Bài 9: BÀI TOÁN ƯỚC LƯỢNG TRUNG BÌNH

Vũ Mạnh Tới

Bộ môn Toán-Trường Đại học Thủy lợi

Ngày 17 tháng 5 năm 2024

Đặt vấn đề:

Lí thuyết về suy luận thống kê bao gồm phương pháp đưa ra những kết luận, khái quát về tổng thể thông qua việc phân tích những thông tin thu được từ mẫu ngẫu nhiên được chọn từ tổng thể đó.
Suy luận thống kê được chia thành hai mảng chính:

- Ước lượng;
- Kiểm định giả thuyết.

- Một ứng cử viên tổng thống muốn ước lượng tỉ lệ cử tri ủng hộ ông ta trong cả nước bằng cách lấy ý kiến của 200 cử tri, được chọn ngẫu nhiên và độc lập từ các nơi khác nhau. Tỉ lệ trong mẫu ủng hộ ông ta làm cơ sở để đưa ra ước lượng về tỉ lệ ủng hộ ông ta trong cả nước. $\Longrightarrow$ Bài toán ước lượng tham số.
- Khi ông ta tuyên bố tỉ lệ ủng hộ ông ta trên cả nước là 70%, thì vấn đề đặt ra đối với ứng cử viên đối lập là cần tìm hiểu xem lời phát biểu trên có cơ sở hay không. Lúc này, ta không ước lượng tham số mà phải đưa ra quyết định chấp nhận hay không chấp nhận lời tuyên bố trên. Lại phải dựa vào lý thuyết về mẫu và các số liệu thực nghiệm để đưa ra quyết định với độ chính xác có thể xác định được $\Longrightarrow$ Kiểm định giả

9.1. Các phương pháp ước lượng truyền thống
9.1.1. Ước lượng điểm

a) Định nghĩa:

Ước lượng điểm về một tham số chung của tổng thể θ nào đó là một giá trị đơn $\hat{\theta}$ của một thống kê $\hat{\Theta}$.

Ví dụ:

- Giá trị trung bình mẫu $\overline{x}$ của thống kê $\overline{X}$, là một ước lượng điểm của trung bình tổng thể μ.
- Giá trị $\hat{p} = \dfrac{x}{n}$ của thống kê $\hat{P}$ là một ước lượng điểm của tỷ lệ p của toàn tổng thể.

Định nghĩa: Cho trước $\alpha \in (0; 1)$. Khoảng tin cậy cho tham số tổng thể θ với độ tin cậy $(1-\alpha)\,100\%$ là khoảng $(\Theta_{\mathbf{L}}; \Theta_{\mathbf{U}})$ nếu như

$$P\left(\Theta_{\mathbf{L}} < \theta < \Theta_{\mathbf{U}}\right) = 1 - \alpha$$

$1 - \alpha$ được gọi là hệ số tin cậy.
Ngoài ra còn có kiểu ước lượng hiệu quả và ước lượng không chệch (tự đọc).

9.2.1. Khoảng tin cậy cho kỳ vọng μ, khi σ đã biết

Bài toán: Lấy ra một mẫu cỡ n từ một tổng thể cho giá trị trung bình mẫu $\overline{x}$. Khi đó

- Ước lượng điểm cho trung bình tổng thể μ là $\overline{x}$.
- Ước lượng khoảng: Dự đoán khoảng giá trị mà μ có thể nhận từ số liệu mẫu được không? Tức là ta cần tìm khoảng tin cậy $(1-\alpha)100\%$ cho trung bình tổng thể μ.

Phân ra hai trường hợp khi tổng thể:

- đã biết phương sai tổng thể σ^2,
- chưa biết phương sai tổng thể σ^2.

Ta xét khoảng tin cậy $(1-\alpha)\,100\%$ cho trung bình μ của tổng thể khi đã biết độ lệch chuẩn tổng thể σ.

Theo định lí giới hạn trung tâm: $Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(z; 0; 1)$.

Ta gọi $z_{\alpha/2}$ sao cho

$$P\left(Z > z_{\alpha/2}\right) = \frac{\alpha}{2} \Longleftrightarrow 1 - P\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right) - P\left(Z < -z_{\alpha/2}\right) = \frac{\alpha}{2}$$

$$\Leftrightarrow P\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right) = 1 - \alpha$$

Do $Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ thay vào ta được

$$P\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Định lí: (Công thức khoảng tin cậy)
Nếu $\overline{x}$ là trung bình của một mẫu ngẫu nhiên cỡ n được lấy từ tổng thể với σ^2 đã biết, thì khoảng tin cậy $(1-\alpha)100\%$ cho trung bình μ xác định như sau

$$\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Trong đó $z_{\alpha/2}$ thỏa mãn $P(Z > z_{\alpha/2}) = \frac{\alpha}{2}$.

