170 \end{cases}$: KĐ 1 phía, $\mu_0 = 170$.
- **Mức ý nghĩa:** $\alpha = 0.01$.
- **Giá trị Thống kê kiểm định:**

$$t_0 = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}} = \frac{170.85-170}{\frac{7.7543}{\sqrt{100}}} \approx 1.0962.$$

- **Miền bác bỏ:** Bác bỏ H_0 nếu $t_0 > t_{\alpha;n-1}$.
Ta có $\alpha = 0.01$
$\Rightarrow t_{\alpha;n-1} = t_{0.01;99} \approx z_{0.99} \approx 2.33.$
- **So sánh và kết luận:**
Ta có: $t_0 > t_{\alpha;n-1}$
$\Leftrightarrow 1.0962 > 2.33$ (sai)
$\Rightarrow$ Chưa đủ cơ sở để bác bỏ $H_0 : \mu \leq 170$.
Kết luận: Với mức ý nghĩa 1%, trọng lượng trứng trung bình không lớn hơn 170 g.

d) Có ý kiến cho rằng 50% số trứng thuộc loại một. Hãy kiểm định ý kiến trên với mức ý nghĩa 1%.
Ta có: $n = 100$, $y = 24 + 14 + 6 = 44 \Rightarrow$ tỷ lệ mẫu $\hat{p} = \dfrac{y}{n} = \dfrac{44}{100} = 0.44$.
- **GTKĐ:** $\begin{cases} H_0 : p = 0.5 \\ H_1 : p \neq 0.5 \end{cases}$: KĐ 2 phía, $p_0 = 50\% = 0.5$.
- **Mức ý nghĩa:** $\alpha = 1\% = 0.01$.
- **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{0.44 - 0.5}{\sqrt{\dfrac{0.5(1-0.5)}{100}}} = -1.2.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $|z_0| > z_{1-\alpha/2}$.
Ta có $\alpha = 0.01$
$\Rightarrow z_{1-\alpha/2} = z_{0.995} \approx 2.58.$
• **So sánh và kết luận:**
Ta có: $|z_0| > z_{1-\alpha/2}$

$|-1.2| > 2.58$

$\Leftrightarrow 1.2 > 2.58$ (sai)

$\Rightarrow$ Chưa đủ cơ sở để bác bỏ $H_0 : p = 0.5$.
Kết luận: Với mức ý nghĩa 1%, ý kiến cho rằng 50% số trứng thuộc loại một không được chấp nhận.

□

Câu 2 (2 điểm). Hai chất xúc tác có thể được sử dụng trong một phản ứng hóa 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 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*)

✍ **Lời giải.**

Gọi X_1 (%) là hiệu suất phản ứng khi sử dụng chất xúc tác 1

X_2 (%) là hiệu suất phản ứng khi sử dụng chất xúc tác 2.

Theo đề bài $X_1 \sim N(\mu_1; \sigma_1^2)$; $X_2 \sim N(\mu_2; \sigma_2^2)$
$\sigma_1 = \sigma_2$ chưa biết.
Mẫu 1: $n_1 = 12$; $\bar{x}_1 = 86$ và $s_1 = 3$.
Mẫu 2: $n_2 = 15$; $\bar{x}_2 = 89$ và $s_2 = 2$.

• **Giả thuyết kiểm định** $\begin{cases} H_0 : \mu_1 \geq \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases}$; $\Rightarrow \mu_1 - \mu_2 = 0$; $\alpha = 0.01$.

• **Giá trị thống kê kiểm định**

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(12-1).3^2 + (15-1).2^2}{12+15-2} = 6.2$$

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}} = \frac{86 - 89 - 0}{\sqrt{\dfrac{6.2}{12} + \dfrac{6.2}{15}}} = -3.1109.$$

• **Miền bác bỏ:** Bác bỏ H_0 khi $t_0 < -t_{\alpha; n_1+n_2-2}$.
Ta có $\alpha = 0.01$
$\Rightarrow -t_{\alpha;n_1+n_2-2} = -t_{0.01;25} = -2.4851.$
• **So sánh và kết luận:**
Ta có: $t_0 < -t_{\alpha;n_1+n_2-2}$

$-3.1109 < -2.4851$ (đúng)

$\Rightarrow$ Bác bỏ $H_0 : \mu_1 \geq \mu_2$.
Kết luận: Với mức ý nghĩa 1% thì có bằng chứng để khẳng định được rằng sử dụng chất xúc tác 2 tạo ra hiệu suất trung bình cao hơn khi dùng chất xúc tác 1.
⁎ Cách 2: Dùng p - giá trị:

$t_0 = -3.1109 \Rightarrow p$-giá trị $= \mathbb{P}(T_{12+15-2} \le t_0) = \mathbb{P}(T_{12+15-2} \le -3.1109) \approx 0.0023$.
Vì $p-$giá trị $\approx 0.0023 < 0.01 = \alpha$ nên bác bỏ $H_0 : \mu_1 \ge \mu_2$.
□

Câu 3 (2 điểm). Hai loại giải pháp khác nhau để đánh bóng thấu kính nội nhãn (được dùng trong mắt người sau phẫu thuật đục thủy tinh thể) đang được đánh giá để sử dụng. Trong 300 thấu kính đã được đánh bóng bằng giải pháp 1 thì có 253 thấu kính không có khuyết tật do đánh bóng. Trong 300 thấu kính khác được đánh bóng bằng giải pháp 2 thì có 196 thấu kính không có khuyết tật do đánh bóng. Có lý do nào để tin rằng hai giải pháp đánh bóng là khác nhau không? Sử dụng $\alpha = 0.05$. $p-$giá trị cho kiểm định này là bao nhiêu?

✍ **Lời giải.**

Gọi Y_1 là số thấu kính được đánh bóng bằng cách sử dụng giải pháp đánh bóng thứ nhất

Y_2 là số thấu kính được đánh bóng bằng cách sử dụng giải pháp đánh bóng thứ hai.

Mẫu 1: $n_1 = 300$; $y_1 = 253 \Rightarrow \hat{p}_1 = \dfrac{y_1}{n_1} = \dfrac{253}{300}$.

Mẫu 2: $n_2 = 300$; $y_2 = 196 \Rightarrow \hat{p}_2 = \dfrac{y_2}{n_2} = \dfrac{196}{300} = \dfrac{49}{75}$.

$$\Rightarrow \hat{p} = \frac{y_1 + y_2}{n_1 + n_2} = \frac{253 + 196}{300 + 300} = \frac{449}{600}$$

• **Giả thuyết kiểm định:** $\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \ne p_2 \end{cases}$; $\quad \Rightarrow \Delta_0 = p_1 - p_2 = 0$; $\alpha = 0.05$.

• **Giá trị thống kê kiểm định**

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2 - \Delta_0}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} = \frac{\dfrac{253}{300} - \dfrac{49}{75} - 0}{\sqrt{\dfrac{449}{600}\left(1 - \dfrac{449}{600}\right)\left(\dfrac{1}{300} + \dfrac{1}{300}\right)}} = 5.3621$$

• **Miền bác bỏ:** Bác bỏ H_0 khi $|z_0| > z_{1-\alpha/2}$.

Ta có $\alpha = 0.05$

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

• **So sánh và kết luận:**

Ta có: $\quad |z_0| > z_{1-\alpha/2}$

$|5.3621| > z_{1-\alpha/2}$

$5.3621 > 1.96 \quad$ (đúng)

$\Rightarrow$ bác bỏ H_0.

Kết luận: Với mức ý nghĩa 5% thì hai giải pháp đánh bóng có tỷ lệ kính khuyết tật khác nhau.

⋇ p - giá trị:

$z_0 = 5.3621 \Rightarrow p$-giá trị $= 2(1 - \Phi(|5.3621|)) \approx 2(1-1) = 0$.
□

Câu 4 (2 điểm). Dữ liệu bên dưới mô tả về trọng lượng của các điếu thuốc lá x (g) được sản xuất từ các nhà máy khác nhau và hàm lượng nicotin y (mg) trong mỗi điếu thuốc:

x	15.8	14.9	9.0	4.5	15.0	17.0	8.6	12.0	4.1	16.0
y	0.957	0.886	0.852	0.911	0.889	0.919	0.969	1.118	0.946	1.094

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

b) Dự đoán hàm lượng nicotin của một điếu thuốc có trọng lượng 11 g.

✍ **Lời giải.**
Giả sử X, Y thỏa mãn mô hình hồi quy tuyến tính đơn biến.

a) Tìm phương trình hồi quy tuyến tính đơn y theo x:
Từ bảng dữ liệu, dùng máy cầm tay, ta ước lượng được: $\hat{\beta}_0 \approx 0.9184$; $\hat{\beta}_1 \approx 3.0563 \cdot 10^{-3} \approx 0.0031$.
Phương trình hồi quy tuyến tính đơn y theo x là $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 0.9184 + 0.0031x$.
⁂ Giải thích ý nghĩa của $\hat{\beta}_1$ nhận được:
Vì $\hat{\beta}_1 = 0.0031 > 0$ nên khi trọng lượng của điếu thuốc lá tăng 1 g và các yếu tố khác không đổi thì hàm lượng nicotin trong điếu thuốc tăng 0.0031 mg.

b) Dự đoán hàm lượng nicotin của một điếu thuốc có trọng lượng 11 g.
Điếu thuốc có trọng lượng 11 g $\Rightarrow x_0 = 11$

$$\Rightarrow \hat{y}_0 = 0.9184 + 0.0031 \cdot 11 = 0.9525.$$

Vậy nếu một điếu thuốc có trọng lượng 11 g thì hàm lượng nicotin là 0.9525 mg.

□

XÁC SUẤT THỐNG KÊ

ĐỀ CUỐI KỲ II NĂM HỌC 2021-2022 (CNTT Hệ CLC)

Thời gian làm bài: 90 phút

Đề số: 2

Câu 1 (2 điểm). Trong một nhà máy sản xuất linh kiện điện tử, biết rằng tuổi thọ của một loại chip Led (sử dụng trong các bóng đèn Led) có phân phối chuẩn với tuổi thọ trung bình là 70000 giờ và độ lệch chuẩn 1500 giờ.

a) Những chip Led có tuổi thọ trên 71500 giờ được phân loại là sản phẩm loại I. Tính tỷ lệ sản phẩm loại I.

b) Chọn ngẫu nhiên 15 chip Led trong một dây chuyền có rất nhiều sản phẩm, tính xác suất chọn được ít nhất 2 chip Led loại I.

✍ **Lời giải.**
Gọi X (giờ) là tuổi thọ của chip Led.
Theo đề bài, ta có $X \sim N(70000; 1500^2) \Rightarrow \mu = 70000; \sigma = 1500$.

a) Những chip Led có tuổi thọ trên 71500 giờ được phân loại là sản phẩm loại I. Tính tỷ lệ sản phẩm loại I:

$$\mathbb{P}(X > 71500) = \mathbb{P}\left(\frac{X - 70000}{1500} > \frac{71500 - 70000}{1500}\right) = \mathbb{P}(Z > 1) \approx 0.1587.$$

b) Chọn ngẫu nhiên 15 chip Led trong một dây chuyền có rất nhiều sản phẩm, tính xác suất chọn được ít nhất 2 chip Led loại I.
Gọi Y là số chip Led loại I trong 15 chip.
Khi đó $Y \sim B(15; 0.1587)$.
Xác suất chọn được ít nhất 2 chip Led loại I trong 15 chip là

$$\mathbb{P}(Y \geq 2) = 1 - \mathbb{P}(Y < 2) = 1 - \mathbb{P}(Y \leq 1) = 1 - \sum_{x=0}^{1} \mathrm{C}_{15}^{x} \cdot 0.1587^x \cdot (1 - 0.1587)^{15-x} \approx 0.7133.$$

□

Câu 2 (1.5 điểm). Một cái cân điện tử luôn hiển thị kết quả bằng cân nặng đúng của vật thể cộng với sai số ngẫu nhiên. Giả sử rằng sai số ngẫu nhiên này có phân phối chuẩn với kỳ vọng bằng 0 và độ lệch chuẩn là 0.23 (kg). Kết quả cân nặng của một vật thể trong 10 lần đo từ cân điện tử này như sau

$$5.09 \quad 5.33 \quad 4.61 \quad 5.04 \quad 5.36 \quad 5.22 \quad 5.22 \quad 5.34 \quad 5.23 \quad 4.72$$

a) Tìm khoảng tin cậy 95% cho cân nặng trung bình của vật thể này.

b) Người ta cần cân vật thể này ít nhất bao nhiêu lần để sai số ước lượng của khoảng tin cậy 95% cho cân nặng trung bình của vật thể không quá 0.09 (kg).

✍ **Lời giải.**

Gọi X (kg) là cân nặng của vật thể.
Theo đề bài: $X \sim N(\mu; 0.23^2)$
$\sigma = 0.23$: đã biết.
Từ số liệu đề bài, ta có: $n = 10$; $\bar{x} = 5.116$.

a) Tìm khoảng tin cậy 95% cho cân nặng trung bình của vật thể này.
• **Độ tin cậy:** 95% $\Rightarrow \alpha = 0.05$.
$\Rightarrow z_{1-\alpha/2} = z_{0.975} \approx 1.96$.
• **Sai số:** $\epsilon = z_{1-\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}$
$= 1.96 \cdot \dfrac{0.23}{\sqrt{10}}$
≈ 0.1426.
• **KTC** 95% **cho trung bình** μ **là:**

$$\begin{aligned} \bar{x} - \epsilon \leq & \quad \mu \quad \leq \bar{x} - \epsilon \\ \Leftrightarrow 5.116 - 0.1426 \leq & \quad \mu \quad \leq 5.116 + 0.1426 \\ \Leftrightarrow 4.9734 \leq & \quad \mu \quad \leq 5.2586 \end{aligned}$$

$\Rightarrow \mu \in [4.9734; 5.2586]$.

b) Người ta cần cân vật thể này ít nhất bao nhiêu lần để sai số ước lượng của khoảng tin cậy 95% cho cân nặng trung bình của vật thể không quá 0.09 (kg).
• **Độ tin cậy:** 95% $\Rightarrow \alpha = 0.05$.
$\Rightarrow z_{1-\alpha/2} = z_{0.975} \approx 1.96$.
Để $\epsilon \leq 0.09$ thì

$$n \geq \left(\frac{z_{1-\alpha/2} \cdot \sigma}{\epsilon_0}\right)^2 = \left(\frac{1.96 \cdot 0.23}{0.09}\right)^2 \approx 25.0890$$

$\Rightarrow n \geq 26$.
Vậy người ta cần cân vật thể này ít nhất 26 lần để sai số ước lượng của khoảng tin cậy 95% cho cân nặng trung bình của vật thể không quá 0.09 (kg).

□

Câu 3 (2.5 điểm). Các tác giả của bài báo "Adjuvant Radiotherapy an Chemotherapy in Node-Positive Premenopausal Women with Breast Cancer" (New Engl. J. of Med., 1997: 956-962) báo cáo các kết quả nghiên cứu của họ về một thí nghiệm được thiết kế để so sánh hiệu quả của 2 phương pháp điều trị bệnh ung thư: chỉ điều trị bằng hóa trị (phương pháp I) và điều trị kết hợp giữa hóa trị và xạ trị (phương pháp II). Trong số 181 bệnh nhân được điều trị bằng phương pháp I, có 89 bệnh nhân sống sót ít nhất 15 năm, trong khi 107 trong số 186 bệnh nhân mà được điều trị bằng phương pháp II sống sót ít nhất 15 năm.

a) Tìm khoảng tin cậy 95% cho tỷ lệ bệnh nhân sống sót ít nhất 15 năm khi điều trị bằng phương pháp I.

b) Có ý kiến cho rằng phương pháp điều trị kết hợp giữa hóa trị và xạ trị có hiệu quả hơn là chỉ điều trị bằng hóa trị. Dựa trên kết quả thí nghiệm, ta có đủ bằng chứng ủng hộ ý kiến trên không với mức ý nghĩa $\alpha = 1\%$?

✍ Lời giải.

a) Tìm khoảng tin cậy 95% cho tỷ lệ bệnh nhân sống sót ít nhất 15 năm khi điều trị bằng phương pháp I.

Gọi Y_1 là số bệnh nhân sống sót ít nhất 15 năm khi điều trị bằng phương pháp I.

Ta có $y_1 = 89$; $n_1 = 181 \Rightarrow$ tỷ lệ mẫu $\hat{p}_1 = \dfrac{y_1}{n_1} = \dfrac{89}{181}$.

- **Độ tin cậy:** $95\% \Rightarrow \alpha = 0.05$.
$\Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.975} \approx 1.96$.
- **Sai số:** $\epsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1}}$

$$= 1.96 \cdot \sqrt{\frac{\frac{89}{181}\left(1 - \frac{89}{181}\right)}{181}}$$
$$\approx 0.0728.$$

- **KTC 95% cho tỷ lệ p là**

$$\begin{aligned} \hat{p}_1 - \epsilon \leq \ & p_1 \ \leq \hat{p}_1 + \epsilon \\ \Leftrightarrow \frac{89}{181} - 0.0728 \leq \ & p_1 \ \leq \frac{89}{181} + 0.0728 \\ \Leftrightarrow 0.4189 \leq \ & p_1 \ \leq 0.5645 \end{aligned}$$

$\Rightarrow p_1 \in [0.4189; 0.5645]$.

b) Có ý kiến cho rằng phương pháp điều trị kết hợp giữa hóa trị và xạ trị có hiệu quả hơn là chỉ điều trị bằng hóa trị. Dựa trên kết quả thí nghiệm, ta có đủ bằng chứng ủng hộ ý kiến trên không với mức ý nghĩa $\alpha = 1\%$?

Gọi Y_1 là số bệnh nhân sống sót ít nhất 15 năm khi điều trị bằng phương pháp I;

Y_2 là số bệnh nhân sống sót ít nhất 15 năm khi điều trị bằng phương pháp II.

Mẫu 1: $n_1 = 181$; $y_1 = 89 \Rightarrow \hat{p}_1 = \dfrac{y_1}{n_1} = \dfrac{89}{181}$.

Mẫu 2: $n_2 = 186$; $y_2 = 107 \Rightarrow \hat{p}_2 = \dfrac{y_2}{n_2} = \dfrac{107}{186}$.

$$\Rightarrow \hat{p} = \frac{y_1 + y_2}{n_1 + n_2} = \frac{89 + 107}{181 + 186} = \frac{196}{367}$$

- **Giả thuyết kiểm định:** $\begin{cases} H_0 : p_1 \geq p_2 \\ H_1 : p_1 < p_2 \end{cases}$; $\quad \Rightarrow \Delta_0 = p_1 - p_2 = 0$; $\alpha = 0.01$.
- **Giá trị thống kê kiểm định**

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2 - \Delta_0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\frac{89}{181} - \frac{107}{186} - 0}{\sqrt{\frac{196}{367}\left(1 - \frac{196}{367}\right)\left(\frac{1}{181} + \frac{1}{186}\right)}} = -1.6043.$$

• **Miền bác bỏ:** Bác bỏ H_0 khi $z_0 < -z_{1-\alpha}$.
Ta có $\alpha = 0.01$
$\Rightarrow -z_{1-\alpha} = -z_{0.99} = -2.33$.
• **So sánh và kết luận:**
Ta có: $\quad z_0 < -z_{1-\alpha}$

$-1.6043 < -2.33 \quad$ (sai)

$\Rightarrow$ chưa đủ cơ sở để bác bỏ $H_0 : p_1 \geq p_2$.
Kết luận: Với mức ý nghĩa 1% thì ta chưa đủ bằng chứng ủng hộ ý kiến cho rằng phương pháp điều trị kết hợp giữa hóa trị và xạ trị có hiệu quả hơn là chỉ điều trị bằng hóa trị.

□

Câu 4 (4 điểm). Một công ty sản xuất xe ô tô chế tạo ra một dòng xe ô tô **Z**, công ty kiểm tra xe bằng cách cho 30 xe chạy trên cùng 1 quãng đường 100 km và đo lượng xăng sử dụng X (Đv: lít), kết quả cho ở bảng sau

X (lít)	7.0	7.1	7.2	7.3	7.4	7.5
Số xe	5	7	6	6	3	3

Giả sử lượng xăng tiêu hao là đại lượng ngẫu nhiên có phân phối chuẩn.

a) Lời khẳng định: "Lượng xăng tiêu hao trung bình của dòng xe này là 7.2 lít (trên 100 km)" có được chấp nhận hay không? (Mức ý nghĩa $\alpha = 5\%$)

b) Những xe có lượng xăng tiêu hao từ 7.3 lít trở lên được xếp vào loại không đạt tiêu chuẩn về nhiên liệu. Giám đốc công ty khẳng định rằng tỷ lệ xe không đạt tiêu chuẩn về nhiên liệu tối đa bằng 10%. Với dữ liệu khảo sát đã cho, ta có đủ bằng chứng để bác bỏ ý kiến trên không? $\alpha = 1\%$.

c) Khảo sát lượng xăng tiêu hao Y (Đv: lít/100 km) trên một mẫu gồm 25 xe của dòng xe **W** cùng hãng, tính được $\bar{y} = 7.35$ (lít/100 km) và $s_Y = 0.2$. Có ý kiến cho rằng dòng xe **Z** tiêu hao ít nhiên liệu hơn dòng xe **W**. Với mức ý nghĩa $\alpha = 1\%$, hãy kiểm định ý kiến trên. Giả sử phương sai của X và Y bằng nhau.

✍ Lời giải.

a) Lời khẳng định: "Lượng xăng tiêu hao trung bình của dòng xe này là 7.2 lít (trên 100 km)" có được chấp nhận hay không? (Mức ý nghĩa $\alpha = 5\%$)
Theo đề bài $X \sim N(\mu; \sigma^2)$
σ^2: chưa biết.
Từ bảng số liệu đề bài, ta tính được: $n = 30$; $\bar{x} \approx 7.2133$; $s \approx 0.1570$.
• **Giả Thuyết KĐ:** $\begin{cases} H_0 : \mu = 7.2 \\ H_1 : \mu \neq 7.2 \end{cases}$: $\quad$ KĐ 2 phía, $\mu_0 = 7.2$.
• **Mức ý nghĩa:** $\alpha = 0.05$.
• **Giá trị Thống kê kiểm định:**

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{7.2133 - 7.2}{\frac{0.1570}{\sqrt{30}}} \approx 0.4640.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $|t_0| > t_{\alpha/2;n-1}$.
Ta có $\alpha = 0.05$
$\Rightarrow t_{\alpha/2;n-1} = t_{0.025;29} = 2.045$.
• **So sánh và kết luận:**

Ta có: $\quad |t_0| > t_{\alpha/2;n-1}$

$|0.4640| > 2.045$

$\Leftrightarrow 0.4640 > 2.045 \quad$ (sai)

$\Rightarrow$ Chưa đủ cơ sở để bác bỏ $H_0 : \mu = 7.2$.

Kết luận: Với mức ý nghĩa 5% thì lượng xăng tiêu hao trung bình của dòng xe này là 7.2 lít (trên 100 km).

b) Những xe có lượng xăng tiêu hao từ 7.3 lít trở lên được xếp vào loại không đạt tiêu chuẩn về nhiên liệu. Giám đốc công ty khẳng định rằng tỷ lệ xe không đạt tiêu chuẩn về nhiên liệu tối đa bằng 10%. Với dữ liệu khảo sát đã cho, ta có đủ bằng chứng để bác bỏ ý kiến trên không? $\alpha = 1\%$.

Gọi Y_1 là số xe không đạt tiêu chuẩn.

Ta có: $n = 30$, $y_1 = 6 + 3 + 3 = 12 \Rightarrow$ tỷ lệ mẫu $\hat{p} = \dfrac{y_1}{n} = \dfrac{12}{30} = 0.4$.

• **GTKĐ:** $\begin{cases} H_0 : p \leq 0.1 \\ H_1 : p > 0.1 \end{cases}$: $\qquad$ KĐ 1 phía, $p_0 = 10\% = 0.1$.

• **Mức ý nghĩa:** $\alpha = 1\% = 0.01$.

• **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{0.4 - 0.1}{\sqrt{\dfrac{0.1(1-0.1)}{30}}} \approx 5.4772.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $z_0 > z_{1-\alpha}$.

Ta có $\alpha = 0.01$

$\Rightarrow z_{1-\alpha} = z_{0.99} \approx 2.33$.

• **So sánh và kết luận:**

Ta có: $\quad z_0 > z_{1-\alpha}$

$\Leftrightarrow 5.4772 > 2.33 \quad$ (đúng)

$\Rightarrow$ bác bỏ $H_0 : p \leq 0.1$.

Kết luận: Với mức ý nghĩa 1%, ta có đủ bằng chứng để bác bỏ ý kiến trên của vị giám đốc.

c) Khảo sát lượng xăng tiêu hao Y (Đv: lít/100 km) trên một mẫu gồm 25 xe của dòng xe **W** cùng hãng, tính được $\bar{y} = 7.35$ (lít/100 km) và $s_Y = 0.2$. Có ý kiến cho rằng dòng xe **Z** tiêu hao ít nhiên liệu hơn dòng xe **W**. Với mức ý nghĩa $\alpha = 1\%$, hãy kiểm định ý kiến trên. Giả sử phương sai của X và Y bằng nhau.

Theo đề bài $X \sim N(\mu_X; \sigma_X^2)$; $Y \sim N(\mu_Y; \sigma_Y^2)$

$\sigma_X = \sigma_Y$ chưa biết.

Mẫu 1: $n_X = 30$; $\bar{x} \approx 7.2133$ và $s_X \approx 0.1570$.

Mẫu 2: $n_Y = 35$; $\bar{y} = 7.35$ và $s_Y = 0.2$.

• **Giả thuyết kiểm định** $\begin{cases} H_0 : \mu_X \geq \mu_Y \\ H_1 : \mu_X < \mu_Y \end{cases}$; $\qquad \Rightarrow \mu_X - \mu_Y = 0$; $\alpha = 0.01$.

• **Giá trị thống kê kiểm định**

$$s_p^2 = \frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2} = \frac{(30-1)\cdot 0.1570^2 + (25-1)\cdot 0.2^2}{30 + 25 - 2} \approx 0.0316$$

$$t_0 = \frac{\bar{x} - \bar{y} - (\mu_X - \mu_Y)}{\sqrt{\dfrac{s_p^2}{n_X} + \dfrac{s_p^2}{n_Y}}} = \frac{7.2133 - 7.35 - 0}{\sqrt{\dfrac{0.0316}{30} + \dfrac{0.0316}{25}}} = -2.8397.$$

• **Miền bác bỏ:** Bác bỏ H_0 khi $t_0 < -t_{\alpha;n_X+n_Y-2}$.

Ta có $\alpha = 0.01$

$\Rightarrow -t_{\alpha;n_X+n_Y-2} = -t_{0.01;53} \approx -z_{0.99} \approx -2.33.$

• **So sánh và kết luận:**

Ta có: $t_0 < -t_{\alpha;n_X+n_Y-2}$

$-2.8397 < -2.33$ (đúng)

$\Rightarrow$ Bác bỏ $H_0 : \mu_X \geq \mu_Y$.

Kết luận: Với mức ý nghĩa 1% thì dòng xe **Z** tiêu hao ít nhiên liệu hơn dòng xe **W**.

□

XÁC SUẤT THỐNG KÊ

ĐỀ CUỐI KỲ I NĂM HỌC 2021-2022

Thời gian làm bài: 90 phút

Đề số: 1

Câu 1. Trong cấu tạo của một loại dây thừng, người ta quan tâm đến hàm lượng nylon X (Đv: %) có ảnh hưởng như thế nào đến lực căng Y (Đv: psi) (là lực kéo tối đa trước khi sợi dây bị đứt). Số liệu bên dưới cho kết quả đo của 10 sợi dây với hàm lượng nylon khác nhau

Hàm lượng nylon X	7	12	17	22	32	42	47	52	52	62
Lực căng Y	169	249	289	329	349	404	459	519	529	559

(*Làm tròn các đáp án đến 4 chữ số sau dấu chấm thập phân*)

a) Tìm phương trình đường thẳng hồi quy ước lượng $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ ($\hat{\beta}_0$: hệ số chặn hay hệ số tự do; $\hat{\beta}_1$: hệ số góc).

b) Tính hệ số tượng quan r_{XY} giữa X và Y.

c) Nếu một sợi dây có hàm lượng nylon bằng 45% thì giá trị dự báo cho lực căng của sợi dây bằng bao nhiêu?

✍ Lời giải.

Giả sử X, Y thỏa mãn mô hình hồi quy tuyến tính đơn.

a) Tìm phương trình đường thẳng hồi quy ước lượng $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ ($\hat{\beta}_0$: hệ số chặn hay hệ số tự do; $\hat{\beta}_1$: hệ số góc).
Từ bảng dữ liệu, dùng máy cầm tay, ta ước lượng được: $\hat{\beta}_0 \approx 154.4151$; $\hat{\beta}_1 \approx 6.6981$.
Phương trình hồi quy tuyến tính đơn y theo x là $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 154.4151 + 6.6981x$.

b) Tính hệ số tượng quan r_{XY} giữa X và Y.
Ta có $r_{XY} \approx 0.9836$. (bấm máy)

c) Nếu một sợi dây có hàm lượng nylon bằng 45% thì giá trị dự báo cho lực căng của sợi dây bằng bao nhiêu?
Sợi dây có hàm lượng nylon bằng 45% $\Rightarrow x_0 = 45$

$$\Rightarrow \hat{y}_0 = 154.4151 + 6.6981 \cdot 45 = 455.8296.$$

Nếu một sợi dây có hàm lượng nylon bằng 45% thì giá trị dự báo cho lực căng của sợi dây bằng 455.8296 psi.

□

Câu 2. Trong một nhà máy sản xuất vỏ bao bì, các kỹ sư áp dụng một phương pháp sản xuất mới với mục đích làm giảm tỷ lệ bao bì hỏng hoặc kém chất lượng. Để kiểm tra hiệu quả của phương pháp sản xuất mới này trước khi đi vào áp dụng chính thức, các kỹ sư so sánh giữa 2 phân xưởng sản xuất: phân xưởng I vẫn áp dụng phương pháp sản xuất cũ và phân xưởng II được thử nghiệm với phương pháp sản xuất mới. Đối với 123 bao bì được sản xuất ở phân xưởng I, có 14 bao bì kém chất lượng hoặc hỏng. Ở phân xưởng II, các kỹ sư thấy có 10 bao bì hỏng/kém chất lượng trong 194 bao bì được sản xuất. Với mức ý nghĩa $\alpha = 0.01$, ta có đủ bằng chứng để kết luận rằng phương pháp sản xuất mới có hiệu quả hơn phương pháp sản xuất cũ không?

✍ **Lời giải.**

Gọi Y_1 là số bao bì kém chất lượng hoặc bị hỏng ở phân xưởng I;

Y_2 là số bao bì kém chất lượng hoặc bị hỏng ở phân xưởng II.

