ƯỚC LƯỢNG THỐNG KÊ

27 October 2015

GV : Đinh Công Khải – FETP

Môn: Các Phương Pháp Định Lượng

1. Tóm tắt các nội dung đã học

- ***Tổng thể và mẫu***: Làm thế nào để suy luận các tham số của tổng thể dựa trên thông tin chứa trong mẫu?

Lấy mẫu
ngẫu nhiên

Tổng thể
N (Cỡ)
μ (Trung bình)/
p (Tỷ lệ)
σ (Độ lệch chuẩn)

Mẫu
n
$\overline{x}$
$\overline{p}$
s

•Ước lượng
•Kiểm định giả thuyết

1. Tóm tắt các nội dung đã học (tt)

- ***Thống kê mô tả***
- ***Xác xuất và phân phối xác xuất*** ➔ cơ chế để thực hiện thống kê suy luận từ mẫu.
- ***Chọn mẫu*** và ***Định lý giới hạn trung tâm***: "Một mẫu ngẫu nhiên gồm n quan sát được chọn từ *một tổng thể không chuẩn tắc* có trung bình là μ và độ lệch chuẩn là σ, nếu n lớn, thì *phân phối mẫu của trung bình mẫu sẽ có phân phối xấp xỉ chuẩn tắc* với trung bình là μ và độ lệch chuẩn $\sigma/\sqrt{n}$ "

2. Ước lượng các tham số của tổng thể

- Có 2 loại ước lượng:
- *Ước lượng điểm* của một tham số tổng thể là cách thức tính toán một giá trị đơn lẻ của tham số tổng thể dựa trên dữ liệu mẫu.
- *Ước lượng khoảng* của một tham số tổng thể là cách thức tính toán 2 giá trị dựa trên dữ liệu mẫu, từ đó tạo nên một khoảng được kỳ vọng chứa tham số thống kê của tổng thể.

2. Ước lượng các tham số thống kê của tổng thể

- Các yêu cầu cần có của ước lượng
- *Không bị chệch:* Ước lượng của một tham số tổng thể không chệch nếu trung bình của phân phối mẫu bằng với giá trị đúng của tham số đó.
- *Phương sai của phân phối mẫu càng nhỏ càng tốt* (đảm bảo cho các ước lượng gần với giá trị đúng của tham số với một xác xuất cao)
- *Sai số ước lượng* (error of estimation): khoảng cách giữa giá trị ước lượng và giá trị đúng của tham số được ước lượng.
- *Hệ số tin cậy (confidence coefficient)*: Xác suất mà khoảng tin cậy bao quanh tham số được ước lượng.

3. Ước lượng cho mẫu lớn

- **Ước lượng điểm**

- Giả sử chúng ta có một ước lượng không chệch với tham số tổng thể mà phân phối mẫu của nó tuân theo phân phối chuẩn hoặc xấp xỉ chuẩn.
- Với xác xuất là 95%, sai số ước lượng sẽ không vượt quá 1,96 lần sai số chuẩn của ước lượng (*biên sai số – margin of error*).

3. Ước lượng cho mẫu lớn

- **Ước lượng khoảng**

- Ước lượng khoảng được xây dựng để cho khi lấy mẫu lặp lại nhiều lần thì một tỷ lệ lớn các khoảng này sẽ bao quanh tham số tổng thể mà chúng ta đang quan tâm. Tỷ lệ này là *hệ số tin cậy (confidence coefficient).* Khoảng được tạo ra được gọi là *khoảng tin cậy (confidence interval).*

- Một khoảng tin cậy mẫu lớn với hệ số tin cậy *(1-α)*100%* dựa trên một ước lượng không bị chệch có phân phối chuẩn được tính như sau

Ước lượng điểm $\pm z_{\alpha/2}$ * Sai số chuẩn của ước lượng

➔ (giới hạn tin cậy dưới, giới hạn tin cậy trên)

