

ÔN TẬP XÁC SUẤT THỐNG KÊ ỨNG DỤNG

(MATH132901)

CHƯƠNG 3: BIẾN NGẪU NHIÊN RỜI RẠC VÀ PHÂN PHỐI XÁC SUẤT

3.1 Biến ngẫu nhiên (đại lượng ngẫu nhiên): 2 loại ĐLNN (BNN)

- *BNN rời rạc*: tập giá trị của nó là hữu hạn hay vô hạn đếm được.

- *BNN liên tục*: tập giá trị của nó lấp đầy một khoảng nào đó trên trục số và xác suất của một điểm bất kì luôn là 0.

3.2 Phân phối xác suất của biến ngẫu nhiên rời rạc:

a/ **Hàm phân phối xác suất của biến rời rạc X (hay hàm khối lượng xác suất)** là

$$p(x) = P(X = x) \text{ với } p(x) \geq 0 ; \sum_{i=1}^n p(x_i) = 1.$$

Bảng phân phối xác suất:

X	x_1	...	x_n
$p(x_i)$	$p(x_1)$	...	$p(x_n)$

b/ **Hàm phân phối tích lũy (CDF) :** kí hiệu $F(x)$

(cả biến rời rạc và biến liên tục đều có hàm phân phối xác suất tích lũy)

$$F(x) = P(X \leq x)$$

Chú ý: với X là biến rời rạc thì $F(x) = P(X \leq x) = \sum_{y \leq x} p(y) ; \forall x \in \mathbb{R}$

3.3 Kỳ vọng và phương sai:

a/ **Kỳ vọng (trung bình) :**

$$E(X) = \mu_x = \sum_{x \in D} x \cdot p(x) = x_1 p(x_1) + x_2 p(x_2) + \dots + x_n p(x_n)$$

Tính chất:

$$1/ E(c) = c, c = \text{const}$$

$$2/ E(c.X) = c.E(X)$$

$$3/ E(X + Y) = E(X) + E(Y)$$

$$4/ E(X.Y) = E(X).E(Y) \text{ nếu như X và Y độc lập.}$$

$$5 / E[h(X)] = \sum_{x \in D} h(x) \cdot p(x)$$

b/ Phương sai: $V(X) = \sigma^2 = E(X^2) - (EX)^2$ với $E(X^2) = \sum_{x \in D} x^2 p(x)$.

Độ lệch chuẩn: $\sigma = \sqrt{V(X)}$

Chú ý: $V(X) = \sum_D (x - EX) \cdot p(x) = E[(X - EX)^2] = E(X^2) - (EX)^2$

Tính chất :

1/ $V(X) \geq 0$

2/ $V(c) = 0$, $c = \text{const}$

3/ $V(cX) = c^2 V(X)$

4/ $V(X + Y) = V(X) + V(Y)$ nếu X, Y độc lập

3.4-3.6 Các phân phối xác suất (biến rời rạc)

1/ Phân phối nhị thức: Kí hiệu: $X \sim \text{Bin}(n, p)$

X là số phép thử xảy ra trong n phép thử, xs của mỗi phép thử xảy ra là p.(mỗi phép thử chỉ có hai kết quả).

Hàm khối lượng xác suất của X kí hiệu là $b(x, n, p)$:

$$b(x; n; p) = P(X = x) = \begin{cases} C_n^x p^x (1-p)^{n-x} & , \forall x = 0, 1, 2, \dots, n \\ 0 & , \text{khác} \end{cases} \quad (\text{công thức Bernoulli})$$

Hàm phân phối tích lũy của X là $B(x; n; p)$:

$$B(x; n; p) = P(X \leq x) = \sum_{y=0}^x b(y; n; p) \quad ; x = 0, 1, \dots, n$$

Tính chất:
$$\begin{cases} EX = \mu = np \\ V(X) = \sigma^2 = npq \quad (\text{vì } q = 1 - p) \end{cases}$$

2/ Phân phối siêu bội: X có phân phối siêu bội thì

$$P(X = x) = h(x; n, M, N) = \frac{C_M^x \cdot C_{N-M}^{n-x}}{C_N^n} \quad \text{và} \quad \boxed{EX = n \cdot \frac{M}{N} \quad ; \quad V(X) = n \cdot p \cdot (1-p) \cdot \frac{N-n}{N-1}}$$

3/ Phân phối nhị thức âm: (khác với trong sách)

Y là số phép thử thực hiện cho đến khi nào có r phép thử xuất hiện thì dừng.

$$P(Y = n) = C_{n-1}^{r-1} \cdot p^{r-1} \cdot (1-p)^{n-r} \cdot p \quad \text{for } n \geq r.$$

Tính chất: $EY = \frac{r}{p} \quad ; \quad V(Y) = \frac{r(1-p)}{p^2}$

4/ Phân phối Poisson: Kí hiệu: $X \sim P(\mu)$ tham số $\mu > 0$

X là số phép thử xảy ra.

$$p(x; \mu) = P(X = x) = \frac{e^{-\mu} \cdot \mu^x}{x!} \quad (x = 0, 1, 2, \dots)$$

Tính chất: $EX = V(X) = \mu$

Nhận biết pp Poisson:

- $X \sim \text{Bin}(n, p)$, biết n và p . Trong đó n lớn ($n > 50$); p nhỏ; tích $np = \mu < 5$. Xem $X \sim P(\mu)$
- $X \sim \text{Bin}(n, p)$, chưa biết n và p mà chỉ biết trung bình $np = \mu$. Xem $X \sim P(\mu)$

CHƯƠNG 4: BIẾN NGẪU NHIÊN LIÊN TỤC và PHÂN PHỐI XÁC SUẤT

4.1 Hàm mật độ của biến ngẫu nhiên liên tục (pdf): kí hiệu f_x

X là biến ngẫu nhiên liên tục có hàm mật độ thỏa

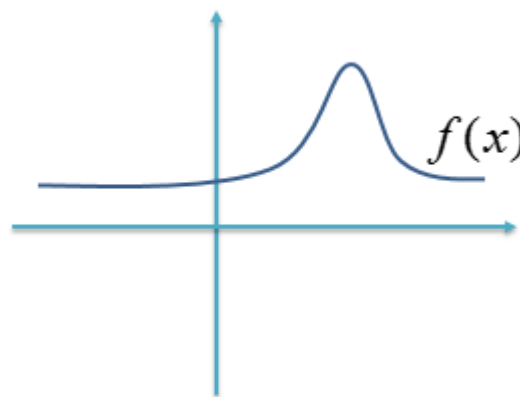
1/ $f(u) \geq 0, \forall u \in R$

2/ $\int_{-\infty}^{+\infty} f(u) du = 1$

3/

$$\begin{aligned} P(a \leq X \leq b) &= P(a \leq X < b) \\ &= P(a < X \leq b) \\ &= P(a < X < b) = \int_a^b f(u) du \end{aligned}$$

4/ $P(X = a) = 0$



4.2 Hàm phân phối tích lũy và các đặc trưng

a/ Hàm phân phối tích lũy $F(x)$ của biến ngẫu nhiên liên tục X là :

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(y) dy$$

Chú ý: $F'(x) = f(x)$

Tính chất: $P(X > a) = 1 - F(a)$; $P(a \leq X \leq b) = F(b) - F(a)$

b/ **Phân vị của phân phối liên tục:** Cho p là một số thuộc khoảng $[0;1]$. **Phân vị thứ (100p)** của một hàm phân phối của một biến ngẫu nhiên liên tục X , kí hiệu là $\eta(p)$

được xác định bởi $p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(y) dy$

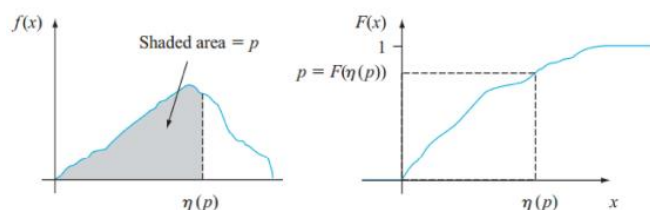


Figure 4.10 The (100p)th percentile of a continuous distribution

c/ **Trung vị** của một phân phối liên tục, kí hiệu μ , là phân vị thứ 50, sao cho μ thoả mãn $F(\mu) = 0.5$. Hay trung vị là giá trị chia phân phối thành 2 phần bằng nhau.

d/ **Kì vọng (trung bình):** $\mu_x = E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$.

Đặc biệt $\mu_{h(X)} = E[h(X)] = \int_{-\infty}^{\infty} h(x) \cdot f(x) dx$

e/ **Phương sai:** $\sigma_x^2 = V(X) = E(X^2) - (EX)^2 = \int_{-\infty}^{\infty} x^2 \cdot f(x) dx - \left(\int_{-\infty}^{\infty} x \cdot f(x) dx \right)^2$

Chú ý: Có thể dùng công thức $V(X) = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) dx$

f/ **Độ lệch chuẩn:** $\sigma = \sqrt{V(X)}$.