Mẫu 1: $n_1 = 123$; $y_1 = 14 \Rightarrow \hat{p}_1 = \dfrac{y_1}{n_1} = \dfrac{14}{123}$.

Mẫu 2: $n_2 = 194$; $y_2 = 10 \Rightarrow \hat{p}_2 = \dfrac{y_2}{n_2} = \dfrac{10}{194} = \dfrac{5}{97}$.

$$\Rightarrow \hat{p} = \frac{y_1 + y_2}{n_1 + n_2} = \frac{14 + 10}{123 + 194} = \frac{24}{317}$$

• **Giả thuyết kiểm định:** $\begin{cases} H_0 : p_1 \le p_2 \\ H_1 : p_1 > p_2 \end{cases}$; $\quad \Rightarrow \Delta_0 = p_1 - p_2 = 0$; $\alpha = 0.01$.

• **Giá trị thống kê kiểm định**

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2 - \Delta_0}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} = \frac{\dfrac{14}{123} - \dfrac{5}{97} - 0}{\sqrt{\dfrac{24}{317}\left(1 - \dfrac{24}{317}\right)\left(\dfrac{1}{123} + \dfrac{1}{194}\right)}} = 2.0425.$$

• **Miền bác bỏ:** Bác bỏ H_0 khi $z_0 > z_{1-\alpha}$.

Ta có $\alpha = 0.01$

$\Rightarrow z_{1-\alpha} = z_{0.99} = 2.33$.

• **So sánh và kết luận:**

Ta có: $\quad z_0 > z_{1-\alpha}$

$\quad 2.0425 > 2.33 \quad$ (sai)

$\Rightarrow$ chưa đủ cơ sở để bác bỏ $H_0 : p_1 \le p_2$.

Kết luận: Với mức ý nghĩa 1% thì ta chưa đủ bằng chứng để kết luận rằng phương pháp sản xuất mới có hiệu quả hơn phương pháp sản xuất cũ. □

Câu 3. 8 bóng đèn dây tóc được chọn ngẫu nhiên và ghi nhận tuổi thọ (đơn vị: giờ) như sau: 41.1, 39.9, 39.2, 41.5, 41.2, 40.6, 40.7, 40.6. Giả sử rằng tuổi thọ của các bóng đèn tuân theo phân phối chuẩn.

a) Tìm khoảng tin cậy 92% cho tuổi thọ trung bình của các bóng đèn dây tóc.

b) Giả sử rằng tuổi thọ của các bóng đèn tuân theo phân phối chuẩn với độ lệch tiêu chuẩn là 3 (giờ). Người ta cần khảo sát ít nhất bao nhiêu bóng đèn để sai số ước lượng của khoảng tin cậy 92% cho tuổi thọ trung bình của các bóng đèn là không quá 0.5 (giờ).

✍ **Lời giải.**

Gọi X (giờ) là tuổi thọ của bóng đèn dây tóc.

Theo đề bài: $X \sim N(\mu; \sigma^2)$ σ^2: chưa biết.

Từ số liệu đề bài, ta có: $n = 8$; $\bar{x} = 40.6$; $s \approx 0.7445$.

a) Tìm khoảng tin cậy 92% cho tuổi thọ trung bình của các bóng đèn dây tóc.

• **Độ tin cậy:** $92\% \Rightarrow \alpha = 0.08$.

$\Rightarrow z_{1-\alpha/2} = z_{0.96} \approx 1.75$.

• **Sai số:** $\epsilon = z_{1-\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}$

$= 1.75 \cdot \dfrac{0.7445}{\sqrt{8}}$

≈ 0.4606.

• **KTC** 92% **cho trung bình** μ **là:**

$$\begin{aligned} \bar{x} - \epsilon \leq \ & \mu && \leq \bar{x} - \epsilon \\ \Leftrightarrow 40.6 - 0.4606 \leq \ & \mu && \leq 40.6 + 0.4606 \\ \Leftrightarrow 40.1394 \leq \ & \mu && \leq 41.0606 \end{aligned}$$

$\Rightarrow \mu \in [40.1394; 41.0606]$.

b) Giả sử rằng tuổi thọ của các bóng đèn tuân theo phân phối chuẩn với độ lệch tiêu chuẩn là 3 (giờ). Người ta cần khảo sát ít nhất bao nhiêu bóng đèn để sai số ước lượng của khoảng tin cậy 92% cho tuổi thọ trung bình của các bóng đèn là không quá 0.5 (giờ).

Theo đề bài $\sigma = 3$ đã biết.

Để $\epsilon \leq 0.5$ thì

$$n \geq \left(\frac{z_{1-\alpha/2} \cdot \sigma}{\epsilon_0}\right)^2 = \left(\frac{1.75 \cdot 3}{0.5}\right)^2 = 110.25$$

$\Rightarrow n \geq 111$.

Người ta cần khảo sát ít nhất 111 bóng đèn để sai số ước lượng của khoảng tin cậy 92% cho tuổi thọ trung bình của các bóng đèn là không quá 0.5 (giờ).

□

Câu 4. Một công ty sản xuất trò chơi điện tử thực hiện một nghiên cứu về thời gian mà người chơi hoàn thành các cấp độ trong một trò chơi mà họ vừa phát hành. Họ ghi nhận rằng thời gian để hoàn thành cấp độ I là một biến ngẫu nhiên X_1 có phân phối chuẩn với trung bình 48.1 phút và độ lệch chuẩn 15.1 phút.

a) Tính xác suất để người chơi cần nhiều hơn 50.1 phút để hoàn thành cấp độ I của trò chơi này.

b) Nhà sản xuất cần tính ngưỡng thời gian t_0 sao cho 90.15% người chơi sẽ cần ít hơn t_0 phút để hoàn thành cấp độ I của trò chơi này. Tính t_0.

c) Giả sử trò chơi gồm 3 cấp độ, trong đó thời gian để hoàn thành các cấp độ II và III lần lượt là các biến ngẫu nhiên $X_2 \sim N(55, 14.5^2)$ và $X_3 \sim N(56.8, 10.5^2)$. Giả sử rằng X_1, X_2 và X_3 là các biến ngẫu nhiên độc lập. Tính trung bình và độ lệch chuẩn cho tổng thời gian để người chơi hoàn thành cả 3 cấp độ trò chơi này.

✍ Lời giải.

a) Tính xác suất để người chơi cần nhiều hơn 50.1 phút để hoàn thành cấp độ I của trò chơi này.

Theo đề bài: $X_1 \sim N(48.1; 15.1^2) \Rightarrow \mu_1 = 48.1; \sigma_1 = 15.1$.

Xác suất để người chơi cần nhiều hơn 50.1 phút để hoàn thành cấp độ I của trò chơi này là

$$\mathbb{P}(X_1 > 50.1) = \mathbb{P}\left(\frac{X_1 - 48.1}{15.1} > \frac{50.1 - 48.1}{15.1}\right) = \mathbb{P}\left(Z > \frac{20}{151}\right) \approx 0.4473.$$

b) Nhà sản xuất cần tính ngưỡng thời gian t_0 sao cho 90.15% người chơi sẽ cần ít hơn t_0 phút để hoàn thành cấp độ I của trò chơi này. Tính t_0.

Cần tìm t_0 để $\mathbb{P}(X_1 < t_0) = 90.15\%$.

Ta có

$$\begin{aligned} \mathbb{P}(X_1 < t_0) &= 90.15\% \\ \Leftrightarrow \mathbb{P}\left(Z < \frac{t_0 - 48.1}{15.1}\right) &= 0.9015 \\ \Rightarrow \frac{t_0 - 48.1}{15.1} &\approx 1.2901 \\ \Leftrightarrow t_0 &\approx 67.5805. \end{aligned}$$

c) Giả sử trò chơi gồm 3 cấp độ, trong đó thời gian để hoàn thành các cấp độ II và III lần lượt là các biến ngẫu nhiên $X_2 \sim N(55, 14.5^2)$ và $X_3 \sim N(56.8, 10.5^2)$. Giả sử rằng X_1, X_2 và X_3 là các biến ngẫu nhiên độc lập. Tính trung bình và độ lệch chuẩn cho tổng thời gian để người chơi hoàn thành cả 3 cấp độ trò chơi này.

Lưu ý: Nếu $X_1 \sim N(\mu_1; \sigma_1^2)$; $X_2 \sim N(\mu_2; \sigma_2^2)$; $\ldots$; $X_k \sim N(\mu_k; \sigma_k^2)$ và X_i độc lập X_j với mọi $i \neq j$. Xét

$$X = a_1X_1 + a_2X_2 + \cdots + a_kX_k$$

khi đó $X \sim N(\mu; \sigma^2)$ trong đó

- $\mu = a_1\mu_1 + a_2\mu_2 + \cdots + a_k\mu_k$.
- $\sigma^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_k^2\sigma_k^2$.

Theo đề bài: $X_1 \sim N(48.1; 15.1^2) \Rightarrow \mu_1 = 48.1; \sigma_1 = 15.1$

$X_2 \sim N(55; 14.5^2) \Rightarrow \mu_2 = 55; \sigma_2 = 14.5$

$X_3 \sim N(56.8; 10.5^2) \Rightarrow \mu_3 = 56.8; \sigma_3 = 10.5$

Tổng thời gian để người chơi hoàn thành cả 3 cấp độ trò chơi là

$$X = X_1 + X_2 + X_3.$$

Khi đó $X \sim N(\mu; \sigma^2)$ trong đó

- $\mu = \mu_1 + \mu_2 + \mu_3 = 48.1 + 55 + 56.8 = 159.9$.
- $\sigma^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 = 15.1^2 + 14.5^2 + 10.5^2 = 548.51$
  $\Rightarrow \sigma \approx 23.4203$.

□

Câu 5. Trong sản xuất chất bán dẫn, khắc hóa chất ướt thường được sử dụng để loại bỏ silic từ mặt sau của tấm wafer trước khi kim loại hóa. Tỷ lệ ăn mòn (etch) là một đặc tính quan trọng trong quá trình này và được biết là tuân theo phân phối chuẩn. Hai phương pháp khắc khác nhau đã được so sánh bằng cách sử dụng hai mẫu ngẫu nhiên gồm 8 tấm wafer cho mỗi dung dịch. Gọi X và Y lần lượt là tỷ lệ ăn mòn tương ứng với phương pháp khắc I và II. Kết quả quan sát được như sau

Phương pháp I (x_i)	11.1	11.2	11.4	11.1	11.7	11.4	11.2	11.1
Phương pháp II (y_i)	9.1	8.9	8.9	9.1	8.9	9.2	9	9.3

Giả sử rằng phương sai của hai tổng thể bằng nhau. Với mức ý nghĩa 8%, hỏi dữ liệu trên có hỗ trợ cho tuyên bố rằng "tỷ lệ ăn mòn trung bình là giống nhau cho cả hai phương pháp" không?

✍ **Lời giải.**

Theo đề bài $X \sim N(\mu_X; \sigma_X^2)$; $Y \sim N(\mu_Y; \sigma_Y^2)$

$\sigma_X = \sigma_Y$ chưa biết.

Mẫu 1: $n_X = 8$; $\bar{x} = 11.275$ và $s_X^2 = 0.045$.
Mẫu 2: $n_Y = 8$; $\bar{y} = 9.05$ và $s_Y^2 \approx 0.0229$.

• **Giả thuyết kiểm định** $\begin{cases} H_0 : \mu_X = \mu_Y \\ H_1 : \mu_X \neq \mu_Y \end{cases}$; $\Rightarrow \mu_X - \mu_Y = 0$; $\alpha = 8\% = 0.08$.

• **Giá trị thống kê kiểm định**

$$s_p^2 = \frac{(n_X-1)s_X^2 + (n_Y-1)s_Y^2}{n_X + n_Y - 2} = \frac{(8-1)\cdot 0.045 + (8-1)\cdot 0.0229}{8+8-2} \approx 0.0340$$

$$t_0 = \frac{\bar{x} - \bar{y} - (\mu_X - \mu_Y)}{\sqrt{\frac{s_p^2}{n_X} + \frac{s_p^2}{n_Y}}} = \frac{11.275 - 9.05 - 0}{\sqrt{\frac{0.0340}{8} + \frac{0.0340}{8}}} = 24.1335.$$

• **Miền bác bỏ:** Bác bỏ H_0 khi $|t_0| > t_{\alpha/2;n_X+n_Y-2}$.
Ta có $\alpha = 0.08$
$\Rightarrow t_{\alpha/2;n_X+n_Y-2} = t_{0.04;14} = 1.89$.

• **So sánh và kết luận:**
Ta có: $|t_0| > t_{\alpha/2;n_X+n_Y-2}$
$|24.1335| > t_{\alpha/2;n_X+n_Y-2}$
$24.1335 > 1.89$ (đúng)
$\Rightarrow$ Bác bỏ $H_0 : \mu_X = \mu_Y$.
Kết luận: Với mức ý nghĩa 8% thì tỷ lệ ăn mòn trung bình của hai phương pháp là khác nhau. □

XÁC SUẤT THỐNG KÊ

ĐỀ CUỐI KỲ I NĂM HỌC 2021-2022

Thời gian làm bài: 90 phút

Đề số: 2

Câu 1. Một cái cân điện tử luôn hiển thị kết quả bằng cân nặng đúng của vật thể cộng với sai số ngẫu nhiên. Giả sử rằng sai số ngẫu nhiên này có phân phối chuẩn với kỳ vọng bằng 0 và độ lệch chuẩn là 0.24 (kg). Kết quả cân nặng của một vật thể trong 8 lần đo từ cân điện tử này như sau

5.39 4.3 4.89 4.93 4.8 4.87 4.66 4.74

a) Tìm khoảng tin cậy 95% cho cân nặng trung bình của vật thể này.

b) Người ta cần cân vật thể này ít nhất bao nhiêu lần để sai số ước lượng của khoảng tin cậy 95% cho cân nặng trung bình của vật thể không quá 0.18 (kg).

✍ **Lời giải.**
Gọi X (kg) là cân nặng của vật thể.
Theo đề bài: $X \sim N(\mu; 0.24^2)$
$\sigma = 0.24$: đã biết.
Từ số liệu đề bài, ta có: $n = 8$; $\bar{x} = 4.8225$.

a) Tìm khoảng tin cậy 95% cho cân nặng trung bình của vật thể này.
• **Độ tin cậy:** 95% $\Rightarrow \alpha = 0.05$.
$\Rightarrow z_{1-\alpha/2} = z_{0.975} \approx 1.96$.
• **Sai số:** $\epsilon = z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$
$= 1.96 \cdot \frac{0.24}{\sqrt{8}}$
≈ 0.1663.

• **KTC** 95% **cho trung bình** μ **là:**

$$\begin{aligned} \bar{x} - \epsilon \leq &\quad \mu \quad \leq \bar{x} - \epsilon \\ \Leftrightarrow 4.8225 - 0.1663 \leq &\quad \mu \quad \leq 4.8225 + 0.1663 \\ \Leftrightarrow 4.6562 \leq &\quad \mu \quad \leq 4.9888 \end{aligned}$$

$\Rightarrow \mu \in [4.6562; 4.9888]$.

b) Người ta cần cân vật thể này ít nhất bao nhiêu lần để sai số ước lượng của khoảng tin cậy 95% cho cân nặng trung bình của vật thể không quá 0.18 (kg).
• **Độ tin cậy:** $95\% \Rightarrow \alpha = 0.05$.
$\Rightarrow z_{1-\alpha/2} = z_{0.975} \approx 1.96$.
Để $\epsilon \leq 0.18$ thì

$$n \geq \left(\frac{z_{1-\alpha/2} \cdot \sigma}{\epsilon_0}\right)^2 = \left(\frac{1.96 \cdot 0.24}{0.18}\right)^2 \approx 6.8295$$

$\Rightarrow n \geq 7$.
Vậy người ta cần cân vật thể này ít nhất 7 lần để sai số ước lượng của khoảng tin cậy 95% cho cân nặng trung bình của vật thể không quá 0.18 (kg).

□

Câu 2. Người ta cần nghiên cứu mối quan hệ giữa lượng thức uống có cồn đã sử dụng (x) và nồng độ cồn trong máu (y) của một người. Trong một thí nghiệm, 20 người có thể trạng tương tự được cho dùng ngẫu nhiên một lượng đồ uống có cồn nhất định và kiểm tra nồng độ cồn trong máu của những người này sau 1 giờ. Biết

$$\sum_i^{20} x_i = 78; \sum_i^{20} x_i^2 = 362; \sum_i^{20} y_i = 213; \sum_i^{20} y_i^2 = 2633.052; \sum_i^{20} x_i y_i = 969.36.$$

a) Tìm phương trình đường thẳng hồi quy ước lượng $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ dự đoán nồng độ cồn trong máu theo số lượng thức uống có cồn.

b) Tính các tổng bình phương: SST, SSR, SSE.

c) Tính hệ số xác định R^2.

✍ Lời giải.

a) Tìm phương trình đường thẳng hồi quy ước lượng $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ dự đoán nồng độ cồn trong máu theo số lượng thức uống có cồn.
Ta có $n = 20$ và

- $S_{xy} = \sum_{i=1}^n x_i y_i - \dfrac{\left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right)}{n} = 969.36 - \dfrac{78 \cdot 213}{20} = 138.66.$
- $S_{xx} = \sum_{i=1}^n x_i^2 - \dfrac{\left(\sum_{i=1}^n x_i\right)^2}{n} = 362 - \dfrac{78^2}{20} = 57.8.$

$\Rightarrow \hat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}} = \dfrac{138.66}{57.8} \approx 2.3990.$

- $\bar{x} = \dfrac{1}{n}\sum_{i=1}^n x_i = \dfrac{1}{20} \cdot 78 = 3.9$
- $\bar{y} = \dfrac{1}{n}\sum_{i=1}^n y_i = \dfrac{1}{20} \cdot 213 = 10.65$

$\Rightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 10.65 - 2.3990 \cdot 3.9 \approx 1.2939.$
$\Rightarrow \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 1.2939 + 2.3990x.$
Vậy $\hat{y} = 1.2939 + 2.3990x$

b) Tính các tổng bình phương: SST, SSR, SSE.

- $SST = \sum_i^n y_i^2 - \dfrac{\left(\sum_{i=1}^n y_i\right)^2}{n} = 2633.052 - \dfrac{213^2}{20} = 364.602$
- $SSR = \hat{\beta}_1 S_{xy} = 2.3990 \cdot 138.66 \approx 324.3257$
- $SSE = SST - SSR = 364.602 - 324.3257 = 40.2763$

c) Tính hệ số xác định R^2.

$$R^2 = \frac{SSR}{SST} = \frac{324.3257}{364.602} \approx 0.8895.$$

□

Câu 3. Trong một nhà máy sản xuất linh kiện điện tử, các kỹ sư cân nhắc lựa chọn giữa hai loại nhựa để tiến hành sản xuất linh kiện. Một yếu tố quan trọng được cân nhắc là sức chịu lực phá hủy (đơn vị: psi) của từng loại nhựa. Đối với loại nhựa 1, các kỹ sư kiểm tra trên một mẫu cỡ $m = 13$, tính được sức chịu lực phá hủy trung bình là $\bar{x} = 163.25$ và độ lệch chuẩn mẫu $s_x = 1.1$. Thử nghiệm trên một mẫu cỡ $n = 14$ đối với loại nhựa 2, tính được sức chịu lực phá hủy trung bình là $\bar{y} = 163.5$ và độ lệch chuẩn mẫu $s_y = 1.95$. Với mức ý nghĩa $\alpha = 3\%$, ta có thể khẳng định rằng sức chịu lực phá hủy của hai loại nhựa này là như nhau hay không? Giả sử rằng sức chịu lực phá hủy của hai loại nhựa tuân theo phân phối chuẩn và có phương sai khác nhau.

✍ **Lời giải.**

Gọi X_1 : sức chịu lực phá hủy của nhựa loại 1.

X_2 : sức chịu lực phá hủy của nhựa loại 2.

Mẫu 1: $n_1 = 13$; $\bar{x}_1 = 163.25$; $s_1 = 1.1$.
Mẫu 2: $n_2 = 14$; $\bar{x}_2 = 163.5$; $s_2 = 1.95$.
$\sigma_1^2 \neq \sigma_2^2$: chưa biết.

• **Giả Thuyết KĐ:** $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$: KĐ 2 phía, $\Delta_0 = \mu_1 - \mu_2 = 0$.

• **Mức ý nghĩa:** $\alpha = 0.03$.

• **Giá trị Thống kê kiểm định:**

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{163.25 - 163.5 - 0}{\sqrt{\dfrac{1.1^2}{13} + \dfrac{1.95^2}{14}}} \approx -0.4140.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $|t_0| > t_{\alpha/2;v}$ trong đó

$$v = \left\lfloor \frac{\left[(s_1^2/n_1) + (s_2^2/n_2)\right]^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}} \right\rfloor \approx \lfloor 20.7915 \rfloor = 20.$$

$\alpha = 0.03 \Rightarrow t_{\alpha/2;v} = t_{0.015;20} = 2.34.$

• **So sánh và kết luận:**

Ta có: $|t_0| > t_{\alpha/2;v}$

$\Leftrightarrow |-0.4140| > 2.34$

$\Leftrightarrow 0.4140 > 2.34$ (sai)

$\Rightarrow$ chưa đủ cơ sở để bác bỏ $H_0 : \mu_1 = \mu_2$.

Kết luận: Với mức ý nghĩa 3%, ta có thể khẳng định rằng sức chịu lực phá hủy của hai loại nhựa này là như nhau.

□

Câu 4. Các tác giả của bài báo "Adjuvant Radiotherapy an Chemotherapy in Node-Positive Premenopausal Women with Breast Cancer" (New Engl. J. of Med., 1997: 956-962) báo cáo các kết quả nghiên cứu của họ về một thí nghiệm được thiết kế để so sánh hiệu quả của 2 phương pháp điều trị bệnh ung thư: chỉ điều trị bằng hóa trị (phương pháp 1) và điều trị kết hợp giữa hóa trị và xạ trị (phương pháp 2). Trong số 183 bệnh nhân được điều trị bằng phương pháp 1, có 90 bệnh nhân sống sót ít nhất 15 năm, trong khi 108 trong số 173 bệnh nhân mà được điều trị bằng phương pháp 2 sống sót ít nhất 15 năm. Có ý kiến cho rằng phương pháp điều trị kết hợp giữa hóa trị và xạ trị có hiệu quả hơn là chỉ điều trị bằng hóa trị. Dựa trên kết quả thí nghiệm, ta có đủ bằng chứng ủng hộ ý kiến trên không với mức ý nghĩa $\alpha = 2\%$?

✍ **Lời giải.**

Gọi Y_1 là số bệnh nhân sống sót ít nhất 15 năm khi điều trị bằng phương pháp 1;

Y_2 là số bệnh nhân sống sót ít nhất 15 năm khi điều trị bằng phương pháp 2.

Mẫu 1: $n_1 = 183$; $y_1 = 90 \Rightarrow \hat{p}_1 = \dfrac{y_1}{n_1} = \dfrac{90}{183} = \dfrac{30}{61}$.

Mẫu 2: $n_2 = 173$; $y_2 = 108 \Rightarrow \hat{p}_2 = \dfrac{y_2}{n_2} = \dfrac{108}{173}$.

$$\Rightarrow \hat{p} = \frac{y_1 + y_2}{n_1 + n_2} = \frac{90 + 108}{183 + 173} = \frac{99}{178}$$

• **Giả thuyết kiểm định:** $\begin{cases} H_0 : p_1 \geq p_2 \\ H_1 : p_1 < p_2 \end{cases}$; $\Rightarrow \Delta_0 = p_1 - p_2 = 0$; $\alpha = 2\% = 0.02$.

• **Giá trị thống kê kiểm định**

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2 - \Delta_0}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} = \frac{\dfrac{30}{61} - \dfrac{108}{173} - 0}{\sqrt{\dfrac{99}{178}\left(1 - \dfrac{99}{178}\right)\left(\dfrac{1}{183} + \dfrac{1}{173}\right)}} = -2.5145.$$

• **Miền bác bỏ:** Bác bỏ H_0 khi $z_0 < -z_{1-\alpha}$.

Ta có $\alpha = 0.02$

$\Rightarrow -z_{1-\alpha} = -z_{0.98} = -2.05$.

• **So sánh và kết luận:**

Ta có: $z_0 < -z_{1-\alpha}$

$-2.5145 < -2.05$ (đúng)

$\Rightarrow$ bác bỏ $H_0 : p_1 \geq p_2$.

Kết luận: Với mức ý nghĩa 2% thì ta có đủ bằng chứng ủng hộ ý kiến cho rằng phương pháp điều trị kết hợp giữa hóa trị và xạ trị có hiệu quả hơn là chỉ điều trị bằng hóa trị. □

Câu 5. Một công ty muốn so sánh hiệu quả của hai loại nhiên liệu cho một chiếc ô tô. 5 chiếc xe giống nhau sẽ chạy 1000 km mỗi chiếc, hai chiếc sử dụng loại nhiên liệu thứ nhất và ba chiếc chạy loại thứ hai. Gọi X_1 và X_2 là hiệu suất nhiên liệu quan sát được đối với loại thứ nhất và Y_1, Y_2 và Y_3 là hiệu suất đối với loại thứ hai. Giả sử tất cả các biến này độc lập với nhau và $X_i \sim N(21, 8)$ với $i = 1, 2$ và $Y_j \sim N(18, 11)$ với $j = 1, 2, 3$. Gọi W là biến ngẫu nhiên mới được cho bởi:

$$W = \frac{X_1 + X_2}{2} - \frac{Y_1 + Y_2 + Y_3}{3}.$$

a) Tìm kỳ vọng của W.

b) Tìm phương sai của W.

c) Tìm $\mathbb{P}(W \leq 0)$.

✍ **Lời giải.**

Lưu ý: Nếu $X_1 \sim N(\mu_1; \sigma_1^2)$; $X_2 \sim N(\mu_2; \sigma_2^2)$; $\ldots$; $X_k \sim N(\mu_k; \sigma_k^2)$ và X_i độc lập X_j với mọi $i \neq j$. Xét

$$X = a_1X_1 + a_2X_2 + \cdots + a_kX_k$$

khi đó $X \sim N(\mu; \sigma^2)$ trong đó

- $\mu = a_1\mu_1 + a_2\mu_2 + \cdots + a_k\mu_k$.
- $\sigma^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_k^2\sigma_k^2$.

Theo đề bài: $X_i \sim N(\mu_{X_i}; \sigma_{X_i}^2)$ với $\mu_{X_i} = 21; \sigma_{X_i} = \sqrt{8} = 2\sqrt{2}, \forall i = 1, 2$.

$Y_j \sim N(\mu_{Y_j}; \sigma_{Y_j}^2)$ với $\mu_{Y_j} = 18; \sigma_{Y_j} = \sqrt{11}, \forall j = 1, 2, 3$.

Ta có

$$W = \frac{X_1 + X_2}{2} - \frac{Y_1 + Y_2 + Y_3}{3} = \frac{1}{2}X_1 + \frac{1}{2}X_2 - \frac{1}{3}Y_1 - \frac{1}{3}Y_2 - \frac{1}{3}Y_3.$$

a) Tìm kỳ vọng của W.

$$\begin{aligned}\mu_W = \mathbb{E}(W) &= \mathbb{E}\left(\frac{1}{2}X_1 + \frac{1}{2}X_2 - \frac{1}{3}Y_1 - \frac{1}{3}Y_2 - \frac{1}{3}Y_3\right) \\ &= \frac{1}{2}\mu_{X_1} + \frac{1}{2}\mu_{X_2} - \frac{1}{3}\mu_{Y_1} - \frac{1}{3}\mu_{Y_2} - \frac{1}{3}\mu_{Y_3} \\ &= \frac{1}{2}\cdot 21 + \frac{1}{2}\cdot 21 - \frac{1}{3}\cdot 18 - \frac{1}{3}\cdot 18 - \frac{1}{3}\cdot 18 \\ &= 3.\end{aligned}$$

b) Tìm phương sai của W.

$$\begin{aligned}\sigma_W^2 = Var(W) &= Var\left(\frac{1}{2}X_1 + \frac{1}{2}X_2 - \frac{1}{3}Y_1 - \frac{1}{3}Y_2 - \frac{1}{3}Y_3\right) \\ &= \left(\frac{1}{2}\right)^2\cdot 8 + \left(\frac{1}{2}\right)^2\cdot 8 + \left(-\frac{1}{3}\right)^2\cdot 11 + \left(-\frac{1}{3}\right)^2\cdot 11 + \left(-\frac{1}{3}\right)^2\cdot 11 \\ &= \frac{23}{3}.\end{aligned}$$

c) Tìm $\mathbb{P}(W \leq 0)$.

Ta có $W \sim N\left(3; \frac{23}{3}\right)$, do đó

$$\mathbb{P}(W \leq 0) = \mathbb{P}\left(\frac{W-3}{\sqrt{\frac{23}{3}}} \leq \frac{0-3}{\sqrt{\frac{23}{3}}}\right) = \mathbb{P}\left(Z \leq -\frac{3\sqrt{69}}{23}\right) \approx 0.1393.$$

□

Câu 1 (4 điểm). Đem cân một số trái cây T vừa thu hoạch, ta được kết quả sau

Trọng lượng (gam)	205	215	225	235	245
Số trái	12	17	20	18	15

Giả sử rằng trọng lượng của trái cây T có phân phối chuẩn.

a) Tìm khoảng ước lượng cho trọng lượng trung bình của trái cây với độ tin cậy 95%.

b) Nếu muốn sai số ước lượng không quá 2 gam ở độ tin cậy 96% thì phải quan sát ít nhất bao nhiêu trái?

c) Những trái cây có trọng lượng ≥ 230 gam được xếp vào trái loại I. Hãy ước lượng tỷ lệ trái cây loại I với độ tin cậy 97%.

d) Người ta dự đoán rằng những trái cây T trên có trọng lượng trung bình 230 gam. Hãy cho biết dự đoán trên có chính xác không ở mức ý nghĩa 2%.

