

II. PHẦN TỰ LUẬN

Bài 21. Theo dõi cân nặng của các con heo ở một trang trại theo tháng, người ta quan sát được rằng trọng lượng của các con heo sau 4 tháng là biến ngẫu nhiên tuân theo quy luật phân phối chuẩn với cân nặng trung bình là 55 (kg) và độ lệch chuẩn là 5 (kg).

Câu 1. (0.5 điểm) Xác định mức cân nặng sau 4 tháng sao cho có 70% của tổng số các con heo ở trang trại này có cân nặng lớn hơn mức cân nặng này.

Lời giải

Gọi X là cân nặng của con heo ở trang trại. X là BNN liên tục có phân phối chuẩn với cân nặng trung bình là $\mu_X = 55$ và độ lệch chuẩn là $\sigma_X = 5$. Gọi x_0 mức cân nặng sau 4 tháng sao cho 70 phần trăm của tổng số các con heo ở trang trại này có cân nặng lớn hơn mức cân nặng x_0 . Khi đó ta có

$$\begin{aligned}P(X > x_0) &= 70\% \Rightarrow P(X \leq x_0) = 100\% - 70\% \\&\Rightarrow P\left(\frac{X - \mu_X}{\sigma_X} \leq \frac{x_0 - \mu_X}{\sigma_X}\right) = 30\% = 0.3 \\&\Rightarrow \frac{x_0 - \mu_X}{\sigma_X} \approx -0.53 \Rightarrow x_0 \approx -0.53 \times \sigma_X + \mu_X = -0.53 \times 5 + 55 \approx 52.35\end{aligned}$$

□

Câu 2. (0.5 điểm) Tìm xác suất để trong 500 con heo được chọn từ trang trại này, có từ 100 đến 120 con heo có cân nặng sau 4 tháng lớn hơn 58 (kg).

Lời giải

Gọi X là cân nặng của con heo trong trang trại. X là BNN liên tục có phân phối chuẩn với cân nặng trung bình là $\mu_X = 55$ và độ lệch chuẩn là $\sigma_X = 5$ nên xác suất để con heo được chọn từ trang trại có cân nặng lớn hơn 58 là

$$P(X > 58) = P\left(\frac{X - \mu_X}{\sigma_X} > \frac{58 - \mu_X}{\sigma_X}\right) \approx 0.2743$$

Gọi BNN Y là số con heo có cân nặng lớn hơn 58 trong 500 con heo được chọn. Xác suất để con heo có cân nặng lớn hơn 58 là $p \approx 0.2743$. Xác suất để trong 500 sinh viên được chọn, có từ 100 đến 120 con heo có cân nặng lớn hơn 58 là

$$P(100 \leq Y \leq 120) = \sum_{k=100}^{120} C_{500}^k p^k (1-p)^{500-k}$$

Xác suất này có thể được xấp xỉ bởi BNN có phân phối chuẩn chuẩn tắc với hiệu chỉnh liên tục ± 0.5 như sau:

$$P(100 \leq Y \leq 120) \approx P\left(\frac{100 - 0.5 - 500 \times p}{\sqrt{500 \times p(1-p)}} \leq \frac{Y - 500 \times p}{\sqrt{500 \times p(1-p)}} \leq \frac{120 + 0.5 - 500 \times p}{\sqrt{500 \times p(1-p)}}\right) \approx 0.0475.$$

□

Bài 22. Nồng độ bụi (đơn vị: mg/m^3) tại ba giao lộ ở TPHCM được ghi nhận tại 5 thời điểm khác nhau của năm 2020 và được thể hiện trong bảng sau. Giả thiết rằng các số liệu thỏa mãn mô hình Anova.

Địa điểm	Nồng độ bụi				
Giao lộ 1	0.9	1.2	0.8	0.6	0.7
Giao lộ 2	0.9	0.8	0.8	0.7	0.8
Giao lộ 3	0.5	1.1	0.6	0.5	0.6

Câu 3. (0.5 điểm) Tìm khoảng tin cậy cho sự khác biệt trung bình của nồng độ bụi ở giao lộ 1 và 2 với độ tin cậy 95%.

