XÁC SUẤT THỐNG KÊ VÀ ỨNG DỤNG

LÝ THUYẾT ƯỚC LƯỢNG

Buổi học 7

LÝ THUYẾT ƯỚC LƯỢNG

1. ƯỚC LƯỢNG KHÔNG CHỆCH VÀ ƯỚC LƯỢNG CÓ HIỆU QUẢ
2. ƯỚC LƯỢNG ĐIỂM VÀ ƯỚC LƯỢNG KHOẢNG. ĐỘ TIN CẬY
3. ƯỚC LƯỢNG KHOẢNG TIN CẬY CỦA CÁC THAM SỐ TỔNG THỂ
4. ƯỚC LƯỢNG HỢP LÝ TỐI ĐA

ƯỚC LƯỢNG KHÔNG CHỆCH VÀ ƯỚC LƯỢNG CÓ HIỆU QUẢ

- Một thống kê được gọi là *ước lượng không chệch* của tham số tổng thể nếu trung bình hay kỳ vọng của thống kê đó bằng với tham số tổng thể. Giá trị tương ứng của thống kê được gọi là *ước lượng không chệch* của tham số tổng thể.
- **VÍ DỤ 6.1.** Trung bình $\bar{X}$ và phương sai $\hat{S}^2$ là các ước lượng không chệch của trung bình tổng thể μ và phương sai σ^2 nên các giá trị $\bar{x}$ và $\hat{s}^2$ đều là các ước lượng không chệch.

  Tuy nhiên, $\hat{S}$ mới thực sự là ước lượng chệch của σ vì tổng quát ta có $E(\hat{S}) \neq \sigma$

ƯỚC LƯỢNG KHÔNG CHỆCH VÀ ƯỚC LƯỢNG CÓ HIỆU QUẢ

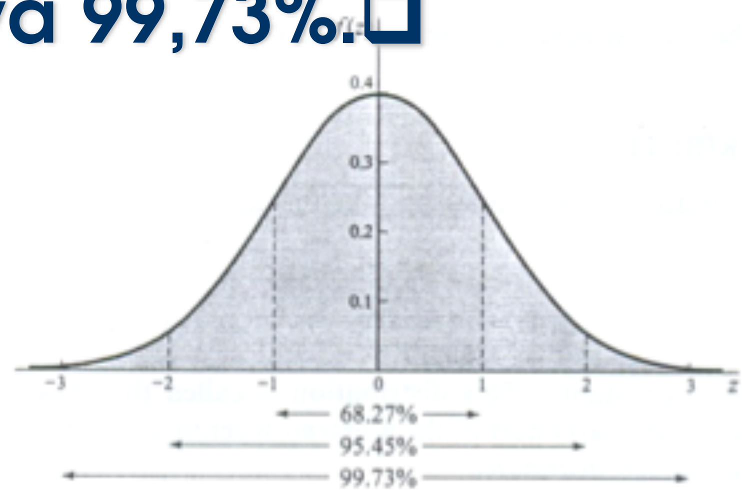
- Nếu luật phân phối mẫu của hai thống kê có cùng trung bình thì thống kê nào ứng với phương sai nhỏ hơn thì thống kê đó được gọi là *ước lượng hiệu quả hơn*. Trong thực hành ta mong muốn được cả hai tiêu chuẩn trên, nhưng điều đó khó có thể xảy ra.
- **VÍ DỤ 6.2.** Với một tổng thể có phân phối chuẩn, trung bình và trung vị mẫu sẽ có cùng giá trị là trung bình, được gọi chung là trung bình của tổng thể. Tuy nhiên, phương sai của các trung bình mẫu thì nhỏ hơn phương sai của các trung vị mẫu nên trung bình cho ta một ước ước lượng hiệu quả hơn trung vị.

ƯỚC LƯỢNG ĐIỂM VÀ KHOẢNG

- Một ước lượng tham số của tổng thể được cho bởi một con số thì được gọi là *ước lượng điểm* của tham số đó. Một ước lượng tham số của tổng thể được cho bởi hai con số mà tham số của tổng thể có thể nằm giữa hai số đó được gọi là *ước lượng khoảng* của tham số.
- VÍ DỤ 6.3. Nếu ta nói khoảng cách là 1,60m thì ta nói về ước lượng điểm. Còn nếu ta nói khoảng cách là $1,60 \pm 0,03$ m, nghĩa là khoảng cách từ 1,57m đến 1,63 m, thì ta đã nói về ước lượng khoảng.

KHOẢNG TIN CẬY CUA TỔNG THỂ

- Gọi μ_s và σ_s lần lượt là trung bình và độ lệch chuẩn (sai số chuẩn) của phân phối thống kê mẫu S. Nếu phân phối mẫu S có thể xấp xỉ phân phối chuẩn (điều này có thể được nếu kích thước mẫu $n \geq 30$), ta có thể hy vọng tìm được S nằm trong khoảng $\mu_S - \sigma_S$ đến $\mu_S + \sigma_S$; từ $\mu_S - 2\sigma_S$ đến $\mu_S + 2\sigma_S$; từ $\mu_S - 3\sigma_S$ đến $\mu_S + 3\sigma_S$ độ khoảng 68,27%, 95,45%, và 99,73%.