✍ **Lời giải.**
Theo đề bài: $T \sim N(\mu; \sigma^2)$
σ^2: chưa biết.
Từ bảng số liệu đề bài, ta có: $n = 82$; $\bar{x} \approx 225.8537$; $s \approx 13.2592$.

a) Tìm khoảng ước lượng cho trọng lượng trung bình của trái cây với độ tin cậy 95%.
• **Độ tin cậy:** 95% $\Rightarrow \alpha = 0.05$.
$\Rightarrow t_{\alpha/2;n-1} = t_{0.025;81} \approx z_{0.975} \approx 1.96$.
• **Sai số:** $\epsilon = t_{\alpha/2;n-1} \cdot \dfrac{s}{\sqrt{n}}$
$= 1.96 \cdot \dfrac{13.2592}{\sqrt{82}}$
≈ 2.8699.
• **KTC** 95% **cho trung bình** μ **là:**

$$\begin{aligned} \bar{x} - \epsilon \leq & \quad \mu \quad \leq \bar{x} - \epsilon \\ \Leftrightarrow 225.8537 - 2.8699 \leq & \quad \mu \quad \leq 225.8537 + 2.8699 \\ \Leftrightarrow 222.9838 \leq & \quad \mu \quad \leq 228.7236 \end{aligned}$$

$\Rightarrow \mu \in [222.9838; 228.7236]$.

b) Nếu muốn sai số ước lượng không quá 2 gam ở độ tin cậy 96% thì phải quan sát ít nhất bao nhiêu trái?
• **Độ tin cậy:** 96% $\Rightarrow \alpha = 0.04$.
$\Rightarrow z_{1-\alpha/2} = z_{0.98} \approx 2.05$.
Để $\epsilon \leq 2$ thì

$$n \geq \left(\frac{z_{1-\alpha/2} \cdot s}{\epsilon_0}\right)^2 = \left(\frac{2.05 \cdot 13.2592}{2}\right)^2 \approx 184.7066$$

$\Rightarrow n \geq 185$.
Vậy nếu muốn sai số ước lượng không quá 2 gam ở độ tin cậy 96% thì phải quan sát ít nhất 185 trái.

c) Những trái cây có trọng lượng ≥ 230 gam được xếp vào trái loại I. Hãy ước lượng tỷ lệ trái cây loại I với độ tin cậy 97%.

Gọi Y là số trái cây loại I.

Ta có: $n = 82$, $y = 18 + 15 = 33 \Rightarrow$ tỷ lệ mẫu $\hat{p} = \dfrac{y}{n} = \dfrac{33}{82}$.

• **Độ tin cậy:** $97\% \Rightarrow \alpha = 0.03$.
$\Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.985} \approx 2.17$.

• **Sai số:** $\epsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

$$= 2.17 \cdot \sqrt{\frac{\frac{33}{82}\left(1 - \frac{33}{82}\right)}{82}}$$

≈ 0.1175.

• **KTC** 97% **cho tỷ lệ** p **là**

$$\begin{aligned} \hat{p} - \epsilon \leq \ & p \ \leq \hat{p} + \epsilon \\ \Leftrightarrow \frac{33}{82} - 0.1175 \leq \ & p \ \leq \frac{33}{82} + 0.1175 \\ \Leftrightarrow 0.2849 \leq \ & p \ \leq 0.5199 \end{aligned}$$

$\Rightarrow p \in [0.2849; 0.5199]$.

d) Người ta dự đoán rằng những trái cây T trên có trọng lượng trung bình 230 gam. Hãy cho biết dự đoán trên có chính xác không ở mức ý nghĩa 2%.

σ^2: chưa biết.

Ta có: $n = 82$; $\bar{x} \approx 225.8537$; $s \approx 13.2592$.

• **Giả Thuyết KĐ:** $\begin{cases} H_0 : \mu = 230 \\ H_1 : \mu \neq 230 \end{cases}$: KĐ 2 phía, $\mu_0 = 230$.

• **Mức ý nghĩa:** $\alpha = 2\% = 0.02$.

• **Giá trị Thống kê kiểm định:**

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{225.8537 - 230}{\frac{13.2592}{\sqrt{82}}} \approx -2.8317.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $|t_0| > t_{\alpha/2;n-1}$.

Ta có $\alpha = 0.02$

$\Rightarrow t_{\alpha/2;n-1} = t_{0.01;81} \approx z_{0.99} \approx 2.33$.

• **So sánh và kết luận:**

Ta có: $|t_0| > t_{\alpha/2;n-1}$

$\Leftrightarrow |-2.8317| > 2.33$

$\Leftrightarrow 2.8317 > 2.33$ (đúng)

$\Rightarrow$ bác bỏ $H_0 : \mu = 230$.

Kết luận: Với mức ý nghĩa 2%, những trái cây T trên có trọng lượng trung bình không phải là 230 gam.

□

Câu 2 (2 điểm). Vào mùa đông của đại dịch cúm, bố mẹ của 2000 bé đã được các nhà nghiên cứu khảo sát tại một công ty dược phẩm nổi tiếng để xác định liệu thuốc mới của công ty có hiệu quả sau hai ngày hay không. Trong số 120 bé bị cúm và được cho dùng thuốc, 29 bé khỏi bệnh trong hai ngày. Trong 280 bé bị cúm những không được cho dùng thuốc mới, có 56 bé phục hồi trong hai ngày. Hỏi có dấu hiệu có ý nghĩa nào ủng hộ lời tuyên bố của công ty về hiệu quả của thuốc hay không với mức ý nghĩa 5%? (*Yêu cầu dùng cả 2 phương pháp: miền bác bỏ và p−giá trị*)

✍ Lời giải.

Gọi Y_1 : số bé uống thuốc mới khỏi bệnh trong hai ngày.

Y_2 : số bé không uống thuốc mới khỏi bệnh trong hai ngày.

Mẫu 1: $n_1 = 120$; $y_1 = 29 \Rightarrow \hat{p}_1 = \dfrac{y_1}{n_1} = \dfrac{29}{120}$.

Mẫu 2: $n_2 = 280$; $y_2 = 56 \Rightarrow \hat{p}_2 = \dfrac{y_2}{n_2} = \dfrac{56}{280} = 0.2$.

$\Rightarrow \hat{p} = \dfrac{y_1 + y_2}{n_1 + n_2} = \dfrac{29 + 56}{120 + 280} = 0.2125$.

• **GTKĐ:** $\begin{cases} H_0 : p_1 \le p_2 \\ H_1 : p_1 > p_2 \end{cases}$: KĐ 1 phía, $p_0 = p_1 - p_2 = 0$.

• **Mức ý nghĩa:** $\alpha = 0.05$.

• **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} = \frac{\dfrac{29}{120} - 0.2 - 0}{\sqrt{0.2125(1-0.2125)\left(\dfrac{1}{120} + \dfrac{1}{280}\right)}} \approx 0.93352.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $z_0 > z_{1-\alpha}$.

Ta có $\alpha = 0.05$

$\Rightarrow z_{1-\alpha} = z_{0.95} = 1.64485$.

• **So sánh và kết luận:**

Ta có: $z_0 > z_{1-\alpha}$

$\Leftrightarrow 0.93352 > 1.64485$ (sai)

$\Rightarrow$ Chưa đủ cơ sở để bác bỏ $H_0 : p_1 \le p_2$.

Kết luận: Với mức ý nghĩa 5%, tỷ lệ bé khỏi bệnh khi uống thuốc mới sau hai ngày không nhiều hơn bé không uống thuốc mới, hay thuốc mới không hiệu quả như tuyên bố của công ty.

⁕ Cách 2: Dùng p-giá trị:

Ta có

$$p - \text{giá trị} = 1 - \Phi(z_0) = 1 - \Phi(0.93352) \approx 0.8247.$$

Vì p−giá trị $= 0.8247 > 0.05 = \alpha$ nên chưa đủ cơ sở để bác bỏ $H_0 : p_1 \le p_2$. □

Câu 3 (2 điểm). Một nghiên cứu được thực hiện để xem việc tăng nồng độ cơ chất có tác dụng đáng kể đến tốc độ của một phản ứng hóa học hay không. Với một nồng độ cơ chất 1.5 mol/l, phản ứng được chạy 15 lần, với tốc độ trung bình 7.5 micromoles mỗi 30 phút và độ lệch chuẩn 1.5. Với một nồng độ cơ chất 2.0 mol/l, 12 lần chạy được thực hiện, thu được tốc độ trung bình 8.8 micromoles mỗi 30 phút và độ lệch chuẩn 1.2. Có lý do để tin rằng việc tăng nồng độ cơ chất làm tăng tốc độ trung bình của phản ứng 0.5 micromoles mỗi 30 phút hay không? Sử dụng mức ý nghĩa 0.01 và giả sử rằng các tổng thể là xấp xỉ chuẩn với các phương sai bằng nhau.

✍ Lời giải.

Gọi X_1 (micromoles) là tốc độ phản ứng ứng với nồng độ cơ chất 1.5 mol/l;

X_2(micromoles) là tốc độ phản ứng ứng với nồng độ cơ chất 2 mol/l;.

Theo đề bài $X_1 \sim N(\mu_1; \sigma_1^2)$; $X_2 \sim N(\mu_2; \sigma_2^2)$

$\sigma_1 = \sigma_2$ chưa biết.

Mẫu 1: $n_1 = 15$; $\bar{x}_1 = 7.5$ và $s_1 = 1.5$.

Mẫu 2: $n_2 = 12$; $\bar{x}_2 = 8.8$ và $s_2 = 1.2$.

• **Giả thuyết kiểm định** $\begin{cases} H_0 : \mu_1 - \mu_2 = -0.5 \\ H_1 : \mu_1 - \mu_2 \ne -0.5 \end{cases}$; $\Rightarrow \Delta_0 = \mu_1 - \mu_2 = -0.5$; $\alpha = 0.01$.

• **Giá trị thống kê kiểm định**

$$s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = \frac{(15-1)\cdot 1.5^2+(12-1)\cdot 1.2^2}{15+12-2} = 1.8936$$

$$t_0 = \frac{\bar{x}_1-\bar{x}_2-\Delta_0}{\sqrt{\dfrac{s_p^2}{n_1}+\dfrac{s_p^2}{n_2}}} = \frac{7.5-8.8-(-0.5)}{\sqrt{\dfrac{1.8936}{12}+\dfrac{1.8936}{15}}} = -1.5011.$$

• **Miền bác bỏ:** Bác bỏ H_0 khi $|t_0| > t_{\alpha/2;n_1+n_2-2}$.
Ta có $\alpha = 0.01$
$\Rightarrow t_{\alpha/2;n_1+n_2-2} = t_{0.005;25} = 2.787.$
• **So sánh và kết luận:**
Ta có: $|t_0| > t_{\alpha/2;n_1+n_2-2}$
$1.5011 > 2.787$ (sai)
$\Rightarrow$ chưa đủ cơ sở để bác bỏ $H_0 : \mu_1-\mu_2 = -0.5$.
Kết luận: Với mức ý nghĩa 1% thì không có lý do để tin rằng việc tăng nồng độ cơ chất làm tăng tốc độ trung bình của phản ứng 0.5 micromoles mỗi 30 phút. □

Câu 4 (2 điểm). Một nghiên cứu về khối lượng đường bị biến đổi trong một quá trình nào đó ở các nhiệt độ khác nhau. Dữ liệu được mã hóa và ghi lại như sau

Nhiệt độ (x)	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
Đường bị biến đổi (y)	8.1	7.8	8.5	9.8	9.5	8.9	8.6	10.2	9.3	9.2	10.5

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

b) Ước lượng khối lượng trung bình của đường bị biến đổi được tạo ra khi nhiệt độ được mã hóa là 1.75.

✍ Lời giải.

a) Tìm phương trình hồi quy tuyến tính đơn y theo x:
Từ bảng dữ liệu đề bài, tiến hành sử dụng máy tính cầm tay, ta tính được: $\hat{\beta}_0 \approx 6.4136$; $\hat{\beta}_1 \approx 1.8091$

$$\Rightarrow \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 6.4136 + 1.8091x.$$

⁎ Giải thích ý nghĩa $\hat{\beta}_1$ nhận được:
Vì $\hat{\beta}_1 \approx 1.8091 > 0$ nên khi nhiệt độ thay đổi 1 đơn vị và các yếu tố khác không thay đổi thì đường bị biến đối 1.8091 đơn vị.

b) Ước lượng khối lượng trung bình của đường bị biến đổi được tạo ra khi nhiệt độ được mã hóa là 1.75.
Khi nhiệt độ được mã hóa là 1.75 $\Rightarrow x_0 = 1.75$

$$\hat{y}_0 = 6.4136 + 1.8091 \cdot 1.75 \approx 9.5795.$$

Khi nhiệt độ được mã hóa là 1.75 thì khối lượng trung bình của đường bị biến đổi được tạo ra 9.5795.

□

Câu 1 (4 điểm). Kết quả quan sát về hàm lượng Vitamin C của một loại trái cây cho bởi bảng sau

Hàm lượng Vitamin C (mg)	6	8	10	12	14	16
Số trái	5	10	20	35	25	5

a) Ước lượng hàm lượng Vitamin C trung bình trong một trái cây với độ tin cậy 90%.

b) Những trái có hàm lượng Vitamin C trên 11 mg trở lên là trái loại I. Ước lượng tỷ lệ trái loại I với độ tin cậy 92%.

c) Có ý kiến rằng hàm lượng Vitamin C trung bình là 10 mg. Với mức ý nghĩa 6%, hãy kiểm tra ý kiến trên.

d) Lời khẳng định: "Tỷ lệ trái cây loại I là 60%" có được chấp nhận hay không với mức ý nghĩa 4%.

✍ **Lời giải.**
Gọi X (mg) là hàm lượng vitamin C trong trái cây trên.
Giả sử $X \sim N(\mu; \sigma^2)$
σ^2: chưa biết.
Từ bảng số liệu đề bài, ta có: $n = 100$; $\bar{x} = 11.6$; $s \approx 2.4288$.

a) Ước lượng hàm lượng Vitamin C trung bình trong một trái cây với độ tin cậy 90%.
- **Độ tin cậy:** $90\% \Rightarrow \alpha = 0.1$.
$\Rightarrow t_{\alpha/2;n-1} = t_{0.05;99} \approx z_{0.95} \approx 1.645$.
- **Sai số:** $\epsilon = t_{\alpha/2;n-1} \cdot \dfrac{s}{\sqrt{n}}$
$= 1.645 \cdot \dfrac{2.4288}{\sqrt{100}}$
≈ 0.3995.
- **KTC** 90% **cho trung bình** μ **là:**

$$\begin{aligned} \bar{x} - \epsilon \leq \quad & \mu \quad \leq \bar{x} - \epsilon \\ \Leftrightarrow 11.6 - 0.3995 \leq \quad & \mu \quad \leq 11.6 + 0.3995 \\ \Leftrightarrow 11.2005 \leq \quad & \mu \quad \leq 11.9995 \end{aligned}$$

$\Rightarrow \mu \in [11.2005; 11.9995]$.

b) Những trái có hàm lượng Vitamin C trên 11 mg trở lên là trái loại I. Ước lượng tỷ lệ trái loại I với độ tin cậy 92%.
Gọi Y là số trái cây loại I.
Ta có: $n = 100$, $y = 35 + 25 + 5 = 65 \Rightarrow$ tỷ lệ mẫu $\hat{p} = \dfrac{y}{n} = \dfrac{65}{100} = 0.65$.
- **Độ tin cậy:** $92\% \Rightarrow \alpha = 0.08$.
$\Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.96} \approx 1.75$.
- **Sai số:** $\epsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$
$= 1.75 \cdot \sqrt{\dfrac{0.65\,(1-0.65)}{100}}$
≈ 0.0835.

• **KTC 92% cho tỷ lệ p là**

$$\begin{aligned} \hat{p}-\epsilon \leq \; & p \; \leq \hat{p}+\epsilon \\ \Leftrightarrow 0.65-0.0835 \leq \; & p \; \leq 0.65+0.0835 \\ \Leftrightarrow 0.5665 \leq \; & p \; \leq 0.7335 \end{aligned}$$

$\Rightarrow p \in [0.5665; 0.7335]$.

c) Có ý kiến rằng hàm lượng Vitamin C trung bình là 10 mg. Với mức ý nghĩa 6%, hãy kiểm tra ý kiến trên.
σ^2: chưa biết.
Ta có: $n = 100$; $\bar{x} = 11.6$; $s \approx 2.4288$.
• **Giả Thuyết KĐ:** $\begin{cases} H_0 : \mu = 10 \\ H_1 : \mu \neq 10 \end{cases}$: KĐ 2 phía, $\mu_0 = 10$.
• **Mức ý nghĩa:** $\alpha = 6\% = 0.06$.
• **Giá trị Thống kê kiểm định:**

$$t_0 = \frac{\bar{x}-\mu_0}{\dfrac{s}{\sqrt{n}}} = \frac{11.6-10}{\dfrac{2.4288}{\sqrt{100}}} \approx 6.5876.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $|t_0| > t_{\alpha/2;n-1}$.
Ta có $\alpha = 0.06$
$\Rightarrow t_{\alpha/2;n-1} = t_{0.03;99} \approx z_{0.97} \approx 1.88$.
• **So sánh và kết luận:**
Ta có: $|t_0| > t_{\alpha/2;n-1}$
$\Leftrightarrow |6.5876| > 1.88$
$\Leftrightarrow 6.5876 > 1.88$ (đúng)
$\Rightarrow$ bác bỏ $H_0 : \mu = 10$.
Kết luận: Với mức ý nghĩa 6%, những trái cây trên có hàm lượng vitamin C trung bình không phải là 10 mg.

d) Lời khẳng định: "Tỷ lệ trái cây loại I là 60%" có được chấp nhận hay không với mức ý nghĩa 4%.
Ta có $n = 100$, $y = 35 + 25 + 5 = 65 \Rightarrow$ tỷ lệ mẫu $\hat{p} = \dfrac{y}{n} = \dfrac{65}{100} = 0.65$.
• **GTKĐ:** $\begin{cases} H_0 : p = 0.6 \\ H_1 : p \neq 0.6 \end{cases}$: KĐ 1 phía, $p_0 = 60\% = 0.6$.
• **Mức ý nghĩa:** $\alpha = 4\% = 0.04$.
• **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\hat{p}-p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{0.65-0.6}{\sqrt{\dfrac{0.6(1-0.6)}{100}}} \approx 1.0206.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $|z_0| > z_{1-\alpha/2}$.
Ta có $\alpha = 0.04$
$\Rightarrow z_{1-\alpha/2} = z_{0.98} \approx 2.05$.
• **So sánh và kết luận:**
Ta có: $|z_0| > z_{1-\alpha/2}$
$\Leftrightarrow |1.0206| > 2.05$
$\Leftrightarrow 1.0206 > 2.05$ (sai)
$\Rightarrow$ chưa đủ cở sở để bác bỏ $H_0 : p = 0.6$.
Kết luận: Với mức ý nghĩa 4% thì tỷ lệ trái cây loại I là 60%.

□

Câu 2 (2 điểm). Trong một nghiên cứu để ước tính tỷ lệ cư dân trong một thành phố nào đó và các vùng ngoại ô của nó ủng hộ việc xây dựng nhà máy năng lượng hạt nhân, người ta thấy rằng 63 trong 100 cư dân thành thị ủng hộ việc xây dựng trong khi chỉ 59 trong 125 cư dân ngoại ô là ủng hộ. Có sự khác biệt giữa tỷ lệ cư dân thành thị và ngoại ô ủng hộ việc xây dựng nhà máy hạt nhân hay không với mức ý nghĩa 5%? (*Yêu cầu dùng cả 2 phương pháp: miền bác bỏ và p−giá trị*)

✍ **Lời giải.**

Gọi Y_1 : số cư dân trong thành phố ủng hộ việc xây dựng nhà máy năng lượng hạt nhân;

Y_2 : số cư dân ở vùng ngoại ô ủng hộ việc xây dựng nhà máy năng lượng hạt nhân.

Mẫu 1: $n_1 = 100$; $y_1 = 63 \Rightarrow \hat{p}_1 = \dfrac{y_1}{n_1} = \dfrac{63}{100} = 0.63$.

Mẫu 2: $n_2 = 125$; $y_2 = 59 \Rightarrow \hat{p}_2 = \dfrac{y_2}{n_2} = \dfrac{59}{125} = 0.472$.

$\Rightarrow \hat{p} = \dfrac{y_1 + y_2}{n_1 + n_2} = \dfrac{63 + 59}{100 + 125} = \dfrac{122}{225}$.

• **GTKĐ:** $\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{cases}$: KĐ 2 phía, $\Delta_0 = p_1 - p_2 = 0$.

• **Mức ý nghĩa:** $\alpha = 0.05$.

• **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2 - \Delta_0}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} = \frac{0.63 - 0.472 - 0}{\sqrt{\dfrac{122}{225}\left(1 - \dfrac{122}{225}\right)\left(\dfrac{1}{100} + \dfrac{1}{125}\right)}} \approx 2.3638.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $|z_0| > z_{1-\alpha/2}$.

Ta có $\alpha = 0.05$

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

• **So sánh và kết luận:**

Ta có: $|z_0| > z_{1-\alpha/2}$

$\Leftrightarrow |2.3638| > 1.96$

$\Leftrightarrow 2.3638 > 1.96$ (đúng)

$\Rightarrow$ bác bỏ $H_0 : p_1 = p_2$.

Kết luận: Với mức ý nghĩa 5%, có sự khác biệt có ý nghĩa giữa tỷ lệ cư dân thành thị và ngoại ô trong ủng hộ việc xây dựng nhà máy hạt nhân.

Cách 2: Dùng p-giá trị:

Ta có $p-$giá trị $= 2\left(1 - \Phi(|2.3638|)\right) = 2\left(1 - 0.99096\right) = 0.01808$.

Vì $p-$giá trị $= 0.01808 < 0.05 = \alpha$

$\Rightarrow$ bác bỏ $H_0 : p_1 = p_2$.

□

Câu 3 (2 điểm). Để tìm ra liệu một loại huyết thanh mới có kìm 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 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 5%, 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.

✍ **Lời giải.**

Gọi X_1 (năm) là thời gian sống của chuột được trị liệu bằng huyết thanh;

X_2 (năm) là thời gian sống của chuột không được trị liệu bằng huyết thanh.

Theo đề bài $X_1 \sim N(\mu_1; \sigma_1^2)$; $X_2 \sim N(\mu_2; \sigma_2^2)$

$\sigma_1^2 = \sigma_2^2$ chưa biết.
Mẫu 1: $n_1 = 5$; $\bar{x}_1 = 2.86$ và $s_1^2 = 3.883$.
Mẫu 2: $n_2 = 4$; $\bar{x}_2 = 2.075$ và $s_2^2 = 1.3625$.

• **Giả thuyết kiểm định** $\begin{cases} H_0 : \mu_1 \leq \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases}$; $\Rightarrow \Delta_0 = \mu_1 - \mu_2 = 0$; $\alpha = 0.05$.

• **Giá trị thống kê kiểm định**

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(5-1)\cdot 3.883 + (4-1)\cdot 1.3625}{5+4-2} \approx 2.8028$$

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} = \frac{2.86 - 2.075 - 0}{\sqrt{\frac{2.8028}{5} + \frac{2.8028}{4}}} \approx 0.6990.$$

• **Miền bác bỏ:** Bác bỏ H_0 khi $t_0 > t_{\alpha;n_1+n_2-2}$.
Ta có $\alpha = 0.05$
$\Rightarrow t_{\alpha;n_1+n_2-2} = t_{0.05;7} = 1.895$.

• **So sánh và kết luận:**
Ta có: $t_0 > t_{\alpha;n_1+n_2-2}$
$\Leftrightarrow 0.6990 > 1.895$ (sai)
$\Rightarrow$ chưa đủ cơ sở để bác bỏ $H_0 : \mu_1 \leq \mu_2$.
Kết luận: Với mức ý nghĩa 5% thì huyết thanh không thể được nói là có hiệu quả. □

Câu 4 (2 điểm). Khối lượng của một hợp chất hóa học (y) hòa tan trong 100 gram nước ở các nhiệt độ khác nhau (x) được ghi lại như sau

x (°C)	0	15	30	45	60	75
y (gram)	8	12	25	31	44	48

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

b) Dự đoán khối lượng hòa tan trong 100 gram nước ở nhiệt độ 50°C.

✍ Lời giải.

a) Tìm phương trình hồi quy tuyến tình đơn y theo x. Giải thích ý nghĩa của $\hat{\beta}_1$ nhận được.
Từ bảng dữ liệu đề bài, tiến hành sử dụng máy tính cầm tay, ta tính được: $\hat{\beta}_0 \approx 6.4286$; $\hat{\beta}_1 \approx 0.5752$

$$\Rightarrow \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 6.4286 + 0.5752x.$$

Giải thích ý nghĩa của $\hat{\beta}_1$ nhận được.
Vì $\hat{\beta}_1 \approx 0.5752 > 0$ nên khi nhiệt độ thay đổi 1°C và các yếu tố khác không thay đổi thì khối lượng của hợp chất hóa học hòa tan trong 100g nước tăng 0.5752 g.

b) Dự đoán khối lượng hòa tan trong 100 gram nước ở nhiệt độ 50°C.
Khi nước ở nhiệt độ 50°C $\Rightarrow x_0 = 50$

$$\hat{y}_0 = 6.4286 + 0.5752 \cdot 50 = 35.1886.$$

Khi nhiệt độ nước là 50°C thì khối lượng hòa tan trung bình trong 100 gram là 35.1886 g.

□

XÁC SUẤT THỐNG KÊ	ĐỀ CUỐI KỲ I NĂM HỌC 2020-2021
	Thời gian làm bài: 90 phút
	Đề số: 1

Câu 1 (3 điểm). Cho biết thu nhập của công nhân trong một nhà máy là biến ngẫu nhiên X có phân phối chuẩn. Khảo sát 26 công nhân

Thu nhập (triệu đồng/tháng)	2.2	2.5	2.7	3	3.2	3.5	4
Số công nhân	4	5	6	4	3	2	2

a) Ước lượng thu nhập trung bình của công nhân trong nhà máy với độ tin cậy 95%.

b) Lập khoảng tin cậy 98% dành cho tỷ lệ công nhân có thu nhập không quá 2.5 triệu đồng/tháng. Nếu muốn có sai số ϵ không quá 0.1 thì cỡ mẫu phải chọn bé nhất là bao nhiêu?

c) Lời khẳng định: "Trên 60% công nhân của nhà máy có thu nhập từ 3 triệu đồng/tháng trở lên" có được chấp nhận hay không? Mức ý nghĩa $\alpha = 3\%$.

✍ **Lời giải.**

Gọi X (triệu đồng) là thu nhập hàng tháng của công nhân trong một nhà máy.
Theo đề bài: $X \sim N(\mu; \sigma^2)$
σ^2: chưa biết.
Từ bảng số liệu đề bài, ta có: $n = 26$; $\bar{x} = 2.85$; $s \approx 0.5054$.

a) Ước lượng thu nhập trung bình của công nhân trong nhà máy với độ tin cậy 95%.

• **Độ tin cậy:** 95% $\Rightarrow \alpha = 0.05$.
$\Rightarrow t_{\alpha/2;n-1} = t_{0.025;25} = 2.060$.

• **Sai số:** $\epsilon = t_{\alpha/2;n-1} \cdot \dfrac{s}{\sqrt{n}}$

$$= 2.060 \cdot \frac{0.5054}{\sqrt{26}}$$
$$\approx 0.2042.$$

• **KTC** 95% **cho trung bình** μ **là:**

$$\begin{aligned} \bar{x} - \epsilon \leq \; & \mu \leq \bar{x} - \epsilon \\ \Leftrightarrow 2.85 - 0.2042 \leq \; & \mu \leq 2.85 + 0.2042 \\ \Leftrightarrow 2.6458 \leq \; & \mu \leq 3.0542 \end{aligned}$$

$\Rightarrow \mu \in [2.6458; 3.0542]$.

b) Lập khoảng tin cậy 98% dành cho tỷ lệ công nhân có thu nhập không quá 2.5 triệu đồng/tháng:
Gọi Y là số công nhân có thu nhập không quá 2.5 triệu đồng/tháng trong nhà máy trên.
Ta có: $n = 26$, $y = 4 + 5 = 9 \Rightarrow$ tỷ lệ mẫu $\hat{p} = \dfrac{y}{n} = \dfrac{9}{26}$.

• **Độ tin cậy:** 98% $\Rightarrow \alpha = 0.02$.
$\Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.99} \approx 2.33$.

• **Sai số:** $\epsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

$$= 2.33 \cdot \sqrt{\frac{\frac{9}{26}\left(1 - \frac{9}{26}\right)}{26}}$$
$$\approx 0.2174.$$

• **KTC** 98% **cho tỷ lệ** p **là**

$$\begin{aligned} \hat{p}-\epsilon \leq & \quad p \quad \leq \hat{p}+\epsilon \\ \Leftrightarrow \frac{9}{26}-0.2174 \leq & \quad p \quad \leq \frac{9}{26}+0.2174 \\ \Leftrightarrow 0.1288 \leq & \quad p \quad \leq 0.5636 \end{aligned}$$

$\Rightarrow p \in [0.1288; 0.5636]$.

⋇ Nếu muốn có sai số ϵ không quá 0.1 thì cỡ mẫu phải chọn bé nhất là bao nhiêu?

Để $\epsilon \leq 0.1$ thì

$$n \geq \left(\frac{z_{1-\frac{\alpha}{2}}}{0.1}\right)^2 \hat{p}(1-\hat{p}) = \left(\frac{2.33}{0.1}\right)^2 \cdot \frac{9}{26}\left(1-\frac{9}{26}\right) \approx 122.8790$$

$\Rightarrow n \geq 123$.

Vậy nếu muốn có sai số ϵ không quá 0.1 thì cỡ mẫu phải chọn ít nhất là 123 công nhân.

c) Lời khẳng định: "Trên 60% công nhân của nhà máy có thu nhập từ 3 triệu đồng/tháng trở lên" có được chấp nhận hay không? Mức ý nghĩa $\alpha = 3\%$.

Gọi T là số công nhân của nhà máy có thu nhập từ 3 triệu đồng/tháng trong nhà máy.

Ta có $n = 26$, $t = 4+3+2+2 = 11 \Rightarrow$ tỷ lệ mẫu $\hat{p} = \dfrac{t}{n} = \dfrac{11}{26}$.

• **GTKĐ:** $\begin{cases} H_0 : p \leq 0.6 \\ H_1 : p > 0.6 \end{cases}$: KĐ 1 phía, $p_0 = 60\% = 0.6$.

• **Mức ý nghĩa:** $\alpha = 3\% = 0.03$.

• **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\hat{p}-p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{\dfrac{11}{26}-0.6}{\sqrt{\dfrac{0.6(1-0.6)}{26}}} \approx -1.8415.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $z_0 > z_{1-\alpha}$.

Ta có $\alpha = 0.03$

$\Rightarrow z_{1-\alpha} = z_{0.97} \approx 1.88$.

• **So sánh và kết luận:**

Ta có: $z_0 > z_{1-\alpha}$

$\Leftrightarrow -1.8415 > 1.88$ (sai)

$\Rightarrow$ chưa đủ cở sở để bác bỏ $H_0 : p \leq 0.6$.

Kết luận: Với mức ý nghĩa 3% thì lời khẳng định trên không được chấp nhận.

□

Câu 2 (1.5 điểm). Giả sử rằng trong 1000 khách hàng được khảo sát có 850 người hài lòng hoặc rất hài lòng với các sản phẩm và dịch vụ của công ty. Gọi p là tỷ lệ người hài lòng hoặc rất hài lòng trong tất cả khách hàng. Kiểm định giả thuyết $H_0 : p = 0.9$ với đối thuyết $H_1 : p \neq 0.9$ với $\alpha = 0.05$. Tìm $p-$giá trị.