4.3 Các phân phối xác suất (biến liên tục):

1/ Phân phối chuẩn:

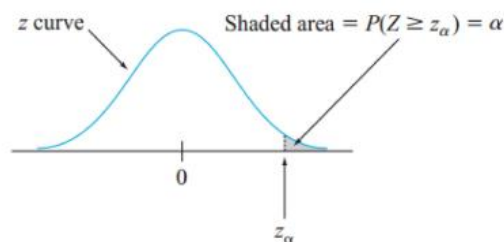
a/ Phân phối chuẩn chuẩn tắc: Kí hiệu: $Z \sim N(0;1)$, $\mu = 0$; $\sigma = 1$

- Hàm mật độ có dạng: $f(z; 0,1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$
- Hàm phân phối của Z là: $P(Z \leq z) = \int_{-\infty}^z f(y; 0,1) dy = \Phi(z)$

- Tính chất:

$P(Z \leq a) = \Phi(a)$ $P(Z \geq b) = 1 - \Phi(b)$ $P(a \leq Z \leq b) = \Phi(b) - \Phi(a)$ $\Phi(z) = \frac{1}{2} \text{ khi } z \geq 4$
--

b/ Phân vị của phân phối chuẩn: z_α biểu thị giá trị trên trục z mà α là diện tích vùng dưới đường cong z nằm bên phải z_α . Khi đó z_α chính là phân vị thứ $100(1-\alpha)$ của phân phối chuẩn.



c/ Phân phối chuẩn: Kí hiệu: $X \sim N(\mu, \sigma^2)$

- Hàm mật độ có dạng: $f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$
- Tính chất: $EX = \mu$; $V(X) = \sigma^2$

Chú ý: Chuyển từ phân phối chuẩn về dạng chuẩn chuẩn tắc :

Với $X \sim N(\mu, \sigma^2)$; ta đặt : $Z = \frac{X - \mu}{\sigma} \sim N(0,1)$

d/ Sự hội tụ của phân phối nhị thức về phân phối chuẩn: $X \sim Bin(n, p)$

Nếu đồ thị của phân phối nhị thức không quá lệch (không quá mất đối xứng)

($np \geq 10$; $n(1-p) \geq 10$) thì có thể xấp xỉ X như một phân phối chuẩn $X \sim N(\mu, \sigma^2)$ bằng

cách đặt $\mu = np$ và $\sigma^2 = npq$ và $P(X \leq x) \approx \Phi\left(\frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right)$.

2/ Phân phối mũ :

- X có phân phối mũ với tham số $\lambda > 0$, thì hàm mật độ của X là

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} ; & x \geq 0 \\ 0 & ; \text{neu khac} \end{cases}.$$

- Hàm phân phối xác suất của X là $F(x; \lambda) = P(X \leq x) = \begin{cases} 1 - e^{-\lambda x} ; & x \geq 0 \\ 0 & ; x < 0 \end{cases}$

- Tính chất: $\mu = \frac{1}{\lambda}$; $\sigma^2 = \frac{1}{\lambda^2}$.

CHƯƠNG 5: PHÂN PHỐI XÁC SUẤT ĐỒNG THỜI VÀ MẪU NGẪU NHIÊN

(đọc sách)

CHƯƠNG 6,7: ƯỚC LƯỢNG

- **Tổng thể:** tập hợp tất cả các phần tử mà từ các phần tử đó ta có thể thu thập, khảo sát những thông tin về các dấu hiệu ta cần nghiên cứu được gọi là *tổng thể* (population).

- **Mẫu ngẫu nhiên:** Cho đại lượng ngẫu nhiên X với quy luật phân phối xác suất nào đó. Một mẫu ngẫu nhiên kích thước n được thành lập từ đại lượng ngẫu nhiên X là n đại lượng ngẫu nhiên độc lập $X_1, X_2, \dots, X_n$, có cùng phân phối xác suất với đại lượng ngẫu nhiên X.

+ Chọn ngẫu không hoàn lại.

+ Chọn mẫu có hoàn lại.

- **Các tham số của tổng thể:**

N : số phần tử của tổng thể

M_A : số phần tử có tính chất A của tổng thể

$p = \frac{M_A}{N}$: tỷ lệ của tổng thể về tính chất

A.

μ : trung bình của tổng thể

σ^2 : phương sai của tổng thể

σ : Độ lệch chuẩn của tổng thể

- Các tham số của mẫu:

n : kích thước mẫu

m_A : số phần tử có tính chất A của mẫu.

$p = \frac{m_A}{n}$: tỷ lệ mẫu về tính chất A.

$\bar{x}$: trung bình mẫu.

s^2 : phương sai của mẫu hiệu chỉnh.

s : Độ lệch chuẩn mẫu hiệu chỉnh (hoặc kí hiệu σ_{n-1})

1. Phương pháp ước lượng điểm: (ước lượng không chệch)

Dùng trung bình mẫu $\bar{x}$ để ước lượng cho trung bình tổng thể μ .

Dùng phương sai mẫu s^2 để ước lượng cho phương sai tổng thể σ^2 .

Dùng tỷ lệ mẫu p để ước lượng cho tỷ lệ tổng thể ρ .

2. Phương pháp ước lượng khoảng:

- **Độ tin cậy** : kí hiệu γ với $\gamma = 1 - \alpha$ (α là mức ý nghĩa)

2.1. Ước lượng khoảng cho trung bình của tổng thể μ :

a/ Khoảng ước lượng đối xứng :

Gọi μ là trung bình của tổng thể (μ chưa biết), ta cần tìm ước lượng khoảng đối xứng của μ với độ tin cậy $\gamma = 1 - \alpha$.

- Trung bình mẫu: $\bar{x}$

- Độ chính xác (sai số) của ước lượng:

TH1 : σ^2 đã biết :

Sai số là

$$\varepsilon = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

TH2 : σ^2 chưa biết , mẫu lớn ($n > 40$):

$$\varepsilon = z_{\alpha/2} \frac{s}{\sqrt{n}}$$

(tra bảng Z)

TH3 : σ^2 chưa biết , mẫu nhỏ

$$\varepsilon = t_{\frac{\alpha}{2}; n-1} \frac{s}{\sqrt{n}}$$

(tra bảng T - Student)

- Khoảng ước lượng cho trung bình : $(\bar{x} - \varepsilon < \mu < \bar{x} + \varepsilon)$

(Xem giới thiệu về phân phối Student trong giáo trình)

b/ Khoảng ước lượng 1 phía :

- Giá trị ước lượng **tối đa** cho trung bình tổng thể μ là:

$$\mu < \bar{x} + z_{\alpha} \frac{s}{\sqrt{n}} ; \quad \mu < \bar{x} + t_{\alpha; n-1} \frac{s}{\sqrt{n}}$$

- Giá trị ước lượng **tối thiểu** cho trung bình tổng thể μ là:

$$\mu > \bar{x} - z_{\alpha} \frac{s}{\sqrt{n}} ; \quad \mu > \bar{x} - t_{\alpha; n-1} \frac{s}{\sqrt{n}}$$

Chú ý:

Với $\phi(z_{\alpha/2}) = \frac{\gamma + 1}{2} = 1 - \frac{\alpha}{2}$; $\phi(z_{\alpha}) = \gamma = 1 - \alpha$ được tra từ bảng Z- chuẩn tắc

c/ Tìm cỡ mẫu, độ tin cậy :

- Tìm cỡ mẫu n : Cho chiều rộng cỡ mẫu $w = 2\varepsilon$, mà $w = 2\varepsilon = 2 \cdot z_{\alpha/2} \frac{s}{\sqrt{n}} \Rightarrow \boxed{n = \frac{4 \cdot z_{\alpha/2}^2 \cdot s^2}{w^2}}$
 - Tìm độ tin cậy khi biết sai số: $\varepsilon = z_{\alpha/2} \frac{s}{\sqrt{n}} \Rightarrow z_{\alpha/2} = \frac{\varepsilon \sqrt{n}}{s} \Rightarrow \phi(z_{\alpha/2}) = 1 - \frac{\alpha}{2} = \frac{\gamma + 1}{2} \Rightarrow \gamma = ?$
- Hoặc $\varepsilon = z_{\alpha/2} \frac{s}{\sqrt{n}} \Rightarrow z_{\alpha} = \frac{\varepsilon \sqrt{n}}{s} \Rightarrow \phi(z_{\alpha}) = ? \Rightarrow \gamma = ?$

2.2. Ước lượng khoảng cho tỷ lệ của tổng thể p :

a/ Khoảng ước lượng đối xứng :

Gọi p là tỷ lệ các phần tử có tính chất A của tổng thể (p chưa biết), ta cần tìm ước lượng khoảng đối xứng của p với độ tin cậy $\gamma = 1 - \alpha$.

