ÔN THI CAO HỌC
MÔN TOÁN KINH TẾ

(GV: Trần Ngọc Hội - 2011)

BÀI GIẢI
PHẦN III: THỐNG KÊ

Bài 1. Để khảo sát chiều cao X của một giống cây trồng, người ta quan sát một mẫu và có kết qủa sau:

X(cm)	95-105	105-115	115-125	125-135	135-145	145-155	155-165
Số cây	10	10	15	30	10	10	15

a) Ước lượng chiều cao trung bình của giống cây trồng trên với độ tin cậy 96%. Với độ tin cậy đó, chiều cao trung bình tối đa của giống cây trồng trên là bao nhiêu? Tối thiểu là bao nhiêu?

b) Nếu muốn ước lượng chiều cao trung bình của giống cây trồng trên với độ tin cậy 99% và độ chính xác 4 cm thì cần phải điều tra thêm bao nhiêu cây nữa?

c) Nếu ước lượng chiều cao trung bình của giống cây trồng trên với độ chính xác 4,58cm thì sẽ đạt được độ tin cậy là bao nhiêu?

d) Một tài liệu thống kê cũ cho rằng chiều cao trung bình của giống cây trồng trên là 127cm. Hãy cho kết luận về tài liệu đó với mức ý nghĩa 1%.

e) Những cây trồng có chiều cao từ 135cm trở lên được gọi là những cây "cao". Hãy ước lượng tỉ lệ cây cao với độ tin cậy 95%. Với độ tin cậy đó, tỉ lệ cây cao đạt giá trị tối đa là bao nhiêu? Tối thiểu là bao nhiêu?

f) Nếu ước lượng tỉ lệ cây cao với độ chính xác 10% thì sẽ đạt được độ tin cậy là bao nhiêu?

g) Nếu ước lượng tỉ lệ cây cao với độ tin cậy 95% và độ chính xác 11% thì cần phải điều tra thêm bao nhiêu cây nữa?

h) Trước đây, tỉ lệ cây cao của loại cây trồng trên là 40%. Các số liệu trên thu thập được sau khi đã áp dụng một kỹ thuật mới. Hãy cho kết luận về kỹ thuật mới với mức ý nghĩa 5%.

i) Những cây trồng có chiều cao từ 105cm đến 125cm được gọi là những cây loại A. Hãy ước lượng chiều cao trung bình của những cây loại A với độ tin cậy 95% (GS X có phân phối chuẩn).

j) Bằng phương pháp mới, sau một thời gian người ta thấy chiều cao trung bình của những cây loại A là 119,5cm. Hãy cho kết luận về phương pháp mới với mức ý nghĩa 1% (GS X có phân phối chuẩn).

k) Giả sử X có phân phối chuẩn. Với độ tin cậy 95%, hãy ước lượng phương sai của X trong hai trường hợp :
α) Biết kỳ vọng của X là 130 cm.
β) Chưa biết kỳ vọng của X.

l) Khi canh tác bình thường thì phương sai của chiều cao X là 300cm^2. Hãy nhận định về tình hình canh tác với mức ý nghĩa 5% (GS X có phân phối chuẩn).

Lời giải

Ta có:

$$n = 100;\ \sum X_i n_i = 13100;\ \sum X_i^2 n_i = 1749000.$$

- Kỳ vọng mẫu của X là $\overline{X} = \frac{1}{n}\sum X_i n_i = 131(\text{cm}).$
- Phương sai mẫu của X là:
$$\hat{S}^2 = \frac{1}{n}\sum X_i^2 n_i - \overline{X}^2 = (18,1384)^2(\text{cm}^2).$$
- Phương sai mẫu hiệu chỉnh của X là:
$$S^2 = \frac{n}{n-1}\hat{S}^2 = (18,2297)^2(\text{cm}^2).$$

a) Ước lượng chiều cao trung bình của giống cây trồng trên với độ tin cậy 96%. Với độ tin cậy đó, chiều cao trung bình tối đa của giống cây trồng trên là bao nhiêu? Tối thiểu là bao nhiêu?

Đây là bài toán ước lượng khoảng cho kỳ vọng μ = M(X) với độ tin cậy γ = 1 – α = 96% = 0,96.

Vì n ≥ 30, σ^2 = D(X) chưa biết nên ta có công thức ước lượng khoảng cho kỳ vọng:

$$(\overline{X} - z_\alpha \frac{S}{\sqrt{n}}; \overline{X} + z_\alpha \frac{S}{\sqrt{n}}),$$

trong đó $\varphi(z_\alpha) = \gamma/2 = 0,96/2 = 0,48$. Tra bảng giá trị hàm Laplace ta được z_α = 2,06. Vậy ước lượng khoảng là:

$$(131 - 2,06\frac{18,2297}{\sqrt{100}}; 131 + 2,06\frac{18,2297}{\sqrt{100}}) = (127,2447; 134,7553).$$

Nói cách khác, với độ tin cậy 96%, chiều cao trung bình của một cây nằm trong khoảng từ 127,2447cm đến 134,7553cm.

Để biết chiều cao trung bình tối đa của giống cây trồng trên là bao nhiêu ta cần ước lượng khoảng bên trái cho kỳ vọng μ = M(X).

Vì n ≥ 30, σ^2 = D(X) chưa biết nên ta có công thức ước lượng khoảng bên trái cho kỳ vọng:

$$(-\infty; \overline{X} + z_{2\alpha}\frac{S}{\sqrt{n}}),$$

trong đó $\alpha = 1 - \gamma = 1 - 0,96 = 0,04$, $\varphi(z_{2\alpha}) = (1 - 2\alpha)/2 = (1 - 2.0,04)/2 = 0,92/2 = 0,46$. Tra bảng giá trị hàm Laplace ta được $z_{2\alpha} = 1,75$. Suy ra chiều cao trung bình tối đa là:

$$\overline{X} + z_{2\alpha}\frac{S}{\sqrt{n}} = 131 + 1,75\frac{18,2297}{\sqrt{100}} = 134,1902(cm).$$

Vậy với độ tin cậy 96%, chiều cao trung bình tối đa của giống cây trồng trên là 134,1902cm.

Để biết chiều cao trung bình tối thiểu của giống cây trồng trên là bao nhiêu ta cần ước lượng khoảng bên phải cho kỳ vọng $\mu = M(X)$.

Vì $n \ge 30$, $\sigma^2 = D(X)$ chưa biết nên ta có công thức ước lượng khoảng bên trái cho kỳ vọng:

$$(\overline{X} - z_{2\alpha}\frac{S}{\sqrt{n}}; +\infty),$$

trong đó $z_{2\alpha} = 1,75$. Suy ra chiều cao lượng trung bình tối thiểu là:

$$\overline{X} - z_{2\alpha}\frac{S}{\sqrt{n}} = 131 - 1,75\frac{18,2297}{\sqrt{100}} = 127,8098(cm).$$

Vậy với độ tin cậy 96%, trọng lượng trung bình tối thiểu của loại vật nuôi trên là 127,8098cm.

b) Nếu muốn ước lượng chiều cao trung bình của giống cây trồng trên với độ tin cậy 99% và độ chính xác 4cm thì cần phải điều tra thêm bao nhiêu cây nữa?

Đây là bài toán xác định cỡ mẫu khi ước lượng kỳ vọng của chỉ tiêu X với độ chính xác $\varepsilon = 4$cm và độ tin cậy $\gamma = 99\% = 0,99$.

Vì $n \ge 30$, $\sigma^2 = D(X)$ chưa biết nên ta có công thức tính độ chính xác của ước lượng:

$$\varepsilon = z_\alpha\frac{S}{\sqrt{n}},$$

trong đó $\varphi(z_\alpha) = \gamma/2 = 0,99/2 = 0,495$. Tra bảng giá trị hàm Laplace ta được $z_\alpha = 2,58$. Suy ra

$$n = \left(\frac{z_\alpha S}{\varepsilon}\right)^2.$$

Thực tế yêu cầu:

$$n \ge \left(\frac{z_\alpha S}{\varepsilon}\right)^2 = \left(\frac{2,58.18,2297}{4}\right)^2 \approx 138,254.$$

Giá trị n nguyên nhỏ nhất thỏa bất đẳng thức trên là $n_1 = 139$. Vì $n_1 = 139 > 100$ (100 là cỡ mẫu đang có) nên ta cần điều tra thêm ít nhất là $139 - 100 = 39$ cây nữa.

c) Nếu ước lượng chiều cao trung bình của giống cây trồng trên với độ chính xác 4,58cm thì sẽ đạt được độ tin cậy là bao nhiêu?

Đây là bài toán xác định độ tin cậy $\gamma = 1 - \alpha$ khi ước lượng kỳ vọng của chỉ tiêu X với độ chính xác $\varepsilon = 4,58$cm.

Vì $n \ge 30$, $\sigma^2 = D(X)$ chưa biết nên ta có công thức tính độ chính xác của ước lượng:

$$\varepsilon = z_\alpha\frac{S}{\sqrt{n}},$$

trong đó $\varphi(z_\alpha) = \gamma/2$. Suy ra

$$z_\alpha = \frac{\varepsilon\sqrt{n}}{S} = \frac{4,58.\sqrt{100}}{18,2297} = 2,5123.$$

Tra bảng giá trị hàm Laplace ta được độ tin cậy là

$$\gamma = 2\varphi(z_\gamma) = 2\varphi(2,5123) = 2\varphi(2,52) = 2.0,4941 = 98,82\%.$$

d) Một tài liệu thống kê cũ cho rằng chiều cao trung bình của giống cây trồng trên là 127cm. Hãy cho kết luận về tài liệu đó với mức ý nghĩa 1%.

Đây là bài toán kiểm định giả thiết về kỳ vọng $\mu = M(X)$ với mức ý nghĩa $\alpha = 1\% = 0,01$:

H_0: $\mu = 127$ với giả thiết đối H_1: $\mu \neq 127$.

Vì $n \ge 30$; σ^2 chưa biết, nên ta có qui tắc kiểm định như sau:

Bước 1: Ta có

$$z = \frac{(\overline{X} - \mu_0)\sqrt{n}}{S} = \frac{(131 - 127)\sqrt{100}}{18,2297} = 2,1942.$$

Bước 2: Tra bảng giá trị hàm Laplace để tìm z_α thoả

$$\varphi(z_\alpha) = (1 - \alpha)/2 = 0,99/2 = 0,495$$

ta được $z_\alpha = 2,58$.

Bước 3: Kiểm định. Vì $|z| = 2,1942 < 2,58 = z_\alpha$ nên ta chấp nhận H_0: $\mu = 127$.

Kết luận: Với mức ý nghĩa 1%, tài liệu cũ về chiều cao trung bình của giống cây trồng trên còn phù hợp với thực tế.

e) Những cây trồng có chiều cao từ 135cm trở lên được gọi là những cây "cao". Hãy ước lượng tỉ lệ cây cao với độ tin cậy 95%. Với độ tin cậy đó, tỉ lệ cây cao đạt giá trị tối đa là bao nhiêu? Tối thiểu là bao nhiêu?