✍ **Lời giải.**

Gọi Y là số người hài lòng hoặc rất hài lòng trong tất cả khách hàng.

Ta có $n = 1000$, $y = 850 \Rightarrow$ tỷ lệ mẫu $\hat{p} = \dfrac{y}{n} = \dfrac{850}{1000} = 0.85$.

• **GTKĐ:** $\begin{cases} H_0 : p = 0.9 \\ H_1 : p \neq 0.9 \end{cases}$: KĐ 2 phía, $p_0 = 0.9$.

• **Mức ý nghĩa:** $\alpha = 0.05$.

• **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{0.85 - 0.9}{\sqrt{\dfrac{0.9(1-0.9)}{1000}}} \approx -5.2705.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $|z_0| > z_{1-\alpha/2}$.
Ta có $\alpha = 0.05$
$\Rightarrow z_{1-\alpha/2} = z_{0.975} \approx 1.96.$
• **So sánh và kết luận:**
Ta có: $|z_0| > z_{1-\alpha/2}$

$\Leftrightarrow |-5.2705| > 1.96$

$\Leftrightarrow 5.2705 > 1.96$ (đúng)

$\Rightarrow$ chưa đủ cở sở để bác bỏ $H_0 : p = 0.9$.
Kết luận: Với mức ý nghĩa 5% thì tỷ lệ người hài lòng hoặc rất hài lòng trong tất cả khách hàng không phải là 90%.
⁎ Tìm p-giá trị:
Ta có $p-$giá trị $= 2\left[1 - \Phi(|z_0|)\right] = 2\left[1 - \Phi(|-5.2705|)\right] = 2\left[1 - \Phi(5.2705)\right] \approx 2\left[1-1\right] = 0.$ □

Câu 3 (1.5 điểm). Một mẫu ngẫu nhiên 500 cư dân trưởng thành của Quận A chỉ ra rằng 385 ủng hộ việc tăng giới hạn tốc độ đường cao tốc lên 75 dặm một giờ và một mẫu khác gồm 400 cư dân trưởng thành của Quận B đã chỉ ra rằng 267 đã ủng hộ giới hạn tốc độ tăng lên. Những dữ liệu này cho thấy có sự khác biệt trong việc hỗ trợ tăng giới hạn tốc độ cho cư dân của hai quận? Sử dụng $\alpha = 0.05$.
✍ **Lời giải.**
Gọi Y_1 : số cư dân Quận A ủng hộ việc tăng giới hạn tốc độ đường cao tốc lên 75 dặm một giờ;

Y_2 : số cư dân Quận B ủng hộ việc tăng giới hạn tốc độ đường cao tốc lên 75 dặm một giờ.

Mẫu 1: $n_1 = 500$; $y_1 = 385 \Rightarrow \hat{p}_1 = \dfrac{y_1}{n_1} = \dfrac{385}{500} = 0.77.$

Mẫu 2: $n_2 = 400$; $y_2 = 267 \Rightarrow \hat{p}_2 = \dfrac{y_2}{n_2} = \dfrac{267}{400} = 0.6675.$

$\Rightarrow \hat{p} = \dfrac{y_1 + y_2}{n_1 + n_2} = \dfrac{385 + 267}{500 + 400} = \dfrac{163}{225}.$

• **GTKĐ:** $\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{cases}$: KĐ 2 phía, $\Delta_0 = p_1 - p_2 = 0.$
• **Mức ý nghĩa:** $\alpha = 0.05$.
• **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2 - \Delta_0}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} = \frac{0.77 - 0.6675 - 0}{\sqrt{\dfrac{163}{225}\left(1 - \dfrac{163}{225}\right)\left(\dfrac{1}{500} + \dfrac{1}{400}\right)}} \approx 3.4199.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $|z_0| > z_{1-\alpha/2}$.
Ta có $\alpha = 0.05$
$\Rightarrow z_{1-\alpha/2} = z_{0.975} = 1.96.$
• **So sánh và kết luận:**
Ta có: $|z_0| > z_{1-\alpha/2}$

$\Leftrightarrow |3.4199| > 1.96$

$\Leftrightarrow 3.4199 > 1.96$ (đúng)

$\Rightarrow$ bác bỏ $H_0 : p_1 = p_2$.
Kết luận: Với mức ý nghĩa 5%, dữ liệu này cho thấy có sự khác biệt trong việc hỗ trợ tăng giới hạn tốc độ cho cư dân của hai quận. □

Câu 4 (2 điểm). Hai máy được sử dụng để rót đầy các chai nhựa. Khối lượng được rót từ hai máy được giả định có phân phối chuẩn với độ lệch chuẩn lần lượt là $\sigma_1 = 0.020$ và $\sigma_2 = 0.025$ ounce. Một nhân viên kỹ thuật cho rằng khối lượng trung bình được rót hai máy là bằng nhau. Hai mẫu ngẫu nhiên từ mỗi máy như sau

Máy 1	16.03	16.01	16.04	15.96	16.05	15.98	16.05	16.02	16.02	15.99
Máy 2	16.02	16.03	15.97	16.04	15.96	16.02	16.01	16.01	15.99	16.00

Kiểm định ý kiến của nhân viên kỹ thuật trên. Sử dụng $\alpha = 0.05$. Tìm $p-$giá trị.

✍ **Lời giải.**

Gọi X_1 (ounce) là khối lượng được rót vào chai từ máy 1;

X_2 (ounce) là khối lượng được rót vào chai từ máy 2.

Theo đề bài $X_1 \sim N(\mu_1; \sigma_1^2)$; $X_2 \sim N(\mu_2; \sigma_2^2)$

$\sigma_1 = 0.020; \sigma_2 = 0.025$: đã biết.

Mẫu 1: $n_1 = 10$; $\bar{x}_1 = 16.015$.

Mẫu 2: $n_2 = 10$; $\bar{x}_2 = 16.005$.

• **Giả thuyết kiểm định** $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$; $\quad \Rightarrow \Delta_0 = \mu_1 - \mu_2 = 0$; $\alpha = 0.05$.

• **Giá trị thống kê kiểm định**

$$z_0 = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \frac{16.015 - 16.005 - 0}{\sqrt{\dfrac{0.020^2}{10} + \dfrac{0.025^2}{10}}} \approx 0.9877.$$

• **Miền bác bỏ:** Bác bỏ H_0 khi $|z_0| > z_{1-\alpha/2}$.

Ta có $\alpha = 0.05$

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

• **So sánh và kết luận:**

Ta có: $\quad |z_0| > z_{1-\alpha/2}$

$\Leftrightarrow |0.9877| > 1.96$

$\Leftrightarrow 0.9877 > 1.96 \quad$ (sai)

$\Rightarrow$ chưa đủ cơ sở để bác bỏ $H_0 : \mu_1 = \mu_2$.

Kết luận: Với mức ý nghĩa 5% thì khối lượng trung bình được rót hai máy là bằng nhau hay ý kiến của nhân viên kỹ thuật trên là đúng.

⁕ Tính p-giá trị:

$$z_0 = 0.9877 \Rightarrow p\text{-giá trị} = 2\,(1 - \Phi(|0.9877|)) = 2\,(1 - 0.8384) = 03232.$$

□

Câu 5 (2 điểm). Các phương pháp hồi quy đã được sử dụng để phân tích dữ liệu từ một nghiên cứu điều tra mối quan hệ giữa nhiệt độ bề mặt đường (x) và độ lún mặt đường (y). Số liệu được tóm tắt như sau

$$n = 20; \quad \sum_{i=1}^{n} y_i = 12.75; \quad \sum_{i=1}^{n} y_i^2 = 8.86; \quad \sum_{i=1}^{n} x_i = 1478;$$
$$\sum_{i=1}^{n} x_i^2 = 143215.8; \quad \sum_{i=1}^{n} x_i y_i = 1083.67$$

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

b) Hãy tiên đoán lượng độ lún mặt đường sẽ quan trắc được khi nhiệt độ bề mặt đường là 85?

✍ **Lời giải.**

a) Tìm phương trình hồi quy tuyến tính đơn y theo x. Giải thích ý nghĩa của $\hat{\beta}_1$ nhận được.
Ta có $n = 20$ và

- $S_{xy} = \sum_{i=1}^{n} x_i y_i - \dfrac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n} = 1083.67 - \dfrac{1478 \cdot 12.75}{20} = 141.445.$
- $S_{xx} = \sum_{i=1}^{n} x_i^2 - \dfrac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = 143215.8 - \dfrac{1478^2}{20} = 33991.6.$

$\Rightarrow \hat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}} = \dfrac{141.445}{33991.6} \approx 0.0042.$

- $\bar{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i = \dfrac{1}{20} \cdot 1478 = 73.9$
- $\bar{y} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} y_i = \dfrac{1}{20} \cdot 12.75 = 0.6375$

$\Rightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 0.6375 - 0.0042 \cdot 73.9 \approx 0.3271.$
$\Rightarrow \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 0.3271 + 0.0042x.$
Vậy $\hat{y} = 0.3271 + 0.0042x.$
Giải thích ý nghĩa của $\hat{\beta}_1$ nhận được.
Vì $\hat{\beta}_1 \approx 0.0042 > 0$ nên khi nhiệt độ bề mặt đường tăng 1 đơn vị thì độ lún mặt đường sẽ tăng gần 0.0042 đơn vị.

b) Hãy tiên đoán lượng độ lún mặt đường sẽ quan trắc được khi nhiệt độ bề mặt đường là 85?
Khi nhiệt độ bề mặt đường là 85 $\Rightarrow x_0 = 85$

$$\Rightarrow \hat{y}_0 = 0.3271 + 0.0042 \cdot 85 = 0.6841.$$

Vậy nhiệt độ bề mặt đường là 85 thì độ lún mặt đường sẽ là khoảng 0.6841 đơn vị.

□

XÁC SUẤT THỐNG KÊ

ĐỀ CUỐI KỲ I NĂM HỌC 2020-2021
Thời gian làm bài: 90 phút
Đề số: 2

Câu 1 (3 điểm). Số liệu thống kê về doanh số bán hàng (Đv: triệu đồng/ngày) của một cửa hàng cho bởi bảng sau

Doanh số (triệu đồng/ngày)	Số ngày	Doanh số (triệu đồng/ngày)	Số ngày
20 − 40	5	80 − 90	15
40 − 50	10	90 − 100	10
50 − 60	20	100 − 110	8
60 − 70	25	110 − 130	3
70 − 80	25		

Giả thiết doanh số bán hàng là đại lượng ngẫu nhiên có phân phối chuẩn.

a) Ước lượng doanh số trung bình của một ngày tại cửa hàng với độ tin cậy 95%.

b) Những ngày có doanh số trên 90 triệu đồng là những ngày bán đắt hàng, Hãy ước lượng tỷ lệ những ngày bán đắt hàng ở cửa hàng này với độ tin cậy 96%.

c) Lời khẳng định: "Tỷ lệ những ngày bán đắt hàng dưới 20%" có được chấp nhận hay không, với mức ý nghĩa $\alpha = 1\%$?

✍ **Lời giải.**

Biến đổi số liệu

Doanh số (triệu đồng/ngày)	Số ngày
30	5
45	10
55	20
65	25
75	25
85	15
95	10
105	8
120	3

$n = 121$, $\bar{x} = 71.28099$, $s = 19.73289$.
Phương sai σ^2: chưa biết.

a) **Ước lượng doanh số trung bình của một ngày tại cửa hàng với độ tin cậy 95%.**

• **Độ tin cậy:** $95\% \Rightarrow \alpha = 5\%$.
$\Rightarrow t_{\alpha/2;n-1} = t_{0.025;120} = 1.98$.

• **Sai số:** $\epsilon = t_{\alpha/2;n-1} \cdot \dfrac{s}{\sqrt{n}}$
$$= 1.98 \cdot \frac{19.73289}{\sqrt{121}}$$
$$\approx 3.55192.$$

• **KTC 95% cho trung bình μ là:**

$$\begin{aligned} \bar{x} - \epsilon \leq \quad & \mu \quad \leq \bar{x} - \epsilon \\ \Leftrightarrow 71.28099 - 3.55192 \leq \quad & \mu \quad \leq 71.28099 + 3.55192 \\ \Leftrightarrow 67.72907 \leq \quad & \mu \quad \leq 74.83291 \end{aligned}$$

$\Rightarrow \mu \in [67.72907; 74.83291]$.

b) **Những ngày có doanh số trên 90 triệu đồng là những ngày bán đắt hàng, Hãy ước lượng tỷ lệ những ngày bán đắt hàng ở cửa hàng này với độ tin cậy 96%.**

Gọi Y là những ngày bán đắt hàng.

Ta có: $n = 121$, $y = 10 + 8 + 3 = 21 \Rightarrow$ tỷ lệ mẫu $\hat{p} = \dfrac{y}{n} = \dfrac{21}{121}$.

• **Độ tin cậy:** $96\% \Rightarrow \alpha = 0.04$.
$\Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.98} = 2.05375$.

• **Sai số:** $\epsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$
$$= 2.05375 \cdot \sqrt{\frac{\frac{21}{121}\left(1 - \frac{21}{121}\right)}{121}}$$
$$\approx 0.07071.$$

• **KTC 96% cho tỷ lệ p là**

$$\begin{aligned} \hat{p} - \epsilon \leq \quad & p \quad \leq \hat{p} + \epsilon \\ \Leftrightarrow \frac{21}{121} - 0.07071 \leq \quad & p \quad \leq \frac{21}{121} + 0.07071 \\ \Leftrightarrow 0.10284 \leq \quad & p \quad \leq 0.24426 \end{aligned}$$

$\Rightarrow p \in [0.10284; 0.24426]$.

c) **Lời khẳng định: "Tỷ lệ những ngày bán đắt hàng dưới 20%" có được chấp nhận hay không, với mức ý nghĩa $\alpha = 1\%$?**

Ta có: $n = 121$, $y = 21 \Rightarrow$ tỷ lệ mẫu $\hat{p} = \dfrac{y}{n} = \dfrac{21}{121}$.

• **GTKĐ:** $\begin{cases} H_0 : p \geq 20\% \\ H_1 : p < 20\% \end{cases}$: KĐ 1 phía, $p_0 = 20\% = 0.2$.

• **Mức ý nghĩa:** $\alpha = 1\% = 0.01$.

• **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{\dfrac{21}{121} - 0.2}{\sqrt{\dfrac{0.2(1-0.2)}{121}}} \approx -0.72727.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $z_0 < -z_{1-\alpha}$.

Ta có $\alpha = 0.01$

$\Rightarrow -z_{1-\alpha} = -z_{0.99} = -2.32635$.

• **So sánh và kết luận:**

Ta có: $z_0 < -z_{1-\alpha}$

$\Leftrightarrow -0.72727 < -2.32635$ (sai)

$\Rightarrow$ Chưa đủ cơ sở để bác bỏ $H_0 : p \geq 20\%$.

Kết luận: Với mức ý nghĩa 1%, tỷ lệ những ngày bán đắt hàng không dưới 20%, hay lời khẳng định không được chấp nhận.

□

Câu 2 (1,5 điểm). Giả sử người ta kiểm tra 500 thành phần máy móc do một nhà máy sản xuất và thấy có 10 thành phần bị loại bỏ. Gọi p là tỷ lệ thành phần bị loại bỏ do nhà máy sản xuất. Kiểm định giả thuyết $H_0 : p = 0{,}03$ với đối thuyết $H_1 : p < 0{,}03$, sử dụng $\alpha = 0{,}05$. Tìm $p-$giá trị.

✍ **Lời giải.**

Ta có: $n = 500$, $y = 10 \Rightarrow$ tỷ lệ mẫu $\hat{p} = \dfrac{y}{n} = \dfrac{10}{500} = 0.02$.

• **GTKĐ:** $\begin{cases} H_0 : p = 0.03 \\ H_1 : p < 0.03 \end{cases}$: KĐ 1 phía, $p_0 = 0.03$.

• **Mức ý nghĩa:** $\alpha = 0.05$.

• **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{0.02 - 0.03}{\sqrt{\dfrac{0.03(1-0.03)}{500}}} \approx -1.31081.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $z_0 < -z_{1-\alpha}$.

Ta có $\alpha = 0.05$

$\Rightarrow -z_{1-\alpha} = -z_{0.95} = -1.64485$.

• **So sánh và kết luận:**

Ta có: $z_0 < -z_{1-\alpha}$

$\Leftrightarrow -1.31081 < -1.64485$ (sai)

$\Rightarrow$ Chưa đủ cơ sở để bác bỏ $H_0 : p = 0.03$.

Kết luận: Với mức ý nghĩa 1%, tỷ lệ thành phần bị loại bỏ do nhà máy sản xuất là 0.03.

• **$p-$giá trị**

$p-$giá trị $= \Phi(z_0) = \Phi(-1.31081) \approx 0.09496$. □

Câu 3 (1,5 điểm). Vào mùa đông của đại dịch cúm, bố mẹ của 2 000 bé đã được các nhà nghiên cứu khảo sát tại một công ty dược phẩm nổi tiếng để xác định liệu thuốc mới của công ty có hiệu quả sau hai ngày hay không. Trong số 120 bé bị cúm và được cho dùng thuốc, 29 bé khỏi bệnh trong hai ngày. Trong 280 bé bị cúm nhưng không được cho dùng thuốc mới, có 56 bé khỏi bệnh trong hai ngày. Hỏi có bằng chứng ủng hộ lời tuyên bố của công ty về hiệu quả của thuốc hay không? Sử dụng $\alpha = 0{,}05$.

✍ **Lời giải.**

Gọi $Y_1:$ số bé uống thuốc mới khỏi bệnh trong hai ngày.

$Y_2:$ số bé không uống thuốc mới khỏi bệnh trong hai ngày.

Mẫu 1: $n_1 = 120;\ y_1 = 29 \Rightarrow \hat{p}_1 = \dfrac{y_1}{n_1} = \dfrac{29}{120}.$

Mẫu 2: $n_2 = 280;\ y_2 = 56 \Rightarrow \hat{p}_2 = \dfrac{y_2}{n_2} = \dfrac{56}{280} = 0.2.$

$\Rightarrow \hat{p} = \dfrac{y_1 + y_2}{n_1 + n_2} = \dfrac{29 + 56}{120 + 280} = 0.2125.$

• **GTKĐ:** $\begin{cases} H_0 : p_1 \leq p_2 \\ H_1 : p_1 > p_2 \end{cases}$: KĐ 1 phía, $p_0 = p_1 - p_2 = 0.$

• **Mức ý nghĩa:** $\alpha = 0.05.$

• **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} = \frac{\dfrac{29}{120} - 0.2 - 0}{\sqrt{0.2125(1 - 0.2125)\left(\dfrac{1}{120} + \dfrac{1}{280}\right)}} \approx 0.93352.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $z_0 > z_{1-\alpha}$.

Ta có $\alpha = 0.05$

$\Rightarrow z_{1-\alpha} = z_{0.95} = 1.64485.$

• **So sánh và kết luận:**

Ta có: $z_0 > z_{1-\alpha}$

$\Leftrightarrow 0.93352 > 1.64485$ (sai)

$\Rightarrow$ Chưa đủ cơ sở để bác bỏ $H_0 : p_1 \leq p_2$.

Kết luận: Với mức ý nghĩa 5%, tỷ lệ bé khỏi bệnh khi uống thuốc mới sau hai ngày không nhiều hơn bé không uống thuốc mới, hay thuốc mới không hiệu quả như tuyên bố của công ty. □

Câu 4 (2 điểm). Đường kính của các thanh thép được sản xuất trên hai máy đúc khác nhau đang được nghiên cứu. Hai mẫu ngẫu nhiên có cỡ mẫu $n_1 = 15$, $n_2 = 17$ được chọn có trung bình và phương sai mẫu $\bar{x}_1 = 8{,}73$; $s_1^2 = 0{,}35$ và $\bar{x}_2 = 8{,}68$; $s_2^2 = 0{,}40$. Giả sử rằng $\sigma_1^2 = \sigma_2^2$ và tổng thể có phân phối chuẩn. Có bằng chứng để khẳng định rằng hai máy sản xuất thanh thép có đường kính trung bình khác nhau không? Sử dụng $\alpha = 0{,}05$ khi đưa ra kết luận này. Tìm $p-$giá trị.

✍ **Lời giải.**

Gọi $X_1:$ Đường kính của các thanh thép được sản xuất trên máy đúc thứ nhất.

$X_2:$ Đường kính của các thanh thép được sản xuất trên máy đúc thứ hai.

Mẫu 1: $n_1 = 15;\ \bar{x}_1 = 8.73;\ s_1^2 = 0.35.$

Mẫu 2: $n_2 = 17;\ \bar{x}_2 = 8.68;\ s_2^2 = 0.40.$

Phương sai $\sigma_1^2 = \sigma_2^2$ chưa biết.

• **GTKĐ:** $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$: KĐ 2 phía, $\mu_0 = \mu_1 - \mu_2 = 0.$

• **Mức ý nghĩa:** $\alpha = 0.05.$

• **Giá trị Thống kê kiểm định:**

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(15-1)0.35 + (17-1)0.40}{15 + 17 - 2} = \frac{113}{300}.$$

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} = \frac{8.73 - 8.68 - 0}{\sqrt{\frac{113/300}{15} + \frac{113/300}{17}}} \approx 0.22998.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $|t_0| > t_{\frac{\alpha}{2};n_1+n_2-2}$.
Ta có $\alpha = 0.05$
$\Rightarrow t_{\frac{\alpha}{2};n_1+n_2-2} = t_{0.025;30} = 2.042$.
• **So sánh và kết luận:**
Ta có: $|t_0| > t_{\frac{\alpha}{2};n_1+n_2-2}$
$\Leftrightarrow |0.22998| > 2.042$
$\Leftrightarrow 0.22998 > 2.042$ (sai)
$\Rightarrow$ Chưa đủ cơ sở để bác bỏ $H_0 : \mu_1 = \mu_2$.
Kết luận: Với mức ý nghĩa 5%, đường kính trung bình của các thanh thép được sản xuất trên máy đúc thứ nhất không khác máy thứ hai.
• **p−giá trị**
Vì bậc tự do $n_1 + n_2 - 2 = 30 \geq 30$ nên ta có thể xấp xỉ p−giá trị bằng phân phối chuẩn.
p−giá trị $\approx 2\left[1 - \Phi(|z_0|)\right] = 2\left[1 - \Phi(|0.22998|)\right] = 2\left[1 - \Phi(0.22998)\right] = 2\,(1 - 0.59095) = 0.8181.$ □

Câu 5 (2 điểm). Điểm thi giữa kỳ (x) và cuối kỳ (y) của một lớp có 9 sinh viên là như sau

x	77	50	71	72	81	94	96	99	67
y	82	66	78	34	47	85	99	99	68

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

b) Dự đoán điểm bài thi cuối kỳ của một sinh viên có điểm giữa kỳ là 85.

✍ **Lời giải.**

a) **Tìm phương trình hồi quy tuyến tính đơn y theo x.**
Từ bảng dữ liệu, dùng máy tính cầm tay, ta ước lượng phương trình hồi quy tuyến tính đơn y theo x là $\hat{y} = 12.06232 + 0.77714x$.
Ý nghĩa hệ số $\hat{\beta}_1$: Ta có $\hat{\beta}_1 = 0.77714 > 0$ nên khi điểm giữa kỳ thay đổi 1 điểm thì điểm cuối kỳ sẽ thay đổi 0.77714 điểm. Cụ thể, khi điểm giữa kỳ tăng 1 điểm thì điểm cuối kỳ sẽ tăng 0.77714 điểm.

b) **Dự đoán điểm bài thi cuối kỳ của một sinh viên có điểm giữa kỳ là** 85.
Điểm giữa kỳ là 85, tức $x = 85 \Rightarrow \hat{y} = 12.06232 + 0.77714 \cdot 85 = 78.11922$ điểm.

□

XÁC SUẤT THỐNG KÊ

ĐỀ CUỐI KỲ I NĂM HỌC 2020-2021
Thời gian làm bài: 90 phút
Đề số: 3

Câu 1 (6 điểm). Đo chiều cao của 100 cây đậu (cm) sau một thời gian trồng ta có được kết quả sau

Chiều cao	5.7	7.5	8.3	9.8	10.1	10.3
Số cây	25	15	21	10	16	13

Biết chiều cao của cây đậu có phân phối chuẩn.

a) Tìm khoảng ước lượng cho chiều cao trung bình của cây đậu với độ tin cậy 95%.

b) Dựa vào bảng số liệu trên, hãy kiểm tra nhận xét rằng chiều cao trung bình của đậu là 6.9 cm với mức ý nghĩa 5%.

c) Những cây đậu có chiều cao trên 6.9 cm được gọi là xếp loại chuẩn. Tìm khoảng ước lượng cho tỷ lệ p của số cây đậu có chiều cao đạt chuẩn với độ tin cậy 95%.

d) Có nhận xét rằng tỷ lệ đậu có chiều cao đạt loại chuẩn là 0.6, dựa vào số liệu trên hãy kiểm tra nhận xét này với mức ý nghĩa 1%.

✍ **Lời giải.**

Gọi X (cm) là chiều cao của cây đậu.
Theo đề bài: $X \sim N(\mu; \sigma^2)$
σ^2: chưa biết.
Từ bảng số liệu đề bài, ta có: $n = 100$; $\bar{x} = 8.228$; $s \approx 1.7536$.

a) Tìm khoảng ước lượng cho chiều cao trung bình của cây đậu với độ tin cậy 95%.
- **Độ tin cậy:** 95% $\Rightarrow \alpha = 5\%$.
$\Rightarrow t_{\alpha/2;n-1} = t_{0.025;99} \approx z_{0.975} \approx 1.96$.
- **Sai số:** $\epsilon = t_{\alpha/2;n-1} \cdot \dfrac{s}{\sqrt{n}}$
$= 1.96 \cdot \dfrac{1.7536}{\sqrt{100}}$
≈ 0.3437.
- **KTC** 95% **cho trung bình** μ **là:**

$$\begin{aligned} \bar{x} - \epsilon \leq {} & \mu \leq \bar{x} - \epsilon \\ \Leftrightarrow 8.228 - 0.3437 \leq {} & \mu \leq 8.228 + 0.3437 \\ \Leftrightarrow 7.8843 \leq {} & \mu \leq 8.5717 \end{aligned}$$

$\Rightarrow \mu \in [7.8843; 8.5717]$.

b) Dựa vào bảng số liệu trên, hãy kiểm tra nhận xét rằng chiều cao trung bình của đậu là 6.9 cm với mức ý nghĩa 5%.
σ^2: chưa biết.
Ta có: $n = 100$; $\bar{x} = 8.228$; $s \approx 1.7536$.
- **Giả Thuyết KĐ:** $\begin{cases} H_0 : \mu = 6.9 \\ H_1 : \mu \neq 6.9 \end{cases}$: KĐ 2 phía, $\mu_0 = 6.9$.
- **Mức ý nghĩa:** $\alpha = 5\% = 0.05$.
- **Giá trị Thống kê kiểm định:**

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{8.228 - 6.9}{\frac{1.7536}{\sqrt{100}}} \approx 7.5730.$$

- **Miền bác bỏ:** Bác bỏ H_0 nếu $|t_0| > t_{\alpha/2;n-1}$.
Ta có $\alpha = 0.05$
$\Rightarrow t_{\alpha/2;n-1} = t_{0.025;99} \approx z_{0.975} \approx 1.96$.
- **So sánh và kết luận:**
Ta có: $|t_0| > t_{\alpha/2;n-1}$
$\Leftrightarrow |7.5730| > 1.96$
$\Leftrightarrow 7.5730 > 1.96$ (đúng)
$\Rightarrow$ bác bỏ $H_0 : \mu = 6.9$.
Kết luận: Với mức ý nghĩa 5% thì chiều cao trung bình của đậu là 6.9 cm.

c) Những cây đậu có chiều cao trên 6.9 cm được gọi là xếp loại chuẩn. Tìm khoảng ước lượng cho tỷ lệ p của số cây đậu có chiều cao đạt chuẩn với độ tin cậy 95%.

Gọi Y là số cây đậu có chiều cao đạt chuẩn.

Ta có: $n = 100$, $y = 15 + 21 + 10 + 16 + 13 = 75 \Rightarrow$ tỷ lệ mẫu $\hat{p} = \dfrac{y}{n} = \dfrac{75}{100} = 0.75$.

• **Độ tin cậy:** $95\% \Rightarrow \alpha = 0.05$.

$\Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.975} \approx 1.96$.

• **Sai số:** $\epsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

$$= 1.96 \cdot \sqrt{\frac{0.75\,(1-0.75)}{100}}$$

$$\approx 0.0849.$$

• **KTC** 95% **cho tỷ lệ** p **là**

$$\begin{aligned} \hat{p} - \epsilon \le \quad & p \quad \le \hat{p} + \epsilon \\ \Leftrightarrow 0.75 - 0.0849 \le \quad & p \quad \le 0.75 + 0.0849 \\ \Leftrightarrow 0.6651 \le \quad & p \quad \le 0.8349 \end{aligned}$$

$\Rightarrow p \in [0.6651; 0.8349]$.

d) Có nhận xét rằng tỷ lệ đậu có chiều cao đạt loại chuẩn là 0.6, dựa vào số liệu trên hãy kiểm tra nhận xét này với mức ý nghĩa 1%.

• **GTKĐ:** $\begin{cases} H_0 : p = 0.6 \\ H_1 : p \neq 0.6 \end{cases}$: KĐ 2 phía, $p_0 = 60\% = 0.6$.

• **Mức ý nghĩa:** $\alpha = 1\% = 0.01$.

• **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{0.75 - 0.6}{\sqrt{\dfrac{0.6(1-0.6)}{100}}} \approx 3.0619.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $|z_0| > z_{1-\alpha/2}$.

Ta có $\alpha = 0.01$

$\Rightarrow z_{1-\alpha/2} = z_{0.995} \approx 2.58$.

• **So sánh và kết luận:**

Ta có: $|z_0| > z_{1-\alpha/2}$

$\Leftrightarrow |3.0619| > 2.58$

$\Leftrightarrow 3.0619 > 2.58$ (đúng)

$\Rightarrow$ bác bỏ $H_0 : p = 0.6$.

Kết luận: Với mức ý nghĩa 1% thì tỷ lệ đậu có chiều cao đạt loại chuẩn không phải là 0.6 hay lời nhận xét trên không đúng.

□

Câu 2 (4 điểm). Một doanh nghiệp sử dụng 3 phương án xử lý hạt giống trên cùng một loại cây và nhận được kết quả như sau

Kết quả / Phương án	Số hạt nảy mầm	Số hạt không nảy mầm
Phương án 1	360	40
Phương án 2	603	97
Phương án 3	490	180

Dựa vào bảng số liệu trên, hãy kiểm tra xem các phương án xử lý có tác dụng như nhau đối với tỷ lệ nảy mầm hay không, với mức ý nghĩa 5%.

