Bài 10: BÀI TOÁN ƯỚC LƯỢNG TỶ LỆ

Vũ Mạnh Tới

Bộ môn Toán-Trường Đại học Thủy lợi

Ngày 21 tháng 5 năm 2024

Bài toán: Lấy ra một mẫu cỡ n ($n \geq 30$) từ một tổng thể thấy có x phần tử có tính chất A ta quan tâm. Khi đó

- Ước lượng điểm cho tỷ lệ p của tổng thể có tính chất A là $\hat{p} = \dfrac{x}{n}$.
- Khoảng tin cậy $(1-\alpha)100\%$ cho p khi cỡ mẫu lớn ($n \geq 30$)

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

với $\hat{q} = 1 - \hat{p}$, $z_{\alpha/2}$ tra A_3 sao cho $P(Z > z_{\alpha/2}) = \alpha/2$.

Sai số (khi biết ước lượng điểm $\hat{p}$)

$$|\hat{p} - p| < \varepsilon = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Cỡ mẫu (khi biết ước lượng điểm $\hat{p}$): Với độ tin cậy $(1-\alpha)\,100\%$, để sai số không vượt quá ε thì cỡ mẫu tối thiểu là:

$$n = \left[\frac{z_{\alpha/2}^2 \hat{p}\hat{q}}{\varepsilon^2}\right] + 1$$

Ví dụ 1: Điều tra ngẫu nhiên 500 gia đình có tivi ở thành phố Hamilton, Canada, thấy rằng có 340 gia đình sử dụng chương trình HBO.

a. Hãy tìm khoảng tin cậy 95% cho tỷ lệ gia đình sử dụng chương trình HBO trong số những gia đình có tivi trong thành phố này.
b. Vơi độ tin cậy 95%, tìm sai số khi dùng $\hat{p}$ làm ước lượng điểm cho p?
c. Muốn sai số không vượt quá 0,02, thì với độ tin cậy 95%, cỡ mẫu tối thiểu phải dùng là bao nhiêu?
d. Nếu thành phố đó có 15000 hộ gia đình có ti vi. Hỏi rằng với độ tin cậy 95% thì có tối đa và tối thiểu bao nhiêu hộ gia đình sử dụng chương trình HBO?

Chú ý:

Ở đây $\hat{p} + \hat{q} = 1$, nên sử dụng bất đẳng thức Cauchy ta có: $\hat{p} + \hat{q} \geq 2\sqrt{\hat{p}\hat{q}} \Rightarrow \hat{p}\hat{q} \leq \left(\dfrac{\hat{p} + \hat{q}}{2}\right)^2 = \dfrac{1}{4}$

Từ đó ta có một công thức nữa sử dụng để tính kích thước mẫu thỏa mãn điều kiện về sai số, mà không cần dựa vào tỷ lệ của một mẫu cho trước nào.

Sai số, cỡ mẫu khi chưa biết $\hat{p}$

Với độ tin cậy $(1-\alpha)\,100\%$ thì

- Sai số: $|\hat{p}-p| < \varepsilon = \dfrac{z_{\alpha/2}}{2\sqrt{n}}$.
- Muốn sai số của ước lượng không vượt quá ε thì cỡ mẫu tối thiểu là:

$$n = \left[\frac{z^2_{\alpha/2}}{4\varepsilon^2}\right] + 1$$

Ví dụ 2: Một mẫu cỡ 50 lấy ra từ một tổng thể. Với độ tin cậy 96% thì sai số mắc phải khi ước lượng cho tỉ lệ p của tổng thể.

Bài toán: Lấy ra hai mẫu từ hai tổng thể. Mẫu 1 có cỡ n_1 và có x_1 phần tử có tính chất A; mẫu 2 có cỡ n_2 và có x_2 phần tử có tính chất A. Khi đó

- Ước lượng điểm cho hiệu tỷ lệ tổng thể $p_1 - p_2$ là $\hat{p}_1 - \hat{p}_2 = \dfrac{x_1}{n_1} - \dfrac{x_2}{n_2}$.
- Khoảng tin cậy $(1-\alpha)100\%$ cho $p_1 - p_2$:

$$\hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < p_1 - p_2 < \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

với $\hat{q}_1 = 1 - \hat{p}_1, \hat{q}_2 = 1 - \hat{p}_2$
và $z_{\alpha/2}$ tra A_3 sao cho $P(Z > z_{\alpha/2}) = \alpha/2$.

Ví dụ 1: Người ta đang xem xét xem có nên thay đổi quy trình sản xuất không bằng cách nghiên cứu quy trình cũ và mới nhằm xem quá trình mới có hiệu quả không. Trong 1500 sản phẩm được sản xuất theo quy trình cũ người ta thấy có 75 phế phẩm, sử dụng quy trình mới thì trong 2000 sản phẩm người ta đếm được 80 phế phẩm. Tìm khoảng tin cậy 90% cho hiệu tỷ lệ phế phẩm của 2 quy trình cũ và mới, từ đó cho biết việc cải tiến có đem lại chất lượng tốt hơn cho dây chuyền sản xuất không?

Ví dụ 2: Vào năm 1999, phỏng vấn ngẫu nhiêu 1500 nam giới thì tỷ lệ hút thuốc lá là 43%. Sau 5 năm vận động, phỏng vấn 1000 người thì thấy tỉ lệ hút thuốc là 38%. Với độ tin cậy 95%, ước lượng tỷ lệ người hút thuốc lá đã giảm trong 5 năm vận động.

Ví dụ 3. Kiểm tra một lượng trứng gà tại một cơ sở ấp trứng tại Bắc Ninh cho kết quả về khối lượng trứng như sau:

Khối lượng (gam)	155	160	165	170	180	185
Số quả	10	15	35	20	14	6

Tìm khoảng tin cậy 96% cho tỷ lệ trứng gà đạt chuẩn. Biết rằng trứng đạt chuẩn nếu khối lượng từ 160g đến 170g.

1. Bài toán ước lượng khoảng cho μ
   1. σ đã biết: (công thức khoảng, sai số, cỡ mẫu) (tra A_3)
   2. σ chưa biết (công thức khoảng, sai số)
      1. Cỡ mẫu nhỏ ($n < 30$) (tra A_4)
      2. Cỡ mẫu lớn ($n \geq 30$) (tra A_3)
2. Bài toán ước lượng khoảng cho $\mu_1 - \mu_2$
   1. Khi σ_1, σ_2 đã biết (tra A_3)
   2. Khi σ_1, σ_2 chưa biết, nhưng biết rằng $\sigma_1 = \sigma_2$
      1. Cỡ mẫu nhỏ $n_1 < 30$ hoặc $n_2 < 30$ hoặc $n_1, n_2 < 30$ (tra A_4)
      2. Cỡ mẫu lớn $n_1 \geq 30, n_2 \geq 30$. (tra A_3)
3. Bài toán ước lượng khoảng cho 1 tỷ lệ, cỡ mẫu lớn ($n \geq 30$) (tra A_3)
   1. Công thức khoảng tin cậy
   2. Sai số và cỡ mẫu khi biết $\hat{p}$ và khi chưa biết $\hat{p}$
   3. Các câu hỏi phụ liên quan
4. Bài toán ước lượng khoảng cho hiệu hai tỷ lệ, cỡ mẫu lớn (tra A_3)

Bài tập

1,3,7,8,14,15,17,19**(299-300)**