Đây là bài toán ước lượng khoảng cho tỉ lệ p các cây cao với độ tin cậy $\gamma = 1 - \alpha = 95\% = 0,95$. Ta có công thức ước lượng khoảng :

$$(F_n - z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}}; F_n + z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}}),$$

trong đó $\varphi(z_\alpha) = \gamma/2 = 0,95/2 = 0,475$. Tra bảng giá trị hàm Laplace ta được $z_\alpha = 1,96$. Trong $n = 100$ cây có $m = 10 + 10 + 15 = 35$ cây có chiều cao từ 135cm trở lên nên tỉ lệ mẫu các cây cao là $F_n = 35/100 = 0,35$. Vậy ước lượng khoảng là:

$$(0,35 - 1,96\sqrt{\frac{0,35(1-0,35)}{100}}; 0,35 + 1,96\sqrt{\frac{0,35(1-0,35)}{100}}) = (25,65\%; 44,35\%).$$

Nói cách khác, với độ tin cậy 95%, tỉ lệ cây cao nằm trong khoảng từ 25,65% đến 44,35%.

Để biết tỉ lệ cây cao đạt giá trị tối đa là bao nhiêu ta cần ước lượng khoảng bên trái cho tỉ lệ p các cây cao. Ta có công thức ước lượng khoảng bên trái cho tỉ lệ:

$$(-\infty; F_n + z_{2\alpha}\sqrt{\frac{F_n(1-F_n)}{n}}),$$

trong đó α = 1−γ = 1 – 0,95 = 0,05, $\varphi(z_{2\alpha})$ = (1– 2α)/2 = (1– 2.0,05)/2 = 0,90/2 = 0,45. Tra bảng giá trị hàm Laplace ta được $z_{2\alpha}$ = 1,65. Suy ra giá trị tối đa của tỉ lệ cây cao là:

$$F_n + z_{2\alpha}\sqrt{\frac{F_n(1-F_n)}{n}} = 0,35 + 1,65\sqrt{\frac{0,35(1-0,35)}{100}} = 0,4287.$$

Vậy với độ tin cậy 95%, giá trị tối đa của tỉ lệ cây cao là 42,87%.

Để biết tỉ lệ cây cao đạt giá trị tối thiểu là bao nhiêu ta cần ước lượng khoảng bên phải cho tỉ lệ p các cây cao. Ta có công thức ước lượng khoảng bên trái cho tỉ lệ:

$$(F_n - z_{2\alpha}\sqrt{\frac{F_n(1-F_n)}{n}}; +\infty),$$

trong đó $z_{2\alpha}$ = 1,65. Suy ra giá trị tối thiểu của tỉ lệ cây cao là:

$$F_n - z_{2\alpha}\sqrt{\frac{F_n(1-F_n)}{n}} = 0,35 - 1,65\sqrt{\frac{0,35(1-0,35)}{100}} = 0,2713.$$

Vậy với độ tin cậy 95%, giá trị tối thiểu của tỉ lệ cây cao là 27,13%.

f) Nếu ước lượng tỉ lệ những những cây "cao" với độ chính xác 10% thì sẽ đạt được độ tin cậy là bao nhiêu?

Đây là bài toán xác định độ tin cậy khi lượng tỉ lệ cây cao với độ chính xác ε = 10% = 0,1.

Ta có công thức tính độ chính xác của ước lượng:

$$\varepsilon = z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}},$$

trong đó $\varphi(z_\alpha) = \gamma/2$. Ta có tỉ lệ mẫu cây cao là: $F_n = 0,35$. Suy ra

$$z_\alpha = \varepsilon\sqrt{\frac{n}{F_n(1-F_n)}} = 0,1.\sqrt{\frac{100}{0,35(1-0,35)}} = 2,0966.$$

Tra bảng giá trị hàm Laplace ta được độ tin cậy là

$$\gamma = 2\varphi(z_\alpha) = 2\varphi(2,0966) = 2\varphi(2,1) = 2.0,4821 = 96,42\%.$$

g) Nếu ước lượng tỉ lệ những những cây "cao" với độ tin cậy 95% và độ chính xác 11% thì cần phải điều tra thêm bao nhiêu cây nữa?

Đây là bài toán xác định cỡ mẫu khi ước lượng tỉ lệ cây cao với độ chính xác ε = 11% = 0,11 và độ tin cậy γ = 1 – α = 95% = 0,95.

Ta có công thức tính độ chính xác của ước lượng:

$$\varepsilon = z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}},$$

trong đó $\varphi(z_\alpha) = \gamma/2 = 0,95/2 = 0,475$. Tra bảng giá trị hàm Laplace ta được z_α = 1,96. Suy ra

$$n = \frac{z_\alpha^2 F_n(1-F_n)}{\varepsilon^2} = \frac{1,96^2.0,35(1-0,35)}{0,11^2} \approx 72,23.$$

Thực tế yêu cầu $n \ge \lceil 72,23 \rceil = 73$. Vì $n_1 = 73 < 100$ (100 là cỡ mẫu đang có) nên ta không cần điều tra thêm cây nào nữa.

h) Trước đây, tỉ lệ cây cao của loại cây trồng trên là 40%. Các số liệu trên thu thập được sau khi đã áp dụng một kỹ thuật mới. Hãy cho kết luận về kỹ thuật mới với mức ý nghĩa 5%.

Đây là bài toán kiểm định giả thiết về tỉ lệ p các cây cao với mức ý nghĩa α = 5% = 0,05:

H_0: p = 40% = 0,4 với giả thiết đối H_1: p ≠ 0,4.

Ta có qui tắc kiểm định như sau:

Bước 1: Ta có

$$z = \frac{(F_n - p_0)\sqrt{n}}{\sqrt{p_0 q_0}} = \frac{(0,35-0,4)\sqrt{100}}{\sqrt{0,4(1-0,4)}} = -1,0206.$$

Bước 2: Tra bảng giá trị hàm Laplace để tìm z_α thoả

$$\varphi(z_\alpha) = (1-\alpha)/2 = 0,95/2 = 0,475$$

ta được z_α = 1,96.

Bước 3: Kiểm định. Vì $|z| = 1,0206 < 1,96 = z_\alpha$ nên ta chấp nhận giả thiết H_0: p = 0,4.

Kết luận: Với mức ý nghĩa 5%, phương pháp mới không có tác dụng làm thay đổi tỉ lệ các cây cao.

i) Đây là bài toán ước lượng khoảng cho kỳ vọng $\mu_A = M(X_A)$ của chiều cao X = X_A của những cây loại A với độ tin cậy γ = 1 – α = 95% = 0,95.

Ta lập bảng số liệu của X_A:

X_{Ai}	110	120
N_{Ai}	10	15

Từ bảng trên ta tính được:

$$n_A = 25; \sum X_{Ai} n_{Ai} = 2900; \sum X_{Ai}^2 n_{Ai} = 337000.$$

Kỳ vọng mẫu của X_A là

$$\overline{X}_A = \frac{1}{n}\sum X_{Ai} n_{Ai} = 116 (cm).$$

Phương sai mẫu của X_A là:

$$\hat{S}_A^2 = \frac{1}{n}\sum X_{Ai}^2 n_{Ai} - \overline{X}_A^2 = (4,8990)^2 (cm^2).$$

Phương sai mẫu hiệu chỉnh của X_A là:

$$S_A^2 = \frac{n_A}{n_A - 1}\hat{S}_A^2 = 5^2(cm^2).$$

Vì $n_A < 30$, X_A có phân phối chuẩn, $\sigma_A{}^2 = D(X_A)$ chưa biết, nên ta có công thức ước lượng khoảng cho kỳ vọng:

$$(\overline{X}_A - t_\alpha^k \frac{S_A}{\sqrt{n_A}}; \overline{X}_A + t_\alpha^k \frac{S_A}{\sqrt{n_A}})$$

trong đó t_α^k được xác định từ bảng phân phối Student với $k = n_A - 1 = 24$ và $\alpha = 1 - \gamma = 1 - 0,95 = 0,05$. Tra bảng phân phối Student ta được $t_\alpha^k = 2,064$.

Vậy ước lượng khoảng là:

$$(116 - 2,064\frac{5}{\sqrt{25}}; 116 + 2,064\frac{5}{\sqrt{25}}) = (113,936; 118,064).$$

Nói cách khác, với độ tin cậy 95%, chiều cao trung bình của cây loại A từ 113,936cm đến 118,064cm.

j) Bằng phương pháp mới, sau một thời gian người ta thấy chiều cao trung bình của những cây loại A là 119,5cm. Hãy cho kết luận về phương pháp mới với mức ý nghĩa 1% (GS X có phân phối chuẩn).

Đây là bài toán kiểm định giả thiết về kỳ vọng $\mu_A = M(X_A)$ của chiều cao $X = X_A$ của các cây loại A với mức ý nghĩa $\alpha = 1\% = 0,01$:

H_0: $\mu_A = 119,5$ với giả thiết đối H_1: $\mu_A \neq 119,5$.

Vì $n_A = 25 < 30$, X_A có phân phối chuẩn, $\sigma^2{}_A = D(X_A)$ chưa biết, nên ta kiểm định như sau:

Bước 1: Ta có

$$z = \frac{(\overline{X}_A - \mu_0)\sqrt{n_A}}{S_A} = \frac{(116 - 119,5)\sqrt{25}}{5} = -3,5.$$

Bước 2: Đặt $k = n_A - 1 = 24$. Tra bảng phân phối Student ứng với $k = 24$ và $\alpha = 0,01$ ta được $t_\alpha^k = 2,797$.

Bước 3: Kiểm định. Vì $|z| = 3,5 > 2,797 = t_\alpha$ nên ta bác bỏ giả thiết H_0: $\mu_A = 119,5$, nghĩa là chấp nhận H_1: $\mu_A \neq 119,5$. Cụ thể, ta nhận định $\mu_A < 119,5$ (vì $\overline{X}_A = 116 < 119,5$).

Kết luận: Với mức ý nghĩa 1%, phương pháp mới có tác dụng làm thay đổi chiều cao trung bình của các cây loại A, theo hướng làm tăng chiều cao trung bình của các cây loại này.

k) Giả sử X có phân phối chuẩn. Với độ tin cậy 95%, hãy ước lượng phương sai của X trong hai trường hợp :

α) Biết kỳ vọng của X là 130cm.

Giả thiết cho ta $\mu = M(X) = 130$. Ta có công thức ước lượng khoảng cho phương sai:

$$\left(\frac{\sum (X_i - \mu)^2 n_i}{\chi^2_{\frac{\alpha}{2}}}; \frac{\sum (X_i - \mu)^2 n_i}{\chi^2_{1-\frac{\alpha}{2}}}\right)$$

Ta lập bảng:

$X_i - \mu$	−30	−20	−10	0	10	20	30
n_i	10	10	15	30	10	10	15

Từ đó ta tìm được cỡ mẫu n = 100; $\sum (X_i - \mu)^2 n_i = 33000$.

Tra bảng phân phối chi bình phương $\chi^2 \sim \chi^2(n)$ với n = 100 bậc tự do ta được:

$$\chi^2_{\frac{\alpha}{2}} = \chi^2_{0,025} = 129,56 \quad \text{và} \quad \chi^2_{1-\frac{\alpha}{2}} = \chi^2_{0,975} = 74,222.$$

Vậy ước lượng khoảng của phương sai là:

$$\left(\frac{33000}{129,56}; \frac{33000}{74,222}\right) = (254,7082; 444,6121)$$

Nói cách khác, với độ tin cậy 95%, phương sai của chiều cao X của giống cây trồng trên từ 254,7082(cm^2) đến 444,6121(cm^2).

β) Chưa biết kỳ vọng của X.