XÁC SUẤT THỐNG KÊ

ĐỀ CUỐI KỲ II NĂM HỌC 2019-2020

Thời gian làm bài: 90 phút

Đề số: 1

Câu 1 (6 điểm). Điều tra năng suất lúa trên diện tích 100 hecta trồng lúa của một vùng, ta thu được bảng số liệu sau

Năng suất (tạ/ha)	41	44	45	46	48	52	54
Diện tích (ha)	10	20	30	15	10	10	5

Giả sử năng suất lúa có phân phối chuẩn.

a) Hãy ước lượng năng suất lúa trung bình của vùng đó với độ tin cậy 97%.

b) Nếu muốn sai số ước lượng của năng suất lúa trung bình không vượt quá 0.5 tạ/ha, với độ tin cậy 98% thì cần điều tra thêm ít nhất bao nhiêu hecta lúa nữa?

c) Những thửa ruộng có năng suất từ 48 tạ/ha trở lên được xem là những thửa có năng suất cao. Hãy ước lượng tỉ lệ diện tích có năng suất cao trong vùng với độ tin cậy 96%.

d) Có tài liệu cho biết năng suất lúa trung bình là 47 tạ/ha. Giá trị này có phù hợp với mẫu quan trắc không? (Kết luận với mức ý nghĩa 4%).

✍ **Lời giải.**

Gọi X (tạ/ha) là năng suất lúa trên diện tích 1 hecta.
Theo đề bài: $X \sim N(\mu; \sigma^2)$
σ^2: chưa biết.
Từ bảng số liệu đề bài, ta có: $n = 100$; $\bar{x} = 46$; $s \approx 3.3029$.

a) Hãy ước lượng năng suất lúa trung bình của vùng đó với độ tin cậy 97%.
• **Độ tin cậy:** $97\% \Rightarrow \alpha = 3\%$.
$\Rightarrow t_{\alpha/2;n-1} = t_{0.015;99} \approx z_{0.985} \approx 2.17$.
• **Sai số:** $\epsilon = t_{\alpha/2;n-1} \cdot \dfrac{s}{\sqrt{n}}$
$= 2.17 \cdot \dfrac{3.3029}{\sqrt{100}}$
≈ 0.7167.
• **KTC** 97% **cho trung bình** μ **là:**

$$\begin{aligned} \bar{x} - \epsilon \leq \ & \mu \ \leq \bar{x} - \epsilon \\ \Leftrightarrow 46 - 0.7167 \leq \ & \mu \ \leq 46 + 0.7167 \\ \Leftrightarrow 45.2833 \leq \ & \mu \ \leq 46.7167 \end{aligned}$$

$\Rightarrow \mu \in [45.2833; 46.7167]$.

b) Nếu muốn sai số ước lượng của năng suất lúa trung bình không vượt quá 0.5 tạ/ha, với độ tin cậy 98% thì cần điều tra thêm ít nhất bao nhiêu hecta lúa nữa?
• **Độ tin cậy:** $98\% \Rightarrow \alpha = 2\%$.

$\Rightarrow z_{1-\alpha/2} = z_{0.99} \approx 2.33.$
Để $\epsilon \le 0.5$ thì

$$n \ge \left(\frac{z_{1-\alpha/2} \cdot s}{0.5}\right)^2 = \left(\frac{2.33 \cdot 3.3029}{0.5}\right)^2 \approx 236.8987$$

$\Rightarrow n \ge 237.$
Vậy nếu muốn sai số ước lượng của năng suất lúa trung bình không vượt quá 0.5 tạ/ha, với độ tin cậy 98% thì cần điều tra **thêm** ít nhất $237 - 100 = 137$ hecta lúa nữa.

c) Những thửa ruộng có năng suất từ 48 tạ/ha trở lên được xem là những thửa có năng suất cao. Hãy ước lượng tỉ lệ diện tích có năng suất cao trong vùng với độ tin cậy 96%.
Gọi Y là số thửa ruộng có năng suất cao.
Ta có: $n = 100$, $y = 10 + 10 + 5 = 25 \Rightarrow$ tỷ lệ mẫu $\hat{p} = \dfrac{y}{n} = \dfrac{25}{100} = 0.25.$

• **Độ tin cậy:** $96\% \Rightarrow \alpha = 0.06.$
$\Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.97} \approx 1.88.$

• **Sai số:**
$$\epsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.88 \cdot \sqrt{\frac{0.25\,(1-0.25)}{100}} \approx 0.0814.$$

• **KTC 96% cho tỷ lệ p là**

$$\begin{aligned} \hat{p} - \epsilon \le \quad & p \quad \le \hat{p} + \epsilon \\ \Leftrightarrow 0.25 - 0.0814 \le \quad & p \quad \le 0.25 + 0.0814 \\ \Leftrightarrow 0.1686 \le \quad & p \quad \le 0.3314 \end{aligned}$$

$\Rightarrow p \in [0.1686; 0.3314].$

d) Có tài liệu cho biết năng suất lúa trung bình là 47 tạ/ha. Giá trị này có phù hợp với mẫu quan trắc không? (Kết luận với mức ý nghĩa 4%).
σ^2: chưa biết.
Ta có: $n = 100$; $\bar{x} = 46$; $s \approx 3.3029.$

• **Giả Thuyết KĐ:** $\begin{cases} H_0 : \mu = 47 \\ H_1 : \mu \neq 47 \end{cases}$: KĐ 2 phía, $\mu_0 = 47.$

• **Mức ý nghĩa:** $\alpha = 4\% = 0.04.$

• **Giá trị Thống kê kiểm định:**

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{46 - 47}{\frac{3.3029}{\sqrt{100}}} \approx -3.0276.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $|t_0| > t_{\alpha/2;n-1}$.
Ta có $\alpha = 0.04$
$\Rightarrow t_{\alpha/2;n-1} = t_{0.02;99} \approx z_{0.98} \approx 2.05.$

• **So sánh và kết luận:**
Ta có:
$$\begin{aligned} & |t_0| > t_{\alpha/2;n-1} \\ \Leftrightarrow\ & |-3.0276| > 2.05 \\ \Leftrightarrow\ & 3.0276 > 2.05 \quad \text{(đúng)} \end{aligned}$$
$\Rightarrow$ bác bỏ $H_0 : \mu = 47.$
Kết luận: Với mức ý nghĩa 4% thì năng suất lúa trung bình là 47 tạ/ha không phù hợp với mẫu quan trắc.

□

Câu 2 (4 điểm). Một nhà máy có 3 phân xưởng cùng sản xuất một loại sản phẩm. Chất lượng sản phẩm được chia thành 3 loại. Kiểm tra, phân loại ngẫu nhiên một số sản phẩm từ lô sản phẩm của 3 phân xưởng ta có số liệu sau

Phân xưởng \ Chất lượng	Loại I	Loại II	Loại III
PX I	70	25	5
PX II	80	20	10
PX III	60	15	5

Với mức ý nghĩa 5%, chất lượng sản phẩm của 3 phân xưởng có như nhau không?

XÁC SUẤT THỐNG KÊ

ĐỀ CUỐI KỲ II NĂM HỌC 2019-2020

Thời gian làm bài: 90 phút

Đề số: 2

Câu 1. Khảo sát 500 người trưởng thành ở Việt Nam thì thấy có 24 người không biết nhóm máu của mình.

a) Ước lượng tỷ lệ người trưởng thành ở Việt Nam không biết nhóm máu của mình với độ tin cậy 95%.

b) Nếu muốn ước lượng tỷ lệ người trưởng thành ở Việt Nam không biết nhóm máu của mình với độ tin cậy 99% và sai số tối đa là 0.02 thì cần khảo sát thêm ít nhất bao nhiêu người?

✍ **Lời giải.**

a) Ước lượng tỷ lệ người trưởng thành ở Việt Nam không biết nhóm máu của mình với độ tin cậy 95%.

Gọi Y là số người trưởng thành ở Việt Nam không biết nhóm máu của mình.

Ta có: $n = 500$, $y = 24 \Rightarrow$ tỷ lệ mẫu $\hat{p} = \dfrac{y}{n} = \dfrac{24}{500} = 0.048$.

• **Độ tin cậy:** $95\% \Rightarrow \alpha = 0.05$.

$\Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.975} \approx 1.96$.

• **Sai số:** $\epsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

$$= 1.96 \cdot \sqrt{\frac{0.048\,(1-0.048)}{500}}$$

≈ 0.0187.

• **KTC** 95% **cho tỷ lệ** p **là**

$$\begin{aligned} \hat{p}-\epsilon \leq\ & p \leq \hat{p}+\epsilon \\ \Leftrightarrow 0.048-0.0187 \leq\ & p \leq 0.048+0.0187 \\ \Leftrightarrow 0.0293 \leq\ & p \leq 0.0667 \end{aligned}$$

$\Rightarrow p \in [0.0293; 0.0667]$.

b) Nếu muốn ước lượng tỷ lệ người trưởng thành ở Việt Nam không biết nhóm máu của mình với độ tin cậy 99% và sai số tối đa là 0.02 thì cần khảo sát thêm ít nhất bao nhiêu người?

• **Độ tin cậy:** $99\% \Rightarrow \alpha = 0.01$.

$\Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.995} \approx 2.58$.
Để $\epsilon \leq 0.02$ thì

$$n \geq \left(\frac{z_{1-\frac{\alpha}{2}}}{0.02}\right) \hat{p}(1-\hat{p}) = \left(\frac{2.58}{0.02}\right) \cdot 0.048(1-0.048) = 760.427136$$

$\Rightarrow n \geq 761$.
Vậy nếu muốn ước lượng tỷ lệ người trưởng thành ở Việt Nam không biết nhóm máu của mình với độ tin cậy 99% và sai số tối đa là 0.02 thì cần khảo sát **thêm** ít nhất $761 - 500 = 261$ người nữa.

□

Câu 2. Một trung tâm khám chữa bệnh tuyên bố rằng thời gian trung bình một bệnh nhân chờ khám không quá 20 phút. Một cuộc khảo sát ngẫu nhiên 35 bệnh nhân cho thấy thời gian chờ khám trung bình là 24.77 phút và độ lệch chuẩn là 7.26 phút. Giả sử thời gian chờ khám là đại lượng ngẫu nhiên có phân phối chuẩn. Dựa vào dữ liệu khảo sát, hãy kiểm tra tuyên bố của phòng khám đó có đúng không với mức ý nghĩa 3%.

✍ **Lời giải.**

Gọi X (phút) là thời gian một bệnh nhân chờ khám tại trung tâm.
Theo đề bài: $X \sim N(\mu; \sigma^2)$
σ^2: chưa biết.
Ta có: $n = 35$; $\bar{x} = 24.77$; $s \approx 7.26$.

• **Giả Thuyết KĐ:** $\begin{cases} H_0 : \mu \leq 20 \\ H_1 : \mu > 20 \end{cases}$: KĐ 1 phía, $\mu_0 = 20$.

• **Mức ý nghĩa:** $\alpha = 3\% = 0.03$.

• **Giá trị Thống kê kiểm định:**

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{24.77 - 20}{\frac{7.26}{\sqrt{35}}} \approx 3.8870.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $t_0 > t_{\alpha;n-1}$.
Ta có $\alpha = 0.03$
$\Rightarrow t_{\alpha;n-1} = t_{0.03;34} \approx z_{0.97} \approx 1.88$.

• **So sánh và kết luận:**

Ta có: $t_0 > t_{\alpha;n-1}$

$\Leftrightarrow 3.8870 > 1.88$ (đúng)

$\Rightarrow$ bác bỏ $H_0 : \mu \leq 20$.

Kết luận: Với mức ý nghĩa 3% thì thời gian trung bình một bệnh nhân chờ khám hơn 20 phút hay tuyên bố của phòng khám đó không đúng. □

Câu 3. Đường kính của các thanh thép được sản xuất trên hai máy đúc khác nhau đang được nghiên cứu. Hai mẫu ngẫu nhiên có cỡ mẫu $n_1 = 15$, $n_2 = 17$ được chọn có trung bình và phương sai mẫu là $\bar{x}_1 = 8.73$, $s_1^2 = 0.35$ và $\bar{x}_2 = 8.68$, $s_2^2 = 0.40$. Giả sử rằng $\sigma_1^2 = \sigma_2^2$ và quan trắc lấy có phân phối chuẩn. Với độ tin cậy 95%, có bằng chứng để khẳng định rằng hai máy sản xuất thanh thép có đường kính trung bình khác nhau hay không?

✍ **Lời giải.**

Gọi X_1 là đường kính của các thanh thép được sản xuất từ máy đúc thứ nhất;

X_2 là đường kính của các thanh thép được sản xuất từ máy đúc thứ hai.

Theo đề bài $X_1 \sim N(\mu_1; \sigma_1^2)$; $X_2 \sim N(\mu_2; \sigma_2^2)$
$\sigma_1^2 = \sigma_2^2$ chưa biết.
Mẫu 1: $n_1 = 15$; $\bar{x}_1 = 8.73$ và $s_1^2 = 0.35$.

Mẫu 2: $n_2 = 17$; $\bar{x}_2 = 8.68$ và $s_2^2 = 0.40$.

• **Giả thuyết kiểm định** $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$; $\Rightarrow \Delta_0 = \mu_1 - \mu_2 = 0$; $\alpha = 0.05$.

• **Giá trị thống kê kiểm định**

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(15-1)\cdot 0.35 + (17-1)\cdot 0.40}{15+17-2} = \frac{113}{300}$$

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}} = \frac{8.73 - 8.68 - 0}{\sqrt{\dfrac{\frac{113}{300}}{15} + \dfrac{\frac{113}{300}}{17}}} \approx 0.2300.$$

• **Miền bác bỏ:** Bác bỏ H_0 khi $|t_0| > t_{\alpha/2;n_1+n_2-2}$.
Ta có $\alpha = 0.05$
$\Rightarrow t_{\alpha/2;n_1+n_2-2} = t_{0.025;30} = 2.042$.

• **So sánh và kết luận:**
Ta có: $|t_0| > t_{\alpha/2;n_1+n_2-2}$
$\Leftrightarrow |0.2300| > 2.042$
$\Leftrightarrow 0.2300 > 2.042$ (sai)
$\Rightarrow$ chưa đủ cơ sở để bác bỏ $H_0 : \mu_1 = \mu_2$.
Kết luận: Với mức ý nghĩa 5% thì không có bằng chứng để khẳng định rằng hai máy sản xuất thanh thép có đường kính trung bình khác nhau. □

Câu 4. Bảng số liệu sau cung cấp thông tin về chiều cao và số nhánh của 10 cây vừng.

Chiều cao (X-cm)	10	15	20	22	30	35	40	45	50	55
Số nhánh (Y-nhánh)	12	16	20	25	35	40	45	50	52	60

a) Viết phương trình hồi quy biểu diễn mối liên hệ của Y theo X có dạng như sau: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$ và nếu ý nghĩa của hệ số $\hat{\beta}_1$.

b) Dự đoán số nhánh của một cây vừng có chiều cao 60 cm.

✍ **Lời giải.**

a) Viết phương trình hồi quy biểu diễn mối liên hệ của Y theo X có dạng như sau: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$
Từ bảng dữ liệu, sử dụng máy tính cầm tay, ta tính được $\hat{\beta}_0 \approx 0.8130$; $\hat{\beta}_1 \approx 1.0772$

$$\Rightarrow \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 0.8130 + 1.0772x.$$

Nêu ý nghĩa của hệ số $\hat{\beta}_1$.
Vì $\hat{\beta}_1 \approx 1.0772 > 0$ nên khi chiều cao của cây tăng 1 cm và các yếu tố khác không đổi thì số nhánh sẽ tăng gần 1.0772 nhánh.

b) Dự đoán số nhánh của một cây vừng có chiều cao 60 cm.
Cây vừng có chiều cao 60 cm $\Rightarrow x_0 = 60$

$$\hat{y} = 0.8130 + 1.0772 \cdot 60 = 65.445.$$

Vậy nếu một cây vừng có chiều cao 60 cm thì sẽ có khoảng 65 nhánh.

□

Câu 1. Đem cân một số trái cây vừa thu hoạch, ta được kết quả sau

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

a) Tìm khoảng ước lượng của trọng lượng trung bình của trái cây với độ tin cậy 0.99.

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

✍ **Lời giải.**

Gọi X (g) là trọng lượng của trái cây vừa thu hoạch.
Giả sử: $X \sim N(\mu; \sigma^2)$
σ^2: chưa biết.
Biến đổi số liệu

X (g)	205	215	225	235	245
Số trái	12	17	20	18	15

Ta có: $n = 82$; $\bar{x} = 225.8537$; $s \approx 13.2592$.

a) Tìm khoảng ước lượng của trọng lượng trung bình của trái cây với độ tin cậy 0.99.
• **Độ tin cậy:** $99\% \Rightarrow \alpha = 0.01$.
$\Rightarrow t_{\alpha/2;n-1} = t_{0.005;81} \approx z_{0.995} \approx 2.58$.
• **Sai số:** $\epsilon = t_{\alpha/2;n-1} \cdot \dfrac{s}{\sqrt{n}}$

$$= 2.58 \cdot \frac{13.2592}{\sqrt{82}}$$
$$\approx 3.7777.$$

• **KTC** 99% **cho trung bình** μ **là:**

$$\begin{aligned} \bar{x} - \epsilon \leq & \quad \mu \quad \leq \bar{x} - \epsilon \\ \Leftrightarrow 225.8537 - 3.7777 \leq & \quad \mu \quad \leq 225.8537 + 3.7777 \\ \Leftrightarrow 222.076 \leq & \quad \mu \quad \leq 229.6314 \end{aligned}$$

$\Rightarrow \mu \in [222.076; 229.6314]$.

b) Nếu muốn sai số ước lượng không quá $E = 2$ g ở độ tin cậy 99% thì phải quan sát ít nhất bao nhiêu trái?
• **Độ tin cậy:** $99\% \Rightarrow \alpha = 0.01$.
$\Rightarrow z_{1-\alpha/2} = z_{0.995} \approx 2.58$.
Để $\epsilon \leq 2$ thì

$$n \geq \left(\frac{z_{1-\alpha/2} \cdot s}{2}\right)^2 = \left(\frac{2.58 \cdot 13.2592}{2}\right)^2 \approx 292.5594$$

$\Rightarrow n \geq 293$.
Vậy nếu muốn sai số ước lượng không quá $E = 2$ g ở độ tin cậy 99% thì phải quan sát ít nhất 293 trái.

□

Câu 2. Một loại thuốc mới đem điều trị cho 50 người bị bệnh B, kết quả có 40 người khỏi bệnh.

a) Ước lượng tỷ lệ khỏi bệnh p nếu dùng thuốc đó điều trị với độ tin cậy 0.95.

b) Nếu khoảng ước lượng của tỉ lệ khỏi bệnh có dung sai (sai số ước lượng) $E = 0.02$ thì độ tin cậy của khoảng là bao nhiêu?

✍ **Lời giải.**

a) Ước lượng tỷ lệ khỏi bệnh p nếu dùng thuốc đó điều trị với độ tin cậy 0.95.
Gọi Y là số người uống thuốc khỏi bệnh.
Ta có: $n = 50$, $y = 40 \Rightarrow$ tỷ lệ mẫu $\hat{p} = \dfrac{y}{n} = \dfrac{40}{50} = 0.8$.

• **Độ tin cậy:** $95\% \Rightarrow \alpha = 0.05$.
$\Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.975} \approx 1.96$.

• **Sai số:** $\epsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$
$= 1.96 \cdot \sqrt{\dfrac{0.8\,(1-0.8)}{50}}$
≈ 0.1109.

• **KTC** 95% **cho tỷ lệ** p **là**

$$\begin{aligned}
\hat{p}-\epsilon \leq \quad & p \quad \leq \hat{p}+\epsilon\\
\Leftrightarrow 0.8-0.1109 \leq \quad & p \quad \leq 0.8+0.1109\\
\Leftrightarrow 0.6891 \leq \quad & p \quad \leq 0.9109
\end{aligned}$$

$\Rightarrow p \in [0.6891; 0.9109]$.

b) Nếu khoảng ước lượng của tỉ lệ khỏi bệnh có dung sai (sai số ước lượng) $E = 0.02$ thì độ tin cậy của khoảng là bao nhiêu?
Nếu $\epsilon = 0.02$ thì

$$\begin{aligned}
\epsilon &= z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\\
\Leftrightarrow 0.02 &= z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\\
\Leftrightarrow z_{1-\frac{\alpha}{2}} &= \frac{0.02}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}\\
\Leftrightarrow z_{1-\frac{\alpha}{2}} &= \frac{0.02}{\sqrt{\frac{0.8(1-0.8)}{50}}}\\
\Leftrightarrow z_{1-\frac{\alpha}{2}} &= \frac{\sqrt{2}}{4}.
\end{aligned}$$

Mà $\mathbb{P}(Z \leq z_{1-\frac{\alpha}{2}}) = 1 - \dfrac{\alpha}{2}$ do đó

$$\begin{aligned}
\mathbb{P}\left(Z \leq \frac{\sqrt{2}}{4}\right) &= 1 - \frac{\alpha}{2} \\
\Leftrightarrow \Phi\left(\frac{\sqrt{2}}{4}\right) &= 1 - \frac{\alpha}{2} \\
\Rightarrow 0.6382 &= 1 - \frac{\alpha}{2} \\
\Rightarrow \alpha &= 0.7236.
\end{aligned}$$

$\Rightarrow$ Độ tin cậy $= (1-\alpha)100\% = (1-0.7236)100\% = 27.64\%$.
Nếu khoảng ước lượng của tỉ lệ khỏi bệnh có dung sai (sai số ước lượng) $E = 0.02$ thì độ tin cậy của khoảng là 27.64%.

□

Câu 3. Một máy đóng gói các sản phẩm có khối lượng 1 kg. Nghi ngờ máy hoạt động không bình thường, người ta chọn ra một mẫu ngẫu nhiên gồm 100 sản phẩm thì thấy như sau

Khối lượng	0.95	0.97	0.99	1.01	1.03	1.05
Số gói	9	31	40	15	3	2

Với mức ý nghĩa 5%, hãy kết luận về nghi ngờ trên.

✍ **Lời giải.**

Gọi X (kg) là khối lượng của mỗi sản phẩm.
Theo đề bài: $X \sim N(\mu; \sigma^2)$
σ^2: chưa biết.
Ta có: $n = 100$; $\bar{x} = 1.1746$; $s \approx 1.3390$.

• **Giả Thuyết KĐ:** $\begin{cases} H_0 : \mu = 1 \\ H_1 : \mu \neq 1 \end{cases}$: KĐ 2 phía, $\mu_0 = 1$.

• **Mức ý nghĩa:** $\alpha = 5\% = 0.05$.

• **Giá trị Thống kê kiểm định:**

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{1.1746 - 1}{\frac{1.3390}{\sqrt{100}}} \approx 1.30396.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $|t_0| > t_{\alpha/2;n-1}$.
Ta có $\alpha = 0.05$
$\Rightarrow t_{\alpha/2;n-1} = t_{0.025;99} \approx z_{0.975} \approx 1.96$.

• **So sánh và kết luận:**

Ta có: $|t_0| > t_{\alpha/2;n-1}$

$\Leftrightarrow |1.30396| > 1.96$

$\Leftrightarrow 1.30396 > 1.96$ (sai)

$\Rightarrow$ chưa đủ cơ sở để bác bỏ $H_0 : \mu = 1$.
Kết luận: Với mức ý nghĩa 5% thì khối lượng của mỗi sản phẩm vẫn là 1 kg hay nghi ngờ máy hoạt động không bình thường là không đúng. □

Câu 4. Một bài báo trên tạp chí y khoa Anh "So sánh điều trị sỏi thận bằng phẫu thuật phẫu thuật, cắt bỏ sỏi thận, và Lithotrips sóng bổ sung," (1986, Vol. 292, pp. 879–882)] thấy rằng tác động qua da (PN) có tỷ lệ thành công trong việc loại bỏ sỏi thận của 289 trong số 350 bệnh nhân. Phương pháp truyền thống đạt hiệu quả 78%. Có bằng chứng gì cho thấy tỉ lệ thành công của PN lớn hơn so với truyền thống với mức ý nghĩa $\alpha = 1\%$? Tìm $p-$giá trị.

✍ **Lời giải.**

Gọi Y là số bệnh nhân loại bỏ sỏi thận thành công bằng phương pháp PN.

Ta có: $n = 350$, $y = 289 \Rightarrow$ tỷ lệ mẫu $\hat{p} = \dfrac{y}{n} = \dfrac{289}{350}$.

• **GTKĐ:** $\begin{cases} H_0 : p \leq 0.78 \\ H_1 : p > 0.78 \end{cases}$: KĐ 1 phía, $p_0 = 78\% = 0.78$.

• **Mức ý nghĩa:** $\alpha = 1\% = 0.01$.

• **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{\dfrac{289}{350} - 0.78}{\sqrt{\dfrac{0.78(1-0.78)}{350}}} \approx 0.9032.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $z_0 > z_{1-\alpha}$.

Ta có $\alpha = 0.01$

$\Rightarrow z_{1-\alpha} = z_{0.99} \approx 2.33$.

• **So sánh và kết luận:**

Ta có: $z_0 > z_{1-\alpha}$

$\Leftrightarrow 0.9032 > 2.33$ (sai)

$\Rightarrow$ chưa đủ cơ sở để bác bỏ $H_0 : p \leq 0.78$.

Kết luận: Với mức ý nghĩa 1% thì tỉ lệ thành công của PN không lớn hơn 0.78 hay không có bằng chứng cho thấy tỉ lệ thành công của PN lớn hơn so với truyền thống.

⋇ Tính p-giá trị:

$$p - \text{giá trị} = 1 - \Phi(z_0) = 1 - \Phi(0.9032) \approx 1 - 0.8168 = 0.1832.$$

□

Câu 5. Ô nhiễm không khí có liên quan đến việc giảm cân ở trẻ sơ sinh. Trong một nghiên cứu được công bố trên Tạp chí của Hiệp hội Y khoa Hoa Kỳ, các nhà nghiên cứu đã kiểm tra tỷ lệ trẻ sơ sinh nhẹ cân được sinh ra từ các bà mẹ tiếp xúc với liều lượng bồ hóng và tro nặng trong vụ tấn công của Trung tâm Thương mại Thế giới ngày 11/9/2001. Có 182 đứa bé sinh ra từ những bà mẹ này, 15 đứa được xếp loại có trọng lượng thấp. Trong số 2300 trẻ sinh ra trong cùng một khoảng thời gian ở New York ở một bệnh viện khác, 92 đứa được phân loại là có trọng lượng thấp. Có bằng chứng cho thấy rằng các bà mẹ tiếp xúc ô nhiễm có tỷ lệ trẻ sơ sinh nhẹ cân cao hơn không? Mức ý nghĩa $\alpha = 5\%$.

✍ **Lời giải.**

Gọi Y_1 : số trẻ sơ sinh có trọng lượng thấp được sinh từ người mẹ tiếp xúc ổ nhiễm.

Y_2 : số trẻ sơ sinh có trọng lượng thấp được sinh từ người mẹ không tiếp xúc ổ nhiễm.

Mẫu 1: $n_1 = 182$; $y_1 = 15 \Rightarrow \hat{p}_1 = \dfrac{y_1}{n_1} = \dfrac{15}{182}$.

Mẫu 2: $n_2 = 2300$; $y_2 = 92 \Rightarrow \hat{p}_2 = \dfrac{y_2}{n_2} = \dfrac{92}{2300} = 0.04$.

$\Rightarrow \hat{p} = \dfrac{y_1 + y_2}{n_1 + n_2} = \dfrac{15 + 92}{182 + 2300} = \dfrac{107}{2482}$.

• **GTKĐ:** $\begin{cases} H_0 : p_1 \leq p_2 \\ H_1 : p_1 > p_2 \end{cases}$: KĐ 1 phía, $p_0 = p_1 - p_2 = 0$.

• **Mức ý nghĩa:** $\alpha = 0.05$.

• **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{\frac{15}{182} - 0.04 - 0}{\sqrt{\frac{107}{2482}\left(1-\frac{107}{2482}\right)\left(\frac{1}{182}+\frac{1}{2300}\right)}} \approx 2.71221.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $z_0 > z_{1-\alpha}$.
Ta có $\alpha = 0.05$
$\Rightarrow z_{1-\alpha} = z_{0.95} = 1.64485$.
• **So sánh và kết luận:**
Ta có: $z_0 > z_{1-\alpha}$

$\Leftrightarrow 2.71221 > 1.64485$ (đúng)

$\Rightarrow$ bác bỏ $H_0 : p_1 \le p_2$.
Kết luận: Với mức ý nghĩa 5%, có bằng chứng cho thấy rằng các bà mẹ tiếp xúc ô nhiễm có tỷ lệ trẻ sơ sinh nhẹ cân cao hơn. □

Câu 6. Bệnh tiểu đường và béo phì là những vấn đề sức khỏe nghiêm trọng ở Hoa Kỳ và phần lớn các nước phát triển. Đo lượng mỡ cơ thể của một người là một cách để theo dõi tiến độ kiểm soát cân nặng, nhưng đo chính xác nó phải sử dụng đến thiết bị X-quang đắt tiền hoặc nhúng cơ thể xuống một hồ bơi. Thay vào đó, chỉ số khối cơ thể (BMI) thường được sử dụng làm đại diện cho mỡ cơ thể vì nó dễ đo: BMI = khối lượng (kg)/(chiều cao (m))^2 = 703 khối lượng (lb)/(chiều cao (in))^2. Trong một nghiên cứu của 250 người đàn ông tại Đại học Bingham Young, cả BMI (X) và mỡ cơ thể (Y) được đo lường. Các nhà nghiên cứu đã tìm thấy các thống kê tóm tắt sau:

$$\sum_{i=1}^{n} x_i = 6322.28; \quad \sum_{i=1}^{n} x_i^2 = 162674.18; \quad \sum_{i=1}^{n} y_i = 4757.90;$$

$$\sum_{i=1}^{n} y_i^2 = 107679.27; \quad \sum_{i=1}^{n} x_i y_i = 125471.10$$

a) Viết phương trình hồi quy biểu diễn mối liên hệ của Y theo X có dạng như sau: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$ và nếu ý nghĩa của hệ số $\hat{\beta}_1$.

b) Sử dụng đường thẳng hồi quy, hãy tiên đoán lượng mỡ cơ thể của một người đàn ông sẽ được quan trắc nếu có chỉ số BMI là 30?

XÁC SUẤT THỐNG KÊ

ĐỀ CUỐI KỲ I NĂM HỌC 2019-2020
Thời gian làm bài: 90 phút
Đề số: 1

Câu 1 (3 điểm). Người ta đo nồng độ 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

Giả sử rằng nồng độ ion Na^+ có phân phối chuẩn.

a) Tính trung bình mẫu và phương sai mẫu.