Lời giải

Ta có $a = 3, n = 5, \alpha = 0.05$, và $y_{1\cdot} = \sum_{j=1}^n y_{1j} = 4.2; \bar{y}_{1\cdot} = \frac{y_{1\cdot}}{n} = 0.84; y_{2\cdot} = \sum_{j=1}^n y_{2j} = 4; \bar{y}_{2\cdot} = \frac{y_{2\cdot}}{n} = 0.8;$
 $y_{3\cdot} = \sum_{j=1}^n y_{3j} = 3.3; \bar{y}_{3\cdot} = \frac{y_{3\cdot}}{n} = 0.66; SSE = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i\cdot})^2 = 0.484; MSE = \frac{SSE}{a(n-1)} = 0.0403$. Khi đó,
 95% = $100(1 - \alpha)\%$ khoảng tin cậy cho $\mu_1 - \mu_2$ là

$$\begin{aligned} \bar{y}_1 - \bar{y}_2 - t_{\alpha/2, a(n-1)} \sqrt{\frac{2MSE}{n}} &\leq \mu_1 - \mu_2 \leq \bar{y}_1 - \bar{y}_2 + t_{\alpha/2, a(n-1)} \sqrt{\frac{2MSE}{n}} \\ \Leftrightarrow 0.84 - 0.8 - 2.179 \sqrt{\frac{2 \times 0.0403}{5}} &\leq \mu_1 - \mu_2 \leq 0.84 - 0.8 + 2.179 \sqrt{\frac{2 \times 0.0403}{5}} \\ \Leftrightarrow -0.2366 &\leq \mu_1 - \mu_2 \leq 0.3167 \end{aligned}$$

□

Câu 4. (1.5 điểm) Hãy dùng phương pháp ANOVA để so sánh nồng độ bụi trong không khí ở ba giao lộ trên với mức ý nghĩa 5%?

Lời giải

Ta có $a = 3, n = 5, N = a \times n = 15, \alpha = 0.05$, và $y_{1\cdot} = \sum_{j=1}^n y_{1j} = 4.2; \bar{y}_{1\cdot} = \frac{y_{1\cdot}}{n} = 0.84; y_{2\cdot} = \sum_{j=1}^n y_{2j} = 4;$
 $\bar{y}_{2\cdot} = \frac{y_{2\cdot}}{n} = 0.8; y_{3\cdot} = \sum_{j=1}^n y_{3j} = 3.3; \bar{y}_{3\cdot} = \frac{y_{3\cdot}}{n} = 0.66; y_{\cdot\cdot} = \sum_{i=1}^a \sum_{j=1}^n y_{ij} = 11.5; \bar{y}_{\cdot\cdot} = \frac{y_{\cdot\cdot}}{N} = \frac{23}{30}; SST =$
 $\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{\cdot\cdot})^2 = 0.5733, MST = \frac{SST}{an-1} = 0.041; SSTr = n \sum_{i=1}^a (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 = 0.0893; MSTr = \frac{SSTr}{a-1} =$
 $0.0447; SSE = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i\cdot})^2 = 0.484; MSE = \frac{SSE}{a(n-1)} = 0.0403.$

Giả thuyết $H_0 : \tau_1 = \tau_2 = \tau_3 = 0$

Thống kê kiểm định

$$f_0 = \frac{SSTr/(a-1)}{SSE/[a(n-1)]} = \frac{MSTr}{MSE} \approx 1.1074$$

Mặt khác $f_{\alpha, a-1, a(n-1)} = f_{0.05, 2, 12} = 3.89 > f_0$ nên ta không bác bỏ H_0 .

Kết luận: không có sự khác biệt giữa các trung bình nồng độ bụi tại 03 giao lộ 1,2,3.

□