

Chương Trình Giảng Dạy Kinh tế Fulbright

Học kỳ Thu năm 2011

Các Phương Pháp Phân Tích Định Lượng

Gợi ý giải bài tập 4

LẤY MẪU VÀ ƯỚC LƯỢNG

Ngày Phát: Thứ Hai, 24/10/2011

Ngày Nộp: 8:20 sáng, Thứ Hai, 31/11/2011

Bản in nộp tại Hộp nộp bài tập trong phòng Lab

Bản điện tử gửi lên <http://intranet.fetp.edu.vn:81>

Bài 1: (20 điểm)

Một mẫu có cỡ $n = 50$ được chọn ngẫu nhiên từ một tổng thể. Giả sử tổng thể có trị trung bình $\mu = 200$ và độ lệch chuẩn $\sigma = 40$.

- a. Hãy cho biết giá trị kỳ vọng và phương sai của trung bình mẫu \bar{x}

Giá trị kỳ vọng của trung bình mẫu $E(\bar{X}) = \mu = 200$

Phương sai của trung bình mẫu $\sigma_{\bar{X}}^2$:

$$\sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n} = \frac{40^2}{50} = 32$$

- b. Bây giờ, giả sử cỡ mẫu là 60. Hãy tính lại các yêu cầu trong câu a và cho nhận xét về sự giống nhau và khác nhau trong các kết quả của hai cỡ mẫu khác nhau. Giải thích.

Bây giờ $n = 60$

Giá trị kỳ vọng của trung bình mẫu $E(\bar{X}) = \mu = 200$

Phương sai của trung bình mẫu $\sigma_{\bar{X}}^2$:

$$\sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n} = \frac{40^2}{60} \approx 26.667$$

Nhận xét:

Đối với cỡ mẫu 50 và 60, giá trị kỳ vọng của trung bình mẫu là như nhau và bằng giá trị trung bình của tổng thể.

Với cỡ mẫu 60, phương sai của trung bình mẫu sẽ nhỏ hơn phương sai của trung bình mẫu đối với cỡ mẫu 50. Khi kích thước của mẫu ngẫu nhiên càng lớn, các trung bình mẫu có tính ổn định hơn nhờ vào sự triệt tiêu lẫn nhau của các quan sát riêng biệt. Nói cách khác, khi kích thước của mẫu ngẫu nhiên nhỏ, các giá trị rất lớn hoặc rất nhỏ trong mẫu sẽ ảnh hưởng nhiều hơn đến giá trị trung bình mẫu.

hường mạnh lên các trung bình mẫu. Vì vậy mà mẫu ngẫu nhiên có kích thước mẫu càng lớn thì độ lệch chuẩn của phân phối mẫu càng nhỏ.

- c. Nếu Anh/Chị mong muốn độ lệch chuẩn của trung bình mẫu tối đa là 10. Hãy tính cỡ mẫu tối thiểu mà Anh/Chị cần phải lấy.

$$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n} \Rightarrow \text{cỡ mẫu tối thiểu } n = \frac{\sigma_x^2}{\sigma_{\bar{x}}^2} = \frac{1600}{100} = 16$$

Bài 2: (20 điểm)

Vòng đời của một nhãn hiệu bóng đèn Bình Minh có giá trị trung bình 1600 giờ và độ lệch chuẩn là 500 giờ. Cho phân phối xác suất của tổng thể là phân phối chuẩn. Giả sử bạn mua 25 bóng đèn của nhãn hiệu này và xem đây như là một mẫu ngẫu nhiên sản phẩm của nhà sản xuất này.

- a. Hãy tính giá trị trung bình

$$\text{Giá trị kỳ vọng của trung bình mẫu } E(\bar{X}) = \mu = 1600 \text{ (giờ)}$$

- b. Hãy tính độ lệch chuẩn của trung bình mẫu.

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{500}{5} = 100$$

$$\Rightarrow \sigma_{\bar{x}} = 100 \text{ (giờ)}$$

- c. Hãy tính xác suất, trung bình mà nói, 25 bóng đèn này có vòng đời thấp hơn 1200 giờ.

$$P(\bar{x} < 1200) = P(z < (1200-1600)/100) = P(z < -4) = 0.000032$$

Bài 3: (20 điểm)

Để khuyến khích người dân mua hàng ở khu vực trung tâm thành phố, lãnh đạo thành phố dự định xây dựng một khu giữ xe công cộng nằm gần khu trung tâm mua sắm của thành phố. Thành phố dự định sẽ sử dụng phí giữ xe để bù đắp chi phí xây dựng. Khảo sát trong vòng 36 ngày người ta thấy số tiền giữ xe trung bình thu được trong một ngày là \$600 với độ lệch chuẩn là \$42.