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

c) Nếu muốn sai số ước lượng trung bình không quá $E = 1$ với độ tin cậy 0,95 thì phải quan sát mẫu gồm ít nhất mấy người?

✍ **Lời giải.**

Gọi X là nồng độ ion Na^+ trên một số người.

Ta có: $n = 12$; $\bar{x} = 137.83333$; $s = 4.40729$.

a) Trung bình mẫu $\bar{x} = 137.83333$ và phương sai mẫu $s^2 = 4.40729^2 = 19.42424$.

b) **Ước lượng trung bình của tổng thể ở độ tin cậy $0{,}95$.**

• **Độ tin cậy:** $0.95 \Rightarrow \alpha = 0.05$.

$\Rightarrow t_{\alpha/2;n-1} = t_{0.025;11} = 2.201$.

• **Sai số:** $\epsilon = t_{\alpha/2;n-1} \cdot \dfrac{s}{\sqrt{n}}$

$= 2.201 \cdot \dfrac{4.40729}{\sqrt{12}}$

≈ 2.80028.

• **KTC 95% cho trung bình μ là:**

$$\begin{aligned} \bar{x} - \epsilon \leq & \quad \mu \quad \leq \bar{x} - \epsilon \\ \Leftrightarrow 137.83333 - 2.80028 \leq & \quad \mu \quad \leq 137.83333 + 2.80028 \\ \Leftrightarrow 135.03305 \leq & \quad \mu \quad \leq 140.63361 \end{aligned}$$

$\Rightarrow \mu \in [135.03305; 140.63361]$.

c) **Nếu muốn sai số ước lượng trung bình không quá $E = 1$ với độ tin cậy 0.95 thì phải quan sát mẫu gồm ít nhất mấy người?**

• **Độ tin cậy:** $0.95 \Rightarrow \alpha = 0.05$.

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

Ta có: $\epsilon = z_{1-\frac{\alpha}{2}} \cdot \dfrac{s}{\sqrt{n}}$

Để $\epsilon \leq 1 \Rightarrow n \geq \left(\dfrac{z_{1-\frac{\alpha}{2}} \cdot s}{\epsilon_0}\right)^2 = \left(\dfrac{1.96 \cdot 4.40729}{1}\right)^2 \approx 74.62003$.

$\Rightarrow n \geq 75$.

Vậy nếu muốn sai số ước lượng trung bình không quá $E = 1$ với độ tin cậy 0.95 thì phải quan sát mẫu gồm ít nhất 75 người.

□

Câu 2 (1 điểm). Một phần của các mạch tích hợp bị lỗi được tạo ra trong một quá trình quang khắc đang được nghiên cứu. Một thử nghiệm ngẫu nhiên gồm 300 mạch được thử nghiệm, cho thấy có 13 khuyết điểm. Xây dựng khoảng tin cậy 95% cho tỷ lệ mạch khuyết tật được sản xuất bởi công cụ đặc biệt này.

✍ **Lời giải.**

Gọi Y là số mạch khuyết tật được sản xuất bởi công cụ đặc biệt.

Ta có: $n = 300$; $y = 13 \Rightarrow$ tỷ lệ mẫu $\hat{p} = \dfrac{y}{n} = \dfrac{13}{300}$.

• **Độ tin cậy:** $95\% \Rightarrow \alpha = 0.05$.

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

• **Sai số:** $\epsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

$= 1.96 \cdot \sqrt{\dfrac{\frac{13}{300}\left(1 - \frac{13}{300}\right)}{300}}$

≈ 0.02304.

• **KTC** 95% **cho tỷ lệ** p **là**

$$\begin{aligned} \hat{p}-\epsilon \leq \quad & p \quad \leq \hat{p}+\epsilon \\ \Leftrightarrow \frac{13}{300}-0.02304 \leq \quad & p \quad \leq \frac{13}{300}+0.02304 \\ \Leftrightarrow 0.02029 \leq \quad & p \quad \leq 0.06637 \end{aligned}$$

$\Rightarrow p \in [0.02029; 0.06637]$. □

Câu 3 (1 điểm). Một hệ thống tên lửa phản lực sử dụng động cơ đẩy nhiên liệu rắn. Tốc độ cháy của nhiên liệu rắn là một đặc trưng quan trọng của động cơ. Thông số kĩ thuật yêu cầu tốc độ cháy trung bình của thanh nhiên liệu là 50 cm/s. Các kĩ sư biết rằng độ lệch chuẩn của tốc độ cháy là 2 cm/s. Những kĩ sư kiểm nghiệm xác định xác suất của sai lầm loại I hoặc mức ý nghĩa $\alpha = 0{,}05$ và chọn cỡ mẫu là $n = 25$ với trung bình mẫu tốc độ cháy là $\bar{x} = 51{,}3$ cm/s. Kết luận rút ra như thế nào?

✍ **Lời giải.**

Gọi X (cm/s) là tốc độ cháy của nhiên liệu rắn.
$n = 25$; $\bar{x} = 51.3$.
$\sigma = 2$: đã biết.

• **Giả Thuyết KĐ:** $\begin{cases} H_0 : \mu = 50 \\ H_1 : \mu \neq 50 \end{cases}$: KĐ 2 phía, $\mu_0 = 50$.

• **Mức ý nghĩa:** $\alpha = 0.05$.

• **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{51.3-50}{\frac{2}{\sqrt{25}}} \approx 3.25.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $|z_0| > z_{1-\frac{\alpha}{2}}$.

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

• **So sánh và kết luận:**

Ta có: $\quad |z_0| > z_{1-\frac{\alpha}{2}}$

$\Leftrightarrow |3.25| > 1.96$

$\Leftrightarrow 3.25 > 1.96 \quad$ (đúng)

$\Rightarrow$ Bác bỏ $H_0 : \mu = 50$.

Kết luận: Với mức ý nghĩa 5%, tốc độ cháy trung bình của nhiên liệu rắn không phải là 50 cm/s. □

Câu 4 (2 điểm). Hai loại nhựa phù hợp để sử dụng cho một nhà sản xuất linh kiện điện tử. Sức chịu phá hủy của loại nhựa này là quan trọng và được giả sử có phân phối chuẩn. Được biết, $\sigma_1 = \sigma_2 = 1$ psi. Từ một mẫu ngẫu nhiên có kích thước $n_1 = 10$ và $n_2 = 12$, ta có được $\bar{x}_1 = 162.5$ và $\bar{x}_2 = 155.0$. Công ty sẽ không áp dụng nhựa loại 1 trừ khi sức chịu phá hủy trung bình của nó vượt quá nhựa loại 2 ít nhất 10 psi.

Trên cơ sở thông tin đó, ta có nên sử dụng nhựa loại 1? Sử dụng $\alpha = 0.05$ đưa ra câu trả lời.

✍ **Lời giải.**

Gọi X_1 : sức chịu phá hủy của nhựa loại 1.

X_2 : sức chịu phá hủy của nhựa loại 2.

Mẫu 1: $n_1 = 10$; $\bar{x}_1 = 162.5$.
Mẫu 2: $n_2 = 12$; $\bar{x}_2 = 155.0$.
$\sigma_1 = \sigma_2 = 1$ đã biết.

• **Giả Thuyết KĐ:** $\begin{cases} H_0 : \mu_1 - \mu_2 \geq 10 \\ H_1 : \mu_1 - \mu_2 < 10 \end{cases}$: KĐ 1 phía, $\mu_0 = \mu_1 - \mu_2 = 10$.

• **Mức ý nghĩa:** $\alpha = 0.05$.
• **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \frac{162.5 - 155.0 - 10}{\sqrt{\dfrac{1^2}{10} + \dfrac{1^2}{12}}} \approx -5.83874.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $z_0 < -z_{1-\alpha}$.
Ta có $\alpha = 0.05$
$\Rightarrow -z_{1-\alpha} = -z_{0.95} = -1.64485$.
• **So sánh và kết luận:**
Ta có: $\quad z_0 < -z_{1-\alpha}$

$\Leftrightarrow -5.83874 < -1.64485 \quad$ (đúng)

$\Rightarrow$ Bác bỏ $H_0 : \mu_1 - \mu_2 \geq 10$.
Kết luận: Với mức ý nghĩa 5%, sức chịu phá hủy trung bình của nhựa loại 1 không vượt quá nhựa loại 2 ít nhất 10 psi, hay công ty sẽ không áp dụng nhựa loại 1. □

Câu 5 (1 điểm). Trong một nghiên cứu để ước tính tỷ lệ cư dân trong một thành phố nào đó và các vùng ngoại ô của nó ủng hộ việc xây dựng nhà máy năng lượng hạt nhân, người ta thấy rằng 63 trong 100 cư dân thành thị ủng hộ việc xây dựng trong khi chỉ 59 trong 125 cư dân ngoại ô là ủng hộ. Có sự khác biệt có ý nghĩa nào giữa tỷ lệ cư dân thành thị và ngoại ô ủng hộ việc xây dựng nhà máy hạt nhân hay không? Sử dụng $p-$giá trị.

✍ **Lời giải.**
Gọi Y_1 : số cư dân thành thị ủng hộ việc xây dựng nhà máy năng lượng hạt nhân.

Y_2 : số cư dân nông thôn ủng hộ việc xây dựng nhà máy năng lượng hạt nhân.

Mẫu 1: $n_1 = 100$; $y_1 = 63 \Rightarrow \hat{p}_1 = \dfrac{y_1}{n_1} = \dfrac{63}{100} = 0.63$.

Mẫu 2: $n_2 = 125$; $y_2 = 59 \Rightarrow \hat{p}_2 = \dfrac{y_2}{n_2} = \dfrac{59}{125} = 0.472$.

$\Rightarrow \hat{p} = \dfrac{y_1 + y_2}{n_1 + n_2} = \dfrac{63 + 59}{100 + 125} = \dfrac{122}{225}$.

• **GTKĐ:** $\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{cases}$: $\quad$ KĐ 2 phía, $p_0 = p_1 - p_2 = 0$.

• **Mức ý nghĩa:** Giả sử $\alpha = 0.05$.
• **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\hat{p}(1 - \hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} = \frac{0.63 - 0.472 - 0}{\sqrt{\dfrac{122}{225}\left(1 - \dfrac{122}{225}\right)\left(\dfrac{1}{100} + \dfrac{1}{125}\right)}} \approx 2.36377.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $|z_0| > z_{1-\frac{\alpha}{2}}$.
Ta có $\alpha = 0.05$
$\Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$.
• **So sánh và kết luận:**
Ta có: $\quad |z_0| > z_{1-\frac{\alpha}{2}}$

$\Leftrightarrow |2.36377| > 1.96$

$\Leftrightarrow 2.36377 > 1.96 \quad$ (đúng)

$\Rightarrow$ Bác bỏ $H_0 : p_1 = p_2$.
Kết luận: Với mức ý nghĩa 5%, không có sự khác biệt có ý nghĩa nào giữa tỷ lệ cư dân thành thị và ngoại ô ủng hộ việc xây dựng nhà máy hạt nhân.
• **Sử dụng $p-$giá trị**

$p-$giá trị $= 2\,[1 - \Phi(|z_0|)] = 2\,[1 - \Phi(|2.36377|)] = 2\,[1 - \Phi(2.36377)] = 2\,[1 - 0.99095] = 0.0181$.
Ta có: $p-$giá trị $= 0.0181 < 0.05 = \alpha$
$\Rightarrow$ Bác bỏ $H_0 : p_1 = p_2$.
Kết luận: Với mức ý nghĩa 5%, không có sự khác biệt có ý nghĩa nào giữa tỷ lệ cư dân thành thị và ngoại ô ủng hộ việc xây dựng nhà máy hạt nhân.
□

Câu 6 (2 điểm). Bệnh tiểu đường và béo phì là những vấn đề sức khỏe nghiêm trọng ở Hoa Kỳ và phần lớn các nước phát triển. Đo lượng mỡ cơ thể của một người là một cách để theo dõi tiến độ kiểm soát cân nặng, nhưng đo chính xác nó phải sử dụng đến thiết bị X-quang đắt tiền hoặc nhúng cơ thể xuống một hồ bơi. Thay vào đó, chỉ số khối cơ thể (BMI) thường được sử dụng làm đại diện cho mỡ cơ thể vì nó dễ đo: BMI = khối lượng (kg)/(chiều cao (m))^2 = 703 khối lượng (lb)/(chiều cao (in))^2. Trong một nghiên cứu của 250 người đàn ông tại Đại học Bingham Young, cả BMI (X) và mỡ cơ thể (Y) được đo lường. Các nhà nghiên cứu đã tìm thấy các thống kê tóm tắt sau:

$$\sum_{i=1}^{n} x_i = 6322.28; \quad \sum_{i=1}^{n} x_i^2 = 162674.18; \quad \sum_{i=1}^{n} y_i = 4757.90;$$

$$\sum_{i=1}^{n} y_i^2 = 107679.27; \quad \sum_{i=1}^{n} x_i y_i = 125471.10$$

a) Tính những ước lượng bình phương tối thiểu cho hệ số góc và tung độ góc. Vẽ đường thẳng hồi quy.

b) Sử dụng đường thẳng hồi quy, hãy tiên đoán lượng mỡ cơ thể của một người đàn ông sẽ được quan trắc nếu có chỉ số BMI là 30?

✍ Lời giải.
Ta có

$$n = 250; \quad \sum_{i=1}^{n} x_i = 6322.28; \quad \sum_{i=1}^{n} x_i^2 = 162674.18; \quad \sum_{i=1}^{n} y_i = 4757.90;$$

$$\sum_{i=1}^{n} y_i^2 = 107679.27; \quad \sum_{i=1}^{n} x_i y_i = 125471.10$$

a) Tính những ước lượng bình phương tối thiểu cho hệ số góc và tung độ góc.

- $S_{xy} = \sum_{i=1}^{n} x_i y_i - \dfrac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n} = 125471.10 - \dfrac{6322.28 \cdot 4757.90}{250} = 5147.99595.$
- $S_{xx} = \sum_{i=1}^{n} x_i^2 - \dfrac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = 162674.18 - \dfrac{6322.28^2}{250} = 2789.28241.$

$\Rightarrow \hat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}} = \dfrac{5147.99595}{2789.28241} \approx 1.84563.$

- $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i = \dfrac{1}{250} \cdot 6322.28 = 25.28912$
- $\bar{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i = \dfrac{1}{250} \cdot 4757.90 = 19.0316$

$\Rightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 19.0316 - 25.28912 \cdot 1.84563 \approx -27.64276.$
$\Rightarrow \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = -27.64276 + 1.84563x.$
Vậy $\hat{y} = -27.64276 + 1.84563x$

b) Sử dụng đường thẳng hồi quy, hãy tiên đoán lượng mỡ cơ thể của một người đàn ông sẽ được quan trắc nếu có chỉ số BMI là 30?
Chỉ số BMI là 30, tức $x = 30 \Rightarrow \hat{y} = -27.64276 + 1.84563 \cdot 30 = 27.72614$.

□

XÁC SUẤT THỐNG KÊ

ĐỀ CUỐI KỲ I NĂM HỌC 2019-2020
Thời gian làm bài: 90 phút
Đề số: 2

Câu 1 (3 điểm). Quan sát tuổi thọ X (giờ) của một số bóng đèn do xí nghiệp A sản xuất, ta ghi nhận

X	1000	1100	1200	1300	1400	1500	1600	1700	1800
N	10	14	16	17	18	16	16	12	9

a) Tính trung bình mẫu và độ lệch tiêu chuẩn mẫu.

b) Ước lượng tuổi thọ trung bình của bóng đèn ở độ tin cậy 0,95.

c) Nếu muốn sai số ước lượng trung bình không quá $E = 30$ với độ tin cậy 0,95 thì phải quan sát mẫu gồm ít nhất mấy bóng đèn?

✍ Lời giải.
Ta có: $n = 128$; $\bar{x} = 1391.40625$; $s = 234.44602$.

a) Trung bình mẫu $\bar{x} = 1391.40625$; độ lệch tiêu chuẩn mẫu $s = 234.44602$.

b) Ước lượng tuổi thọ trung bình của bóng đèn ở độ tin cậy 0,95.
• **Độ tin cậy:** $0.95 \Rightarrow \alpha = 0.05$.
$\Rightarrow t_{\alpha/2;n-1} = t_{0.025;127} = 1.96$.
• **Sai số:** $\epsilon = t_{\alpha/2;n-1} \cdot \dfrac{s}{\sqrt{n}}$

$$= 1.96 \cdot \frac{234.44602}{\sqrt{128}}$$
$$\approx 40.61570.$$

• **KTC 95% cho trung bình μ là:**

$$\begin{aligned} \bar{x} - \epsilon \leq \ & \mu \leq \bar{x} - \epsilon \\ \Leftrightarrow 1391.40625 - 40.61570 \leq \ & \mu \leq 1391.40625 + 40.61570 \\ \Leftrightarrow 1350.79055 \leq \ & \mu \leq 1432.02195 \end{aligned}$$

$\Rightarrow \mu \in [1350.79055; 1432.02195]$.

c) Nếu muốn sai số ước lượng trung bình không quá $E = 30$ với độ tin cậy 0,95 thì phải quan sát mẫu gồm ít nhất mấy bóng đèn?
• **Độ tin cậy:** $0.95 \Rightarrow \alpha = 0.05$.
$\Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$.
Ta có: $\epsilon = z_{1-\frac{\alpha}{2}} \cdot \dfrac{s}{\sqrt{n}}$
Để $\epsilon \leq 30 \Rightarrow n \geq \left(\dfrac{z_{1-\frac{\alpha}{2}} \cdot s}{\epsilon}\right)^2 = \left(\dfrac{1.96 \cdot 234.44602}{30}\right)^2 \approx 234.61478$.
$\Rightarrow n \geq 235$.
Vậy nếu muốn sai số ước lượng trung bình không quá $E = 30$ với độ tin cậy 0,95 thì phải quan sát mẫu gồm ít nhất 235 bóng đèn.

□

Câu 2 (1 điểm). Bộ Giao thông Vận tải muốn khảo sát người dân để xác định tỷ lệ người dân muốn tăng giới hạn tốc độ từ 40 km/giờ lên 50 km/giờ. Họ cần khảo sát bao nhiêu người nếu họ muốn tự tin rằng sai số ước lượng không quá 0,05 với độ tin cậy 99%?

✍ **Lời giải.**

• **Độ tin cậy:** $99\% \Rightarrow \alpha = 0.01$.

$\Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.995} = 2.57583$.

Ta có: $\epsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

Để $\epsilon \leq 0.05 \Rightarrow n \geq \left(\dfrac{z_{1-\frac{\alpha}{2}}}{\epsilon}\right)^2 \cdot 0.25 = \left(\dfrac{1.96}{0.05}\right)^2 \cdot 0.25 \approx 384.16$.

$\Rightarrow n \geq 385$.

Vậy họ cần khảo sát 385 người nếu họ muốn tự tin rằng sai số ước lượng không quá 0,05 với độ tin cậy 99%. □

Câu 3 (1 điểm). Hàm lượng natri của hai mươi hộp bắp hữu cơ 300 gram được xác định. Dữ liệu (tính bằng miligam) như sau:

131.15	130.69	130.91	129.54	129.64	128.77	130.72	128.33	128.24	129.65
130.14	129.29	128.71	129.00	129.39	130.42	129.53	130.12	129.78	130.92

Giả sử rằng hàm lượng natri có phân phối chuẩn. Bạn hãy kiểm định xem giá trị trung bình có khác 130 milligram với $\alpha = 0.05$?

✍ **Lời giải.**

Gọi X (milligram) là hàm lượng natri của mỗi hộp bắp hữu cơ 300 gram.

Ta có: $n = 20$; $\bar{x} = 129.747$; $s = 0.87643$.

Phương sai σ^2 : chưa biết.

• **Giả Thuyết KĐ:** $\begin{cases} H_0 : \mu = 130 \\ H_1 : \mu \neq 130 \end{cases}$: KĐ 2 phía, $\mu_0 = 130$.

• **Mức ý nghĩa:** $\alpha = 0.05$.

• **Giá trị Thống kê kiểm định:**

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{129.747 - 130}{\frac{0.87643}{\sqrt{20}}} \approx -1.29098.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $|t_0| > t_{\frac{\alpha}{2};n-1}$.

Ta có $\alpha = 0.05$

$\Rightarrow t_{\frac{\alpha}{2};n-1} = t_{0.025;19} = 2.093$.

• **So sánh và kết luận:**

Ta có: $|t_0| > t_{\frac{\alpha}{2};n-1}$

$\Leftrightarrow |-1.29098| > 2.093$

$\Leftrightarrow 1.29098 > 2.093$ (sai)

$\Rightarrow$ Chưa đủ cơ sở để bác bỏ $H_0 : \mu = 130$.

Kết luận: Với mức ý nghĩa 5%, hàm lượng natri trung bình của mỗi hộp bắp hữu cơ 300 gram là 130 milligram. □

Câu 4 (1 điểm). Hai loại máy ép phun khác nhau được sử dụng để tạo thành các bộ phận bằng nhựa. Một phần được coi là khiếm khuyết nếu nó bị co rút quá mức hoặc bị đổi màu. Hai mẫu ngẫu nhiên, mỗi mẫu có kích thước 300, được chọn thì có 15 bộ phận bị lỗi được tìm thấy trong mẫu từ máy 1 và 8 bộ phận bị lỗi được tìm thấy trong mẫu từ máy 2. Có hợp lý để kết luận rằng cả hai máy sản xuất cùng một tỉ lệ các bộ phận lỗi, sử dụng $\alpha = 0.05$?

✍ **Lời giải.**

Gọi $Y_1:$ số bộ phận bị lỗi trong mẫu từ máy 1.

$Y_2:$ số bộ phận bị lỗi trong mẫu từ máy 2.

Mẫu 1: $n_1 = 300$; $y_1 = 15 \Rightarrow \hat{p}_1 = \dfrac{y_1}{n_1} = \dfrac{15}{300} = 0.05$.

Mẫu 2: $n_2 = 300$; $y_2 = 8 \Rightarrow \hat{p}_2 = \dfrac{y_2}{n_2} = \dfrac{8}{300} = \dfrac{2}{75}$.

$\Rightarrow \hat{p} = \dfrac{y_1 + y_2}{n_1 + n_2} = \dfrac{15+8}{300+300} = \dfrac{23}{600}$.

• **GTKĐ:** $\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{cases}$: KĐ 2 phía, $p_0 = p_1 - p_2 = 0$.

• **Mức ý nghĩa:** $\alpha = 0.05$.

• **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} = \frac{0.05 - \dfrac{2}{75} - 0}{\sqrt{\dfrac{23}{600}\left(1 - \dfrac{23}{600}\right)\left(\dfrac{1}{300} + \dfrac{1}{300}\right)}} \approx 1.48841.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $|z_0| > z_{1-\frac{\alpha}{2}}$.

Ta có $\alpha = 0.05$

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

• **So sánh và kết luận:**

Ta có: $|z_0| > z_{1-\frac{\alpha}{2}}$

$\Leftrightarrow |1.48841| > 1.96$

$\Leftrightarrow 1.48841 > 1.96$ (sai)

$\Rightarrow$ Chưa đủ cơ sở để bác bỏ $H_0 : p_1 = p_2$.

Kết luận: Với mức ý nghĩa 5%, cả hai máy sản xuất cùng một tỉ lệ các bộ phận lỗi. □

Câu 5 (2 điểm). Hai công thức khác nhau của nhiên liệu động cơ ôxy hóa đang được thử nghiệm để nghiên cứu số octane của chúng. Phương sai chỉ số octane của công thức thứ nhất $\sigma_1^2 = 1.5$ và công thức thứ hai $\sigma_2^2 = 1.2$. Hai mẫu ngẫu nhiên có cỡ mẫu $n_1 = 15$ và $n_2 = 20$ được nghiên cứu có chỉ số octane trung bình lần lượt là $\bar{x}_1 = 89.6$ và $\bar{x}_2 = 92.5$. Với giả sử có phân phối chuẩn.
Nếu công thức 2 tạo ra một số octane cao hơn so với công thức 1, nhà sản xuất muốn phát hiện nó. Xây dựng và kiểm định giả thuyết thích hợp sử dụng $\alpha = 0.05$ và tính $p-$value.

✍ **Lời giải.**

Gọi $X_1:$ số octane được tạo ra từ công thức 1.

$X_2:$ số octane được tạo ra từ công thức 2.

Mẫu 1: $n_1 = 15$; $\bar{x}_1 = 89.6$; $\sigma_1^2 = 1.5$

Mẫu 2: $n_2 = 20$; $\bar{x}_2 = 92.5$; $\sigma_2^2 = 1.2$

Phương sai $\sigma_1^2; \sigma_2^2$: đã biết.

• **Giả thuyết KĐ:** $\begin{cases} H_0 : \mu_1 \geq \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases}$: KĐ 1 phía, $\mu_0 = \mu_1 - \mu_2 = 0$.

• **Mức ý nghĩa:** $\alpha = 0.05$.

• **Giá trị Thống kê kiểm định:**

$$z_0 = \frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \frac{89.6 - 92.5 - 0}{\sqrt{\dfrac{1.5^2}{15} + \dfrac{1.2^2}{20}}} \approx -6.15491.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $z_0 < -z_{1-\alpha}$.

Ta có $\alpha = 0.05$
$\Rightarrow -z_{1-\alpha} = -z_{0.95} = -1.64485.$
• So sánh và kết luận:
Ta có: $$z_0 < -z_{1-\alpha}$$
$$\Leftrightarrow -6.15491 < -1.64485 \quad \text{(đúng)}$$
$\Rightarrow$ Bác bỏ $H_0 : \mu_1 \geq \mu_2$.
Kết luận: Với mức ý nghĩa 5%, số octane trung bình được tạo ra từ công thức 2 cao hơn công thức 1.
• p−giá trị
p−giá trị $= \Phi(z_0) = \Phi(-6.15491) \approx 3.771 \cdot 10^{-10} \approx 0.$ □

Câu 6 (2 điểm). Một bài báo trong Nghiên cứu Bê tông "Đặc tính bề mặt gần bê tông: Tính thấm nội tại" (1989, Tập 41) trình bày dữ liệu về cường độ nén x và độ thấm nội tại y của các hỗn hợp bê tông và phương pháp xử lý khác nhau. Số liệu được tóm tắt như sau:

$$n = 14; \quad \sum_{i=1}^{n} y_i = 572; \quad \sum_{i=1}^{n} y_i^2 = 23530; \quad \sum_{i=1}^{n} x_i = 43;$$
$$\sum_{i=1}^{n} x_i^2 = 157.42; \quad \sum_{i=1}^{n} x_i y_i = 1697.80$$

a) Tính những ước lượng bình phương tối thiểu cho hệ số góc và tung độ góc. Vẽ đồ thị của đường thẳng hồi quy.

b) Sử dụng đường thẳng hồi quy, hãy tiên đoán lượng độ thấm nội sẽ quan trắc được khi cường độ nén là $x = 4.3$?

✍ Lời giải.

Ta có

$$n = 14; \quad \sum_{i=1}^{n} y_i = 572; \quad \sum_{i=1}^{n} y_i^2 = 23530; \quad \sum_{i=1}^{n} x_i = 43;$$
$$\sum_{i=1}^{n} x_i^2 = 157.42; \quad \sum_{i=1}^{n} x_i y_i = 1697.80$$

a) Tính những ước lượng bình phương tối thiểu cho hệ số góc và tung độ góc.

- $S_{xy} = \sum_{i=1}^{n} x_i y_i - \dfrac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n} = 1697.80 - \dfrac{43 \cdot 572}{14} = -\dfrac{2067}{35}.$
- $S_{xx} = \sum_{i=1}^{n} x_i^2 - \dfrac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = 157.42 - \dfrac{43^2}{14} = \dfrac{4436}{175}.$

$$\Rightarrow \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{-\dfrac{2067}{35}}{\dfrac{4436}{175}} \approx -2.32980.$$

- $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i = \dfrac{1}{14} \cdot 43 = \dfrac{43}{14}$
- $\bar{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i = \dfrac{1}{14} \cdot 572 = \dfrac{286}{7}$

$\Rightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = \dfrac{286}{7} - \dfrac{43}{14} \cdot (-2.32980) \approx 48.01296.$
$\Rightarrow \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 48.01296 - 2.32980x.$
Vậy $\hat{y} = 48.01296 - 2.32980x.$

b) **Sử dụng đường thẳng hồi quy, hãy tiên đoán lượng độ thấm nội sẽ quan trắc được khi cường độ nén là $x = 4.3$?**
Cường độ nén là $x = 4.3 \Rightarrow \hat{y} = 48.01296 - 2.32980 \cdot 4.3 = 37.99482.$

□

XÁC SUẤT THỐNG KÊ

ĐỀ CUỐI KỲ I NĂM HỌC 2018-2019
Thời gian làm bài: 90 phút
Đề số: 1

Câu 1 (7 điểm). Gọi X (giờ) là thời gian tự học hàng ngày của sinh viên, khảo sát 120 sinh viên trường Đại học KHXHNV. Kết quả cho bởi bảng sau:

Thời gian tự học (giờ)	1	2	3	4	5	6	7	8
Số sinh viên	13	18	14	23	15	16	17	4

Giả sử thời gian tự học của sinh viên có phân phối chuẩn.

a) Ước lượng thời gian tự học trung bình của sinh viên trường KHXHNV với độ tin cậy 98%. (1.5đ)

b) Với độ tin cậy 95%, ước lượng tỉ lệ những sinh viên có thời gian tự học trên 5 giờ mỗi ngày. (1.5đ)

c) Khảo sát thời gian tự học của 90 sinh viên trường Đại học Kinh tế

Thời gian tự học (giờ)	1	2	3	4	5	6
Số sinh viên	7	8	17	24	20	14

Có ý kiến cho rằng thời gian tự học của sinh viên trường KHXHNV lớn hơn sinh viên trường Kinh tế. Với mức ý nghĩa 5%, hãy kiểm định ý kiến trên. (2đ)

d) So sánh tỷ lệ những sinh viên có thời gian tự học trên 5 giờ mỗi ngày giữa hai trường KHXHNV và Kinh tế. ($\alpha = 1\%$) (2đ)

✍ **Lời giải.**

a) Ước lượng thời gian tự học trung bình của sinh viên trường KHXHNV với độ tin cậy 98%.
Gọi X_1 là thời gian tự học hàng ngày của sinh viên trường Đại học KHXHNV. (đv: giờ)
Ta có: $n_1 = 120$; $\bar{x}_1 = 4.2083$; $s_1 = 2.0370$.
Phương sai σ^2 chưa biết.
• **Độ tin cậy:** $98\% \Rightarrow \alpha = 0{,}02$
$\Rightarrow t_{\alpha/2;n-1} = t_{0.01;119} = 2.326$
• **Sai số:** $\epsilon = t_{\alpha/2;n-1} \cdot \dfrac{s_1}{\sqrt{n_1}}$
$= 2.326 \cdot \dfrac{2.0370}{\sqrt{120}}$
$\approx 0.43252.$
• **Khoảng tin cậy** 98% **cho** μ **là:**

$$\begin{aligned}\bar{x} - \epsilon \le &\ \mu \le \bar{x} + \epsilon\\ \Leftrightarrow 4.2083 - 0.43252 \le &\ \mu \le 4.2083 + 0.43252\\ \Leftrightarrow 3.77578 \le &\ \mu \le 4.64082\end{aligned}$$

$\Rightarrow \mu \in [3.77578; 4.64082]$.

b) Với độ tin cậy 95%, ước lượng tỉ lệ những sinh viên có thời gian tự học trên 5 giờ mỗi ngày.
Gọi Y_1 là số sinh viên Đại học KHXHNV có thời gian tự học trên 5 giờ mỗi ngày.
Ta có $y_1 = 16 + 17 + 4 = 37$; $n_1 = 120 \Rightarrow \hat{p} = \dfrac{y_1}{n_1} = \dfrac{37}{120}$.