KHOẢNG TIN CẬY CỦA TỔNG THỂ

- Một phát biểu tương đương, ta tin rằng có thể tìm được μ_s trong khoảng $S \pm k\sigma_S$ (k = 1, 2, 3) đúng đến 68,27%, 95,45%, và 99,73%. Chính vì lý do đó, ta gọi khoảng trên là *khoảng tin cậy* 68,27%, 95,45%, và 99,73% dùng để ước lượng μ_s (thống kê S là ước lượng không chệch). Các số $k\sigma_S$ (k = 1, 2, 3) được gọi là độ chính xác của ước lượng.
- Tương tự, $S \pm 1{,}96\sigma_S$ và $S \pm 2{,}58\sigma_S$ là *độ tin cậy* 95% và 99% của μ_s, ký hiệu *độ tin cậy* $1-\alpha$. Các số 1,96, 2,58, ... được gọi là *các giá trị tới hạn*, ký hiệu z_α.

Nếu ta có độ tin cậy, ta có thể tìm được giá trị tới hạn và ngược lại, nếu ta có giá trị tới hạn, ta có thể tìm được độ tin cậy.

KHOẢNG TIN CẬY CỦA TỔNG THỂ

- Trong Bảng 6-1 ta cho các giá trị của tương ứng với các độ tin cậy khác nhau, được sử dụng trong thực hành. Với các độ tin cậy không có trong bảng này, các giá trị của z_α có thể tra bảng phân phối chuẩn trong Phụ lục II.

Bảng 6–1

Độ tin cậy	99%	98%	96%	95%	90%	80%	50%
z_α	2,58	2,33	2,05	1,96	1,645	1,28	0,6745

- Trường hợp thống kê có phân phối mẫu khác với phân phối chuẩn (phân phối χ^2, t, F), ta có thể sửa đổi cho phù hợp để tìm khoảng tin cậy của các ước lượng.

ƯỚC LƯỢNG KHOẢNG TIN CẬY CỦA CÁC THAM SỐ TỔNG THỂ

3.1. KHOẢNG TIN CẬY CỦA TRUNG BÌNH

3.2. KHOẢNG TIN CẬY CỦA TỶ LỆ

3.3. KHOẢNG TIN CẬY CỦA HIỆU VÀ TỔNG

3.4. KHOẢNG TIN CẬY CỦA PHƯƠNG SAI CÓ PHÂN PHỐI CHUẨN

3.5. KHOẢNG TIN CẬY TỶ LỆ PHƯƠNG SAI

KHOẢNG TIN CẬY CỦA TRUNG BÌNH

1. TRƯỜNG HỢP MẪU LỚN (n ≥ 30). Khoảng tin cậy trung bình của tổng thể có 2 trường hợp:

❑ Chọn mẫu từ tổng thể vô hạn hay chọn mẫu có hoàn lại từ tổng thể hữu hạn, ta có

$$\bar{X} \pm z_{\alpha} \frac{\sigma}{\sqrt{n}} \qquad (1)$$

❑ Chọn mẫu không hoàn lại từ tổng thể hữu hạn có kích thước N, ta có

$$\bar{X} \pm z_{\alpha} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \qquad (2)$$

❑ Tổng quát, nếu chưa biết độ lệch chuẩn của tổng thể σ, ta thay σ bởi S hay $\hat{S}$.

KHOẢNG TIN CẬY CỦA TRUNG BÌNH

2. TRƯỜNG HỢP MẪU NHỎ (n < 30) VÀ TỔNG THỂ CÓ PHÂN PHỐI CHUẨN.

Khoảng tin cậy trung bình của tổng thể

$$\bar{X} \pm t_{\alpha} \frac{\hat{S}}{\sqrt{n}} \qquad (3)$$

Trong đó $t_{\alpha}(n - 1)$ bậc tự do, tra từ bảng III

- So sánh hai biểu thức (3) và (1) cho thấy, đối với mẫu nhỏ ta thay z_{α} bằng t_{α}. Với mẫu $n \geq 30$, z_{α} và t_{α} bằng nhau.
- Chú ý, thuận lợi trong lý thuyết chọn mẫu nhỏ là ta có thể dùng độ lệch chuẩn mẫu S thay cho σ của tổng thể thường chưa biết.

KHOẢNG TIN CẬY CỦA TỶ LỆ

Với mẫu có kích thước n ≥ 30 được chọn từ một tổng thể có phân phối nhị thức. Khoảng tin cậy P của tỷ lệ tổng thể là $f \pm z_{\alpha}\sigma_f$, trong đó *f* là tỷ lệ thành công trong mẫu.