Khi chưa biết $\mu = M(X)$, ta có công thức ước lượng khoảng cho phương sai:

$$\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}}; \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}}\right).$$

Tra bảng phân phối chi bình phương $\chi^2 \sim \chi^2(n-1)$ với $n - 1 = 99 \approx 100$ bậc tự do ta được:

$$\chi^2_{\frac{\alpha}{2}} = \chi^2_{0,025} = 129,56 \quad \text{và} \quad \chi^2_{1-\frac{\alpha}{2}} = \chi^2_{0,975} = 74,222.$$

Vậy ước lượng khoảng của phương sai là:

$$\left(\frac{99.(18,2297)^2}{129,56}; \frac{99.(18,2297)^2}{74,222}\right) = (253,9354; 443,2631)$$

Nói cách khác, với độ tin cậy 95%, phương sai của chiều cao X của giống cây trồng trên từ 253,9354(cm^2) đến 443,2631(cm^2).

l) Khi canh tác bình thường thì phương sai của chiều cao X là 300cm^2. Hãy nhận định về tình hình canh tác với mức ý nghĩa 5% (GS X có phân phối chuẩn).

Đây là bài toán kiểm định giả thiết về phương sai $\sigma^2 = D(X)$, X có phân phối chuẩn, với mức ý nghĩa $\alpha = 5\% = 0,05$:

H_0: $\sigma^2 = 300$ với giả thiết đối H_1: $\sigma^2 \neq 300$.

Bước 1: Ta có:

$$z = \frac{(n-1)S^2}{\sigma_0^2} = \frac{99.(18,2297)^2}{300} = 109,6662$$

Bước 2: Tra bảng phân phối chi bình phương $\chi^2 \sim \chi^2(k)$ với $k = n - 1 = 99 \approx 100$ bậc tự do, ta tìm được $\chi^2_{\frac{\alpha}{2}} = \chi^2_{0,025} = 129,56$ và $\chi^2_{1-\frac{\alpha}{2}} = \chi^2_{0,975} = 74,222$.

Bước 3: Kiểm định:

Vì $\chi^2_{1-\frac{\alpha}{2}} = 74,222 \le z = 109,6662 \le 129,56 = \chi^2_{\frac{\alpha}{2}}$ nên ta chấp nhận giả thiết H_0: $\sigma^2 = 300$.

Kết luận: Với mức ý nghĩa 5%, tình hình canh tác là bình thường.

Bài 2. Để nghiên cứu nhu cầu của một loại hàng ở một khu vực, người ta khảo sát 400 hộ gia đình. Kết quả như sau:

Nhu cầu (kg/tháng/hộ)	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8
Số hộ	10	35	86	132	78	31	18	10

Cho biết trong khu vực có 4000 hộ.

a) Ước lượng nhu cầu trung bình về mặt hàng này của toàn khu vực trong một năm với độ tin cậy 95%.

b) Khi ước lượng nhu cầu trung bình về mặt hàng này của toàn khu vực trong một năm, nếu ta muốn đạt được độ tin cậy 99% và độ chính xác là 4,8tấn thì cần khảo sát ở ít nhất bao nhiêu hộ gia đình?

Lời giải

Gọi X(kg) là nhu cầu của một hộ về loại hàng trên trong một tháng. Ta có:

Xi	0,5	1,5	2,5	3,5	4,5	5,5	6,5	7,5
ni	10	35	86	132	78	31	18	10

$$n = 400; \ \sum X_i n_i = 1448; \ \sum X_i^2 n_i = 6076.$$

- Kỳ vọng mẫu của X là $\overline{X} = \frac{1}{n}\sum X_i n_i = 3,62.$
- Phương sai mẫu của X là:

$$\hat{S}^2 = \frac{1}{n}\sum X_i^2 n_i - \overline{X}^2 = (1,4442)^2.$$

- Phương sai mẫu hiệu chỉnh của X là:

$$S^2 = \frac{n}{n-1}\hat{S}^2 = (1,4460)^2.$$

a) Ước lượng nhu cầu trung bình về mặt hàng này của toàn khu vực trong một năm với độ tin cậy 95%.

Trước hết ta ước lượng nhu cầu trung bình về mặt hàng này của một hộ trong khu vực trong một tháng với độ tin cậy 95%.

Ôn thi Cao học – Toán kinh tế – Bài giải Thống kê Trần Ngọc Hội

Đây là bài toán ước lượng khoảng cho kỳ vọng $\mu = M(X)$ với độ tin cậy $\gamma = 1 - \alpha = 95\% = 0,95$.

Vì $n \ge 30$, $\sigma^2 = D(X)$ chưa biết nên ta có công thức ước lượng khoảng cho kỳ vọng:

$$(\overline{X} - z_\alpha \frac{S}{\sqrt{n}}; \overline{X} + z_\alpha \frac{S}{\sqrt{n}}),$$

trong đó $\varphi(z_\alpha) = \gamma/2 = 0,95/2 = 0,475$. Tra bảng giá trị hàm Laplace ta được $z_\alpha = 1,96$. Vậy ước lượng khoảng là:

$$(3,62 - 1,96\frac{1,4460}{\sqrt{400}}; 3,62 + 1,96\frac{1,4460}{\sqrt{400}}) = (3,4783;\ 3,7617).$$

Nói cách khác, với độ tin cậy 95%, nhu cầu trung bình về mặt hàng này của một hộ trong khu vực trong một tháng nằm trong khoảng từ 3,4783kg đến 3,7617kg. Xét 4000 hộ trong một năm 12 tháng, ta có các nhu cầu tương ứng là:

$3,4783 \times 4000 \times 12 = 166958,4$kg = 166,9584tấn;

$3,7617 \times 4000 \times 12 = 180561,6$kg = 180,5616tấn.

Kết luận: Với độ tin cậy 95%, nhu cầu trung bình về mặt hàng này của toàn khu vực trong một năm nằm trong khoảng từ 166,9584tấn đến 180,5616tấn.

b) Khi ước lượng nhu cầu trung bình về mặt hàng này của toàn khu vực trong một năm, nếu ta muốn đạt được độ tin cậy 99% và độ chính xác là 4,8tấn thì cần khảo sát ở ít nhất bao nhiêu hộ gia đình?

Khi ước lượng nhu cầu trung bình về mặt hàng này của toàn khu vực trong một năm với độ tin cậy 99% và độ chính xác là 4,8 tấn= 4800kg, nghĩa là ta ước lượng nhu cầu trung bình về mặt hàng này của một hộ trong một tháng với độ tin cậy $\gamma = 1 - \alpha = 0,99$ và độ chính xác $\varepsilon = 4800/(4000\times12) = 0,1$kg. Như vậy, ta đưa về bài toán xác định cỡ mẫu khi ước lượng kỳ vọng của chỉ tiêu X với độ chính xác $\varepsilon = 0,1$ và độ tin cậy $\gamma = 1 - \alpha = 99\% = 0,99$.

Vì $n \ge 30$, $\sigma^2 = D(X)$ chưa biết nên ta có công thức tính độ chính xác của ước lượng:

$$\varepsilon = z_\alpha \frac{S}{\sqrt{n}},$$

trong đó $\varphi(z_\alpha) = \gamma/2 = 0,99/2 = 0,495$. Tra bảng giá trị hàm Laplace ta được $z_\alpha = 2,58$. Suy ra

$$n = \left(\frac{z_\alpha S}{\varepsilon}\right)^2 = \left(\frac{2,58 \times 1,4460}{0,1}\right)^2 \approx 1391,8.$$

Thực tế yêu cầu $n \ge \lceil 1391,8 \rceil = 1392$. Vậy cần khảo sát ít nhất là 1392 hộ gia đình.

Bài 3. Để khảo sát chỉ tiêu X của một loại sản phẩm của xí nghiệp I, người ta quan sát một mẫu trong kho và có kết qủa sau:

X(cm)	11-15	15-19	19-23	23-27	27-31	31-35	35-39
Số sphẩm	8	9	20	16	16	13	18

a) Những sản phẩm có chỉ tiêu X từ 19cm trở xuống được gọi là những sản phẩm loại B. Hãy ước lượng tỉ lệ sản phẩm loại B với độ tin cậy 92%.

b) Giả sử trong kho có 1000 sản phẩm loại B. Hãy ước lượng số sản phẩm trong kho với độ tin cậy 92%.

c) Giả sử trong kho có 10.000 sản phẩm. Hãy ước lượng số sản phẩm loại B có trong kho với độ tin cậy 92%.

d) Giả sử trong kho để lẫn 1000 sản phẩm của xí nghiệp II và trong 100 sản phẩm lấy từ kho có 9 sản phẩm của xí nghiệp II. Hãy ước lượng số sản phẩm của xí nghiệp I có trong kho với độ tin cậy 82%.

Lời giải

a) Những sản phẩm có chỉ tiêu X từ 19cm trở xuống được gọi là những sản phẩm loại B. Hãy ước lượng tỉ lệ sản phẩm loại B với độ tin cậy 92%.

Đây là bài toán ước lượng khoảng cho tỉ lệ p các sản phẩm loại B với độ tin cậy $\gamma = 1-\alpha = 92\% = 0{,}92$. Ta có công thức ước lượng khoảng :

$$\left(F_n - z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}}; F_n + z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}}\right),$$

trong đó $\varphi(z_\alpha) = \gamma/2 = 0{,}92/2 = 0{,}46$. Tra bảng giá trị hàm Laplace ta được $z_\alpha = 1{,}75$. Mặt khác, trong n =100 sản phẩm có m = 17 sản phẩm loại B nên tỉ lệ mẫu sản phẩm loại B là $F_n = 0{,}17$. Vậy ước lượng khoảng là:

$$\left(0{,}17 - 1{,}75\sqrt{\frac{0{,}17(1-0{,}17)}{100}};\ 0{,}17 + 1{,}75\sqrt{\frac{0{,}17(1-0{,}17)}{100}}\right) = (10{,}43\%;\ 23{,}57\%).$$

Nói cách khác, với độ tin cậy 92%, tỉ lệ sản phẩm loại B nằm trong khoảng từ 10,43% đến 23,57%.

b) Giả sử trong kho có 1000 sản phẩm loại B. Hãy ước lượng số sản phẩm trong kho với độ tin cậy 92%.

Khi trong kho có 1000 sản phẩm loại B, gọi N là số sản phẩm có trong kho, ta có tỉ lệ sản phẩm loại B là 1000/N. Theo kết quả câu a), với độ tin cậy 92%, tỉ lệ các sản phẩm loại B từ 10,43% đến 23,57%, do đó:

$$\begin{aligned}
10{,}43\% \le \frac{1000}{N} \le 23{,}57\% &\Leftrightarrow \frac{10{,}43}{100} \le \frac{1000}{N} \le \frac{23{,}57}{100}\\
&\Leftrightarrow \frac{100.1000}{23{,}57} \le N \le \frac{100.1000}{10{,}430}\\
&\Leftrightarrow 4242{,}68 \le N \le 9587{,}73\\
&\Leftrightarrow 4243 \le N \le 9587
\end{aligned}$$

Vậy với độ tin cậy 92%, ta ước lượng trong kho có từ 4243 đến 9587 sản phẩm.

c) Giả sử trong kho có 10.000 sản phẩm. Hãy ước lượng số sản phẩm loại B có trong kho với độ tin cậy 92%.

Khi trong kho có 10.000 sản phẩm loại, gọi M là số sản phẩm loại B có trong kho, ta có tỉ lệ sản phẩm loại B là M/10.000. Theo kết quả câu a), với độ tin cậy 92%, tỉ lệ các sản phẩm loại B từ 10,43% đến 23,57%, do đó:

$$\begin{aligned}
10{,}43\% \le \frac{M}{10.000} \le 23{,}57\% &\Leftrightarrow 10{,}43\% \times 10.000 \le M \le 23{,}57\% \times 10.000\\
&\Leftrightarrow 1043 \le M \le 2357
\end{aligned}$$

Vậy với độ tin cậy 92%, ta ước lượng trong kho có từ 1043 đến 2357 sản phẩm loại B.

d) Giả sử trong kho để lẫn 1000 sản phẩm của xí nghiệp II và trong 100 sản phẩm lấy từ kho có 9 sản phẩm của xí nghiệp II. Hãy ước lượng số sản phẩm của xí nghiệp I có trong kho với độ tin cậy 82%.