- **Độ tin cậy:** $95\% \Rightarrow \alpha = 0.05$
$\Rightarrow z_{1-\alpha/2} = z_{0.975} = 1.96$
- **Sai số:** $\epsilon = z_{1-\alpha/2}.\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} = 1.96.\sqrt{\dfrac{\dfrac{37}{120}.\left(1-\dfrac{37}{120}\right)}{120}} = 0.0826$
- **Khoảng tin cậy** 95% **cho tỷ lệ** p **là:**

$$\begin{aligned} \hat{p} - \epsilon \leq \ & p \ \leq \hat{p} + \epsilon \\ \Leftrightarrow \frac{37}{120} - 0.0826 \leq \ & p \ \leq \frac{37}{120} + 0.0826 \\ \Leftrightarrow 0.2257 \leq \ & p \ \leq 0.3909 \end{aligned}$$

$\Rightarrow p \in [0.2257; 0.3909]$.

c) Gọi X_2 là thời gian tự học hàng ngày của sinh viên trường Đại học Kinh tế. (đv: giờ)
Ta có: $n_2 = 90$; $\bar{x}_2 = 3.9333$; $s_2 = 1.4519$
Mẫu 1: $n_1 = 120$; $\bar{x}_1 = 4.2083$ và $s_1^2 = 4.1495$.
Mẫu 2: $n_2 = 90$; $\bar{x}_2 = 3.9333$ và $s_2^2 = 2.1079$.
Mọi thông tin về phương sai là chưa biết $\Rightarrow$ kiểm định phương sai.
⋇ **Kiểm định phương sai**

- Giả thuyết kiểm định: $\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$: KĐ 2 phía.
- Mức ý nghĩa: $\alpha = 0{,}05$.
- Giá trị thống kê kiểm định:

$$f_0 = \frac{s_1^2}{s_2^2} = \frac{4.1495}{2.1079} = 1.96855.$$

Với $\alpha = 0.05 \Rightarrow f_{\alpha/2;n_1-1;n_2-1} = f_{0.025;119;89} = 1.51$.
Ta có: $f_0 = 1.96855 > 1.51$ (đúng)
$\Rightarrow$ bác bỏ $H_0 : \sigma_1^2 = \sigma_2^2$.
Với mức ý nghĩa 5%: $\sigma_1^2 \neq \sigma_2^2$.

- **Giả thuyết kiểm định:** $\begin{cases} H_0 : \mu_1 \leq \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases}$; $\Rightarrow \mu_1 - \mu_2 = 0$.
- **Mức ý nghĩa:** $\alpha = 0{,}05$.
- **Giá trị thống kê kiểm định:**

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{4.2083 - 3.9333 - 0}{\sqrt{\dfrac{4.1495}{120} + \dfrac{2.1079}{90}}} \approx 1.1419.$$

- **Miền bác bỏ:** Bác bỏ H_0 khi $t_0 > t_{\alpha;v}$ với

$$v = \left\lfloor \frac{[(S_1^2/n_1) + (S_2^2/n_2)]^2}{\dfrac{(S_1^2/n_1)^2}{n_1 - 1} + \dfrac{(S_2^2/n_2)^2}{n_2 - 1}} \right\rfloor = \left\lfloor \frac{[(4.1495/120) + (2.1079/90)]^2}{\dfrac{(4.1495/120)^2}{119} + \dfrac{(2.1079/90)^2}{89}} \right\rfloor \approx \lfloor 207.50872 \rfloor = 207.$$

$\Rightarrow t_{\alpha;v} = t_{0.05;207} = 1.645.$

• **So sánh và kết luận:**

Ta có: $t_0 > t_{\alpha;v}$

$\Leftrightarrow 1.1419 > 1.645 \quad$ (sai)

$\Rightarrow$ Chưa đủ cơ sở để bác bỏ $H_0 : \mu_1 \leq \mu_2$.

Kết luận: Với mức ý nghĩa 5% thì thời gian tự học trung bình của sinh viên trường Đại học KHXHNV không lớn hơn sinh viên trường Đại học Kinh tế.

d) So sánh tỷ lệ những sinh viên có thời gian tự học trên 5 giờ mỗi ngày giữa hai trường KHXHNV và Kinh tế. ($\alpha = 1\%$)

Gọi Y_2 là số sinh viên Đại học Kinh tế có thời gian tự học trên 5 giờ mỗi ngày.

Mẫu 1: $n_1 = 120$; $y_1 = 37 \Rightarrow \hat{p}_1 = \dfrac{y_1}{n_1} = \dfrac{37}{120}$.

Mẫu 2: $n_2 = 90$; $y_2 = 14 \Rightarrow \hat{p}_2 = \dfrac{y_2}{n_2} = \dfrac{14}{90} = \dfrac{7}{45}$. $\Rightarrow \hat{p} = \dfrac{y_1 + y_2}{n_1 + n_2} = \dfrac{37 + 14}{120 + 90} = \dfrac{17}{70}$.

• **Giả thuyết kiểm định:** $\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{cases}$; KĐ 2 phía, $p_1 - p_2 = 0$.

• **Mức ý nghĩa:** $\alpha = 0.01$.

• **Giá trị thống kê kiểm định**

$$z_0 = \frac{\hat{p_1} - \hat{p_2} - (p_1 - p_2)}{\sqrt{\hat{p}(1 - \hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} = \frac{\dfrac{37}{120} - \dfrac{7}{45} - 0}{\sqrt{\dfrac{17}{70}\left(1 - \dfrac{17}{70}\right)\left(\dfrac{1}{120} + \dfrac{1}{90}\right)}} \approx 2.5550$$

• **Miền bác bỏ:** Bác bỏ H_0 khi $|z_0| > z_{1-\alpha/2}$.

$\alpha = 0.01 \Rightarrow z_{1-\alpha/2} = z_{0.995} = 2.57$.

• **So sánh và kết luận:**

Ta có $|z_0| > z_{1-\alpha/2}$

$\Leftrightarrow |2.5550| > 2.57$

$\Leftrightarrow 2.5550 > 2.57 \quad$ (sai)

$\Rightarrow$ chưa đủ cơ sở để bác bỏ $H_0 : p_1 = p_2$.

Kết luận: Với mức ý nghĩa 1% thì tỷ lệ những sinh viên có thời gian tự học trên 5 giờ mỗi ngày giữa hai trường là như nhau.

□

Câu 2 (3 điểm). Trong cấu tạo một loại dây thừng, người ta quan tâm đến hàm lượng nylon x (đv: %) ảnh hưởng như thế nào đến lực căng y (đv: 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 biểu diễn mối liên hệ của x vào y dưới dạng: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$. (2đ)

b) Giải thích ý nghĩa của hệ số β_1 nhận được và dự đoán lực căng của một sợi dây có hàm lượng nylon bằng 45. (1đ)

✍ **Lời giải.**

a) **Tìm phương trình hồi quy tuyến tính biểu diễn mối liên hệ của x vào y dưới dạng: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$.**
Từ bảng dữ liệu, dùng máy tính cầm tay, ta ước lượng phương trình hồi quy tuyến tính đơn y theo x là $\hat{y} = 176.86170 + 6.90957x$.

b) **Giải thích ý nghĩa của hệ số β_1 nhận được:**
Ta có $\hat{\beta}_1 = 6.90957 > 0$ nên khi hàm lượng nylon thay đổi 1% thì lực căng sẽ thay đổi 0.77714 psi. Cụ thể, khi hàm lượng nylon tăng 1% thì lực căng sẽ tăng 0.77714 psi.
Dự đoán lực căng của một sợi dây có hàm lượng nylon bằng 45
Hàm lượng nylon bằng 45, tức $x = 45 \Rightarrow \hat{y} = 176.86170 + 6.90957 \cdot 45 = 487.79235$ psi.

□

XÁC SUẤT THỐNG KÊ

ĐỀ CUỐI KỲ I NĂM HỌC 2018-2019
Thời gian làm bài: 90 phút
Đề số: 2

Câu 1 (7 điểm). Thực hiện một khảo sát xã hội về số tiền chi trả cho các hoạt động vui chơi giải trí trong 1 tháng của 400 thanh niên tại TP.HCM người ta thu được bảng sau:

Số tiền (USD)	50 – 80	80 – 120	120 – 160	160 – 200	200 – 220	220 – 250
Số người	50	80	100	80	60	30

Giả thiết số tiền phải bỏ ra cho các hoạt động vui chơi giải trí trong một tháng của một thanh niên tại TP.HCM là một đại lượng ngẫu nhiên phân phối theo qui luật chuẩn.

a) Ước lượng số tiền trung bình một thanh niên phải bỏ ra với độ tin cậy 95%. (1.5đ)

b) Những thanh niên bỏ ra trên 200 USD/tháng cho các hoạt động vui chơi là những thanh niên khá giả. Hãy ước lượng tỉ lệ những thanh niên khá giả với độ tin cậy 97%. Nếu muốn sai số ≤ 0.01 thì phải khảo sát thêm bao nhiêu thanh niên? (2.5đ)

c) Một nhà nghiên cứu xã hội cho rằng cứ 100 thanh niên ở TP.HCM thì có 30 người thuộc diện khá giả, trong khi nhà thống kê lại tỏ ra nghi ngờ và họ cho rằng con số này thực sự phải nhỏ hơn con số thống kê do nhà nghiên cứu ngày đưa ra. Vậy theo các bạn, ý kiến nào đúng đắn với mức ý nghĩa $\alpha = 5\%$. (1.5đ)

d) Lời khẳng định: "Tỷ lệ thanh niên có thu nhập hạn chế là 50%" có được chấp nhận hay không, mức ý nghĩa 1%. Biết rằng một thanh niên được gọi là có thu nhập hạn chế nếu số tiền bỏ ra cho hoạt động vui chơi dưới 120 USD/tháng. (1.5đ)

✍ **Lời giải.**
Ta có

Số tiền (USD)	65	100	140	180	210	235
Số người	50	80	100	80	60	30

Gọi X_1 là số tiền chi trả cho hoạt động vui chơi giải trí trong 1 tuần của thanh niên tại Tp.HCM. (đv: USD)

a) Ta có $n = 400$; $\bar{x} = 148.25$; $s = 51.90$.
Phương sai σ^2: chưa biết.
• **Độ tin cậy:** 95% $\Rightarrow \alpha = 0.05. \Rightarrow t_{\alpha/2;n-1} = t_{0.025;399} = 1.96$

• **Sai số:** $\epsilon = t_{\alpha/2;n-1} \cdot \dfrac{s}{\sqrt{n}}$

$= 1.96 \cdot \dfrac{51.90}{\sqrt{400}}$

$\approx 5.0862.$

• **Khoảng tin cậy** 95% **cho** μ **là:**

$$\begin{aligned} \bar{x} - \epsilon \leq \quad & \mu \quad \leq \bar{x} + \epsilon \\ \Leftrightarrow 148.25 - 5.0862 \leq \quad & \mu \quad \leq 148.25 + 5.0862 \\ \Leftrightarrow 143.1638 \leq \quad & \mu \quad \leq 153.3362 \end{aligned}$$

$\Rightarrow \mu \in [143.1638; 153.3362].$

b) Gọi Y là số thanh niên khá giả.

Ta có $y = 60 + 30 = 90$, $n = 400 \Rightarrow \hat{p} = \dfrac{y}{n} = \dfrac{90}{400} = 0{,}225.$

• **Độ tin cậy:** $97\% \Rightarrow \alpha = 0.03$

$\Rightarrow z_{1-\alpha/2} = z_{0.985} = 2.17.$

• **Sai số:** $\epsilon = z_{1-\alpha/2} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} = 2.17 \cdot \sqrt{\dfrac{0{,}225(1-0{,}225)}{400}} \approx 0{,}0453.$

• **KTC** 97% **cho tỷ lệ** p **là**

$$\begin{aligned} \hat{p} - \epsilon \leq \quad & p \quad \leq \hat{p} + \epsilon \\ \Leftrightarrow 0{,}225 - 0{,}0453 \leq \quad & p \quad \leq 0{,}225 + 0{,}0453 \\ \Leftrightarrow 0.1797 \leq \quad & p \quad \leq 0.2703 \end{aligned}$$

$\Rightarrow p \in [0.1797; 0.2703].$

Ta có $\epsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

Để $\epsilon \leq 0.01$ thì

$$n \geq \left(\frac{z_{1-\alpha/2}}{\epsilon}\right)^2 \cdot \hat{p}(1-\hat{p}) = \left(\frac{2.17}{0.01}\right)^2 \cdot 0{,}225(1-0{,}225) \approx 8211.14$$

Để $\epsilon \leq 0.01$ thì phải khảo sát ít nhất 8212 thanh niên.

Vậy để $\epsilon \leq 0.01$ thì phải khảo sát thêm $8212 - 400 = 7812$ thanh niên.

c) Gọi Y_1 là số thanh niên khá giả trong Tp.HCM.

Ta có $y_1 = 60 + 30 = 90$; $n = 400 \Rightarrow \hat{p}_1 = \dfrac{y_1}{n} = \dfrac{90}{400} = 0.225$

• **Giả thuyết kiểm định:** $\begin{cases} H_0 : p_1 \geq \dfrac{30}{100} \\ H_1 : p_1 < \dfrac{30}{100} \end{cases}$; KĐ 1 phía; $p_0 = \dfrac{30}{100} = 0.3.$

• **Mức ý nghĩa:** $\alpha = 0.05.$

• **Giá trị thống kê kiểm định**

$$z_0 = \frac{\hat{p}_1 - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{0.225 - 0.3}{\sqrt{\dfrac{0.3(1-0.3)}{400}}} = -3.2751.$$

• **Miền bác bỏ:** Bác bỏ H_0 khi $z_0 < -z_{1-\alpha}$ Ta có $\alpha = 0.05$

$\Rightarrow -z_{1-\alpha} = -z_{0.95} = -1.65.$

• **So sánh và kết luận:**

Ta có: $\quad z_0 < -z_{1-\alpha}$

$\Leftrightarrow -3.2751 < -1.65 \quad$ (đúng)

$\Rightarrow$ Bác bỏ $H_0 : p \geq \dfrac{30}{100}$.

Kết luận: Với mức ý nghĩa 5% thì tỷ lệ người thuộc diện khá giả ít hơn 30% hay ý kiến nhà thống kê đúng.

d) Gọi Y_2 là số thanh niên có thu nhập thấp trong Tp.HCM.

Ta có $y_2 = 310$; $n = 400 \Rightarrow \hat{p}_2 = \dfrac{y_2}{n} = \dfrac{310}{400} = 0.775$

Giả thuyết kiểm định: $\begin{cases} H_0 : p_2 = 0.5 \\ H_1 : p_2 \neq 0.5 \end{cases}$; $\quad$ KĐ 2 phía; $p_0 = 0.5$.

- **Mức ý nghĩa:** $\alpha = 0.01$.
- **Giá trị thống kê kiểm định**

$$z_0 = \frac{\hat{p}_2 - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{0.775 - 0.5}{\sqrt{\dfrac{0.5(1-0.5)}{400}}} = 11$$

- **Miền bác bỏ:** Bác bỏ H_0 khi $|z_0| > z_{1-\alpha/2}$.

Ta có $\alpha = 0.01$

$\Rightarrow z_{1-\alpha/2} = z_{0.995} = 2.57$.

- **So sánh và kết luận:**

Ta có: $\quad |z_0| > z_{1-\alpha/2}$

$\Leftrightarrow |11| > 2.57$

$\Leftrightarrow 11 > 2.57 \quad$ (đúng)

$\Rightarrow$ Bác bỏ $H_0 : p_2 = 0.5$.

Kết luận: Với mức ý nghĩa 1% thì tỷ lệ thanh niên có thu nhập hạn chế không phải là 50% hay lời khẳng định trên là không đúng.

□

Câu 2 (3 điểm). Bảng số liệu bên dưới mô tả về chỉ số khối cơ thể x (Body Mass Index - BMI) và huyết áp y của 8 người được chọn ngẫu nhiên. BMI được tính bởi công thức (cân nặng (kg)/chiều cao (m))^2.

BMI	20.3	22.0	26.4	28.2	31	32.6	17.6	19.4
Huyết áp	116	110	131	136	144	138	122	115

a) Tìm phương trình hồi quy tuyến tính biểu diễn mối liên hệ của x vào y dưới dạng: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$. (2đ)

b) Giải thích ý nghĩa của hệ số β_1 nhận được và dự đoán huyết áp của một người có BMI bằng 30. (1đ)

✍ **Lời giải.**

a) **Tìm phương trình hồi quy tuyến tính biểu diễn mối liên hệ của x vào y dưới dạng: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$.**

Từ bảng dữ liệu, dùng máy tính cầm tay, ta ước lượng phương trình hồi quy tuyến tính đơn y theo x là $\hat{y} = 79.24835 + 1.91399x$.

b) **Giải thích ý nghĩa của hệ số β_1 nhận được**
Ta có $\hat{\beta}_1 = 1.91399 > 0$ nên khi chỉ số cơ thể thay đổi 1 đơn vị thì huyết áp người đó sẽ thay đổi 1.91399 đơn vị. Cụ thể, khi chỉ số cơ thể tăng 1 đơn vị thì huyết áp người đó sẽ tăng 1.91399 đơn vị.
Dự đoán huyết áp của một người có BMI bằng 30
Với BMI bằng 30, tức $x = 30 \Rightarrow \hat{y} = 79.24835 + 1.91399 \cdot 30 = 136.66805$.

□

XÁC SUẤT THỐNG KÊ

ĐỀ CUỐI KỲ I NĂM HỌC 2017-2018

Thời gian làm bài: 90 phút

Đề số: 1

Câu 1. Các sản phẩm được sản xuất trong một dây chuyền. Để thực hiện kiểm tra chất lượng, mỗi giờ người ta rút ngẫu nhiên không hoàn lại 10 sản phẩm từ một hộp có 25 sản phẩm. Quá trình sản xuất được báo cáo là đạt yêu cầu nếu có không quá một sản phẩm là thứ phẩm.

a) Nếu tất cả các hộp được kiểm tra đều chứa chính xác hai thứ phẩm, thì xác suất quá trình sản xuất được báo cáo đạt yêu cầu ít nhất 7 lần trong một ngày làm việc 8 giờ là bao nhiêu?

b) Sử dụng phân phối Poisson để xấp xỉ xác suất được tính trong câu a).

c) Biết rằng lần kiểm tra chất lượng cuối cùng trong câu a), quá trình sản xuất được báo cáo đạt yêu cầu. Hỏi xác suất mẫu 10 sản phẩm tương ứng không chứa thứ phẩm là bao nhiêu?

✍ **Lời giải.**

a) Gọi X là số thứ phẩm trong 10 sản phẩm được kiểm tra.
Khi đó $X(\Omega) = \{0; 1; 2\}$.
Ta có

$$\begin{aligned}\mathbb{P}(X=0) &= \frac{C_{23}^{10}}{C_{25}^{10}} = 0.35\\ \mathbb{P}(X=1) &= \frac{C_2^1 C_{23}^9}{C_{25}^{10}} = 0.5\end{aligned}$$

Xác suất quá trình sản xuất được báo cáo đạt yêu cầu là

$$\mathbb{P}(X \le 1) = \mathbb{P}(X=0) + \mathbb{P}(X=1) = 0.35 + 0.5 = 0.85.$$

Gọi Y là số lần quá trình sản xuất được báo cáo đạt yêu cầu trong 8 lần kiểm tra. Ta có $Y \sim B(8; 0.85)$.
Xác suất quá trình sản xuất được báo cáo đạt yêu cầu ít nhất 7 lần trong một ngày làm việc 8 giờ là

$$\begin{aligned}\mathbb{P}(Y \le 7) &= \sum_{x=7}^{8} \mathbb{P}(X = x)\\ &= \sum_{x=7}^{8} C_8^x \cdot 0.85^x \cdot (1-0.85)^{8-x}\\ &\approx 0.6572\end{aligned}$$

b) Sử dụng phân phối Poisson để xấp xỉ xác suất được tính trong câu a).

Gọi Z là số lần quá trình sản xuất được báo cáo không đạt yêu cầu trong 8 lần kiểm tra. Ta có $Z \sim B(8; 0.15)$. Khi đó $Z \sim P(np = 8 \cdot 0.15 = 1.2)$.

Do đó

$$\mathbb{P}(Y \leq 7) = \mathbb{P}(Z \leq 1) = \sum_{x=0}^{1} \mathbb{P}(Z = x) = \sum_{x=0}^{1} \frac{e^{-1.2} \cdot 1.2^x}{x!} \approx 0.6626.$$

c) Biết rằng lần kiểm tra chất lượng cuối cùng trong câu a), quá trình sản xuất được báo cáo đạt yêu cầu. Hỏi xác suất mẫu 10 sản phẩm tương ứng không chứa thứ phẩm là bao nhiêu?

$$\begin{aligned}
\mathbb{P}(X = 0 | X \leq 1) &= \frac{\mathbb{P}(X = 0, X \leq 1)}{\mathbb{P}(X \leq 1)} \\
&= \frac{\mathbb{P}(X = 0)}{\mathbb{P}(X \leq 1)} \\
&= \frac{0.35}{0.85} \\
&= \frac{7}{17}.
\end{aligned}$$

□

Câu 2. Khảo sát về thu nhập của một số người làm việc ở một công ty, người ta thu được các số liệu cho ở bảng sau:

Thu nhập (*triệu đồng/tháng*)	5	5.5	6	6.5	7	7.5	8	9	10
Số người	6	12	20	34	30	16	12	10	4

Giả sử thu nhập là đại lượng ngẫu nhiên có phân phối chuẩn.

a) Giả sử thu nhập là đại lượng ngẫu nhiên có phân phối chuẩn. Hãy ước lượng cho thu nhập trung bình của người làm trong công ty với độ tin cậy 90%.

b) Nếu muốn sai số ước lượng thu nhập trung bình của một người ở công ty không vượt quá 200 000 đồng/tháng ở độ tin cậy 99% thì cần khảo sát thêm ít nhất bao nhiêu người nữa?

c) Nếu nói rằng thu nhập trung bình của một người ở công ty là 7 triệu đồng/tháng thì có đáng tin cậy không? Hãy kết luận với mức ý nghĩa 5%.

✍ **Lời giải.**

a) **Giả sử thu nhập là đại lượng ngẫu nhiên có phân phối chuẩn. Hãy ước lượng cho thu nhập trung bình của người làm trong công ty với độ tin cậy 90%.**

Gọi X (triệu đồng) là thu nhập của một người làm trong công ty.

Ta có $n = 144$; $\bar{x} \approx 6.8958$; $s = 1.0901$.

Phương sai σ^2: chưa biết.

- **Độ tin cậy:** 90% $\Rightarrow \alpha = 0.1. \Rightarrow t_{\alpha/2;n-1} = t_{0.05;143} = 1.645$
- **Sai số:** $\epsilon = t_{\alpha/2;n-1} \cdot \dfrac{s}{\sqrt{n}}$

$$= 1.645 \cdot \frac{1.0901}{\sqrt{144}}$$

$$\approx 0.1494.$$

- **Khoảng tin cậy 90% cho μ là:**

$$\begin{aligned}
\bar{x} - \epsilon \leq\ & \mu \leq \bar{x} + \epsilon \\
\Leftrightarrow 6.8958 - 0.1494 \leq\ & \mu \leq 6.8958 + 0.1494 \\
\Leftrightarrow 6.7464 \leq\ & \mu \leq 7.0452
\end{aligned}$$

$\Rightarrow \mu \in [6.7464; 7.0452]$.

b) Nếu muốn sai số ước lượng thu nhập trung bình của một người ở công ty không vượt quá 200 000 đồng/tháng ở độ tin cậy 99% thì cần khảo sát thêm ít nhất bao nhiêu người nữa?

Đổi đơn vị: 200 000 đồng/tháng $= 0.2$ triệu đồng/tháng.

• **Độ tin cậy:** $0.99 \Rightarrow \alpha = 0.01$.

$\Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.995} = 2.575$.

Ta có: $\epsilon = z_{1-\frac{\alpha}{2}} \cdot \dfrac{s}{\sqrt{n}}$

Để $\epsilon \leq 1 \Rightarrow n \geq \left(\dfrac{z_{1-\frac{\alpha}{2}} \cdot s}{\epsilon_0}\right)^2 = \left(\dfrac{2.575 \cdot 1.0901}{0.2}\right)^2 \approx 196.9823$.

$\Rightarrow n \geq 197$.

Vậy nếu muốn sai số ước lượng trung bình không quá 0.2 với độ tin cậy 0.99 thì phải quan sát thêm ít nhất $197 - 144 = 53$ người nữa.

c) Nếu nói rằng thu nhập trung bình của một người ở công ty là 7 triệu đồng/tháng thì có đáng tin cậy không? Hãy kết luận với mức ý nghĩa 5%.

σ^2: chưa biết.

• **Giả Thuyết KĐ:** $\begin{cases} H_0 : \mu = 7 \\ H_1 : \mu \neq 7 \end{cases}$: KĐ 2 phía, $\mu_0 = 7$.

• **Mức ý nghĩa:** $\alpha = 0.05$.

• **Giá trị Thống kê kiểm định:**

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{6.8958 - 7}{\frac{1.0901}{\sqrt{144}}} \approx -1.1471.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $|t_0| > t_{\frac{\alpha}{2};n-1}$.

Ta có $\alpha = 0.05$

$\Rightarrow t_{\frac{\alpha}{2};n-1} = t_{0.025;143} = 1.96$.

• **So sánh và kết luận:**

Ta có: $|t_0| > t_{\frac{\alpha}{2};n-1}$

$\Leftrightarrow |-1.1471| > 1.96$

$\Leftrightarrow 1.1471 > 1.96$ (sai)

$\Rightarrow$ Chưa đủ cơ sở để bác bỏ $H_0 : \mu = 7$.

Kết luận: Với mức ý nghĩa 5%, thu nhập trung bình của một người ở công ty là 7 triệu đồng/tháng là đáng tin cậy.

□

XÁC SUẤT THỐNG KÊ

ĐỀ CUỐI KỲ I NĂM HỌC 2017-2018

Thời gian làm bài: 90 phút

Đề số: 2

Câu 1. Một công ty kỹ thuật thiết lập một bài kiểm tra năng lực khi các ứng viên nộp đơn để thực tập. Thời gian hoàn thành bài kiểm tra có phân phối chuẩn với trung bình 40,5 phút và độ lệch chuẩn 7,5 phút. Các ứng viên hoàn thành bài thi ít hơn 30 phút được chấp nhận thực tập ngay. Những người hoàn thành bài thi từ 30 đến 36 phút được yêu cầu làm thêm một bài kiểm tra khác. Mọi ứng viên khác sẽ bị từ chối.

a) Với một ứng viên được chọn ngẫu nhiên, tính xác suất để người này được

i) chấp nhận thực tập ngay;

ii) yêu cầu làm thêm một bài kiểm tra khác.

b) Cho biết một ứng viên được chọn ngẫu nhiên không bị từ chối sau bài kiểm tra đầu tiên này, tính xác suất ứng viên được chấp nhận thực tập ngay.

c) Vào một dịp nào đó có 100 ứng viên. Sử dụng một xấp xỉ phù hợp để tính xác suất có nhiều hơn 25 ứng viên được yêu cầu làm thêm một bài kiểm tra khác.

✍ **Lời giải.**

a) Gọi X là thời gian một ứng viên hoàn thành bài kiểm tra. Ta có $X \sim N(40.5; 7.5^2)$ Xác suất ta phải tìm để ứng viên này được

i) chấp nhận thực tập ngay

$$\mathbb{P}(X < 30) = \mathbb{P}\left(\frac{X - 40.5}{7.5} < \frac{30 - 40.5}{7.5}\right) = \mathbb{P}\left(Z < -1.4\right) = \Phi(-1.4) = 1 - \Phi(1.4) \approx 0.0808.$$

ii) yêu cầu làm thêm một bài kiểm tra khác

$$\begin{aligned} \mathbb{P}(30 \leq X \leq 36) &= \mathbb{P}\left(\frac{30 - 40.5}{7.5} \leq \frac{X - 40.5}{7.5} \leq \frac{36 - 40.5}{7.5}\right) \\ &= \mathbb{P}\left(-1.4 \leq Z \leq -0.6\right) \\ &\approx 0.1935. \end{aligned}$$

b) Xác suất để một ứng viên không bị từ chối là

$$\mathbb{P}(X \leq 36) = \mathbb{P}(X < 30) + \mathbb{P}(30 \leq X \leq 36) = 0.0808 + 0.1935 = 0.2743.$$

Nếu một ứng viên được chọn ngẫu nhiên không bị từ chối sau bài kiểm tra đầu tiên này thì xác suất ứng viên được chấp nhận thực tập ngay là

$$\mathbb{P}(X < 30 | X \leq 30) = \frac{\mathbb{P}\left[(X < 30) \cap (x \leq 36)\right]}{\mathbb{P}(X \leq 36)} = \frac{\mathbb{P}(X < 30)}{\mathbb{P}(X \leq 36)} = \frac{0.0808}{0.2743} \approx 0.2946.$$

c) Gọi Y là số ứng viên được yêu cầu làm thêm một bài kiểm tra.
Khi đó $X \sim B(100; 0.1935)$.
Xác suất có nhiều hơn 25 ứng viên được yêu cầu làm thêm một bài kiểm tra khác là

$$\begin{aligned} \mathbb{P}(Y > 25) = 1 - \mathbb{P}(Y \leq 25) \approx 1 - \mathbb{P}(Y \leq 25 + 0.5) &\approx 1 - \mathbb{P}\left(Z \leq \frac{25 + 0.5 - 100 \cdot 0.1935}{\sqrt{100 \cdot 0.1935 \cdot (1 - 0.1935)}}\right) \\ &\approx 1 - \mathbb{P}\left(Z \leq 1.5568\right) \\ &\approx 1 - 0.9402 \\ &= 0.0598. \end{aligned}$$

□

Câu 2. Điều tra năng suất lúa trên diện tích 100 hecta trồng lúa của một vùng, ta thu được bảng số liệu sau:

Năng suất (*tạ/ha*)	41	44	45	46	48	52	54
Diện tích (*ha*)	10	20	30	15	10	10	5

Giả sử năng suất lúa có phân phối chuẩn.

a) Hãy ước lượng năng suất lúa trung bình của vùng đó với độ tin cậy 95%.

b) Nếu muốn sai số ước lượng của năng suất lúa trung bình không vượt quá 0.5 tạ/ha, với độ tin cậy 99% thì cần điều tra thêm ít nhất bao nhiêu hecta lúa nữa?

c) Những thửa ruộng có năng suất từ 48 tạ/ha trở lên được xem là những thửa có năng suất cao. Hãy ước lượng tỷ lệ diện tích có năng suất cao trong vùng với độ tin cậy 90%.

d) Có tài liệu cho biết năng suất lúa trung bình là 47 tạ/ha. Giá trị này có phù hợp với mẫu quan sát không? Hãy kết luận với mức ý nghĩa 5%.