4. Ước lượng cho mẫu lớn về số trung bình tổng thể μ

- **Ước lượng điểm của trung bình tổng thể μ**
  - Ước lượng điểm: $\bar{x}$
  - Biên sai số: $z_{\alpha/2} * \sigma_{\bar{x}} = z_{\alpha/2} * \sigma / \sqrt{n}$
- **Ước lượng khoảng tin cậy (1-α)100% cho mẫu lớn đối với μ**

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Trong đó: * n là cỡ mẫu

* σ là độ lệch chuẩn của tổng thể (nếu chưa biết σ có thể sử dụng một ước lượng xấp xỉ là độ lệch chuẩn của mẫu s nếu cỡ mẫu là lớn (n>= 30)

4. Ước lượng cho mẫu lớn về số trung bình tổng thể μ

- Ví dụ: Một công ty được thuê để ước lượng trung bình lãi suất trái phiếu kỳ hạn 5 năm của các công ty có phát hành trái phiếu đặt tại thị trường A. Một mẫu ngẫu nhiên gồm n=100 công ty được chọn trong thị trường này và lãi suất trái phiếu được thu thập cho từng công ty. Trung bình và độ lệch chuẩn của 100 lãi suất trái phiếu lần lượt là 12%/năm và 0.5.

- *Hãy ước lượng trung bình lãi suất và biên sai số cho các trái phiếu 5 năm của các công ty ở thị trường A?*

- *Tìm khoảng tin cậy 95% cho trung bình lãi suất trái phiếu?*

5. Ước lượng cho mẫu nhỏ về số trung bình tổng thể μ

- Khi *cỡ mẫu nhỏ và σ chưa biết* chúng ta có thể sử dụng phân phối xác xuất Student t.
- **Ước lượng điểm cho mẫu nhỏ**
  - Ước lượng điểm: $\bar{X}$
  - Biên sai số: $t_{\alpha/2} * s / \sqrt{n}$
- **Ước lượng khoảng tin cậy (1-α)100% cho mẫu nhỏ đối với μ**

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Trong đó s là độ lệch chuẩn của mẫu và $s/\sqrt{n}$ sai số chuẩn của trung bình mẫu

5. Ước lượng cho mẫu nhỏ về số trung bình tổng thể μ

- Ví dụ: Các biến phí chủ yếu là lao động khiến cho chi phí xây nhà thay đổi từ đơn vị nhà ở này sang đơn vị nhà ở khác. Một công ty xây dựng nhà tiêu chuẩn cần làm ra một mức lợi nhuận bình quân vượt quá $8500 mỗi căn nhà nhằm đạt được mục tiêu lợi nhuận hàng năm. Các khoản lợi nhuận tính trên mỗi căn nhà cho 5 căn nhà mà công ty xây dựng gần đây là $8.760, $6.370, $9.620, $8.200, và $10.350.

  Câu hỏi: Tìm khoảng tin cậy 95% cho lợi nhuận trung bình một căn nhà ở mà công ty đã xây dựng?

6. Ước lượng sự khác biệt giữa 2 số trung bình

- Vấn đề: Có 2 tổng thể 1 và 2 với các tham số thống kê lần lượt như sau: μ_1, σ_1^2 và μ_2, σ_2^2 ➔ Ước lượng $(\mu_1 - \mu_2)$?
- Lấy mẫu ngẫu nhiên gồm n_1 đại lượng từ tổng thể 1 và n_2 đại lượng từ tổng thể 2. Hai mẫu này có các trị thống kê lần lượt như sau:

$$\bar{x}_1, \quad s_1^2 \quad và \quad \bar{x}_2, \quad s_2^2$$

- Các đặc trưng phân phối mẫu của $\bar{x}_1 - \bar{x}_2$ như sau
- Nếu các tổng thể không có phân phối chuẩn thì phân phối mẫu của $\bar{x}_1 - \bar{x}_2$ là phân phối xấp xỉ chuẩn khi n_1 và n_2 là lớn (theo Định lý Giới hạn trung tâm)

6. Ước lượng sự khác biệt giữa 2 số trung bình

- Trung bình và độ lệch chuẩn của $\bar{x}_1 - \bar{x}_2$ là

$$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- Nếu các tổng thể có phân phối chuẩn thì phân phối mẫu của $\bar{x}_1 - \bar{x}_2$ cũng sẽ có phân phối chuẩn mà không quan tâm đến cỡ mẫu.