Trước hết ta ước lượng tỉ lệ sản phẩm của xí nghiệp I có trong kho với độ tin cậy 82%. Ta có công thức ước lượng khoảng :

$$\left(F_n - z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}}; F_n + z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}}\right)$$

trong đó $\varphi(z_\alpha) = \gamma/2 = 0{,}82/2 = 0{,}41$. Tra bảng giá trị hàm Laplace ta được $z_\alpha = 1{,}34$. Mặt khác, theo giả thiết, trong n =100 sản phẩm có 9 sản phẩm của xí nghiệp II tức là có 91 sản phẩm của xí nghiệp I, nên tỉ lệ mẫu sản phẩm của xí nghiệp I là $F_n = 91/100 = 0{,}91$. Vậy ước lượng khoảng là:

$$\left(0{,}91 - 1{,}34\sqrt{\frac{0{,}91(1-0{,}91)}{100}}; 0{,}91 + 1{,}34\sqrt{\frac{0{,}91(1-0{,}91)}{100}}\right) = (87{,}17\%; 94{,}83\%).$$

Nói cách khác, với độ tin cậy 82%, tỉ lệ sản phẩm của xí nghiệp I nằm trong khoảng từ 87,17% đến 94,83%.

Bây giờ gọi N là số sản phẩm của xí nghiệp I có trong kho. Khi đó:

- Tổng số sản phẩm có trong kho là N + 1000.
- Tỉ lệ sản phẩm của xí nghiệp I có trong kho là N/(N+1000).

Theo kết quả trên, với độ tin cậy 82%, tỉ lệ sản phẩm của xí nghiệp I có trong kho nằm trong khoảng từ 87,17% đến 94,83%, do đó:

$$\begin{aligned}
87{,}17\% \le \frac{N}{N+1000} \le 94{,}83\% &\Leftrightarrow 87{,}17\% \le \frac{N}{N+1000} \le 94{,}83\%\\
&\Leftrightarrow 87{,}17\% \le 1 - \frac{1000}{N+1000} \le 94{,}83\%\\
&\Leftrightarrow 5{,}17\% \le \frac{1000}{N+1000} \le 12{,}83\%\\
&\Leftrightarrow \frac{1000}{12{,}83\%}\text{-}1000 \le N \le \frac{1000}{5{,}17\%}\text{-}1000\\
&\Leftrightarrow 6794{,}23 \le N \le 18342{,}36\\
&\Leftrightarrow 6795 \le N \le 18342
\end{aligned}$$

Vậy với độ tin cậy 82%, ta ước lượng số sản phẩm của xí nghiệp I có trong kho nằm trong khoảng từ 6795 đến 18342.

Bài 4. Trái cây của một chủ hàng được đựng trong các sọt, mỗi sọt 100 trái. Người ta kiểm tra 50 sọt thì thấy có 450 trái không đạt tiêu chuẩn.

a) Ước lượng tỉ lệ trái không đạt tiêu chuẩn của lô hàng trên với độ tin cậy 95%.

b) Nếu muốn ước lượng tỉ lệ trái không đạt tiêu chuẩn với độ chính xác 0,5% thì sẽ đạt được độ tin cậy là bao nhiêu?
c) Nếu muốn ước lượng tỉ lệ trái không đạt tiêu chuẩn với độ chính xác 1% và độ tin cậy 99% thì phải điều tra thêm ít nhất bao nhiêu sọt nữa?

Lời giải

Số trái trong 100 sọt là 50×100 = 5000. Do đó:

- Cỡ mẫu n = 5000.
- Số trái không đạt tiêu chuẩn là: m = 450.
- Tỉ lệ mẫu các trái không đạt tiêu chuẩn là:

$$F_n = m/n = 450/5000 = 0{,}09.$$

a) Ước lượng tỉ lệ trái không đạt tiêu chuẩn của lô hàng trên với độ tin cậy 95%.

Đây là bài toán ước lượng khoảng cho tỉ lệ p các trái không đạt tiêu chuẩn với độ tin cậy $\gamma = 1-\alpha = 95\% = 0{,}95$.

Ta có công thức ước lượng khoảng:

$$(F_n - z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}}; F_n + z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}})$$

trong đó $\varphi(z_\alpha) = (1-\alpha)/2 = \gamma/2 = 0{,}95/2 = 0{,}475$. Tra bảng giá trị hàm Laplace ta được $z_\alpha = 1{,}96$. Vậy ước lượng khoảng là:

$$(0,09 - 1,96\sqrt{\frac{0,09(1-0,09)}{5000}}; 0,09 + 1,96\sqrt{\frac{0,09(1-0,09)}{5000}}) = (8,21\%; 9,79\%)$$

Nói cách khác, với độ tin cậy 95%, tỉ lệ trái không đạt tiêu chuẩn từ 8,21% đến 9,79%.

b) Nếu muốn ước lượng tỉ lệ trái không đạt tiêu chuẩn với độ chính xác 0,5% thì sẽ đạt được độ tin cậy là bao nhiêu?

Yêu cầu của bài tóan: Xác định độ tin cậy $\gamma = 1-\alpha$.

Giả thiết: - Ước khỏang cho tỉ lệ trái không đạt tiêu chuẩn.
- Độ chính xác $\varepsilon = 0{,}5\% = 0{,}005$.

Ta có công thức tính độ chính xác của ước lượng:

$$\varepsilon = z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}}$$

trong đó $\varphi(z_\alpha) = \gamma/2$. Suy ra

$$z_\alpha = \varepsilon\sqrt{\frac{n}{F_n(1-F_n)}} = 0,005.\sqrt{\frac{5000}{0,09(1-0,09)}} = 1,24.$$

Tra bảng giá trị hàm Laplace ta được độ tin cậy là:

$$\gamma = 2\varphi(z_\alpha) = 2\varphi(1,24) = 2.0,3925 = 79,5\%.$$

Vậy độ tin cậy đạt được là 79,5%.

c) Nếu muốn ước lượng tỉ lệ trái không đạt tiêu chuẩn với độ chính xác 1% và độ tin cậy 99% thì phải điều tra thêm ít nhất bao nhiêu sọt nữa?

Yêu cầu của bài tóan: Xác định cỡ mẫu.

Giả thiết: - Ước khỏang cho tỉ lệ trái không đạt tiêu chuẩn.
- Độ chính xác $\varepsilon = 1\% = 0{,}01$.
- Độ tin cậy $\gamma = 1-\alpha = 99\% = 0{,}99$.

Ta có công thức tính độ chính xác của ước lượng:

$$\varepsilon = z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}}$$

trong đó $\varphi(z_\alpha) = (1-\alpha)/2 = 0{,}99/2 = 0{,}495$. Tra bảng giá trị hàm Laplace ta được $z_\alpha = 2{,}58$. Suy ra

$$n = \frac{z_\alpha^2 F_n(1-F_n)}{\varepsilon^2} = \frac{2,58^2.0,09(1-0,09)}{0,01^2} \approx 5451,6.$$

Thực tế yêu cầu $n \ge \lceil 5451,6 \rceil = 5452$. Vì $n_1 = 5452 > 5000$ (5000 là cỡ mẫu đang có) nên ta cần điều tra thêm ít nhất là 5452 – 5000 = 452 trái, nghĩa là khoảng 5 sọt nữa.

Bài 5. Để biết số lượng cá trong hồ lớn người ta bắt lên 2000 con đánh dấu xong rồi thả chúng xuống hồ. Sau đó người ta bắt lên 400 con và thấy có 80 con được đánh dấu.
a) Với độ tin cậy 95%, hãy ước lượng số cá có trong hồ.
b) Ước lượng số cá tối đa có trong hồ với độ tin cậy 96%.
c) Ước lượng số cá tối thiểu có trong hồ với độ tin cậy 94%.

Lời giải

Gọi N là số cá có trong hồ. Khi đó tỉ lệ cá được đánh dấu có trong hồ là p = 2000/N.

Với mẫu thu được, ta có:

- Cỡ mẫu n = 400.
- Số con được đánh dấu trong mẫu là: m = 80.
- Tỉ lệ mẫu con được đánh dấu là:

$$F_n = m/n = 80/400 = 0{,}2.$$

a) Với độ tin cậy 95%, hãy ước lượng số cá có trong hồ.

Trước hết ta ước lượng khoảng cho tỉ lệ p các con được đánh dấu với độ tin cậy $\gamma = 1-\alpha = 95\% = 0{,}95$.

Ta có công thức ước lượng khoảng:

$$(F_n - z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}}; F_n + z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}}),$$

trong đó $\varphi(z_\alpha) = (1-\alpha)/2 = \gamma/2 = 0{,}95/2 = 0{,}475$. Tra bảng giá trị hàm Laplace ta được $z_\alpha = 1{,}96$. Vậy ước lượng khoảng là:

$$(0,2 - 1,96\sqrt{\frac{0,2(1-0,2)}{400}}; 0,2 + 1,96\sqrt{\frac{0,2(1-0,2)}{400}}) = (16,08\%; 23,92\%)$$

Như vậy, với độ tin cậy 95%, tỉ lệ con được đánh dấu nằm trong khoảng từ 16,08% đến 23,92%, do đó:

$$16,08\% \le \frac{2000}{N} \le 23,92\% \Leftrightarrow \frac{2000}{23,92\%} \le N \le \frac{2000}{16,08\%}$$

$$\Leftrightarrow 8361,20 \le N \le 12437,81$$

$$\Leftrightarrow 8362 \le N \le 12437$$

Vậy với độ tin cậy 95%, ta ước lượng số cá có trong hồ khoảng từ 8362 đến 12437 con.

b) Ước lượng số cá tối đa có trong hồ với độ tin cậy 96%.

Số cá tối đa có trong hồ tương ứng với giá trị tối thiểu của tỉ lệ con được đánh dấu. Do đó trước hết ta ước lượng khoảng bên phải cho tỉ lệ p các con được đánh dấu với độ tin cậy $\gamma = 1- \alpha = 96\% = 0,96$ ($\alpha = 0,04$).

Ta có công thức ước lượng khoảng bên phải:

$$(F_n - z_{2\alpha}\sqrt{\frac{F_n(1-F_n)}{n}}; +\infty)$$

trong đó $\varphi(z_{2\alpha}) = (1- 2\alpha)/2 = \gamma/2 = 0,92/2 = 0,46$. Tra bảng giá trị hàm Laplace ta được $z_{2\alpha} = 1,75$. Suy ra giá trị tối thiểu của tỉ lệ con được đánh dấu là:

$$F_n - z_{2\alpha}\sqrt{\frac{F_n(1-F_n)}{n}} = 0,2 - 1,75\sqrt{\frac{0,2(1-0,2)}{400}} = 0,165.$$

Như vậy, với độ tin cậy 96%, ta có

$$\frac{2000}{N} \ge 0,165 \Leftrightarrow N \le \frac{2000}{0,165} = 12121,2.$$

Vậy với độ tin cậy 96%, số cá tối đa có trong hồ là 12121.

c) Ước lượng số cá tối thiểu có trong hồ với độ tin cậy 94%.