- Tìm tỷ lệ mẫu $p = \frac{m_A}{n}$
- Sai số ước lượng (độ chính xác) : $\varepsilon = z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ (tra bảng Z)
- Khoảng ước lượng (khoảng tin cậy): $\hat{p} - \varepsilon < p < \hat{p} + \varepsilon$

b/ Khoảng ước lượng 1 phía :

- Giá trị ước lượng **tối đa** cho tỉ lệ tổng thể p là:

$$p < \hat{p} + z_{\alpha} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- Giá trị ước lượng **tối thiểu** cho tỉ lệ tổng thể p là:

$$p > \hat{p} - z_{\alpha} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

c/ Tìm cỡ mẫu, độ tin cậy :

- Tìm cỡ mẫu n :

Cho chiều rộng cỡ mẫu $w = 2\varepsilon$, mà $n \approx 4 z_{\alpha/2}^2 \frac{\hat{p}(1-\hat{p})}{w^2}$

- Tìm độ tin cậy khi biết sai số:

$$\varepsilon = z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \Rightarrow z_{\alpha/2} = \dots \Rightarrow \phi(z_{\alpha/2}) = \dots \Rightarrow \gamma = ?$$

2.3. Ước lượng khoảng cho phương sai của tổng thể σ^2 :

a/ Khoảng ước lượng đối xứng :

Gọi σ^2 là phương sai của tổng thể (σ^2 chưa biết), ta cần tìm ước lượng khoảng đối xứng của σ^2 với độ tin cậy $\gamma = 1 - \alpha$. Ta có :

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2; n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-(\alpha/2); n-1}}$$

Với $\chi^2_{\alpha/2}$ và $\chi^2_{1-(\alpha/2)}$ là giá trị của ĐLNN có phân phối Chi bình phương $\chi^2 \sim \chi^2(n-1)$ (xem giáo trình)

b/ Khoảng ước lượng 1 phía :

- Giá trị ước lượng **tối đa** cho phương sai tổng thể σ^2 là:

$$\sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha; n-1}}$$

- Giá trị ước lượng **tối thiểu** cho phương sai tổng thể σ^2 là:

$$\sigma^2 > \frac{(n-1)s^2}{\chi^2_{\alpha; n-1}}$$

c/ Khoảng ước lượng cho độ lệch chuẩn σ :

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2; n-1}}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_{1-(\alpha/2); n-1}}}$$

Ví dụ 1: (ULTB) Trọng lượng (kg/con) của một số con heo ở thời kì xuất chuồng là

Trọng lượng	65-85	85 – 95	95 – 105	105 – 115	115-135
Số con	8	40	60	42	10

Hãy tìm khoảng tin cậy đối xứng 87% cho trọng lượng trung bình của loại heo trên . Biết trọng lượng 1 con heo được chọn ngẫu nhiên trong trại có phương sai là 225 kg².

Ví dụ 2: (ULTB) Khảo sát chi tiêu X (triệu đồng/tháng) của một số người chọn ngẫu nhiên từ vùng A có thống kê sau : (Biết X có phân phối chuẩn)

X	3,2-3,7	3,7-4,2	4,2-4,7	4,7-5,2	5,2-5,7	5,7- 6,2	6,2-6,7
n_i	23	33	55	73	45	22	18

Tìm khoảng tin cậy đối xứng 98% cho chi tiêu trung bình mỗi người dân vùng A. đs: (4,750 ; 4,976)

Ví dụ 3: (ULTB) Quan sát mức hao phí của 25 xe máy thuộc cùng một loại xe, chạy trên cùng một quãng đường, người ta thu được kết quả

Mức xăng (l)	1,9 – 2,1	2,1 – 2,3	2,3 – 2,5	2,5 – 2,7
Số xe	5	9	8	3

Hãy tìm ước lượng trung bình tối đa với độ tin cậy 99% cho mức xăng hao phí trung bình của loại xe trên. Đs: 2,3669 lít

Ví dụ 4: (ULTB) Quan sát 100 công nhân trong một xí nghiệp người ta tính được năng suất trung bình của một công nhân ở mẫu này là: $\bar{x} = 12$ sản phẩm/ngày và phương sai mẫu hiệu chỉnh là 25.

Muốn ước lượng năng suất trung bình của một công nhân trong xí nghiệp với độ tin cậy 99% và độ chính xác $\varepsilon = 0,8$ thì cần quan sát năng suất của bao nhiêu công nhân nữa?

Ví dụ 5: (ULTB) Mức tiêu thụ X của mỗi hộ gia đình vùng A trong mùa khô năm nay có phân phối chuẩn. Điều tra 1 số hộ gia đình vùng A :

X(kwh/t)	65-115	115-165	165-215	215-265	265-315	315-365	365-415	415-465
Số hộ	24	36	75	94	97	125	84	75

Nếu muốn ước lượng mức tiêu thụ điện trung bình các hộ vùng A trong mùa khô năm nay với độ chính xác 10 kwh/tháng thì độ tin cậy bằng ? đs: 99%

Ví dụ 6: (ULTL) Công ty M kiểm tra ngẫu nhiên 1200 sản phẩm do ca sáng sản xuất thấy có 45 sản phẩm không đạt chuẩn . Tính tỷ lệ sản phẩm đạt chuẩn tối đa do ca sáng sản xuất với độ tin cậy 99%.

Ví dụ 7: (ULTL) Đo chiều dài 1 số sản phẩm do nhà máy A sản xuất:

X(cm)	53,8	53,81	53,82	53,83	53,84	53,85	53,86	53,87
Số sp (n_i)	9	14	30	47	40	33	15	12

Biết X có phân phối chuẩn. Tìm khoảng ước lượng đối xứng 98% cho tỉ lệ sản phẩm có chiều dài trên 53,84 cm. đs (22,45 % ; 37,55%)

Ví dụ 8: (ULTL) Phòng vấn 400 người ở một khu vực thấy 240 người ủng hộ dự luật A.

1/ Với độ tin cậy 95%, hãy ước lượng tỷ lệ người ủng hộ dự luật A. đs (0,5520; 0,6480)

2/ Nếu độ chính xác là 0,057 khi ước lượng tỉ lệ đối xứng thì độ tin cậy là bao nhiêu?

Ví dụ 9: (ULPS) Độ dày của bản kim loại tuân theo luật phân phối chuẩn. Đo 10 bản kim loại người ta tính được phương sai hiệu chỉnh của mẫu là 0,1367. Hãy xác định khoảng tin cậy 95% cho phương sai của độ dày đó. Đs (0,064; 0,456)

BÀI TẬP

Bài 1: Quan sát 100 công nhân trong một xí nghiệp người ta tính được năng suất trung bình của một công nhân ở mẫu này là: $\bar{x} = 12$ sản phẩm/ngày và phương sai mẫu 25. Ước lượng năng suất trung bình của một công nhân trong xí nghiệp với độ tin cậy 99%. Đs (10,7; 13,3)

Bài 2: Khảo sát chi tiêu X (triệu đồng/tháng) của một số người vùng A có thống kê sau (Biết X có phân phối chuẩn)

X	3,2-3,7	3,7-4,2	4,2-4,7	4,7-5,2	5,2-5,7	5,7- 6,2	6,2-6,7
n_i	23	33	55	73	45	22	18

a/ Tìm khoảng tin cậy 95% cho tỉ lệ người có chi tiêu trên 5,7 triệu đồng/ tháng vùng A.

b/ Biết vùng A có 50 000 người. Tìm số người ở vùng A có chi tiêu trên 5,7 triệu đồng/tháng với độ tin cậy 95%

c/ Biết vùng A có 10 000 người chi tiêu trên 5,7 triệu đồng/ tháng . Tìm số người vùng A có với độ tin cậy 95%

Bài 3: Khảo sát năng suất lúa thu được bảng số liệu sau:

Năng suất (tấn/ha)	5,1	5,4	5,5	5,6	5,8	6,2	6,4
DT có NS tương ứng (ha)	10	20	30	15	10	10	5

Tìm ước lượng trung bình tối thiểu cho năng suất lúa trung bình ở vùng đó với độ tin cậy 95%

Bài 4: Đo chiều dài 1 số sản phẩm do nhà máy A sản xuất, có thống kê sau: (X có phân phối chuẩn).

Chiều dài X(cm)	53,80	53,81	53,82	53,83	53,84	53,85	53,86	53,87
Số sp (n_i)	9	14	30	47	40	33	15	12

1/ Tìm khoảng tin cậy 95% cho chiều dài trung bình các sản phẩm do nhà máy A sản xuất.

2/ Tìm khoảng tin cậy 98% cho tỉ lệ sản phẩm có chiều dài trên 53,84 cm.

Bài 5: Công ty M có 3000 đại lý, cho tiến hành điều tra ngẫu nhiên một số đại lý của mình và thu được bảng số liệu sau (X là doanh số, đơn vị: triệu đồng/tháng), biết X có phân phối chuẩn.