- a. Những giả thiết nào cần được đảm bảo để có thể sử dụng những số liệu thống kê cho việc suy diễn?

Các điều kiện cần phải được đảm bảo là lấy mẫu độc lập và các quan sát được chọn ngẫu nhiên.

- b. Hãy ước lượng khoảng tin cậy 95% cho doanh thu trung bình mà khu giữ xe này có thể tạo ra trong một tháng.

$$\text{Với độ tin cậy 95\%, } Z_{\alpha/2} = Z_{0.025} = 1.96$$

$$\text{Ước lượng điểm: } \bar{x} = 600 \text{ (\$)}$$

$$\text{Sai số ước lượng: } Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}} = 1.96 * \frac{42}{\sqrt{36}} = 13.72 (\$)$$

Khoảng tin cậy 95% cho doanh thu trung bình là $600 \pm 13.72 (\$)$, hay [586.28; 613.72] (\$).

Dựa trên những thông tin từ mẫu, với độ tin cậy 95%, chúng ta đoán chắc rằng thu nhập trung bình một ngày của khu giữ xe công cộng sẽ nằm trong khoảng [586.28; 613.72], hay trong một tháng là [\$17588.4, \$18411.6]

- c. Những nhà tư vấn cho dự án này dự đoán doanh thu giữ xe trung bình là vào khoảng \$630 một ngày. Dựa trên khoảng tin cậy bạn hãy cho biết các nhà tư vấn có đúng không? Tại sao?

Ta thấy $630 \notin [586.28; 613.72]$, tức là giá trị dự báo của các nhà tư vấn nằm ngoài khoảng tin cậy 95% mà chúng ta đã tính toán dựa trên thông tin của mẫu, do đó các nhà tư vấn đã dự báo chưa chính xác về thu nhập trung bình ngày của khu giữ xe.

- d. Chúng ta cần thu thập thông tin trong bao nhiêu ngày để có được khoảng tin cậy 95% trong việc ước lượng giá trị trung bình đúng trong vòng \$10?

$$z_{\alpha/2} * \frac{\sigma}{\sqrt{n}} = B$$

$$n = \frac{z_{\alpha/2}^2 * \sigma^2}{B^2} = \frac{1.96^2 * 42^2}{10^2} = 67.77$$

→ Nên thu thập thông tin trong 68 ngày.

Bài 4: (20 điểm)

Vào năm 201X, Bộ Giáo dục công bố một kết quả điều tra trên diện rộng về việc tốt nghiệp phổ thông trung học. Các nhà nghiên cứu đã liên hệ với hơn 25,000 người ở độ tuổi 24 để xem họ đã tốt nghiệp phổ thông trung học chưa. 85% trong 12,460 đàn ông và 88% trong 12,700 phụ nữ được hỏi trả lời rằng họ đã tốt nghiệp.

- a. Hãy xác định khoảng tin cậy 95% cho sự khác biệt về tỷ lệ tốt nghiệp giữa nam giới và nữ giới.

Khoảng tin cậy 95% = 0.95 = 1 - α \Rightarrow $\alpha = 0.05 \Rightarrow z_{0.025} = 1.96$.

Sai số ước lượng:

$$z_{0.025} * \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = 1.96 * \sqrt{\frac{0.88 * 0.12}{12700} + \frac{0.85 * 0.15}{12460}} \approx 0.00844$$

Khoảng tin cậy 95% cho sự khác biệt về tỉ lệ tốt nghiệp giữa nam giới và nữ giới là:

$$\hat{p}_1 - \hat{p}_2 \pm 0.00844 = (0.88-0.85) \pm 0.00844,$$

hay [0.021559; 0.038441] = [2.16%; 3.84%].

- b. Giải thích kết quả tính toán của bạn.

Sự khác biệt về tỷ lệ tốt nghiệp giữa phụ nữ và đàn ông nằm trong khoảng 2.16% đến 3.84% với độ tin cậy là 95%. [hay ở độ tin cậy 95%, tỷ lệ phụ nữ tốt nghiệp cao hơn so với tỷ lệ nam tốt nghiệp từ 2.16 (điểm) % đến 3.84 (điểm) %]

- c. Kết quả này có phải là bằng chứng xác thực giúp chúng ta kết luận rằng nữ giới có khả năng tốt nghiệp phổ thông trung học nhiều hơn nam giới không?