Số cá tối thiểu có trong hồ tương ứng với giá trị tối đa của tỉ lệ con được đánh dấu. Do đó trước hết ta ước lượng khoảng bên trái cho tỉ lệ p các con được đánh dấu với độ tin cậy $\gamma = 1- \alpha = 94\% = 0,94$ ($\alpha = 0,06$).

Ta có công thức ước lượng khoảng bên trái:

$$(-\infty; F_n + z_{2\alpha}\sqrt{\frac{F_n(1-F_n)}{n}}),$$

trong đó $\varphi(z_{2\alpha}) = (1 - 2\alpha)/2 = \gamma/2 = 0,88/2 = 0,44$. Tra bảng giá trị hàm Laplace ta được $z_{2\alpha} = 1,56$. Suy ra giá trị tối đa của tỉ lệ con được đánh dấu là:

$$F_n + z_{2\alpha}\sqrt{\frac{F_n(1-F_n)}{n}} = 0,2 + 1,56\sqrt{\frac{0,2(1-0,2)}{400}} = 0,2312.$$

Như vậy, với độ tin cậy 94%, ta có

$$\frac{2000}{N} \le 0,2312 \Leftrightarrow N \ge \frac{2000}{0,2312} = 8650,5.$$

Vậy với độ tin cậy 94%, số cá tối thiểu có trong hồ là 8651.

Bài 6. Cho các số liệu như Bài 1.

a) Giả sử trung bình tiêu chuẩn của chiều cao X là 125cm. Có thể khẳng định rằng việc canh tác làm tăng chiều cao trung bình của giống cây trồng trên với mức ý nghĩa 1% hay không?

b) Giả sử trung bình tiêu chuẩn của chiều cao X là 134cm. Có thể khẳng định rằng việc canh tác làm giảm chiều cao trung bình của giống cây trồng trên với mức ý nghĩa 2% hay không?

c) Sau khi áp dụng phương pháp canh tác mới, người ta thấy chiều cao trung bình của các cây loại A là 114cm. Hãy kết luận xem phương pháp mới có làm giảm chiều cao trung bình của các cây loại A hay không với mức ý nghĩa 3% (Giả sử X có phân phối chuẩn) .

d) Trước đây, chiều cao trung bình của các cây loại A là 120cm. Các số liệu trên thu thập được sau khi đã áp dụng một kỹ thuật mới. Hãy kết luận xem kỹ thuật mới có làm giảm chiều cao trung bình của các cây loại A hay không với mức ý nghĩa 2% (Giả sử X có phân phối chuẩn).

e) Sau khi áp dụng một phương pháp sản xuất, người ta thấy tỉ lệ các cây loại A là 35%. Hãy kết luận xem phương pháp mới có làm tăng tỉ lệ các cây loại A lên hay không với mức ý nghĩa 2%.

f) Theo tài liệu thống kê, tỉ lệ cây loại A là 20%. Hãy xét xem hiện nay việc canh tác có làm tăng tỉ lệ cây loại A hay không với mức ý nghĩa 5%?

g) Theo tài liệu cũ, phương sai của chiều cao X là 250cm^2. Với mức ý nghĩa 5%, xét xem hiện tại chiều cao của cây trồng có biến động hơn so với trước đây hay không (GS X có phân phối chuẩn)?

h) Trước đây, phương sai của chiều cao X là 350cm^2. Các số liệu trên thu thập được sau khi đã áp dụng một kỹ thuật mới. Với mức ý nghĩa 5%, hãy xét xem kỹ thuật mới có làm chiều cao của giống cây trồng trên ít biến động hơn hay không (GS X có phân phối chuẩn)?

Lời giải

Ta có:

- Cỡ mẫu là n = 100.
- Kỳ vọng mẫu của X là

$$\overline{X} = \frac{1}{n}\sum X_i n_i = 131(\text{cm}).$$

- Phương sai mẫu của X là

$$\hat{S}^2 = \frac{1}{n}\sum X_i^2 n_i - \overline{X}^2 = (18,1384)^2(\text{cm}^2).$$

- Phương sai mẫu hiệu chỉnh của X là

$$S^2 = \frac{n}{n-1}\hat{S}^2 = (18,2297)^2(\text{cm}^2).$$

a) Giả sử trung bình tiêu chuẩn của chiều cao X là 125cm. Có thể khẳng định rằng việc canh tác làm tăng chiều cao trung bình của giống cây trồng trên với mức ý nghĩa 1% hay không?

Đây là bài toán kiểm định giả thiết về kỳ vọng $\mu = M(X)$ với mức ý nghĩa $\alpha = 1\% = 0,01$:

H$_0$: $\mu = 125$ với giả thiết đối **H$_1$**: $\mu > 125$.

Vì $n \geq 30$; $\sigma^2 = D(X)$ chưa biết, nên ta kiểm định như sau:

Bước 1: Ta có

$$z = \frac{(\bar{X}-\mu_0)\sqrt{n}}{S} = \frac{(131-125)\sqrt{100}}{18,2297} = 3,2913.$$

Bước 2: Tra bảng giá trị hàm Laplace để tìm $z_{2\alpha}$ thoả $\varphi(z_{2\alpha}) = (1-2\alpha)/2 = 0,98/2 = 0,49$ ta được $z_{2\alpha} = 2,33$.

Bước 3: Kiểm định. Vì $z = 3,2913 > 2,33 = z_{2\alpha}$ nên ta bác bỏ giả thiết H$_0$: $\mu=125$, nghĩa là chấp nhận H$_1$: $\mu > 125$.

Kết luận: Với mức ý nghĩa 1%, có thể kết luận rằng việc canh tác làm tăng chiều cao trung bình của giống cây trồng trên.

b) Giả sử trung bình tiêu chuẩn của chiều cao X là 134cm. Có thể khẳng định rằng việc canh tác làm giảm chiều cao trung bình của giống cây trồng trên với mức ý nghĩa 2% hay không?

Đây là bài toán kiểm định giả thiết về kỳ vọng $\mu = M(X)$ với mức ý nghĩa $\alpha = 2\% = 0,02$:

H$_0$: $\mu = 134$ với giả thiết đối **H$_1$**: $\mu < 134$.

Vì $n \geq 30$; $\sigma^2 = D(X)$ chưa biết, nên ta kiểm định như sau:

Bước 1: Ta có

$$z = \frac{(\bar{X}-\mu_0)\sqrt{n}}{S} = \frac{(131-134)\sqrt{100}}{18,2297} = -1,6457.$$

Bước 2: Tra bảng giá trị hàm Laplace để tìm $z_{2\alpha}$ thoả $\varphi(z_{2\alpha}) = (1-2\alpha)/2 = 0,96/2 = 0,48$ ta được $z_{2\alpha} = 2,06$.

Bước 3: Kiểm định. Vì $-z = 1,6457 < 2,06 = z_{2\alpha}$ nên ta chấp nhận giả thiết H$_0$: $\mu = 134$.

Kết luận: Với mức ý nghĩa 2%, không thể kết luận rằng việc canh tác làm giảm chiều cao trung bình của giống cây trồng trên.

c) Sau khi áp dụng phương pháp canh tác mới, người ta thấy chiều cao trung bình của các cây loại A là 114cm. Hãy kết luận xem phương pháp mới có làm giảm chiều cao trung bình của các cây loại A hay không với mức ý nghĩa 3% (Giả sử X có phân phối chuẩn).

Đây là bài toán kiểm định giả thiết về kỳ vọng $\mu_A = M(X_A)$ của chỉ tiêu $X = X_A$ của các cây loại A với mức ý nghĩa $\alpha = 3\% = 0,03$:

H$_0$: $\mu_A = 114$ với giả thiết đối **H$_1$**: $\mu_A > 114$.

Ta lập bảng số liệu của X_A:

X_{Ai}	110	120
N_{Ai}	10	15

Từ bảng trên ta tính được:

$$n_A = 25; \sum X_{Ai}n_{Ai} = 2900; \sum X_{Ai}{}^2 n_{Ai} = 337000.$$

- Kỳ vọng mẫu của X_A là

$$\bar{X}_A = \frac{1}{n}\sum X_{Ai}n_{Ai} = 116(\text{cm}).$$

- Phương sai mẫu của X_A là:

$$\hat{S}_A^2 = \frac{1}{n}\sum X_{Ai}{}^2 n_{Ai} - \bar{X}_A{}^2 = (4,8990)^2(\text{cm}^2).$$

- Phương sai mẫu hiệu chỉnh của X_A là:

$$S_A{}^2 = \frac{n_A}{n_A-1}\hat{S}_A^2 = 5^2(\text{cm}^2).$$

Vì $n_A < 30$, X_A có phân phối chuẩn, $\sigma^2_A = D(X_A)$ chưa biết, nên ta kiểm định như sau:

Bước 1: Ta có

$$z = \frac{(\bar{X}_A-\mu_0)\sqrt{n_A}}{S_A} = \frac{(116-114)\sqrt{25}}{5} = 2.$$

Bước 2: Đặt $k = n_A - 1 = 24$. Tra bảng phân phối Student ứng với $k = 24$ và $2\alpha = 0,06$ ta được $t_{2\alpha} = 1,974$.

Bước 3: Kiểm định. Vì $z = 2 > 1,974 = t_{2\alpha}$ nên ta bác bỏ giả thiết H$_0$: $\mu_A = 114$, nghĩa là chấp nhận H$_1$: $\mu_A > 114$.

Kết luận: Với mức ý nghĩa 3%, phương pháp mới làm giảm chiều cao trung bình của các cây loại A.

d) Trước đây, chiều cao trung bình của các cây loại A là 120cm. Các số liệu trên thu thập được sau khi đã áp dụng một kỹ thuật mới. Hãy kết luận xem kỹ thuật mới có làm giảm chiều cao trung bình của các cây loại A hay không với mức ý nghĩa 2% (Giả sử X có phân phối chuẩn).

Đây là bài toán kiểm định giả thiết về kỳ vọng $\mu_A = M(X_A)$ của chỉ tiêu $X = X_A$ của các cây loại A với mức ý nghĩa $\alpha = 2\% = 0,02$:

H$_0$: $\mu_A = 120$ với giả thiết đối **H$_1$**: $\mu_A < 120$.

Vì $n_A < 30$, X_A có phân phối chuẩn, $\sigma^2_A = D(X_A)$ chưa biết, nên ta kiểm định như sau:

Bước 1: Ta có $z = \frac{(\bar{X}_A-\mu_0)\sqrt{n_A}}{S_A} = \frac{(116-120)\sqrt{25}}{5} = -4.$

Bước 2: Đặt $k = n_A - 1 = 24$. Tra bảng phân phối Student ứng với $k = 24$ và $2\alpha = 0,04$ ta được $t_{2\alpha} = 2,1715$.

Bước 3: Kiểm định. Vì $-z = 4 > 2,1715 = t_{2\alpha}$ nên ta bác bỏ giả thiết H$_0$: $\mu_A = 120$, nghĩa là chấp nhận H$_1$: $\mu_A < 120$.

Kết luận: Với mức ý nghĩa 2%, kỹ thuật mới làm giảm chiều cao trung bình của các cây loại A.

e) Sau khi áp dụng một phương pháp sản xuất, người ta thấy tỉ lệ cây loại A là 35%. Hãy kết luận xem phương pháp mới có làm tăng tỉ lệ cây loại A lên hay không với mức ý nghĩa 2%.

Đây là bài toán kiểm định giả thiết về tỉ lệ p các sản phẩm loại A với mức ý nghĩa $\alpha = 2\% = 0,02$:

H_0: p = 35% = 0,35 với giả thiết đối H_1: p < 0,35.