- **Trường hợp chọn mẫu từ tổng thể vô hạn hay chọn mẫu có hoàn lại từ tổng thể hữu hạn**

$$f \pm z_{\alpha}\sqrt{\frac{f(1-f)}{n}} \qquad (4)$$

- **Trường hợp chọn mẫu không hoàn lại từ tổng thể hữu hạn có kích thước N là**

$$f \pm z_{\alpha}\sqrt{\frac{f(1-f)}{n}}\sqrt{\frac{N-n}{N-1}} \qquad (5)$$

KHOẢNG TIN CẬY CỦA HIỆU VÀ TỔNG

Nếu S_1 và S_2 là hai thống kê mẫu với phân phối mẫu xấp xỉ phân phối chuẩn, khoảng tin cậy của hiệu hai tham số tổng thể tương ứng lần lượt với với S_1 và S_2 là

$$S_1 - S_2 \pm z_\alpha \sigma_{S_1 - S_2} = S_1 - S_2 \pm z_\alpha \sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2} \quad (6)$$

Khoảng tin cậy của tổng hai tham số tổng thể

$$S_1 + S_2 \pm z_\alpha \sigma_{S_1 + S_2} = S_1 + S_2 \pm z_\alpha \sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2} \quad (7)$$

điều kiện là các mẫu độc lập.

Ví dụ, khoảng tin cậy của hiệu hai trung bình tổng thể, trong trường hợp các tổng thể vô hạn và biết độ lệch chuẩn σ_1, σ_2 được cho bởi biểu thức (8)

KHOẢNG TIN CẬY CỦA HIỆU VÀ TỔNG

$$\overline{X}_1 - \overline{X}_2 \pm z_\alpha \sigma_{\overline{X}_1 - \overline{X}_2} = \overline{X}_1 - \overline{X}_2 \pm z_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad (8)$$

trong đó n_1 và n_2 lần lượt là kích thước mẫu của hai mẫu được chọn từ hai tổng thể.

Tương tự, khoảng tin cậy hiệu của hai tỷ lệ tổng thể, trong đó các tổng thể đều vô hạn, được cho bởi

$$f_1 - f_2 \pm z_\alpha \sqrt{\frac{f_1(1-f_1)}{n_1} + \frac{f_2(1-f_2)}{n_2}} \qquad (9)$$

trong đó f_1, f_2 là tỷ lệ và n_1, n_2 là kích thước của hai mẫu được chọn từ hai tổng thể.

KHOẢNG TIN CẬY CỦA PHƯƠNG SAI CÓ PHÂN PHỐI CHUẨN

Lưu ý biểu thức $nS^2/\sigma^2 = (n-1)\hat{S}^2/\sigma^2$ có phân phối chi bình phương với n–1 bậc tự do cho ta khoảng tin cậy của σ^2 hay σ.

với α khá bé ta có thể tìm $\chi^2_{\alpha/2}(n-1)$ và $\chi^2_{1-\alpha/2}(n-1)$ sao cho $P(\chi^2_{1-\alpha/2} < \chi^2 < \chi^2_{\alpha/2}) = 1-\alpha$

Ước lượng cho σ^2 có thể nằm trong khoảng

$$\frac{nS^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{nS^2}{\chi^2_{1-\alpha/2}} \quad \text{hay} \quad \frac{(n-1)\hat{S}^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)\hat{S}^2}{\chi^2_{1-\alpha/2}} \quad (10)$$

Ước lượng cho σ có thể nằm trong khoảng

$$\frac{S\sqrt{n}}{\chi_{\alpha/2}} < \sigma < \frac{S\sqrt{n}}{\chi_{1-\alpha/2}} \quad \text{hay} \quad \frac{\hat{S}\sqrt{n}}{\chi_{\alpha/2}} < \sigma < \frac{\hat{S}\sqrt{n-1}}{\chi_{1-\alpha/2}} \quad (11)$$

KHOẢNG TIN CẬY CỦA TỶ LỆ PHƯƠNG SAI

Lưu ý rằng nếu hai mẫu ngẫu nhiên độc lập có kích thước lần lượt là m và n và phương sai là S_1^2 và S_2^2 được chọn từ hai tổng thể có phân phối chuẩn với phương sai lần lượt là σ_1^2 và σ_2^2 thì đại lượng ngẫu nhiên $\dfrac{\hat{S}_1^2/\sigma_1^2}{\hat{S}_2^2/\sigma_2^2}$ có phân phối *F* với m-1, n-1 bậc tự do.

với α khá bé ta có thể tìm được $F_\alpha(m-1)(n-1)$ và $F_{1-\alpha}(m-1)(n-1)$ sao cho $P(F_\alpha < F < F_{1-\alpha}) = 1-\alpha$

Tỷ lệ phương sai của hai tổng thể được cho bởi

$$F_\alpha \le \frac{\hat{S}_1^2/\sigma_1^2}{\hat{S}_2^2/\sigma_2^2} \le F_{1-\alpha} \quad \Leftrightarrow \frac{1}{F_{1-\alpha}}\frac{\hat{S}_1^2}{\hat{S}_2^2} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{1}{F_\alpha}\frac{\hat{S}_1^2}{\hat{S}_2^2} \qquad (12)$$

THE END