X	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60
Số đại lý	7	12	18	27	22	17	13	4

1/ Những đại lý có $X > 45$ triệu đồng/tháng gọi là đại lý có doanh số cao. Hãy ước lượng số đại lý có doanh số cao với độ tin cậy 95%. Đs: (609; 1089)

2/ Hãy ước lượng doanh số trung bình/tháng của các đại lý với độ tin cậy 99%. Đs(37,4 ; 41,6)

Bài 6: Khảo sát mức tiêu thụ điện X của một số hộ gia đình được chọn ngẫu nhiên ở vùng A ta được bảng số liệu sau:

X(kwh/tháng)	50-100	100-150	150-200	200-250	250-300	300-350	350-400	400-450
Số hộ	24	36	55	64	50	35	20	15

1/ Ước lượng mức tiêu thụ điện trung bình của các hộ gia đình ở vùng A với độ tin cậy 99%.

2/ Hộ có mức tiêu thụ điện dưới 100kwh/tháng gọi là hộ có mức tiêu thụ điện thấp. hãy ước lượng số hộ có mức tiêu thụ điện thấp ở vùng A với độ tin cậy 98% .Biết vùng A có 10.000 hộ dân. Đs: (437; 1169)

Bài 7: Khảo sát thu nhập tại một doanh nghiệp có số liệu:

Thu nhập (triệu đồng/tháng)	2-4	4-6	6-8	8-10	10-12
Số lao động	24	10	26	18	22

1/ Nếu dùng số liệu trên để ước lượng thu nhập trung bình của một người với sai số ko quá 0,5 triệu đồng/tháng thì điều tra ít nhất bao nhiêu người, với độ tin cậy 94%. Đs: $n \geq 122$

2/ Nếu dùng số liệu trên để ước lượng tỷ lệ người thu nhập thấp với sai số 1%. Hỏi độ tin cậy của ước lượng này khoảng bao nhiêu? biết người thu nhập thấp có thu nhập từ 6 trđ/tháng trở xuống. Đs: $t_{\gamma/2} = 0,21$.

Bài 8: Có số liệu thống kê về thu nhập (X: triệu đồng/ tháng) của 100 người ở một công ty như sau:

x_i	1-3	3-4	4-5	5-6	6-7	7-8	8-9	9-13
-------	-----	-----	-----	-----	-----	-----	-----	------

n_i	4	10	17	24	25	9	6	5
-------	---	----	----	----	----	---	---	---

1/ Nếu muốn độ chính xác khi ước lượng thu nhập trung bình là 0,25 (triệu đồng/ tháng) và độ tin cậy là 97% thì cần khảo sát bao nhiêu người?

- a/ 134 b/ 348 c/ 273 d/ 413

2/ Nếu muốn ước lượng thu nhập trung bình của các nhân viên ở công ty này có độ chính xác là 0,25 (triệu đồng/ tháng) thì độ tin cậy đạt được bao nhiêu%?

- a/ 76,56% b/ 81,15% c/ 92,34% d/ 79,18%

Bài 9: Muốn biết trong một hồ nước có bao nhiêu cá, người ta bắt lên 1000 con, đánh dấu xong thả lại xuống hồ. Sau một thời gian, người ta bắt lên 200 con và thấy có 30 con cá có đánh dấu của lần bắt trước. Dựa vào kết quả đó, hãy ước lượng số cá trong hồ với độ tin cậy 95%

- a/ (2013; 4950) b/(4513; 7650) c/(5013; 9950) d/ (6013; 9450)

CHƯƠNG 8: KIỂM ĐỊNH DỰA TRÊN MỘT MẪU ĐƠN

1. Một số khái niệm cơ bản:

- **Giả thiết thống kê:** là những giả thiết về các tham số, phân phối xác suất, hoặc tính độc lập của các đại lượng ngẫu nhiên.
- **Kiểm định giả thiết thống kê:** việc tìm ra kết luận bác bỏ hay chấp nhận một giả thiết.
- **Giả thiết không :** giả thiết cần kiểm định. Ký hiệu là H_0 .
- **Giả thiết đối :** một mệnh đề đối lập với H_0 . Ký hiệu là H_a .
- **Kiểm định giả thiết hai phía:** $H_0 : \theta = \theta_0$; $H_a : \theta \neq \theta_0$
(θ là tham số nào đó của ĐLNN mà ta đang nghiên cứu; θ_0 là giá trị đã biết)
- **Kiểm định giả thiết một phía :** $H_0 : \theta = \theta_0$; $H_a : \theta > \theta_0$ (hay $H_a : \theta < \theta_0$)
(một phía bên phải; một phía bên trái)
- **Sai lầm loại 1 và sai lầm loại 2:** Từ bài toán thực tế nêu giả thiết H_0 và H_a .

	Bản chất	Ho đúng	Ho sai
Kiểm tra mẫu và kết luận			

Ho đúng		Sai lầm loại 2
Ho sai	Sai lầm loại 1	

- **Mức ý nghĩa α** : $P(\text{sai lầm loại 1}) \leq \alpha$; $\alpha = 1 - \gamma$

(Xác suất không mắc sai lầm loại 2 là β)

- **Miền bác bỏ** là miền chứa các giá trị làm cho giả thuyết H_0 bị bác bỏ.

Miền chấp nhận là miền chứa các giá trị giúp cho giả thuyết H_0 không bị bác bỏ (tạm chấp nhận hoặc chưa đủ cơ sở để bác bỏ H_0)

2. Kiểm định giả thiết về trung bình tổng thể μ :

Trường hợp 1: X có phân phối chuẩn; σ^2 đã biết.

Giả thiết không: $H_0 : \mu = \mu_0$

Tiêu chuẩn kiểm định: $z = \frac{\bar{x} - \mu_0}{\sigma} \cdot \sqrt{n}$

Giả thiết đối:

$H_a : \mu > \mu_0$ (kiểm định 1 phía bên phải)

$H_a : \mu < \mu_0$ (kiểm định 1 phía bên trái)

$H_a : \mu \neq \mu_0$ (kiểm định 2 phía)

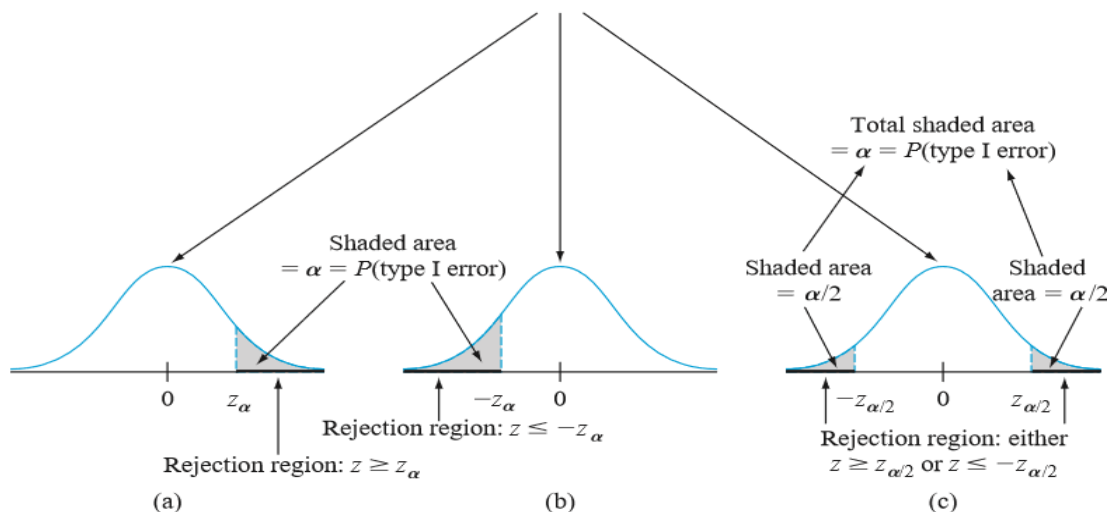
Miền bác bỏ H_0 với mức ý nghĩa α :

$$z \geq z_\alpha$$

$$z \leq -z_\alpha$$

$$z \geq z_{\alpha/2} \text{ hoặc } z \leq -z_{\alpha/2}$$

z curve (probability distribution of test statistic Z when H_0 is true)



Trường hợp 2: Mẫu lớn ($n > 40$), X có phân phối chuẩn; σ^2 chưa biết.

Giả thiết không: $H_0 : \mu = \mu_0$

Tiêu chuẩn kiểm định: $z = \frac{\bar{x} - \mu_0}{s} \cdot \sqrt{n}$

Giả thiết đối:

$H_a : \mu > \mu_0$ (kiểm định 1 phía bên phải)

$H_a : \mu < \mu_0$ (kiểm định 1 phía bên trái)

$H_a : \mu \neq \mu_0$ (kiểm định 2 phía)

Miền bác bỏ H_0 với mức ý nghĩa α :

$z \geq z_\alpha$

$z \leq -z_\alpha$

$|z| \geq z_{\alpha/2}$

Trường hợp 3: Mẫu nhỏ, X có phân phối Student; σ^2 chưa biết.