Ta có tỉ lệ mẫu các cây loại A là $F_n = 25/100 = 0,25$. Ta kiểm định như sau:

Bước 1: Ta có

$$z = \frac{(F_n - p_0)\sqrt{n}}{\sqrt{p_0 q_0}} = \frac{(0,25 - 0,35)\sqrt{100}}{\sqrt{0,35(1-0,35)}} = -2,0966.$$

Bước 2: Tra bảng giá trị hàm Laplace để tìm $z_{2\alpha}$ thoả

$$\varphi(z_{2\alpha}) = (1 - 2\alpha)/2 = 0,96/2 = 0,48$$

ta được $z_{2\alpha} = 2,06$.

Bước 3: Kiểm định. Vì $-z = 2,0966 > 2,06 = z_{2\alpha}$ nên ta bác bỏ giả thiết H_0: p = 0,35, nghĩa là chấp nhận H_1: p < 0,35.

Kết luận: Với mức ý nghĩa 2%, phương pháp mới làm tăng tỉ lệ cây loại A.

f) Một tài liệu thống kê cũ cho rằng tỉ lệ cây loại A là 20%. Hãy xét xem việc canh tác có làm tăng tỉ lệ cây loại A hay không với mức ý nghĩa 5%?

Đây là bài toán kiểm định giả thiết về tỉ lệ p các sản phẩm loại A với mức ý nghĩa $\alpha = 5\% = 0,05$:

H_0: p = 20% = 0,20 với giả thiết đối H_1: p > 0,20.

Ta kiểm định như sau:

Bước 1: Ta có

$$z = \frac{(F_n - p_0)\sqrt{n}}{\sqrt{p_0 q_0}} = \frac{(0,25 - 0,20)\sqrt{100}}{\sqrt{0,20(1-0,20)}} = 1,25.$$

Bước 2: Tra bảng giá trị hàm Laplace để tìm $z_{2\alpha}$ thoả

$$\varphi(z_{2\alpha}) = (1 - 2\alpha)/2 = 0,90/2 = 0,45$$

ta được $z_{2\alpha} = 1,65$.

Bước 3: Kiểm định. Vì $z = 1,25 < 1,65 = z_{2\alpha}$ nên ta chấp nhận giả thiết H_0: p = 0,20.

Kết luận: Với mức ý nghĩa 5%, việc canh tác không làm tăng tỉ lệ các cây loại A.

g) Theo tài liệu cũ, phương sai của chiều cao X là 250cm^2. Với mức ý nghĩa 5%, xét xem hiện tại chiều cao của cây trồng có biến động hơn so với trước đây hay không (GS X có phân phối chuẩn)?

Đây là bài toán kiểm định giả thiết về phương sai $\sigma^2 = D(X)$ với mức ý nghĩa $\alpha = 5\% = 0,05$:

H_0: $\sigma^2 = 250$ với giả thiết đối H_1: $\sigma^2 > 250$.

Ta kiểm định như sau:

Bước 1: Ta có:

$$z = \frac{(n-1)S^2}{\sigma_0^2} = \frac{99.(18,2297)^2}{250} = 131,5995.$$

Bước 2: Tra bảng phân phối chi bình phương $\chi^2 \sim \chi^2(k)$ với k = n – 1 = 99 bậc tự do, ta tìm được $\chi^2_\alpha = \chi^2_{0,05} = 124,3$.

Bước 3: Kiểm định:

Vì $z = 131,5995 > 124,3 = \chi^2_\alpha$ nên ta bác bỏ giả thiết H_0: $\sigma^2 = 250$, nghĩa là chấp nhận H_1: $\sigma^2 > 250$.

Kết luận: Với mức ý nghĩa 5%, tình hình canh tác hiện tại làm chiều cao của cây trồng biến động hơn.

h) Trước đây, phương sai của chiều cao X là 350cm^2. Các số liệu trên thu thập được sau khi đã áp dụng một kỹ thuật mới. Với mức ý nghĩa 5%, hãy xét xem kỹ thuật mới có làm chiều cao của giống cây trồng trên ít biến động hơn hay không (GS X có phân phối chuẩn)?

Đây là bài toán kiểm định giả thiết về phương sai $\sigma^2 = D(X)$ với mức ý nghĩa $\alpha = 5\% = 0,05$:

H_0: $\sigma^2 = 350$ với giả thiết đối H_1: $\sigma^2 < 350$.

Ta kiểm định như sau:

Bước 1: Ta có:

$$z = \frac{(n-1)S^2}{\sigma_0^2} = \frac{99.(18,2297)^2}{250} = 93,9996.$$

Bước 2: Tra bảng phân phối chi bình phương $\chi^2 \sim \chi^2(k)$ với k = n – 1 = 99 bậc tự do, ta tìm được $\chi^2_{1-\alpha} = \chi^2_{0,95} = 77,93$.

Bước 3: Kiểm định:

Vì $z = 93,9996 > 77,93 = \chi^2_{1-\alpha}$ nên ta chấp nhận giả thiết H_0: $\sigma^2 = 350$.

Kết luận: Với mức ý nghĩa 5%, kỹ thuật mới không làm chiều cao của giống cây trồng trên ít biến động hơn.

Bài 7. Để khảo sát đường kính của một chi tiết máy người ta kiểm tra một số sản phẩm của hai nhà máy. Trong kết quả sau đây, X là đường kính của chi tiết máy do nhà máy I sản xuất còn Y là đường kính của chi tiết máy do nhà máy II sản xuất. Những sản phẩm có chi tiết máy nhỏ hơn 19cm được xếp vào loại C.

X(cm)	11-15	15-19	19-23	23-27	27-31	31-35	35-39
Số sản phẩm	9	19	20	26	16	13	18

Y(cm)	13-16	16-19	19-22	22-25	25-28	28-31	31-34
Số sản phẩm	7	9	25	26	18	15	11

a) Có thể kết luận rằng đường kính trung bình của một chi tiết máy do hai nhà máy sản xuất bằng nhau hay không với mức ý nghĩa 1%?
b) Có thể cho rằng đường kính trung bình của một chi tiết máy do nhà máy I sản xuất lớn hơn đường kính trung bình của một chi tiết máy do nhà máy II sản xuất hay không với mức ý nghĩa 5%?

c) Xét xem đường kính trung bình của một chi tiết máy do nhà máy II sản xuất có nhỏ hơn đường kính trung bình của một chi tiết máy do nhà máy I sản xuất hay không với mức ý nghĩa 2%?
d) Với mức ý nghĩa 4%, tỉ lệ sản phẩm loại C do hai nhà máy sản xuất có như nhau không?
e) Với mức ý nghĩa 3%, có thể cho rằng tỉ lệ sản phẩm loại C do nhà máy I sản xuất lớn hơn tỉ lệ sản phẩm loại C do nhà máy II sản xuất hay không?
f) Hãy nhận xét về ý kiến cho rằng tỉ lệ sản phẩm loại C do nhà máy II sản xuất nhỏ hơn tỉ lệ sản phẩm loại C do nhà máy thứ I sản xuất với mức ý nghĩa 5%?

Lời giải

1) Đối với X ta có bảng số liệu:

Xi	13	17	21	25	29	33	37
ni	9	19	20	26	16	13	18

Ta có:

$$n_X = 121; \sum X_i n_{Xi} = 3069; \sum X_i^2 n_{Xi} = 84337.$$

- Kỳ vọng mẫu của X là

$$\overline{X} = \frac{1}{n_X}\sum X_i n_{Xi} = 25,3636(cm).$$

- Phương sai mẫu của X là

$$\hat{S}_X^2 = \frac{1}{n_X}\sum X_i^2 n_{Xi} - \overline{X}^2 = (7,3271)^2(cm^2).$$

- Phương sai mẫu hiệu chỉnh của X là

$$S_X^2 = \frac{n_X}{n_X - 1}\hat{S}_X^2 = (7,3575)^2(cm^2).$$

- Tỉ lệ sản phẩm loại C là

$$F_{Xn} = \frac{m_X}{n_X} = \frac{9+19}{121} = 0,2314.$$

2) Đối với Y ta có bảng số liệu:

Yi	14,5	17,5	20,5	23,5	26,5	29,5	32,5
ni	7	9	25	26	18	15	11

Ta có:

$$n_Y = 111; \sum Y_i n_{Yi} = 2659,5; \sum Y_i^2 n_{Yi} = 66405,75.$$

- Kỳ vọng mẫu của Y là

$$\overline{Y} = \frac{1}{n_Y}\sum Y_i n_{Yi} = 23,9595(cm).$$

- Phương sai mẫu của Y là

$$\hat{S}_Y^2 = \frac{1}{n_Y}\sum Y_i^2 n_{Yi} - \overline{Y}^2 = (4,9188)^2(cm^2).$$

- Phương sai mẫu hiệu chỉnh của Y là

$$S_Y^2 = \frac{n_Y}{n_Y - 1}\hat{S}_Y^2 = (4,9411)^2(cm^2).$$

- Tỉ lệ sản phẩm loại C là

$$F_{Yn} = \frac{m_Y}{n_Y} = \frac{7+9}{111} = 0,1441.$$

a) Có thể kết luận rằng đường kính trung bình của một chi tiết máy do hai nhà máy sản xuất bằng nhau hay không với mức ý nghĩa 1%?

Đây là bài toán kiểm định so sánh hai kỳ vọng với mức ý nghĩa $\alpha = 1\% = 0,01$:

H₀: $\mu_X = \mu_Y$ với giả thiết đối **H₁**: $\mu_X \neq \mu_Y$.

Vì $n_X > 30$; $n_Y > 30$ nên ta kiểm định như sau:

Bước 1: Ta có:

$$z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}} = \frac{25,3636 - 23,9595}{\sqrt{\frac{(7,3575)^2}{121} + \frac{(4,9411)^2}{111}}} = 1,7188.$$

Bước 2: Tra bảng giá trị hàm Laplace để tìm z_α thoả

$$\varphi(z_\alpha) = (1-\alpha)/2 = 0,99/2 = 0,495$$

ta được $z_\alpha = 2,58$.

Bước 3: Kiểm định. Vì $|z| = 1,7188 < 2,58 = z_\alpha$ nên ta chấp nhận giả thiết **H₀**: $\mu_X = \mu_Y$.

Kết luận: Với mức ý nghĩa 1%, có thể xem đường kính trung bình của một chi tiết máy do hai nhà máy sản xuất là bằng nhau.

b) Có thể cho rằng đường kính trung bình của một chi tiết máy do nhà máy thứ I sản xuất lớn hơn đường kính trung bình của một chi tiết máy do nhà máy thứ II sản xuất hay không với mức ý nghĩa 5%?

Đây là bài toán kiểm định so sánh hai kỳ vọng với mức ý nghĩa $\alpha = 5\% = 0,05$:

H₀: $\mu_X = \mu_Y$ với giả thiết đối **H₁**: $\mu_X > \mu_Y$.

Vì $n_X > 30$; $n_Y > 30$ nên ta kiểm định như sau:

Bước 1: Tương tự câu a), ta có:

$$z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}} = 1,7188.$$

Bươc 2: Tra bảng giá trị hàm Laplace để tìm $z_{2\alpha}$ thoả

$$\varphi(z_{2\alpha}) = (1-2\alpha)/2 = 0,90/2 = 0,45$$

ta được $z_{2\alpha} = 1,65$.