✍ **Lời giải.**

a) Hãy ước lượng năng suất lúa trung bình của vùng đó với độ tin cậy 95%.
Gọi X (triệu đồng) là thu nhập của một người làm trong công ty.
Ta có $n = 100$; $\bar{x} = 46$; $s = 3.3029$.
Phương sai σ^2: chưa biết.
• **Độ tin cậy:** 95% $\Rightarrow \alpha = 0.05. \Rightarrow t_{\alpha/2;n-1} = t_{0.025;99} = 1.96$
• **Sai số:** $\epsilon = t_{\alpha/2;n-1} \cdot \dfrac{s}{\sqrt{n}}$

$$= 1.96 \cdot \frac{3.3029}{\sqrt{100}}$$
$$\approx 0.6474.$$

• **Khoảng tin cậy** 95% **cho** μ **là:**

$$\begin{aligned} \bar{x} - \epsilon \leq &\ \mu \leq \bar{x} + \epsilon \\ \Leftrightarrow 46 - 0.6474 \leq &\ \mu \leq 46 + 0.6474 \\ \Leftrightarrow 45.3526 \leq &\ \mu \leq 46.6474 \end{aligned}$$

$\Rightarrow \mu \in [45.3526; 46.6474]$.

b) Nếu muốn sai số ước lượng của năng suất lúa trung bình không vượt quá 0.5 tạ/ha, với độ tin cậy 99% thì cần điều tra thêm ít nhất bao nhiêu hecta lúa nữa?
• **Độ tin cậy:** $0.99 \Rightarrow \alpha = 0.01$.
$\Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.995} = 2.575$.
Ta có: $\epsilon = z_{1-\frac{\alpha}{2}} \cdot \dfrac{s}{\sqrt{n}}$
Để $\epsilon \leq 1 \Rightarrow n \geq \left(\dfrac{z_{1-\frac{\alpha}{2}} \cdot s}{\epsilon_0}\right)^2 = \left(\dfrac{2.575 \cdot 3.3029}{0.5}\right)^2 \approx 289.3479$.
$\Rightarrow n \geq 290$.
Vậy nếu muốn sai số ước lượng trung bình không quá 0.5 với độ tin cậy 0.99 thì phải quan sát thêm ít nhất $290 - 100 = 190$ hecta lúa nữa.

c) Những thửa ruộng có năng suất từ 48 tạ/ha trở lên được xem là những thửa có năng suất cao. Hãy ước lượng tỷ lệ diện tích có năng suất cao trong vùng với độ tin cậy 90%.
Gọi Y là những thửa ruộng có năng suất cao.
Ta có: $n = 100$, $y = 10 + 10 + 5 = 25 \Rightarrow$ tỷ lệ mẫu $\hat{p} = \dfrac{y}{n} = \dfrac{25}{100} = 0.25$.
• **Độ tin cậy:** 90% $\Rightarrow \alpha = 0.1$.
$\Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.95} = 1.645$.

- **Sai số:** $\epsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$
$= 1.645 \cdot \sqrt{\dfrac{0.25\,(1-0.25)}{100}}$
$\approx 0.0712.$
- **KTC** 90% **cho tỷ lệ** p **là**

$$\begin{aligned} \hat{p}-\epsilon \leq \ & p \ \leq \hat{p}+\epsilon \\ \Leftrightarrow 0.25-0.0712 \leq \ & p \ \leq 0.25+0.0712 \\ \Leftrightarrow 0.1788 \leq \ & p \ \leq 0.3212 \end{aligned}$$

$\Rightarrow p \in [0.1788; 0.3212]$.

d) **Có tài liệu cho biết năng suất lúa trung bình là** 47 **tạ/ha. Giá trị này có phù hợp với mẫu quan sát không? Hãy kết luận với mức ý nghĩa** 5%**.**

σ^2: chưa biết.

- **Giả Thuyết KĐ:** $\begin{cases} H_0 : \mu = 47 \\ H_1 : \mu \neq 47 \end{cases}$: KĐ 2 phía, $\mu_0 = 47$.
- **Mức ý nghĩa:** $\alpha = 0.05$.
- **Giá trị Thống kê kiểm định:**

$$t_0 = \frac{\bar{x}-\mu_0}{\dfrac{s}{\sqrt{n}}} = \frac{46-47}{\dfrac{3.3029}{\sqrt{100}}} \approx -3.0276.$$

- **Miền bác bỏ:** Bác bỏ H_0 nếu $|t_0| > t_{\frac{\alpha}{2};n-1}$.

Ta có $\alpha = 0.05$
$\Rightarrow t_{\frac{\alpha}{2};n-1} = t_{0.025;99} = 1.96$.
- **So sánh và kết luận:**

Ta có: $|t_0| > t_{\frac{\alpha}{2};n-1}$
$\Leftrightarrow |-3.0276| > 1.96$
$\Leftrightarrow 3.0276 > 1.96$ (đúng)
$\Rightarrow$ Bác bỏ $H_0 : \mu = 47$.
Kết luận: Với mức ý nghĩa 5%, năng suất lúa trung bình không phải là 47 tạ/ha hay giá trị này không phù hợp với mẫu quan sát.

□

XÁC SUẤT THỐNG KÊ

ĐỀ CUỐI KỲ I NĂM HỌC 2017-2018

Thời gian làm bài: 90 phút

Đề số: 3

Câu 1. Trọng lượng A của mỗi con bò trong một đàn bò là biến ngẫu nhiên có phân phối chuẩn với kỳ vọng 300 kg và độ lệch chuẩn 50 kg. Chọn ngẫu nhiên một con bò trong chuồng. Tính xác suất để con bò được chọn:

a) Có trọng lượng trên 500 kg.

b) Có trọng lượng từ 250 kg tới 350 kg.

c) Chọn ngẫu nhiên 4 con bò trong đàn bò nói trên. Tính xác suất để có 2 trong 4 con bò nói trên có trọng lượng từ 250 kg tới 350 kg.

✍ Lời giải.
Ta có $A \sim N(300; 50^2) \Rightarrow \mu = 300; \sigma = 50$.

a) $\mathbb{P}(A > 500) = \mathbb{P}\left(\dfrac{A-300}{50} > \dfrac{500-300}{50}\right) = \mathbb{P}(Z > 4) = 1 - \mathbb{P}(Z \leq 4) \approx 1 - 0.99997 = 0.00003$.

b) $$\begin{aligned}\mathbb{P}(250 \leq A \leq 350) &= \mathbb{P}\left(\dfrac{250-300}{50} \leq \dfrac{A-300}{50} \leq \dfrac{350-300}{50}\right)\\ &= \mathbb{P}(-1 \leq Z \leq 1)\\ &= \Phi(1) - \Phi(-1)\\ &= 0.84134 - 0.15866\\ &= 0.6827.\end{aligned}$$

c) Gọi Y là số con bò có trọng lượng từ 250 kg tới 350 kg trong 4 con bò.
Khi đó $Y \sim B(4; 0.6827)$.
$\mathbb{P}(Y = 2) = \mathrm{C}_4^2 \cdot 0.6827^2 \cdot (1 - 0.6827)^{4-2} \approx 0.2816$.

□

Câu 2. Một khảo sát về chiều cao X (cm) của một giống cây trồng người ta quan sát một mẫu và có kết quả như 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.

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

b) Những cây trồng có chiều cao từ 135 cm trở lên được gọi là những cây "cao". Hãy ước lượng tỉ lệ những cây cao với độ tin cậy 95%.

c) Người ta áp dụng phương pháp mới trong việc trồng và chăm sóc cây. Sau một thời gian, khảo sát 100 cây trồng theo phương pháp mới được bảng 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 5%, hãy kiểm định xem phương pháp mới có làm tăng chiều cao trung bình của cây hay không?

d) Có ý kiến cho rằng phương pháp mới làm tăng tỉ lệ cây "cao". Với mức ý nghĩa 5%, hãy kiểm tra ý kiến này.

✍ Lời giải.

a) **Ước lượng chiều cao trung bình của giống cây trồng trên với độ tin cậy 95%.**
Ta có $n = 100$; $\bar{x} = 131$; $s = 18.2297$.
Phương sai σ^2: chưa biết.
• **Độ tin cậy:** $95\% \Rightarrow \alpha = 0.05. \Rightarrow t_{\alpha/2;n-1} = t_{0.025;99} = 1.96$

- **Sai số:** $\epsilon = t_{\alpha/2;n-1} \cdot \dfrac{s}{\sqrt{n}}$

$= 1.96 \cdot \dfrac{18.2297}{\sqrt{100}}$

$\approx 3.5730.$

- **Khoảng tin cậy** 95% **cho** μ **là:**

$$\begin{aligned} \bar{x} - \epsilon \leq & \quad \mu \quad \leq \bar{x} + \epsilon \\ \Leftrightarrow 131 - 3.5730 \leq & \quad \mu \quad \leq 131 + 3.5730 \\ \Leftrightarrow 127.427 \leq & \quad \mu \quad \leq 134.573 \end{aligned}$$

$\Rightarrow \mu \in [127.427; 134.573].$

b) **Những cây trồng có chiều cao từ 135 cm trở lên được gọi là những cây "cao". Hãy ước lượng tỉ lệ những cây cao với độ tin cậy 95%.**
Gọi Y_1 là những cây "cao" theo phương pháp cũ.
Ta có: $n_1 = 100$, $y_1 = 10 + 10 + 15 = 35 \Rightarrow$ tỷ lệ mẫu $\hat{p}_1 = \dfrac{y_1}{n_1} = \dfrac{35}{100} = 0.35.$

- **Độ tin cậy:** $95\% \Rightarrow \alpha = 0.05.$

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

- **Sai số:** $\epsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1}}$

$= 1.96 \cdot \sqrt{\dfrac{0.35\,(1-0.35)}{100}}$

$\approx 0.0935.$

- **KTC** 95% **cho tỷ lệ** p **là**

$$\begin{aligned} \hat{p} - \epsilon \leq & \quad p \quad \leq \hat{p} + \epsilon \\ \Leftrightarrow 0.35 - 0.0935 \leq & \quad p \quad \leq 0.35 + 0.0935 \\ \Leftrightarrow 0.2565 \leq & \quad p \quad \leq 0.4435 \end{aligned}$$

$\Rightarrow p \in [0.2565; 0.4435].$

c) Gọi X_2 (cm) là chiều cao của giống cây theo phương pháp mới.
Mẫu 1: $n_1 = 100$; $\bar{x}_1 = 131$; $s_1 \approx 18.2297$.
Mẫu 2: $n_1 = 100$; $\bar{x}_2 = 130.1$; $s_2 \approx 15.9858$.
Mọi thông tin về phương sai là chưa biết $\Rightarrow$ kiểm định phương sai.
⁎ **Kiểm định phương sai**

- Giả thuyết kiểm định: $\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$: KĐ 2 phía.
- Mức ý nghĩa: $\alpha = 0{,}05$.
- Giá trị thống kê kiểm định:

$$f_0 = \frac{s_1^2}{s_2^2} = \frac{18.2297^2}{15.9858^2} \approx 1.3004.$$

Với $\alpha = 0.05 \Rightarrow f_{\alpha/2;n_1-1;n_2-1} = f_{\alpha/2;n_2-1;n_1-1} = f_{0.025;99;99} = 1.4862.$

Ta có: $f_0 > f_{\alpha/2;n_1-1;n_2-1}$ hoặc $f_0 < \dfrac{1}{f_{\alpha/2;n_2-1;n_1-1}}$

$\Leftrightarrow 1.3004 > 1.4862$ (sai)

$\Leftrightarrow 1.3004 < \dfrac{1}{1.4862} \approx 0.6729$ (sai)

$\Rightarrow$ Chưa đủ cơ sở để bác bỏ $H_0 : \sigma_1^2 = \sigma_2^2$.
Với mức ý nghĩa 5%: $\sigma_1^2 = \sigma_2^2$.
⋇ **Kiểm định trung bình khi phương sai bằng nhau chưa biết**
Mẫu 1: $n_1 = 100$; $\bar{x}_1 = 131$; $s_1 \approx 18.2297$.
Mẫu 2: $n_1 = 100$; $\bar{x}_2 = 130.1$; $s_2 \approx 15.9858$.
Phương sai $\sigma_1^2 = \sigma_2^2$ chưa biết.

• **GTKĐ:** $\begin{cases} H_0 : \mu_1 \geq \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases}$: KĐ 1 phía, $\mu_0 = \mu_1 - \mu_2 = 0$.

• **Mức ý nghĩa:** $\alpha = 0.05$.

• **Giá trị Thống kê kiểm định:**

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(100-1)18.2297^2 + (100-1)15.9858^2}{100+100-2} \approx 293.9339.$$

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}} = \frac{131 - 130.1 - 0}{\sqrt{\dfrac{293.9339}{100} + \dfrac{293.9339}{100}}} \approx 0.3712.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $t_0 < -t_{\alpha;n_1+n_2-2}$.
Ta có $\alpha = 0.05$
$\Rightarrow -t_{\alpha;n_1+n_2-2} = -t_{0.05;198} = -1.645$.

• **So sánh và kết luận:**
Ta có: $t_0 < -t_{\alpha;n_1+n_2-2}$
$\Leftrightarrow 0.3712 < -1.645$ (sai)
$\Rightarrow$ Chưa đủ cơ sở để bác bỏ $H_0 : \mu_1 \geq \mu_2$.
Kết luận: Với mức ý nghĩa 5%, phương pháp mới không làm tăng chiều cao trung bình của cây.

□

XÁC SUẤT THỐNG KÊ

ĐỀ CUỐI KỲ I NĂM HỌC 2017-2018
Thời gian làm bài: 90 phút
Đề số: 4

Câu 1. Cho biến ngẫu nhiên X (phút) thể hiện thời gian chờ mua trà sữa của khách hàng ở một quán nhỏ ven đường và có hàm mật độ là:

$$f_X(x) = \begin{cases} kx^2(6-x) & \text{nếu } x \in [0;6] \\ 0 & \text{nếu } x \in \mathbb{R} \setminus [0;6] \end{cases}$$

trong đó k là một số thực nào đó.

a) Tìm k.

b) Tìm thời gian chờ trung bình của khách.

c) Tìm hàm phân phối tích lũy của biến ngẫu nhiên X.

✍ **Lời giải.**

a) Tìm k.
Vì f_X là hàm mật độ nên $f_X(x) \geq 0$, $\forall x \in \mathbb{R}$ mà $x^2(6-x) \geq 0 \forall x \in [0;6] \Rightarrow k \geq 0$.
Ta có

$$\int_{-\infty}^{+\infty} f_X(x)dx = \int_0^6 kx^2(6-x)dx = k\int_0^6 (6x^2 - x^3)dx = 108k.$$

Mà $\int_{-\infty}^{+\infty} f_X(x)dx = 1 \Leftrightarrow 108k = 1 \Rightarrow k = \dfrac{1}{108}$.

Do đó, hàm mật độ của biến ngẫu nhiên X là

$$f_X(x) = \begin{cases} \dfrac{1}{108}x^2(6-x) & \text{nếu } x \in [0;6] \\ 0 & \text{nếu } x \in \mathbb{R} \setminus [0;6] \end{cases}$$

b) Tìm thời gian chờ trung bình của khách.

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f_X(x)dx = \int_0^6 x \cdot \frac{1}{108}x^2(6-x)dx = \int_0^6 \frac{1}{108}x^3(6-x)dx = \frac{18}{5}.$$

c) Tìm hàm phân phối tích lũy của biến ngẫu nhiên X.

⋇ Nếu $x < 0$, khi đó $\int_{-\infty}^{x} f_X(t)dt = \int_{-\infty}^{x} 0dt = 0$.

⋇ Nếu $0 \leq x < 6$, khi đó

$$\begin{aligned} \int_{-\infty}^{x} f_X(t)dt &= \int_{-\infty}^{0} 0dt + \int_0^x \frac{1}{108}t^2(6-t)dt \\ &= 0 + \frac{1}{108}\left(2t^3 - \frac{t^4}{4}\right)\Big|_0^x \\ &= \frac{1}{108}\left(2x^3 - \frac{x^4}{4}\right). \end{aligned}$$

⋇ Nếu $6 \leq x$, khi đó

$$\begin{aligned} \int_{-\infty}^{x} f_X(t)dt &= \int_{-\infty}^{0} 0dt + \int_0^6 \frac{1}{108}t^2(6-t)dt + \int_6^x 0dt \\ &= 0 + \frac{1}{108}\left(2t^3 - \frac{t^4}{4}\right)\Big|_0^6 + 0 \\ &= 1. \end{aligned}$$

Vậy hàm phân phối tích lũy của biến ngẫu nhiên X là

$$f_X(x) = \begin{cases} 0 & \text{nếu } x < 0 \\ \dfrac{1}{108}\left(2x^3 - \dfrac{x^4}{4}\right) & \text{nếu } 0 \leq x < 6 \\ 1 & \text{nếu } 6 \leq x \end{cases}$$

□

Câu 2. Một nhà nông trồng giống cherry Úc theo phương pháp được các nhà khoa học đề xuất; chính vì thể đường kính của quả cherry tuân theo phân phối chuẩn với trung bình 28 mm và độ lệch chuẩn 2 mm. Một quả cherry được gọi là có size 32 nếu đường kính của quả cherry nằm trong $[30; 32]$. Giả sử các quả cherry có size 32 độc lập nhau.

a) Chọn ngẫu nhiên một quả cherry trong vườn, tìm xác suất nhận được quả cherry có size 32.

b) Tìm số tự nhiên n nhỏ nhất sao cho chọn n quả cherry trong vườn thì xác suất nhận được ít nhất 1 quả cherry size 32 không bé hơn 0.99.

c) Chọn ngẫu nhiên 100 quả cherry trong vườn, hãy chọn mô hình xấp xỉ thích hợp để tìm xác suất có ít nhất 20 quả cherry size 32.

✍ **Lời giải.**

Gọi X (mm) là đường kính của một quả cherry.
Khi đó $X \sim N(28; 2^2) \Rightarrow \mu = 28; \sigma = 2$.

a)
$$\begin{aligned}\mathbb{P}(30 \le X \le 32) &= \mathbb{P}\left(\frac{30-28}{2} \le \frac{X-28}{2} \le \frac{32-28}{2}\right)\\ &= \mathbb{P}(1 \le Z \le 2)\\ &= \Phi(2) - \Phi(1)\\ &= 0.97725 - 0.84134\\ &= 0.13591.\end{aligned}$$

b) Gọi Y là số quả cherry size 32 trong n quả được chọn.
Khi đó $Y \sim B(n; 0.13591)$.

Ta có:
$$\begin{aligned}\mathbb{P}(Y \ge 1) &= 1 - \mathbb{P}(Y < 1)\\ &= 1 - \mathbb{P}(Y = 0)\\ &= 1 - \mathrm{C}_n^0 \cdot 0.13591^0 \cdot (1-0.13591)^{n-0}\\ &= 1 - 0.86409^n.\end{aligned}$$

Theo đề bài
$$\begin{aligned} &\mathbb{P}(Y \ge 1) \ge 0.99\\ \Leftrightarrow\ &1 - 0.86409^n \ge 0.99\\ \Leftrightarrow\ &0.86409^n \ge 0.01\\ \Rightarrow\ &n \ge \log_{0.86409}(0.01)\\ \Leftrightarrow\ &n \ge 31.52534\\ \Rightarrow\ &n \approx 32.\end{aligned}$$

c) Gọi W là số quả cherry size 32 trong 100 quả.
Khi đó $W \sim B(100; 0.13591)$.

$$\begin{aligned}\mathbb{P}(W \ge 20) = \mathbb{P}(W \ge 20 - 0.5) &\approx \mathbb{P}\left(Z \ge \frac{20 - 0.5 - 100 \cdot 0.13591}{\sqrt{100 \cdot 0.13591 \cdot (1 - 0.13591)}}\right)\\ &= \mathbb{P}(Z \ge 1.72428)\\ &= 1 - \mathbb{P}(Z < 1.72428)\\ &= 1 - 0.95767\\ &= 0.04233.\end{aligned}$$

□

Câu 3. Để xem xét tình hình học tập môn xác suất thống kê (XSTK), một giảng viên (GV) đã tiến hành lấy mẫu (2 lần). Biết rằng điểm số sinh viên tuân theo phân phối chuẩn. Điểm sinh viên nam và sinh viên nữ độc lập nhau.

a) Ở lần lấy mẫu thứ nhất, GV này thu thập được thông tin của 20 sinh viên với điểm trung bình 5.5 và độ lệch chuẩn 0.75. Sử dụng thông tin này để ước lượng điểm trung bình của sinh viên với độ tin cậy 95%.

GV này tiếp tục thực hiện lấy mẫu lần thứ hai và thu thập được bảng thông tin sau:

Điểm	$[0,1)$	$[1,2)$	$[2,3)$	$[3,4)$	$[4,5)$	$[5,6)$	$[6,7)$	$[7,8)$	$[8,9)$	$[9,10]$
Nam	4	6	4	6	2	6	5	2	6	4
Nữ	5	10	8	8	5	4	3	2	3	2

(Bảng dữ liệu chứa thông tin: có 4 sinh viên nam và 5 sinh viên nữ có điểm số nằm trong $[0, 1)$, có 6 sinh viên nam và 10 sinh viên nữ có điểm số nằm trong $[1, 2)$,.... Hãy sử dụng thông tin ở lần lấy mẫu thứ hai để giải các câu hỏi (b,c,d)

b) Một sinh viên được gọi là có điểm cao nếu điểm lớn hơn hoặc bằng 7.0. Ước lượng khoảng tin cậy cho tỉ lệ học sinh có điểm cao với độ tin cậy 99%.

c) Dựa theo số liệu điểm môn XSTK trong học kì trước, điểm trung bình của sinh viên **nam** là 5.0. Hãy cho biết giá trị trên có phù hợp với dữ liệu quan sát (điểm sinh viên nam đã thu thập được) hay không với mức ý nghĩa 2%?

d) Có nhiều người cho rằng các bạn nữ chăm học hơn các bạn nam nên điểm trung bình của các bạn nữ cao hơn điểm trung bình của các bạn nam. Tuy nhiên, GV này cho rằng điểm trung bình của các bạn nữ **không cao hơn** điểm trung bình của các bạn nam. Hãy sử dụng dữ liệu quan sát để kiểm tra nhận định về hai điểm trung bình của GV với mức ý nghĩa 2%.

✍ **Lời giải.**

a) Ta có $n = 20$; $\bar{x} = 5.5$; $s = 0.75$.
Phương sai σ^2 chưa biết.
• **Độ tin cậy:** 95% $\Rightarrow \alpha = 0{,}05$
$\Rightarrow t_{\alpha/2;n-1} = t_{0.025;19} = 2.093$
• **Sai số:** $\epsilon = t_{\alpha/2;n-1} \cdot \dfrac{s}{\sqrt{n}}$
$= 2.093 \cdot \dfrac{0.75}{\sqrt{20}}$
$\approx 0.3510.$
• **Khoảng tin cậy** 95% **cho** μ **là:**

$$\begin{aligned} \bar{x} - \epsilon \leq &\quad \mu \quad \leq \bar{x} + \epsilon \\ \Leftrightarrow 5.5 - 0.3510 \leq &\quad \mu \quad \leq 5.5 + 0.3510 \\ \Leftrightarrow 5.149 \leq &\quad \mu \quad \leq 5.851 \end{aligned}$$

$\Rightarrow \mu \in [5.149; 5.851]$.

b) Vì đề bài chỉ xét sinh viên chung (không xét nam hay nữ) nên ta có bảng số liệu biến đổi như sau

Điểm	0.5	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5
Số sinh viên	9	16	12	14	7	10	8	4	9	6

Gọi Y là số sinh viên đạt điểm cao.
Ta có $y = 4 + 9 + 6 = 19$; $n = 95 \Rightarrow \hat{p} = \dfrac{y}{n} = \dfrac{19}{95} = 0.2$.
• **Độ tin cậy:** 99% $\Rightarrow \alpha = 0.01$
$\Rightarrow z_{1-\alpha/2} = z_{0.995} = 2.575$
• **Sai số:** $\epsilon = z_{1-\alpha/2} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} = 2.575 \cdot \sqrt{\dfrac{0.2.\,(1-0.2)}{95}} = 0.1057$
• **Khoảng tin cậy** 99% **cho tỷ lệ** p **là:**

$$\begin{aligned} \hat{p} - \epsilon \leq &\quad p \quad \leq \hat{p} + \epsilon \\ \Leftrightarrow 0.2 - 0.1057 \leq &\quad p \quad \leq 0.2 + 0.1057 \\ \Leftrightarrow 0.0943 \leq &\quad p \quad \leq 0.3057 \end{aligned}$$

$\Rightarrow p \in [0.0943; 0.3057]$.

c) Xét bảng số liệu biến đối cho sinh viên nam như sau

Điểm	0.5	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5
Nam	4	6	4	6	2	6	5	2	6	4

Ta có: Ta có $n = 45$; $\bar{x} = 4.9$; $s \approx 2.9108$.
Phương sai σ^2 chưa biết.

• **Giả Thuyết KĐ:** $\begin{cases} H_0 : \mu = 5.0 \\ H_1 : \mu \neq 5.0 \end{cases}$: KĐ 2 phía, $\mu_0 = 5.0$.

• **Mức ý nghĩa:** $\alpha = 0.02$.

• **Giá trị Thống kê kiểm định:**

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{4.9 - 5.0}{\frac{2.9108}{\sqrt{45}}} \approx -0.2305.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $|t_0| > t_{\frac{\alpha}{2};n-1}$.
Ta có $\alpha = 0.02$
$\Rightarrow t_{\frac{\alpha}{2};n-1} = t_{0.01;44} = 2.326$.

• **So sánh và kết luận:**
Ta có: $|t_0| > t_{\frac{\alpha}{2};n-1}$
$\Leftrightarrow |-0.2305| > 2.326$
$\Leftrightarrow 0.2305 > 2.326$ (sai)
$\Rightarrow$ Chưa đủ cơ sở để bác bỏ $H_0 : \mu = 5.0$.
Kết luận: Với mức ý nghĩa 2%, điểm trung bình của sinh viên nam là 5.0 hay giá trị này phù hợp với mẫu quan sát.

d) Xét bảng số liệu biến đối cho sinh viên nữ như sau

Điểm	0.5	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5
Nữ	5	10	8	8	5	4	3	2	3	2

Ta có $n_2 = 50$; $\bar{x}_2 = 3.78$; $s_2 \approx 2.5398$.
Mẫu 1 (sinh viên nam): $n_1 = 45$; $\bar{x}_1 = 4.9$; $s_1 \approx 2.9108$.
Mẫu 2 (sinh viên nữ): $n_2 = 50$; $\bar{x}_2 = 3.78$; $s_2 \approx 2.5398$.
Mọi thông tin về phương sai là chưa biết $\Rightarrow$ kiểm định phương sai.

⁎ Kiểm định phương sai

• Giả thuyết kiểm định: $\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$: KĐ 2 phía.

• Mức ý nghĩa: $\alpha = 0{,}02$.

• Giá trị thống kê kiểm định:

$$f_0 = \frac{s_1^2}{s_2^2} = \frac{2.9108^2}{2.5398^2} \approx 1.3135.$$

Với $\alpha = 0.02 \Rightarrow f_{\alpha/2;n_1-1;n_2-1} = f_{\alpha/2;n_2-1;n_1-1} = f_{0.01;44;49} = 1.9895$.
$f_{\alpha/2;n_2-1;n_1-1} = f_{0.01;49;44} = 2.0130$.

Ta có: $f_0 > f_{\alpha/2;n_1-1;n_2-1}$ hoặc $f_0 < \dfrac{1}{f_{\alpha/2;n_2-1;n_1-1}}$
$\Leftrightarrow 1.3135 > 1.9895$ (sai)
$\Leftrightarrow 1.3135 < \dfrac{1}{2.0130} \approx 0.4968$ (sai)

$\Rightarrow$ Chưa đủ cơ sở để bác bỏ $H_0 : \sigma_1^2 = \sigma_2^2$.

Với mức ý nghĩa 2%: $\sigma_1^2 = \sigma_2^2$.

⋇ **Kiểm định trung bình khi phương sai bằng nhau chưa biết**

Mẫu 1 (sinh viên nam): $n_1 = 45$; $\bar{x}_1 = 4.9$; $s_1 \approx 2.9108$.

Mẫu 2 (sinh viên nữ): $n_2 = 50$; $\bar{x}_2 = 3.78$; $s_2 \approx 2.5398$.

Phương sai $\sigma_1^2 = \sigma_2^2$ chưa biết.

• **GTKĐ:** $\begin{cases} H_0 : \mu_1 \geq \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases}$: KĐ 1 phía, $\mu_0 = \mu_1 - \mu_2 = 0$.

• **Mức ý nghĩa:** $\alpha = 0.02$.

• **Giá trị Thống kê kiểm định:**

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(45-1)2.9108^2 + (50-1)2.5398^2}{45+50-2} \approx 7.4073.$$

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}} = \frac{4.9 - 3.78 - 0}{\sqrt{\dfrac{7.4073}{45} + \dfrac{7.4073}{50}}} \approx 2.0027.$$

• **Miền bác bỏ:** Bác bỏ H_0 nếu $t_0 < -t_{\alpha;n_1+n_2-2}$.

Ta có $\alpha = 0.02$

$\Rightarrow -t_{\alpha;n_1+n_2-2} = -t_{0.02;93} = -2.0830$.

• **So sánh và kết luận:**

Ta có: $t_0 < -t_{\alpha;n_1+n_2-2}$

$\Leftrightarrow 2.0027 < -2.0830$ (sai)

$\Rightarrow$ Chưa đủ cơ sở để bác bỏ $H_0 : \mu_1 \geq \mu_2$.

Kết luận: Với mức ý nghĩa 2%, điểm trung bình của các bạn nữ **không cao hơn** điểm trung bình của các bạn nam hay nhận định của GV là đúng.

□