Giả thiết không: $H_0 : \mu = \mu_0$

Tiêu chuẩn kiểm định: $t = \frac{\bar{x} - \mu_0}{s} \cdot \sqrt{n}$

Giả thiết đối:

$H_a : \mu > \mu_0$ (kiểm định 1 phía bên phải)

$H_a : \mu < \mu_0$ (kiểm định 1 phía bên trái)

$H_a : \mu \neq \mu_0$ (kiểm định 2 phía)

Miền bác bỏ H_0 với mức ý nghĩa α :

$t \geq t_{\alpha, n-1}$

$t \leq -t_{\alpha, n-1}$

$|t| \geq t_{\alpha/2, n-1}$

3. Kiểm định giả thiết về tỉ lệ tổng thể p : (cho mẫu lớn $n \cdot p_0 \geq 10$; $n(1 - p_0) \geq 10$)

Giả thiết không: $H_0 : p = p_0$

Tiêu chuẩn kiểm định: $z = \frac{(p - p_0)\sqrt{n}}{\sqrt{p_0(1 - p_0)}}$

Giả thiết đối:

$H_a : p > p_0$ (kiểm định 1 phía bên phải)

$H_a : p < p_0$ (kiểm định 1 phía bên trái)

$H_a : p \neq p_0$ (kiểm định 2 phía)

Miền bác bỏ H_0 với mức ý nghĩa α :

$z \geq z_\alpha$

$z \leq -z_\alpha$

$z \geq z_{\alpha/2}$ hoặc $z \leq -z_{\alpha/2}$

Ví dụ 1: (KĐTB) Trọng lượng trung bình khi xuất chuồng ở một trại chăn nuôi gà công nghiệp năm trước là 2,8kg/con. Năm nay người ta sử dụng một loại thức ăn mới. Cân thử 25 con khi xuất chuồng người ta tính được trung bình mẫu $\bar{x} = 3,2 \text{ kg}$ và phương sai mẫu $s^2 = 0,25$. Với mức ý nghĩa 5%, hãy kết luận về tác dụng của loại thức ăn này có thực sự làm tăng trọng lượng trung bình của đàn gà lên hay không? Đs: $t=4$

Ví dụ 2: (KĐTB) Cân thử 25 con khi xuất chuồng người ta tính được trung bình mẫu $\bar{x} = 3,2 \text{ kg}$ và phương sai mẫu $s^2 = 0,25$. Với mức ý nghĩa 5%, nếu trại chăn nuôi báo cáo trọng lượng trung bình khi xuất chuồng là 3,3 kg/con thì chấp nhận được không?

Đs: $t=-1$

Ví dụ 3: (KĐTB) Theo báo cáo trước đây mức tiêu thụ điện trung bình trong một tháng ở khu phố X là 150kwh. Sau khi thực hiện chương trình tiết kiệm điện, kiểm tra ngẫu nhiên một số hộ ở khu phố này về mức tiêu dùng điện trong một tháng được bảng:

(KW/tháng)	100-120	120-140	140-160	160-180
Số hộ	5	11	9	3

Với $\alpha = 5\%$, hãy kiểm định xem mức tiêu thụ có giảm xuống không? Đs: $z=-3,37$

Ví dụ 4: (KĐTL) Chủ cửa hàng cho rằng tỉ lệ hài lòng của khách hàng với cửa tiệm là 90%. Nghi ngờ điều trên, tiến hành phỏng vấn ngẫu nhiên 500 người thì có 400 người hài lòng. Với mức ý nghĩa 5% hãy kiểm định xem nhận xét của chủ cửa hàng là đúng hay sai? Đs: $z=-7,453$

Ví dụ 5: (KĐTL) Người ta tiến hành điều tra ngẫu nhiên 400 người ở vùng A thì thấy có 22 người ở độ tuổi trưởng thành không biết chữ. Với mức ý nghĩa 2%, có thể cho rằng tỷ lệ dân số ở độ tuổi trưởng thành không biết chữ ở vùng này trên 5% hay không?

Đs: $z=0,46$

Ví dụ 6: (KĐTL) Tỉ lệ phế phẩm của một nhà máy trước đây là 8%. Năm nay nhà máy ứng dụng biện pháp kĩ thuật mới. Để nghiên cứu tác dụng của biện pháp kĩ thuật mới, người ta lấy một mẫu gồm 710 sản phẩm để kiểm tra và thấy có 30 phế phẩm. Với $\alpha = 0,02$, cho rằng kĩ thuật mới làm giảm tỉ lệ phế phẩm được không? Đs: $z = -3,707$

BÀI TẬP

Bài 1: Năng suất lúa trung bình trong những vụ trước là 5,5 tấn/ha. Vụ lúa năm nay người ta áp dụng một biện pháp kĩ thuật mới. Điều tra 100 hecta lúa ta có bảng:

Năng suất	Diện tích (ha)	Năng suất (tạ/ha)	Diện tích (ha)
-----------	----------------	-------------------	----------------

(tạ/ha)			
40-45	7	60-65	20
45-50	12	65-70	8
50-55	18	70-75	5
55-60	27	75-85	3

Với mức ý nghĩa 5%, kết luận xem biện pháp kỹ thuật mới có làm tăng năng suất lúa trung bình của vùng này lên không?

Bài 2: Sản phẩm của nhà máy có phân phối chuẩn với khối lượng trung bình quy định 50kg và độ lệch chuẩn là 0,25 kg. Nghi ngờ dây chuyền sản xuất không bình thường nên tiến hành kiểm tra khối lượng một số sản phẩm được

Khối lượng	49	49,5	50	50,5	51
Số sản phẩm	1	4	12	6	2

Theo bạn nghi ngờ trên đúng hay sai ? Với $\alpha = 5\%$. Đs: $z = 1,6$

Bài 3: Quan sát mức chi tiêu nhu yếu phẩm (triệu đồng/ năm) của một hộ thì thu được bảng:

Chi tiêu	4	6	8	10	12
Số hộ	15	16	20	14	15

Những hộ chi tiêu dưới 7 triệu/ tháng là chi tiêu thấp? Trước đây tỉ lệ chi tiêu thấp là 30%. Hãy kiểm định xem tỷ lệ hộ chi tiêu thấp bây giờ đã tăng lên chưa? Với $\alpha = 5\%$.

Đs: $z = 1,708$

Bài 4: Trọng lượng của một loại sản phẩm do một nhà máy sản xuất là đại lượng ngẫu nhiên có phân phối chuẩn với trọng lượng trung bình là 500gr. Sau một thời gian sản xuất người ta nghi ngờ trọng lượng của loại sản phẩm này có xu hướng giảm sút nên tiến hành cân thử 25 sản phẩm và thu được kết quả cho ở bảng sau:

Trọng lượng (gr)	480	485	490	495	500	510
Số sản phẩm	2	3	8	5	3	4

Với mức ý nghĩa 5%, hãy kết luận điều nghi ngờ trên có đúng hay không? Đs: $t = 3,37$

Bài 5: Khảo sát thu nhập của một số người của một công ty, người ta thu được bảng sau:

Thu nhập (triệu đ/năm)	26-32	32-36	36-40	40-44	44-48	48-54	54-60
Số người	8	12	20	25	20	10	5

Nếu công ty báo cáo mức thu nhập bình quân của một người là 3,6 triệu đ/tháng thì có chấp nhận được không? Kết luận với mức ý nghĩa $\alpha = 4\%$.

Bài 6: Một công ty lớn chuyên sản xuất phần mềm máy tính, cho rằng những người làm việc ở công ty này có thu nhập trung bình 5 triệu đồng/tháng. Lấy mẫu ở công ty được bảng:

Thu nhập (triệu đồng/tháng)	3	4	5	8	10
-----------------------------	---	---	---	---	----

Số người	6	7	8	2	2
----------	---	---	---	---	---

Giả sử thu nhập của những người làm việc ở công ty này có phân phối chuẩn. Với mức ý nghĩa 5%, hãy cho nhận xét về thông tin thu nhập trung bình ở trên có đáng tin hay không? Đs: $t = -0,2959$

Bài 7: Trong 2115 trẻ sơ sinh chọn ngẫu nhiên có 1115 bé trai. Với mức ý nghĩa 5% có thể kết luận mất cân đối giới tính không? Đs: $z = 2,483$

Bài 8: Năm trước tỷ lệ đạt giải của đội tuyển Olympic tỉnh là 70%. Sau khi triển khai phương pháp học tập mới, người ta tiến hành khảo sát kết quả của đội tuyển thì 120 em chọn ngẫu nhiên thì thấy có 30 em bị trượt. Hãy kiểm định xem phương pháp mới có mang lại hiệu quả hơn? Với $\alpha = 5\%$. Đs: $z = 1,195$

Bài 9 : Khảo sát tuổi thọ X (đơn vị: tháng) của một số sản phẩm chọn ngẫu nhiên từ công ty A:

X	6-9	9-12	12-15	15-18	18-21	21-24	24-27
n_i	23	33	55	73	57	42	35

1/ Tìm khoảng tin cậy 98% cho tuổi thọ trung bình của sản phẩm công ty A.