Bước 3: Kiểm định. Vì $z = 1,7188 > 1,65 = z_{2\alpha}$ nên ta bác bỏ giả thiết **H₀**: $\mu_X = \mu_Y$, nghĩa là chấp nhận H_1: $\mu_X > \mu_Y$.

Kết luận: Với mức ý nghĩa 5%, có thể xem đường kính trung bình của một chi tiết máy do nhà máy thứ I sản xuất lớn hơn đường kính trung bình của một chi tiết máy do nhà máy thứ II sản xuất.

c) Xét xem đường kính trung bình của một chi tiết máy do nhà máy thứ II sản xuất có nhỏ hơn đường kính trung bình của một chi tiết máy do nhà máy thứ I sản xuất hay không với mức ý nghĩa 2%?

Đây là bài toán kiểm định so sánh hai kỳ vọng với mức ý nghĩa α = 2% = 0,02:

H₀: $\mu_X = \mu_Y$ với giả thiết đối **H₁**: $\mu_X > \mu_Y$.

Vì $n_X > 30$; $n_Y > 30$ nên ta kiểm định như sau:

Bước 1: Tương tự câu a), ta có:

$$z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}} = 1,7188.$$

Bước 2: Tra bảng giá trị hàm Laplace để tìm $z_{2\alpha}$ thoả

$$\varphi(z_{2\alpha}) = (1 - 2\alpha)/2 = 0,96/2 = 0,48$$

ta được $z_{2\alpha} = 2,06$.

Bước 3: Kiểm định. Vì z = 1,7188 < 2,06 = $z_{2\alpha}$ nên ta chấp nhận giả thiết **H₀**: $\mu_X = \mu_Y$.

Kết luận: Với mức ý nghĩa 2%, không thể xem đường kính trung bình của một chi tiết máy do nhà máy thứ II sản xuất nhỏ hơn đường kính trung bình của một chi tiết máy do nhà máy thứ I sản xuất.

d) Với mức ý nghĩa 4%, tỉ lệ sản phẩm loại C do hai nhà máy sản xuất có như nhau không?

Đây là bài toán kiểm định giả thiết so sánh hai tỉ lệ với mức ý nghĩa α = 4% = 0,04:

H₀: $p_1 = p_2$ với giả thiết đối **H₁**: $p_1 \neq p_2$

Ta kiểm định như sau:

Bước 1: Ta có:

$$p_0 = \frac{n_1F_{n1} + n_2F_{n2}}{n_1 + n_2} = \frac{28 + 16}{121 + 111} = 0,1897.$$

$$z = \frac{F_{n1} - F_{n2}}{\sqrt{p_0(1 - p_0)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0,2314 - 0,1441}{\sqrt{0,1897(1 - 0,1897)\left(\frac{1}{121} + \frac{1}{111}\right)}} = 1,6942.$$

Bước 2: Tra bảng giá trị hàm Laplace để tìm z_α thoả

$$\varphi(z_\alpha) = (1 - \alpha)/2 = 0,96/2 = 0,48$$

ta được $z_\alpha = 2,06$.

Bước 3: Kiểm định. Vì |z| = 1,6942 < 2,06 = z_α nên ta chấp nhận giả thiết H_0: $p_1 = p_2$.

Kết luận: Với mức ý nghĩa 4%, có thể xem tỉ lệ sản phẩm loại C do hai nhà máy sản xuất là như nhau.

e) Với mức ý nghĩa 3%, có thể cho rằng tỉ lệ sản phẩm loại C do nhà máy thứ I sản xuất lớn hơn tỉ lệ sản phẩm loại C do nhà máy thứ II sản xuất hay không?

Đây là bài toán kiểm định giả thiết so sánh hai tỉ lệ với mức ý nghĩa α = 3% = 0,03:

H₀: $p_1 = p_2$ với giả thiết đối **H₁**: $p_1 > p_2$

Ta kiểm định như sau:

Bước 1: Tương tự câu d), ta có:

$$z = \frac{F_{n1} - F_{n2}}{\sqrt{p_0(1 - p_0)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = 1,6942.$$

Bước 2: Tra bảng giá trị hàm Laplace để tìm $z_{2\alpha}$ thoả

$$\varphi(z_{2\alpha}) = (1 - 2\alpha)/2 = 0,94/2 = 0,47$$

ta được $z_{2\alpha} = 1,88$.

Bước 3: Kiểm định. Vì z = 1,6942 < 1,88 = $z_{2\alpha}$ nên ta chấp nhận giả thiết H_0: $p_1 = p_2$.

Kết luận: Với mức ý nghĩa 3%, không thể cho rằng tỉ lệ sản phẩm loại C do nhà máy thứ I sản xuất lớn hơn tỉ lệ sản phẩm loại C do nhà máy thứ II sản xuất.

f) Hãy nhận xét về ý kiến cho rằng tỉ lệ sản phẩm loại C do nhà máy thứ II sản xuất nhỏ hơn tỉ lệ sản phẩm loại C do nhà máy thứ I sản xuất với mức ý nghĩa 5%?

Đây là bài toán kiểm định giả thiết so sánh hai tỉ lệ với mức ý nghĩa α = 5% = 0,05:

H₀: $p_1 = p_2$ với giả thiết đối **H₁**: $p_1 > p_2$

Ta kiểm định như sau:

Bước 1: Tương tự câu d), ta có:

$$z = \frac{F_{n1} - F_{n2}}{\sqrt{p_0(1 - p_0)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = 1,6942.$$

Bước 2: Tra bảng giá trị hàm Laplace để tìm $z_{2\alpha}$ thoả

$$\varphi(z_{2\alpha}) = (1 - 2\alpha)/2 = 0,90/2 = 0,45$$

ta được $z_{2\alpha} = 1,65$.

Bước 3: Kiểm định. Vì z = 1,6942 > 1,65 = $z_{2\alpha}$ nên ta bác bỏ giả thiết H_0: $p_1 = p_2$, nghĩa là chấp nhận H_1: $p_1 > p_2$.

Kết luận: Với mức ý nghĩa 5%, có thể chấp nhận ý kiến cho rằng tỉ lệ sản phẩm loại C do nhà máy thứ II sản xuất nhỏ hơn tỉ lệ sản phẩm loại C do nhà máy thứ I sản xuất.

Bài 8. Sản phẩm sản xuất ra trên một dây chuyền tự động được đóng gói theo qui cách 3 sản phẩm/hộp. Với mức ý nghĩa 1%, hãy xét xem số sản phẩm loại I có trong mỗi hộp có phải là ĐLNN có phân phối nhị thức hay không. Biết rằng khi kiểm tra 100 hộp người ta thấy có 75 hộp có 3 sản phẩm loại I, 20 hộp có 2 sản phẩm loại I; 5 hộp có 1 sản phẩm loại I.

Lời giải

Gọi X là số sản phẩm loại I có trong một hộp. Yêu cầu của bài toán là kiểm định giả thiết sau với mức ý nghĩa $\alpha = 1\% = 0{,}01$:

H_0: X có phân phối nhị thức $X \sim B(3,p)$ với p chưa biết

với giả thiết đối:

H_1 : X không có phân phối nhị thức như trên.

Trước hết ta thay p bằng tỉ lệ mẫu số sản phẩm loại I có trong một hộp:

$$p \approx F_n = \frac{75.3 + 20.2 + 5.1}{300} = 0{,}9.$$

Ta tính các $p_i = P(X = i)$ theo công thức Bernoulli:

$$p_i = C_3^i (0{,}9)^i (0{,}1)^{3-i}$$

Cụ thể ta tính được:

$p_0 = 0{,}001$; $p_1= 0{,}027$; $p_2 = 0{,}243$; $p_3=0{,}729$.

Ta lập bảng:

X_i	n_i	p_i	np_i	$(n_i - np_i)^2/np_i$
0	0	0,001	0,1	0,1
1	5	0,027	2,7	1,9593
2	20	0,243	24,3	0,7609
3	75	0,729	72,9	0,0605
Tổng	n = 100			$\chi^2 = 2{,}8807$

Bước 1: Ta có $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} = 2{,}8807.$

Bước 2: Số tham số chưa biết là r = 1 (do p chưa biết). Ta có $k - r - 1 = 4 - 1 - 1 = 2$. Tra bảng phân phối Chi bình phương $\chi^2 \sim \chi^2(2)$ với 2 bậc tự do, ta được: $\chi^2_\alpha = \chi^2_{0,01} = 9{,}21.$

Bước 3: Kiểm định:

Vì $\chi^2 = 2{,}8807 < 9{,}21 = \chi^2_\alpha$ nên ta chấp nhận giả thiết H_0.

Kết luận: Với mức ý nghĩa 1%, có thể cho rằng số sản phẩm loại I có trong một hộp là X có phân phối nhị thức: $X \sim B(3;0{,}9)$.

Bài 9. Qua sát trong một số ngày về số tai nạn giao thông X xảy mỗi ngày ở một thành phố ta được số liệu sau:

Số tai nạn X	0	1	2	3	4	≥ 5
Số ngày	10	32	46	35	20	13

Với mức ý nghĩa 1%, có thể xem số tai nạn giao thông xảy mỗi ngày ở thành phố trên là ĐLNN có phân phối Poisson được hay không?

Lời giải

Bài toán yêu cầu kiểm định giả thiết sau với mức ý nghĩa $\alpha = 1\% = 0{,}01$:

H_0: X có phân phối Poisson $X \sim P(a)$ (a chưa biết)

với giả thiết đối:

H_1 : X không có phân phối Poisson.

Trước hết ta thay a bằng kỳ vọng mẫu

$$a \approx \bar{X} = \frac{1}{n}\sum X_i n_i = 2{,}3974$$

Ta tính các $p_i = P(X = i)$ theo công thức:

$$p_i = \frac{e^{-2,3974}(2{,}3974)^i}{i!}$$

và lập bảng:

X_i	n_i	p_i	np_i	$(n_i - np_i)^2/np_i$
0	10	0,090954	14,188824	1,2366
1	32	0,218053	34,016268	0,1195
2	46	0,261381	40,775436	0,6694
3	35	0,208878	32,584968	0,1790
4	20	0,125191	19,529796	0,0113
(5;+∞)	13	0,095543	14,904708	0,2434
Tổng	n = 156			$\chi^2 = 2{,}4592$

Bước 1: Ta có $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} = 2{,}4592$

Bước 2: Số tham số chưa biết là r = 1 (do a chưa biết). Ta có $k - r - 1 = 6 - 1 - 1 = 4$. Tra bảng phân phối Chi bình phương $\chi^2 \sim \chi^2(4)$ với 4 bậc tự do, ta được: $\chi^2_\alpha = \chi^2_{0,01} = 13{,}28.$

Bước 3: Kiểm định:

Vì $\chi^2 = 2{,}4592 < 13{,}28 = \chi^2_\alpha$ nên ta chấp nhận giả thiết H_0.

Kết luận: Với mức ý nghĩa 1%, X có phân phối Poisson $X \sim P(2{,}3974)$.

Bài 10. Quan sát năng suất X của một giống lúa thử nghiệm trên 100 thửa ruộng ta có kết quả sau:

Năng suất (tấn/ha)	8-9	9-10	10-11	11-12	12-13	13-14	14-15
Số thửa	8	15	21	23	16	9	8

Với mức ý nghĩa 5% có thể xem năng suất X là ĐLNN có phân phối chuẩn hay không?

Lời giải

Bài toán yêu cầu kiểm định giả thiết với mức ý nghĩa $\alpha = 5\% = 0{,}05$:

H_0: X có phân phối chuẩn $X \sim N(\mu,\sigma^2)$ (μ, σ^2 chưa biết)

với giả thiết đối:

H_1 : X không có phân phối chuẩn.