6. Ước lượng sự khác biệt giữa 2 số trung bình

- **Ước lượng điểm của (μ_1- μ_2)**
  - Trị ước lượng $\bar{x}_1 - \bar{x}_2$
  - Biên sai số:

$$1,96\sigma_{\bar{x}_1-\bar{x}_2} = 1.96\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

- **Ước lượng khoảng tin cậy (1-α)100% cho (μ_1- μ_2)**

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

trường hợp σ_1^2 và σ_2^2 chưa biết thì chúng có thể được xấp xỉ bằng s_1^2 và s_2^2 với điều kiện n_1 và $n_2 \geq 30$.

6. Ước lượng sự khác biệt giữa 2 số trung bình

Ví dụ: Một bộ phận cho vay của ngân hàng tìm thấy rằng 57 khoản cho vay mua nhà trong tháng 4 có giá trị trung bình là $78.100 và độ lệch chuẩn là $6.300. Một phân tích về khoản cho vay trong tháng 5 với tổng cộng là 66 khoản, cho thấy giá trị trung bình là $82.700 và độ lệch chuẩn là $7.100. Giả định các khoản cho vay mua nhà đại diện cho các mẫu ngẫu nhiên của những giá trị các hồ sơ xin vay mua nhà được bộ phận dịch vụ cho vay của ngân hàng chấp thuận. Tìm khoảng tin cậy 98% cho sự khác biệt trong mức vay trung bình của các hồ sơ xin vay mua nhà được chấp thuận từ tháng 4 đến tháng 5?

6. Ước lượng sự khác biệt giữa 2 số trung bình

- Trong trường hợp cỡ mẫu nhỏ, hai tổng thể có phân phối chuẩn với các phương sai bằng nhau **($\sigma_1^2 = \sigma_2^2 = \sigma^2$)**

Ước lượng khoảng tin cậy (1-α)100% cho mẫu nhỏ đối với (μ_1- μ_2)

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} * s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

7. Ước lượng một tỷ lệ nhị thức

- Tham số nhị thức của tổng thể: tỷ lệ nhị thức p
- Trị thống kê của mẫu: tỷ lệ mẫu $\hat{p} = \frac{x}{n}$

  trong đó x là số lần thành công trong n lân thử
- Theo CLT, với một mẫu ngẫu nhiên có n quan sát được chọn từ tổng thể nhị thức có tham số p thì phân phối mẫu của tỷ lệ mẫu này như sau
- Trung bình và độ lệch chuẩn của $\hat{p}$

$$E(\hat{p}) = \mu_{\hat{p}} = p$$

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

- Trường hợp n lớn phân phối mẫu của tỷ lệ mẫu sẽ có phân phối xấp xỉ chuẩn tắc. Ước lượng xấp xỉ này là phù hợp nếu $\mu_{\hat{p}} \pm 2\sigma_{\hat{p}}$ từ 0 đến 1; là tốt nếu $\mu_{\hat{p}} \pm 3\sigma_{\hat{p}}$ nằm trong khoảng từ 0 đến 1

7. Ước lượng một tỷ lệ nhị thức

- **Ước lượng điểm cho p**
  - Trị ước lượng: $\hat{p} = \frac{x}{n}$
  - Biên sai số: $1{,}96\sigma_{\hat{p}} = 1{,}96\sqrt{\frac{pq}{n}}$
  - Biên sai số ước lượng: $1{,}96\sigma_{\hat{p}} = 1{,}96\sqrt{\frac{\hat{p}\hat{q}}{n}}$
- Ước lượng khoảng tin cậy (1-α)100% cho p

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

- n phải lớn để phân phối mẫu là phân phối xấp xỉ chuẩn.

7. Ước lượng một tỷ lệ nhị thức

- Ví dụ: Một mẫu ngẫu nhiên gồm n=100 nhà bán buôn mua ống nhựa polyvinyl chỉ ra cho thấy rằng 59 người có kế hoạch gia tăng việc mua hàng của mình trong năm tới. Hãy ước lượng tỷ lệ p của các nhà bán buôn trong tổng thể tất cả các nhà bán buôn ống nhựa polyvinyl mà có kế hoạch gia tăng việc mua hàng của mình trong năm tới và tìm biên sai số. Tìm khoảng tin cậy 95% cho p?