Đs: (16,359;17,697)

2/ Dây chuyền sản xuất công ty A hoạt động bình thường nếu tuổi thọ trung bình của sản phẩm sản xuất ra là 18 tháng. Với mức ý nghĩa 1% hãy xem dây chuyền có hoạt động bình thường không? Đs: $t_{qs} = 3,36$

3/ Công ty A chỉ có lãi khi tỉ lệ sản phẩm phải bảo hành dưới 20%. Có ý kiến đề nghị công ty A bảo hành sản phẩm trong 1 năm. Hãy kết luận về đề nghị này với mức ý nghĩa 5%. Đs: $t_{qs} = 1,07$

Bài 10: Công ty M có 3000 đại lý, cho tiến hành điều tra ngẫu nhiên một số đại lý của mình và thu được bảng số liệu sau (X là doanh số, đơn vị: triệu đồng/tháng), biết X có phân phối chuẩn.

X	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60
Số đại lý	7	12	18	27	22	17	13	4

1/ Những đại lý có $X > 45$ triệu đồng/tháng gọi là đại lý có doanh số cao. Hãy ước lượng số đại lý có doanh số cao với độ tin cậy 95%. Đs: (609;1089)

2/ Có ý kiến cho rằng tỉ lệ đại lý có doanh số cao bằng 1/3 tỉ lệ đại lý có doanh thu còn lại. cho nhận xét về ý kiến này với mức ý nghĩa 1%. Đs: $t_{qs} = 0,83$

3/ Hãy ước lượng doanh số trung bình/tháng của các đại lý với độ tin cậy 99%. Đs(37,4 ; 41,6)

4/ Doanh số trung bình/tháng của các đại lý của công ty N là 35trd/tháng. Hãy so sánh doanh số trung bình/tháng của các đại lý của 2 công ty M,N với mức ý nghĩa 5%. Đs: $t_{qs} = 5,51$

Bài 11: Mức tiêu thụ X của mỗi hộ gia đình vùng A trong mùa khô năm nay có phân phối chuẩn. Điều tra 1 số hộ gia đình vùng A có thống kê sau:

$X(\text{kwh/t})$	65-115	115-165	165-	215-	265-	315-	365-	415-
-------------------	--------	---------	------	------	------	------	------	------

			215	265	315	365	415	465
Số hộ	24	36	75	94	97	125	84	75

1/ Mức tiêu thụ điện trung bình các hộ gia đình vùng A trước là 280 kwh/tháng. Với mức ý nghĩa 2% hãy xét xem mức tiêu thụ điện trung bình các hộ gia đình vùng A năm nay có tăng lên không.

2/ Với mức ý nghĩa 5% so sánh tỉ lệ hộ gia đình có mức tiêu thụ $x \leq 315 \text{ kwh} / t$ với tỉ lệ hộ gia đình có mức tiêu thụ $x > 315 \text{ kwh} / t$ vùng A.

3/ Hộ có $x > 315 \text{ kwh} / t$ là hộ có mức tiêu thụ cao. Hãy ước lượng số hộ có mức tiêu thụ điện cao với độ tin cậy 95%, biết vùng này có 3000 hộ.

4/ Nếu muốn ước lượng mức tiêu thụ điện trung bình các hộ vùng A trong mùa khô năm nay với độ chính xác 10 kwh/tháng thì độ tin cậy bằng ?

Bài 12: P-value nào làm H_0 bị bác bỏ với mức ý nghĩa 0,05?

a/ 0,001 b/ 0,021 c/ 0,078 d/ 0,047 e/ 0,148

Bài 13: Tìm P-value trong các trường hợp sau:

a/ Kiểm định phía phải với $z = 1,42$

b/ Kiểm định 2 phía với $z = 2,1$

c/ Kiểm định phía phải với $z = -1,1$

d/ Kiểm định 2 phía với $z = -0,55$

e/ Kiểm định phải với $df = 8; t=2$

f/ Kiểm định trái với $n = 12; t = -2,5$

g/ Kiểm định 2 phía với $df = 15; t = -1,6$

Bài 14: Quan sát dữ liệu về cường độ của bê tông được mẫu sau:

112,3 97 92,7 86 102 99,2 95,8 103,5 89 86,7 (MPa)

Giả sử bê tông sẽ được sử dụng nếu cường độ trung bình của loại bê tông này lớn hơn 100MPa. Liệu bê tông này có được sử dụng không? Sử dụng kiểm định theo phương pháp p-value

Bài 15: Một mẫu 462 sinh viên trường X có 51 em sử dụng rượu bia thường xuyên. Có thể kết luận chắc chắn rằng 10% tỉ lệ sinh viên sử dụng rượu bia thường xuyên của toàn trường lớn hơn 10% được không? Dùng p-value để đưa ra kết luận.

CHƯƠNG 9: KIỂM ĐỊNH GIẢ THUYẾT 2 MẪU ĐỘC LẬP VỚI MỨC Ý NGHĨA α

TRƯỜNG HỢP	GIẢ THUYẾT H_0	ĐỐI THUYẾT H_a	MIỀN BÁC BỐ GIẢ THUYẾT	P giá trị	Khoảng tin cậy $100(1-\alpha)\%$
1/ Kiểm định hai trung bình của hai tổng thể:					
Tổng thể $N(\mu_1, \sigma_1^2);$ $N(\mu_2, \sigma_2^2)$ $\sigma_1^2; \sigma_2^2$ đã biết	$H_0: \mu_1 - \mu_2 = \Delta_0$ $z = \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$	$\mu_1 - \mu_2 \neq \Delta_0$	$ z \geq z_{\alpha/2}$	$P = 2(1 - \Phi(z))$	$\bar{x} - \bar{y} \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$
		$\mu_1 - \mu_2 > \Delta_0$	$z \geq z_{\alpha}$	$P = (1 - \Phi(z))$	$\left(-\infty; \bar{x} - \bar{y} + z_{\alpha} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}\right)$
		$\mu_1 - \mu_2 < \Delta_0$	$z \leq -z_{\alpha}$	$P = \Phi(z)$	$\left(\bar{x} - \bar{y} - z_{\alpha} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}; \infty\right)$
Cỡ mẫu lớn	$H_0: \mu_1 - \mu_2 = \Delta_0$ $z = \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$	$\mu_1 - \mu_2 \neq \Delta_0$	$ z \geq z_{\alpha/2}$	$P = 2(1 - \Phi(z))$	$\bar{x} - \bar{y} \pm z_{\alpha/2} \sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$
		$\mu_1 - \mu_2 > \Delta_0$	$z \geq z_{\alpha}$	$P = (1 - \Phi(z))$	$\left(-\infty; \bar{x} - \bar{y} + z_{\alpha} \sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}\right)$
		$\mu_1 - \mu_2 < \Delta_0$	$z \leq -z_{\alpha}$	$P = \Phi(z)$	$\left(\bar{x} - \bar{y} + z_{\alpha} \sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}; \infty\right)$
Cỡ mẫu nhỏ, $\sigma_1^2; \sigma_2^2$ chưa biết	$H_0: \mu_1 - \mu_2 = \Delta_0$ $t = \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$	$\mu_1 - \mu_2 \neq \Delta_0$	$ t \geq t_{(\frac{\alpha}{2}; v)}^*$	$P = 2(1 - P(T \leq t))^{**}$	$\bar{x} - \bar{y} \pm t_{(\frac{\alpha}{2}; v)} \sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$
		$\mu_1 - \mu_2 > \Delta_0$	$t \geq t_{(\alpha; v)}$	$P = (1 - P(T \leq t))$	$\left(-\infty; \bar{x} - \bar{y} + t_{(\alpha; v)} \sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}\right)$