Chú ý: Để tính giá trị $z_{\alpha/2}$ ta tra bảng A.3, với giá trị diện tích $P\left(Z < z_{\alpha/2}\right) = 1 - \alpha/2$.

Định lí: (Sai số)

$$|\overline{x} - \mu| < \varepsilon = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Định lí: (Cỡ mẫu)

Để sai số của ước lượng không vượt quá ε khi cỡ mẫu tối thiểu cần dùng là:

$$n = \left[\left(\frac{z_{\alpha/2}\sigma}{\varepsilon}\right)^2\right] + 1.$$

Trong đó $\left[\left(\frac{z_{\alpha/2}\sigma}{\varepsilon}\right)^2\right]$ là phần nguyên của $\left(\frac{z_{\alpha/2}\sigma}{\varepsilon}\right)^2$.

Ví dụ: $[124, 35] = 124$

Ví dụ 1: Hàm lượng kẽm trung bình thu được khi đo ở 35 địa điểm khác nhau trên một dòng sông là 2,5 g/ml. Biết rằng hàm lượng kẽm của dòng sông đó tuân theo phân phối có độ lệch chuẩn là 0,3.

a) Hãy tìm các khoảng tin cậy 95% cho hàm lượng kẽm trung bình của dòng sông đó.

b) Hãy tìm các khoảng tin cậy 98% cho hàm lượng kẽm trung bình của dòng sông đó.

c) Tìm sai số khi ước lượng khoảng 95% cho lượng kẽm trung bình của dòng sông.

d) Nếu ta muốn với độ tin cậy 95% sai số của ước lượng $\overline{x}$ cho μ không vượt quá 0,05, thì cỡ mẫu tối thiểu phải là bao nhiêu?

a. $2.4006 < \mu < 2.5994$; b. $2.3818 < \mu < 2.6182$; c. $|\overline{x} - \mu| \leq 0.0994$; d. $n = 139$.

Định lí:

Nếu $\overline{x}$ và s lần lượt là trung bình và độ lệch chuẩn mẫu của một mẫu ngẫu nhiên được chọn từ tổng thể có phân phối chuẩn (hoặc xấp xỉ chuẩn) với σ chưa biết, cỡ mẫu $n < 30$, khoảng tin cậy $(1-\alpha)100\%$ cho μ xác định bởi

$$\overline{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \overline{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

trong đó $t_{\alpha/2}$ là giá trị của phân phối t (tra A_4) với $v = n-1$ bậc tự do mà $P(T > t_{\alpha/2}) = \alpha/2$.

Sai số:

$$|\overline{x} - \mu| < \varepsilon = t_{\alpha/2}\frac{s}{\sqrt{n}}.$$

9.2.2.2. Khoảng tin cậy cho μ khi σ chưa biết cỡ mẫu $n \geq 30$

Định lí (Công thức khoảng).

$$\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}$$

Với $z_{\alpha/2}$ tra bởi bảng A_3 như phần trước.

Sai số:

$$|\bar{x} - \mu| < \varepsilon = z_{\alpha/2}\frac{s}{\sqrt{n}}.$$

Ví dụ 1 Một mẫu được lấy từ tổng thể có phân phối xấp xỉ chuẩn với số liệu như sau

9,8 10,2 10,4 9,8 10,0 10,2 và 9,6

Hãy tìm khoảng tin cậy 95% cho giá trị trung bình của tổng thể.
ĐS: $9.7384 < \mu < 10.2616$; $s = 0,2828, \overline{x} = 10,0$
Ví dụ 2. 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

Biết rằng khối lượng của trứng tuân theo phân phối xấp xỉ chuẩn. Hãy tìm khoảng tin cậy 98% cho khối lượng trung bình của trứng gà tại cơ sở đó.

ĐS: $165.6225 < \mu < 169.4758$. $(\overline{x} = 167,55; s = 8,2724)$.

Bài toán:

Từ tổng thể 1 (với μ_1) lấy ra mẫu cỡ n_1 cho giá trị trung bình mẫu $\overline{x}_1$;

Từ tổng thể 2 (với μ_2) lấy ra mẫu cỡ n_2 cho giá trị trung bình mẫu $\overline{x}_2$.

Khi đó

- Ước lượng điểm cho $\mu_1 - \mu_2$ là $\overline{x}_1 - \overline{x}_2$.
- Hỏi khoảng tin cậy $(1-\alpha)100\%$ cho $\mu_1 - \mu_2$?

Tùy vào σ_1, σ_2 đã biết hay chưa biết!