Trước hết xấp xỉ:

$$\mu \approx \bar{X} = \frac{1}{n}\sum X_i n_i = 11{,}33;$$

$$\sigma^2 \approx S^2 = \frac{1}{n}\sum X_i^2 n_i - (\bar{X})^2 = (1{,}6640)^2.$$

Ta tính các $p_i = P(x_{i-1} \le X \le x_i)$ theo công thức:

$$p_i = \varphi(\frac{x_i - \mu}{\sigma}) - \varphi(\frac{x_{i-1} - \mu}{\sigma}) = \varphi(\frac{x_i - 11,33}{1,6640}) - \varphi(\frac{x_{i-1} - 11,33}{1,6640})$$

trong đó φ là hàm Laplace, và lập bảng:

X_i	n_i	p_i	np_i	$(n_i - np_i)^2/np_i$
(−∞, 9)	8	0,08076	8,076	0,0007
9-10	15	0,1311	13,11	0,2725
10-11	21	0,20888	20,888	0,0006
11-12	23	0,23468	23,468	0,0093
12-13	16	0,18592	18,592	0,3614
13-14	9	0,10386	10,386	0,1850
(14,+∞)	8	0,0548	5,48	1,1588
Tổng	n = 100			χ^2 =1,9883

Bước 1: Ta có $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} = 1,9883.$

Bước 2: Số tham số chưa biết là r = 2 (do μ, σ^2 chưa biết). Ta có k – r – 1 = 7 – 2 – 1 = 4. Tra bảng phân phối Chi bình phương $\chi^2 \sim \chi^2(4)$ với 4 bậc tự do, ta được: $\chi^2_\alpha = \chi^2_{0,05} = 9,488$.

Bước 3: Kiểm định: Vì χ^2= 1,9883 < 9,488 = χ^2_α nên ta chấp nhận giả thiết H_0.

Kết luận: Với mức ý nghĩa 2%, X có phân phối chuẩn: X ~ N(μ,σ^2) với μ = 11,33; σ = 1,6640.

Bài 11. Bảng số liệu điều tra về tình hình học tập của 10.000 sinh viên của một trường đại học như sau:

	Giỏi	Khá	Trung bình và kém
Nam	1620	2680	2500
Nữ	880	1320	1000

Với mức ý nghĩa 5%, xét xem có sự khác biệt về chất lượng học tập của nam và nữ hay không.

Lời giải

Gọi

- X là giới tính của sinh viên.
- Y là chất lượng học tập của sinh viên.

Bài toán yêu cầu kiểm định giả thiết sau với mức ý nghĩa α = 5% = 0,05:

H_0: X độc lập với Y với giả thiết đối **H_1**: X không độc lập với Y

Ta lập bảng:

Y / X	Giỏi	Khá	Trung bình và kém	Tổng
Nam	1620 $\alpha_{11} = 0,154376$	2680 $\alpha_{12} = 0,264059$	2500 $\alpha_{13} = 0,262605$	6800
Nữ	880 $\alpha_{21} = 0,0968$	1320 $\alpha_{22} = 0,136125$	1000 $\alpha_{23} = 0,089286$	3200
Tổng	2500	4000	3500	n=10000

trong đó α_{ij} được tính theo công thức: $\alpha_{ij} = \frac{(n_{ij})^2}{m_i n_j}$. Cụ thể:

$\alpha_{11} = \frac{1620^2}{6800 \times 2500} = 0,154376$,... (kết quả được ghi chi tiết trong bảng).

Bước 1: Ta có $\chi^2 = n\left(\sum\sum \alpha_{ij} - 1\right) = 32,51.$

Bước 2: Ta có (h – 1)(k – 1) = 2 (do h = 2; k = 3). Tra bảng phân phối chi bình phương χ^2~$\chi^2(2)$ với 2 bậc tự do, ta được: $\chi^2_\alpha = \chi^2_{0,05} = 5,991.$

Bước 3: Kiểm định:

Vì χ^2= 32,51 > 5,991 = χ^2_α nên ta bác bỏ giả thiết H_0, nghĩa là chấp nhận **H_1**: X không độc lập với Y.

Kết luận: Với mức ý nghĩa 5%, có sự khác biệt về chất lượng học tập của nam và nữ.

MỘT SỐ ĐỀ THI CAO HỌC THAM KHẢO
(TRƯỜNG ĐH KINH TẾ-LUẬT, ĐHQG TP HCM)

Bài 1 (2005). Để khảo sát chiều cao của một giống cây trồng, người ta quan sát một mẫu và có kết quả sau:

Chiều cao (cm)	95-105	105-115	115-125	125-135	135-145	145-155	155-165
Số cây	10	10	15	30	10	10	15

a) Ước lượng chiều cao trung bình của giống cây trồng trên với độ tin cậy 95%.
b) Nếu muốn ước lượng chiều cao trung bình của giống cây trồng trên với độ tin cậy 99% và độ chính xác 4 cm thì cần phải điều tra thêm bao nhiêu cây nữa?
c) Những cây trồng có chiều cao từ 135cm trở lên được gọi là những cây "cao". Hãy ước lượng tỉ lệ những cây "cao" với độ tin cậy 99,73%.
d) Giả sử trước khi điều tra, tỉ lệ những cây "cao" là 30%.Số liệu điều tra trên được tiến hành sau khi áp dụng phương pháp trồng trọt mới. Hãy cho kết luận về phương pháp trồng trọt mới với mức ý nghĩa 5%.

Cho biết : $\varphi(1,96) = 0,475; \varphi(2,58) = 0,4951; \varphi(3) = 0,49865$.

Bài 2 (2007). Số liệu thống kê về doanh số bán của một cửa hàng như sau:

Doanh số (triệu đồng/ngày)	35-45	45-55	55-65	65-75	75-85	85-95	95-105	105-115
Số ngày	5	15	5	20	25	15	10	5

a) Ước lượng doanh số bán trung bình của cửa hàng với độ tin cậy 99%.
b) Trước đây doanh số bán trung bình của cửa hàng là 70 triệu đồng/ngày. Số liệu bảng trên thu nhập được sau khi cửa hàng này áp dụng phương pháp bán hàng mới. Hãy cho nhận xét về phương pháp bán hàng mới ấy với mức ý nghĩa 1%.
c) Những ngày có doanh số bán trên 75 triệu đồng là những ngày bán đắt hàng. Hãy ước lượng tỉ lệ những ngày bán đắt hàng ở cửa hàng này với độ tin cậy 95%.

Cho biết:
$\varphi(0,49) = 0,1879, \varphi(0,50) = 0,1915, \varphi(1,96) = 0,475; \varphi(2,58) = 0,4951; \varphi(3) = 0,49865,$
$\varphi(x) = 0,5\ \forall x > 5.$

Bài 3 (2008). Để theo dõi sự phát triển chiều cao của một giống cây trồng trong một nông trại, người ta quan sát một mẫu và có kết quả sau:

Chiều cao(cm)	70-74	74-78	78-82	82-86	86-90	90-94
Số cây	20	10	30	20	10	10

a) Ước lượng chiều cao trung bình của giống cây trồng trên với độ tin cậy 95%.
b) Nếu muốn ước lượng chiều cao trung bình của giống cây trồng trên với độ tin cậy 99% và độ chính xác 1,2 cm thì cần phải điều tra thêm bao nhiêu cây nữa?
c) Những cây có chiều cao từ 82 cm trở lên gọi là những cây cao.Trước đây tỷ lệ những cây cao của loại cây trồng trên là 48%. Số liệu trên đây thu được khi áp dụng phương pháp trồng trọt mới. Hãy cho kết luận về phương pháp trồng trọt mới với mức ý nghĩa 5%.

Cho biết:
$\varphi(0,49) = 0,1879, \varphi(0,50) = 0,1915, \varphi(1,96) = 0,475; \varphi(2) = 0,4772; \varphi(2,58) = 0,4951;$
$\varphi(3) = 0,49865, \varphi(x) = 0,5\ \forall x \geq 5.$

Bài 4 (2009). Để nghiên cứu nhu cầu của một loại hàng ở một khu vực , người ta khảo sát 400 hộ gia đình và có kết quả sau:

Nhu cầu (kg/tháng/hộ)	1 – 2	2 – 3	3 – 4	4 – 5	5 – 6	6 – 7	7 – 8	8 – 9
Số hộ	15	30	85	135	75	30	20	10

Cho biết trong khu vực có 4000 hộ dân và nhu cầu của các tháng trong năm là như nhau.

a) Ước lượng nhu cầu trung bình mặt hàng này trong toàn khu vực trong năm với độ tin cậy 95%.
b) Khi ước lương nhu cầu của mặt hàng này của toàn khu vực trong một năm, nếu ta muốn đạt được độ tin cậy 99% và độ chính xác 4,5 tấn thì cần khảo sát ở ít nhất bao nhiêu hộ gia đình?

Cho biết:
$\varphi(0,35) = 0,1368; \varphi(0,49) = 0,1879; \varphi(0,50) = 0,1915; \varphi(1,70) = 0,4554; \varphi(1,96) = 0,4750;$
$\varphi(2,32) = 0,4898; \varphi(2,58) = 0,4951; \varphi(3) = 0,49865; \varphi(x) = 0,5\ \forall x \geq 5.$

Bài 5 (2010 – Đợt 1). Để khảo sát chỉ tiêu X của một loại sản phẩm, người ta quan sát một mẫu và có kết quả sau

X(cm)	11-13	13-15	15-17	17-19	19-21	21-23	23-25
Số sản phẩm	12	14	30	29	18	16	12

a) Nếu muốn ước lượng giá trị trung bình của chỉ tiêu X với độ chính xác 0,6 cm thì sẽ đạt được độ tin cậy là bao nhiêu?
b) Những sản phẩm có chỉ tiêu X từ 17–23cm là những sản phẩm loại A. Ước lượng tỉ lệ sản phẩm loại A với độ tin cậy 95%. Với độ tin cậy đó, nếu trong kho có 1000 sản phẩm loại A thì tổng số sản phẩm có trong kho khoảng bao nhiêu?

Cho biết:
$\varphi(0,11) = 0,0438; \varphi(0,29) = 0,1141; \varphi(0,9) = 0,3159;$
$\varphi(1,70) = 0,4554;\ \varphi(1,96) = 0,4750; \varphi(2) = 0,4772; \varphi(2,32) = 0,4898, \varphi(2,58) = 0,4951;$
$\varphi(3) = 0,49865; \varphi(x) = 0,5\ \forall x \geq 5.$

Bài 6 (2010 – Đợt 2). Để khảo sát chiều cao X của một giống cây trồng, người ta quan sát một mẫu và có kết quả sau:

X(cm)	95 -105	105 -115	115-125	125-135	135-145	145-155	155-165
Số cây	10	10	15	30	10	10	15

a) Nếu muốn ước lượng chiều cao trung bình của giống cây trồng trên với độ tin cậy 99% và độ chính xác 4 cm thì cần phải điều tra thêm bao nhiêu cây nữa?
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%.

Cho biết:
$\varphi(0,11) = 0,0438; \varphi(0,29) = 0,1141; \varphi(0,9) = 0,3159; \varphi(1,96) = 0,4750;$
$\varphi(2) = 0,4772;\ \varphi(2,32) = 0,4898;\ \varphi(2,58) = 0,4951;\ \varphi(3) = 0,49865;$
$\varphi(x) = 0,5\ \forall x \geq 5.$