		$\mu_1 - \mu_2 < \Delta_0$	$t_{\square} \leq -t_{(\alpha;v)}$	$P = P(T \leq t)$	$\left(\bar{x} - \bar{y} + t_{(\alpha;v)} \sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}; \infty \right)$
2/ Kiểm định hai tỉ lệ về cùng một tính chất của hai tổng thể:					
$H_0: p_1 - p_2 = 0$ $\hat{p}_1 - \hat{p}_2$ $z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p} \cdot \hat{q} \left(\frac{1}{m} + \frac{1}{n} \right)}}$ $\hat{p}_1 = \frac{x}{m}; \hat{p}_2 = \frac{y}{n}; \hat{p} = \frac{x+y}{m+n}$ $\hat{q} = 1 - \hat{p}$	$p_1 \neq p_2$	$ z \geq z_{\alpha/2}$	$P = 2(1 - \Phi(z))$		
	$p_1 > p_2$	$z \geq z_{\alpha}$	$P = (1 - \Phi(z_{\square}))$		
	$p_1 < p_2$	$z_{\square} \leq -z_{\alpha}$	$P = \Phi(z_{\square})$		
3/ Kiểm định hai phương sai của hai tổng thể:					
$\sigma_1^2 = \sigma_2^2$ $f_{\square} = \frac{s_1^2}{s_2^2}$	$\sigma_1^2 \neq \sigma_2^2$	$f \geq F_{(\alpha/2, m-1, n-1)}$ hay $f \leq F_{(1-\alpha/2, m-1, n-1)}$	$P = 2(1 - P(F \leq f))$ (***)		
	$\sigma_1^2 > \sigma_2^2$	$f_0 \geq F_{(\alpha, m-1, n-1)}$	$P = (1 - P(F \leq f))$		
	$\sigma_1^2 < \sigma_2^2$	$f_{\square} \leq F_{(1-\alpha, m-1, n-1)}$	$P = P(F \leq f)$		

$$(*) = \frac{\left(\frac{s_1^2}{m} + \frac{s_2^2}{n} \right)^2}{\frac{(s_1^2/m)^2}{m-1} + \frac{(s_2^2/n)^2}{n-1}}; v \text{ được làm tròn xuống số nguyên gần nhất.}$$

Vd: $m_X = 10$; $n_Y = 10$; $s_X = 0,79$; $s_Y = 3,59 \Rightarrow v = 9,87 \Rightarrow v = 9 \Rightarrow t_{\alpha/2; v} = t_{0,25; 9} = 2,262$

$$(**) T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}}} \quad (***) F = \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2}$$

KIỂM ĐỊNH GIẢ THUYẾT 2 MẪU GHEP CẶP của 1 tổng thể VỚI MỨC Ý NGHĨA α

Thực hiện tương tự như kiểm định trên một mẫu với $X - Y$. Giả thuyết $\mu_D = \Delta_0$.

Khoảng tin cậy của $\mu_1 - \mu_2$ là khoảng tin cậy của μ_D .

CHƯƠNG 12: HỎI QUY TUYẾN TÍNH ĐƠN GIẢN và TƯƠNG QUAN

* HƯỚNG DẪN DÙNG MÁY TÍNH TÍNH TOÁN CÁC BÀI TOÁN HỎI QUY

Máy tính 570 ES:

B1: (gọi cột tần số) Shift Mode  4 1.on

B2: Mode 3.STAT 2.

Nhập dữ liệu rồi nhấn AC để ra ngoài.

B3: Shift 1. 7.Reg 1.A

2. B

3. r: hệ số tương quan

Chú ý phương trình hồi quy tuyến tính là $Y = A + BX$

Ví dụ 1: (KĐ 2 trung bình) Người quản lý công ty quan sát 75 buổi sáng đếm số sản phẩm sản xuất được trong mỗi buổi và tính được trung bình mẫu $\bar{x} = 806$ (sản phẩm/buổi) và độ lệch chuẩn mẫu hiệu chỉnh 185. Quan sát 100 buổi chiều và tính được trung bình mẫu $\bar{y} = 723$ (sản phẩm/ buổi) và độ lệch chuẩn mẫu hiệu chỉnh 164.

a/ Có ý kiến cho rằng làm việc buổi sáng hiệu quả hơn buổi chiều. Hãy cho nhận xét với mức ý nghĩa 1% ?

b/ Hãy tìm khoảng ước lượng cho lượng sản phẩm chênh lệch của hai buổi với độ tin cậy 96%?

Ví dụ 2: (KĐ 2 trung bình) Để nghiên cứu tác dụng của việc bón phân đạm theo công thức A đối với sản lượng của bắp, người ta làm thí nghiệm trên các mảnh đất . Quan sát sản lượng thu được trên các mảnh đối chứng (không bón đạm) và các mảnh có bón phân đạm theo công thức A được bảng sau

Sản lượng (tạ/ha)	55	53	30	37	49
Mảnh đối chứng	4	5	6	3	5

Sản lượng (tạ/ha)	60	58	29	39	47	35
Mảnh bón phân	4	7	3	5	8	3

Hãy cho kết luận về hiệu quả của việc bón phân đạm theo công thức A, với mức ý nghĩa 0.05

Ví dụ 3 : (KĐ 2 tỉ lệ) Kiểm tra chất lượng nón bảo hiểm do 2 nhà máy A, B sản xuất được kết quả sau: trong số 500 nón bảo hiểm của nhà máy A, có 95 nón không đạt tiêu chuẩn. Trong số 400 nón của nhà máy B có 95 nón không đạt tiêu chuẩn. Với mức ý nghĩa 3% hãy cho kết luận về chất lượng nón bảo hiểm của 2 nhà máy A, B.

Ví dụ 4: (KĐ 2 tỉ lệ) Trong 500 sv nam có 45 sv đạt loại giỏi. Trong 400 sv nữ có 50 sv đạt loại giỏi. Với mức ý nghĩa 2%, kết luận tỉ lệ giỏi của nam cao hơn nữ được không? Đs :

$$z = 1,69$$

Ví dụ 5 : (KĐ 2 phương sai) Cho một mẫu A có 28 phần tử với độ lệch chuẩn mẫu là 52,6. Một mẫu B có 26 phần tử, độ lệch chuẩn mẫu là 84,2 . So sánh độ lệch chuẩn tổng thể của hai mẫu trên với mức ý nghĩa 2%? Đs : 0,39 (bảng Fisher)

Ví dụ 6: (Hồi quy) Quan sát việc tổng hợp sinh khối ở một nhà máy từ năng lượng bức xạ mặt trời sau 8 tuần người ta thu được bảng số liệu sau:

Bức xạ mặt trời	30	68	121	217	314	419	536	642
Trọng lượng sinh khối (gram)	17	49	122	220	376	571	648	756

Dựa vào số liệu này có thể dự đoán được trọng lượng sinh khối qua bức xạ mặt trời bằng hàm hồi quy tuyến tính thực nghiệm hay không? Nếu được hãy dự báo xem khi bức xạ mặt trời ở mức 600 thì trung bình sinh khối được sản xuất là bao nhiêu?

Ví dụ 7: (Hồi quy) Để nghiên cứu sự phát triển của một loại cây trồng, người ta tiến hành đo chiều cao Y (m) và đường kính X (cm) của một số cây. Kết quả được ghi ở bảng sau đây:

Y	3	4	5	6	7	8
X						
21	2	5				
23		3	11			
25			8	15	10	
27			4	17	3	
29					7	12

Tìm hệ số tương quan mẫu và phương trình hồi quy tuyến tính mẫu Y theo X.

BÀI TẬP

Bài 1: Khảo sát chiều cao của sinh viên trường A, B ta có

Chiều cao (m)	1,5-1,55	1,55-1,6	1,6-1,65	1,65-1,7	1,7-1,75	1,75-1,8	1,8-1,85	1,85-1,9
Số SV (A)	15	38	56	68	70	56	31	12
Số SV (B)	21	43	60	78	71	58	29	10

- So sánh chiều cao trung bình của sinh viên 2 trường A, B với mức ý nghĩa 3%.
- Có ý kiến cho rằng chiều cao trung bình của sinh viên trường A cao hơn trường B. **Hãy** cho nhận xét với mức ý nghĩa 5%.

Bài 2: Nghiên cứu khả năng chống cảm cúm của Vitamin C, có kết quả sau: Trong số 420 người không uống Vitamin C, có 93 người bị cảm cúm. Trong số 417 người, mỗi ngày uống 1g

Vitamin C/ mỗi người, có 51 người bị cảm cúm. Với mức ý nghĩa 1% có thể cho rằng Vitamin C có khả năng chống cảm cúm hay không? Đs: 3,78

Bài 3: Giả thuyết rằng thời gian sử dụng điện thoại loại A, B có phân phối chuẩn. Quan sát thời gian sử dụng một số điện thoại A, B ta có số liệu

Thời gian (h)	5-6	6-7	7-8	8-9	9-10	10-11	11-12
Số điện thoại A	1	2	5	7	6	3	1
Số điện thoại B	1	3	4	5	3	2	1

Hãy so sánh thời gian sử dụng trung bình của 2 loại điện thoại này với mức ý nghĩa 5%.

Bài 4: Theo dõi thu nhập, chi tiêu (triệu đồng/tháng) của một số hộ gia đình trong vùng A có số liệu.

Thu Nhập	15	18	19	21	23	27	29	19	17	24	22	28	35	38	40
Chi tiêu	12	15	15	17	21	25	22	18	17	21	18	21	30	25	26
Số dư	3	3	4	4	2	2	7	1	0	3	4	7	5	13	14

Giả sử thu nhập, chi tiêu trong 1 tháng của mỗi hộ gia đình là các biến ngẫu nhiên có phân phối chuẩn. a/ Tìm khoảng tin cậy 95% cho số tiền dư trung bình trong một tháng của mỗi hộ gia đình.

b/ Có ý kiến cho rằng số tiền dư trung bình của mỗi hộ trong 1 tháng là 4 triệu đồng. Hãy cho nhận xét về ý kiến này với mức ý nghĩa 5%.