9.3.1. Khoảng tin cậy $(1-\alpha)100\%$ cho $\mu_1 - \mu_2$ đã biết $\sigma_1; \sigma_2$

$$(\overline{x}_1 - \overline{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\overline{x}_1 - \overline{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

trong đó $z_{\alpha/2}$ thỏa mãn $P(Z > z_{\alpha/2}) = \dfrac{\alpha}{2}$ tra A_3.

Chú ý:

- Tổng thể cần có phân phối xấp xỉ chuẩn khi $n_1 < 30$ hoặc $n_2 < 30$.
- Khi $n_1, n_2 \geq 30$ thì tổng thể tùy ý.

Ví dụ: Cho hai tổng thể A và B với độ lệch chuẩn lần lượt là 6 và 8. Một mẫu từ A có cỡ là 50, cho ta giá trị trung bình mẫu là 39. Một mẫu từ B có cỡ là 75, cho ta giá trị trung bình là 42. Tìm khoảng tin cậy 96% cho hiệu trung bình của tổng thể A và tổng thể B.

Đáp số: $-5.5714 < \mu_1 - \mu_2 < -0.4286$

9.3.2. Khoảng tin cậy cho $\mu_1 - \mu_2$, chưa biết $\sigma_1; \sigma_2$ nhưng giả thiết $\sigma_1^2 = \sigma_2^2$

Trường hợp cỡ mẫu nhỏ: $n_1 < 30$ hoặc $n_2 < 30$ hoặc $n_1 < 30, n_2 < 30$.

Nếu $\overline{x_1}$ và $\overline{x_2}$ là trung bình mẫu của các mẫu ngẫu nhiên độc lập cỡ lần lượt là n_1 và n_2 được lấy từ những tổng thể có phân phối chuẩn (hoặc xấp xỉ chuẩn) với các phương sai tổng thể bằng nhau nhưng chưa biết, thì khoảng tin cậy $(1-\alpha)100$ % cho $\mu_1 - \mu_2$ xác định bởi

$$(\overline{x}_1 - \overline{x}_2) - t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\overline{x}_1 - \overline{x}_2) + t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

trong đó $t_{\alpha/2}$ là giá trị của t với bậc tự do $v = n_1 + n_2 - 2$ (tra A_4) sao cho $P(T > t_{\alpha/2}) = \alpha/2$ và $s_p = \sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}$.

Trường hợp cỡ mẫu lớn: tức là $n_1 \geq 30$ và $n_2 \geq 30$.

Công thức khoảng tin cậy $(1-\alpha)100\%$ cho $\mu_1 - \mu_2$ khi σ_1, σ_2 chưa biết mà cỡ mẫu lớn xác định bởi

$$(\overline{x}_1 - \overline{x}_2) - z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\overline{x}_1 - \overline{x}_2) + z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

trong đó $z_{\alpha/2}$ thỏa mãn $P(Z > z_{\alpha/2}) = \dfrac{\alpha}{2}$ (tra A_3).

Ví dụ 1: Cho hai tổng thể A và B. Một mẫu từ A có cỡ là 12, cho giá trị trung bình là $3,11$ và độ lệch tiêu chuẩn $0,771$. Một mẫu từ B có cỡ là 10, cho giá trị trung bình là $2,04$ và phương sai là $0,2007$. Tìm khoảng tin cậy 90% cho hiệu trung bình của tổng thể A và B, giả sử rằng hai tổng thể có phân phối xấp xỉ chuẩn và có phương sai bằng nhau nhưng chưa biết.

Đáp số: $0,5929 < \mu_1 - \mu_2 < 1,5471$

Ví dụ 2. Điểm môn Toán từ một số học sinh của hai trường A và B được thống kê với kết quả

A	7.5	5.0	7.0	7.5	8.0	9.5	9.5	9.0	9.0
B	8.5	9.5	7.5	3.5	4.5	8.5	10.0	9.0	9.5

Biết rằng các tổng thể có phân phối xấp xỉ chuẩn và có phương sai bằng nhau.
Hãy tìm khoảng tin cậy cho hiệu trung bình điểm Toán của hai trường đó với độ tin cậy 97%.
Đáp số:
$\overline{x}_1 = 7,6667; s_1^2 = 2,375; \overline{x}_2 = 7,1667; s_2^2 = 2,1360$
$-0,2080 < \mu_1 - \mu_2 < 1,2080$

Các ý chính

- Các loại ước lượng
- Khoảng tin cậy $(1-\alpha)100\%$ cho 1 trung bình khi phương sai tổng thể đã biết và chưa biết
- Khoảng tin cậy $(1-\alpha)100\%$ cho hiệu 2 trung bình.

Bài tập về nhà: Khoảng tin cậy cho 1 trung bình:
4,6,8,10,13,15**(278-279)**
Khoảng tin cậy cho hiệu hai trung bình:
1,4,6,7,8,9**(290-291)**