8. Ước lượng sự khác biệt giữa 2 tỷ lệ nhị thức

- Có 2 tổng thể nhị thức 1 và 2 với các tham số thống kê lần lượt như sau: p_1 và p_2 ➔ Ước lượng $(p_1 - p_2)$?
- Lấy mẫu ngẫu nhiên gồm n_1 đại lượng từ tổng thể 1 và n_2 đại lượng từ tổng thể 2. Hai mẫu này có các trị thống kê lần lượt như sau:

  $\hat{p}_1$ và $\hat{p}_2$

- Các đặc trưng phân phối mẫu của $\hat{p}_1 - \hat{p}_2$ như sau
- Phân phối mẫu của $\hat{p}_1 - \hat{p}_2$ là phân phối xấp xỉ chuẩn khi n_1 và n_2 là lớn (theo Định lý Giới hạn trung tâm)

8. Ước lượng sự khác biệt giữa 2 tỷ lệ nhị thức

- Trung bình và độ lệch chuẩn của $\hat{p}_1 - \hat{p}_2$ là

$$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$$

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

- Khi sử dụng phân phối chuẩn để ước lượng xấp xỉ các xác suất của nhị thức thì khoảng $(\hat{p}_1 - \hat{p}_2) \pm 2\sigma_{(\hat{p}_1 - \hat{p}_2)}$ phải chứa $\hat{p}_1 - \hat{p}_2$ (khoảng này thay đổi từ -1 đến 1)

8. Ước lượng sự khác biệt giữa 2 tỷ lệ nhị thức

- **Ước lượng điểm của (p_1- p_2)**
  - Trị ước lượng $(\hat{p}_1 - \hat{p}_2)$
  - Biên sai số: $$1{,}96\sigma_{\hat{p}_1-\hat{p}_2} = 1.96\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}$$

- **Ước lượng khoảng tin cậy (1-α)100% cho (p_1- p_2)**

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}$$

n phải đủ lớn để phân phối mẫu của $(\hat{p}_1 - \hat{p}_2)$ có ước lượng xấp xỉ phân phân chuẩn. Khoảng $(\hat{p}_1 - \hat{p}_2) \pm 2\sigma_{(\hat{p}_1-\hat{p}_2)}$ được chứa trong khoảng [-1;1]

8. Ước lượng sự khác biệt giữa 2 tỷ lệ nhị thức

- Ví dụ: Một cuộc điều tra ngân hàng về các khoản chi trả thẻ tín dụng trễ hạn đã tìm thấy tỷ lệ trễ hạn trong 1 tháng đối với 414 chủ doanh nghiệp nhỏ là 5,8% so với 3,6% của 1029 nhà quản lý chuyên nghiệp (professionals). Giả định rằng dữ liệu cho 2 đối tượng sử dụng thẻ này có thể được xem như các mẫu ngẫu nhiên độc lập của những tài khoản hàng tháng đã sử dụng trong khoảng thời gian tương đối dài (1 đến 2 năm). Tìm khoảng tin cậy 95% trong các tỷ lệ về những sự trễ hạn cho 2 loại đối tượng sử dụng thẻ tín dụng này?

9. Chọn cỡ mẫu

- **Quy trình chọn lựa cỡ mẫu**
- Xác định tham số được ước lượng và độ lệch chuẩn của ước lượng điểm
- Chọn B (giới hạn biên sai số) và hệ số tin cậy ($1-\alpha$)
- Giải phương trình

  $z_{\alpha/2}$ * độ lệch chuẩn của số ước lượng = B
- Nếu n nhỏ hơn 30 thì chúng cần dùng $t_{\alpha/2}$ để thay thế $z_{\alpha/2}$ và sử dụng s thay thế cho σ. Quy trình này được lặp đi lặp lại cho đến khi cỡ mẫu không đổi.