Bài 5: Quan sát trọng lượng của 150 trẻ sơ sinh ở thành phố có 100 trẻ có trọng lượng trên 3000 gam; trong 200 trẻ sơ sinh ở nông thôn có 98 trẻ có trọng lượng trên 3000 gam. Hãy so sánh tỷ lệ trẻ sơ sinh có trọng lượng trên 3000 gam ở thành phố và nông thôn với mức ý nghĩa 5%.

Bài 6: Một công ty ấn định giá bán X của một loại sản phẩm tại 10 miền khác nhau. Bảng sau đây cho biết số lượng Y bán được trong một tháng ứng với giá bán:

X	34	35	36	36	35	37	38	40	39	40
Y	6	5,9	5,7	5,7	6,2	6,7	5,6	5,5	5,4	5,2

a) Có thể biểu diễn số lượng theo giá bán bằng phương trình hồi quy tuyến tính không? Vì sao?

b) Viết phương trình đường thẳng hồi quy mẫu của Y theo X.

Bài 7: Giả sử giá trị quan sát trên một mẫu của (X,Y) tuân theo quy luật phân phối chuẩn hai chiều được cho trong bảng sau:

X	1	3	4	6	8	9	11	14
---	---	---	---	---	---	---	----	----

Y	1	2	4	4	5	7	8	9
---	---	---	---	---	---	---	---	---

Viết phương trình đường thẳng hồi quy mẫu của Y theo X. Dự đoán giá trị Y khi X=12.

Đs: $y = 0,6364 x + 0,5455$; $y = 8,1823$

Bài 8: Đo chiều cao X (đơn vị: cm) và trọng lượng Y (kg) của một số học sinh chọn ngẫu nhiên được

X	155	156	158	159	159	160	160	162	164	165
Y	48	47	48	49	48	50	51	51	53	54

Hãy viết hàm hồi quy tuyến tính thực nghiệm của Y theo X và tính hệ số tương quan mẫu giữa X và Y

Bài 7:

Observe a sample (X,Y):

X	1	3	4	6	8	9	11	14
Y	1	2	4	4	5	7	8	9

Find the linear regression equation of X and Y? When X=12, find Y

Answer: $y = 0,6364 x + 0,5455$; $y = 8,1823$

Tham khảo

BT: gồm 16 bài

1/ Một bài báo trong Báo cáo tiêu dùng tháng 11 năm 1983 so sánh các loại pin khác nhau. Tuổi thọ trung bình của pin Duracell Alkaline AA và pin Eveready Energizer Alkaline AA lần lượt là 4,1 giờ và 4,5 giờ. Giả sử đây là tuổi thọ trung bình của tổng thể.

- a. Cho $\bar{x}$ là tuổi thọ trung bình của mẫu 100 pin Duracell và tuổi thọ trung bình của mẫu là 100 pin Eveready. Hãy tìm giá trị trung bình của $\bar{x} - \bar{y}$ (nghĩa là, phân phối của $\bar{x} - \bar{y}$ sẽ tập trung ở đâu?)
- b. Giả sử độ lệch chuẩn của tổng thể là 1,8 giờ cho pin Duracell và 2,0 giờ cho pin Eveready. Với các mẫu được nêu trong phần (a), tìm phương sai và độ lệch chuẩn của thống kê $\bar{x} - \bar{y}$?

2/ Báo cáo thống kê sức khỏe quốc gia ngày 22 tháng 10 năm 2008, bao gồm các thông tin sau đây về chiều cao (in.) cho phụ nữ da trắng không phải gốc Tây Ban Nha:

Age	Sample Size	Sample Mean	Std. Error Mean
20–39	866	64.9	.09
60 and older	934	63.1	.11

a. Tìm khoảng tin cậy với độ tin cậy 95% cho hiệu giữa chiều cao trung bình tổng thể của phụ nữ trẻ và phụ nữ lớn tuổi.

b. Cho μ_1 là chiều cao trung bình tổng thể những người độ tuổi 20–39 và μ_2 là chiều cao tổng thể cho những người độ tuổi từ 60 tuổi trở lên. Kiểm định giả thiết : $H_0 : \mu_1 - \mu_2 = 1$; $H_a : \mu_1 - \mu_2 > 1$ với mức ý nghĩa 0,001, sử dụng phương pháp miền bác bỏ.

c. Giá trị P-value cho kiểm định thực hiện trong (b) là bao nhiêu? Dựa trên giá trị P-value này, có bác bỏ giả thuyết không ở bất kỳ mức ý nghĩa hợp lý nào không? Giải thích.

3/ Cho μ_1 kí hiệu tuổi thọ trung bình thực sự cho một thương hiệu lốp radial cao cấp P205 / 65R15, và μ_2 kí hiệu tuổi thọ trung bình thực sự cho một thương hiệu kinh tế có cùng kích thước. Kiểm định $H_0 : \mu_1 - \mu_2 = 5000$; $H_a : \mu_1 - \mu_2 > 5000$ với mức ý nghĩa 0,01. Cho biết $m = 45, \bar{x} = 42500, s_1 = 2200, n = 45, \bar{y} = 36800, s_2 = 1500$

4/ Những người có hội chứng Reynaud có khuynh hướng bị suy giảm đột ngột tuần hoàn máu ở ngón tay và ngón chân. Trong một thí nghiệm để nghiên cứu mức độ suy giảm này, mỗi đối tượng nhúng một ngón trỏ trong nước rồi đo sản nhiệt (cal / cm² / min). Đối với m=10 đối tượng có hội chứng, sản lượng nhiệt trung bình là $\bar{x} = 0,64$, và cho n=10 đối tượng không có hội chứng, sản lượng trung bình là 2,05. Gọi μ_1, μ_2 kí hiệu cho sản nhiệt trung bình thực sự cho hai đối tượng trên. Giả sử rằng phân phối sản nhiệt của hai loại đối tượng trên có phân phối chuẩn $\sigma_1 = 0,2$; $\sigma_2 = 0,4$.

- Kiểm định giả thiết $H_0 : \mu_1 - \mu_2 = -1$; $H_a : \mu_1 - \mu_2 < -1$ với mức ý nghĩa 0,1 .
- Tính giá trị P-value cho giá trị của Z thu được trong phần (a).
- Tính xác suất sai lầm loại II khi $\mu_1 - \mu_2 = -1,2$?
- Giả sử rằng m=n, cỡ mẫu sẽ là bao nhiêu để $\beta = 0,1$ khi $\mu_1 - \mu_2 = -1,2$?

5/

Một thí nghiệm để so sánh độ bền liên kết sức căng của vữa có thể biến đổi (modified mortar) và vữa không biến đổi cho $\bar{x} = 18,12 \text{ kgf / cm}^2$; $m = 40$ và $y = 16,87 \text{ kgf / cm}^2$; $n = 32$ tương ứng. Gọi μ_1, μ_2 kí hiệu cho độ bền liên kết sức căng thật sự cho vữa có thể biến đổi và vữa không biến đổi, và giả sử độ bền liên kết sức căng của hai biến trên có phân phối chuẩn.

a. Giả sử rằng $\sigma_1 = 1,6$ và $\sigma_2 = 1,4$. Hãy kiểm định

$H_0 : \mu_1 - \mu_2 = 0$; $H_a : \mu_1 - \mu_2 > 0$ với mức ý nghĩa 0,01.

b. Tính xác suất sai lầm loại II của kiểm định câu (a) khi $\mu_1 - \mu_2 = 1$.

c. Giả sử người quan sát sử dụng độ tin cậy 0,05 và $\beta = 0,1$ khi $\mu_1 - \mu_2 = 1$.

Nếu $m=40$ thì giá trị n cần thiết là bao nhiêu?

d. Phân tích và kết luận của câu (a) thay đổi như thế nào $\sigma_1; \sigma_2$ chưa biết nhưng $s_1 = 1,6$ và $s_2 = 1,4$.

6/ Bài viết “Có sự khác biệt về giảng dạy toàn thời gian và bán thời gian?”(Cao đẳng

Giảng dạy, 2009: 23–26) báo cáo rằng cho một mẫu 125 các khóa học do giảng viên dạy toàn thời gian, điểm GPA trung bình là 2.7186 và độ lệch chuẩn là .63342, trong khi đối với mẫu của 88 khóa học được giảng dạy bán thời gian, trung bình và độ lệch chuẩn là 2.8639 và .49241, tương ứng.

Liệu có sự khác biệt điểm GPA trung bình thực sự giữa dạy bán thời gian với dạy toàn thời gian không? Kiểm định với mức ý nghĩa 0,01 bởi một giá trị P-value có được đầu tiên.