Kết quả tính toán về khoảng tin cậy 95% của sự khác biệt tỷ lệ tốt nghiệp giữa nữ và nam cho phép chúng ta kết luận rằng nữ giới có khả năng tốt nghiệp phổ thông trung học nhiều hơn nam giới vì khoảng tin cậy này không chứa giá trị 0.

Bài 5: (20 điểm)

Bạn hãy sử dụng phần mềm Stata, mở file dữ liệu muc7.dta trong bộ dữ liệu VHLSS2008. Giả sử rằng mẫu điều tra được chọn theo phương pháp ngẫu nhiên.

Tùy những giả định, mục tiêu, giới hạn ... của bạn mà có thể có những kết quả tính toán khác. Dưới đây là một số gợi ý.

- a. Bạn hãy ước lượng khoảng tin cậy 95% cho trung bình diện tích nhà ở của hộ dân. (chú ý cảnh loại ra những giá trị bị lỗi)

```
. sum m7c2 if m7c1>=1 & m7c2>0
```

Variable	Obs	Mean	Std. Dev.	Min
m7c2	9180	68.05664	40.2955	6

```
. ci m7c2 if m7c1>=1 & m7c2>0
```

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]
m7c2	9180	68.05664	.4205671	67.23224 68.88105

- b. Bạn hãy ước lượng khoảng tin cậy 95% cho tỷ lệ hộ dân có nhà ở thuộc kiểu kiên cố khép kín

```
. tab m7c3 if m7c1>=1
```

3. Lo ⁱ i nhự @ang ẽ	Freq.	Percent	Cum.
Biết thù	33	0.36	0.36
Kiên cố khép kín	1,113	12.12	12.48
Kiên cố không khép kín	1,221	13.30	25.78
Không kiên cố	5,497	59.87	85.66
Như thiếu & khác	1,317	14.34	100.00

Total | 9,181 100.00

. **cii 9181 1113**

```
Binomial Exact --
Variable | Obs Mean Std. Err. [95%
Conf. Interval]
-----+-----
.1280799 | 9181 .1212286 .0034064 .1146187
```

c. Bạn hãy ước lượng khoảng tin cậy 95% cho tỷ lệ hộ dân có máy vi tính

Có thể bạn tạo ra biến nhị phân, sau đó áp dụng lệnh **proportion**

. **gen comaytinh=(m7c41==1)**

. **tab comaytinh**

```
comaytinh | Freq. Percent Cum.
-----+-----
0 | 8,282 90.13 90.13
1 | 907 9.87 100.00
-----+-----
Total | 9,189 100.0
```

. **proportion comaytinh**

Proportion estimation Number of obs = 9189

```
-----+-----
| Proportion Std. Err. [95% Conf. Interval]
-----+-----
comaytinh |
0 | .901295 .0031117 .8951955 .9073946
1 | .098705 .0031117 .0926054 .1048045
-----+-----
```

Hoặc

. **cii 9189 907**

```
Binomial Exact --
Variable | Obs Mean Std. Err. [95%
Conf. Interval]
```

```
-----+-----
-----
.1049856 |          9189      .098705      .0031115      .0926801
```

d. Nếu tính đến yếu tố trọng số trong điều tra mức sống hộ gia đình, bạn hãy ước lượng lại khoảng tin cậy 95% cho giá trị trung bình, cũng như các tỷ lệ trên.

Sau khi nối hai file. Bạn cần đưa biến trọng số vào câu lệnh.

```
. sum m7c2 if m7c1>=1 & m7c2>0 [aw=wt9]
```

```

      Variable |          Obs          Weight          Mean      Std. Dev.
      Min      Max
-----+-----
      m7c2 |          9180      20942772      68.14664      41.62054
6      600
```

```
. ci m7c2 if m7c1>=1 & m7c2>0 [aw=wt9]
```

```

      Variable |          Obs          Mean      Std. Err.      [95%
      Conf. Interval]
-----+-----
      m7c2 |          9180      68.14664      .4343965      67.29513
68.99816
```

```
gen nhakiencokhepkin=( m7c3==2)
```

```
. tab nhakiencokhepkin
```

```

nhakiencokh |
      epkin |          Freq.          Percent          Cum.
-----+-----
          0 |          8,076          87.89          87.89
          1 |          1,113          12.11          100.00
-----+-----
      Total |          9,189          100.00
```

```
. proportion nhakiencokhepkin [pw=wt9]
```

```
Proportion estimation          Number of obs      =      9189
```

```
-----+-----
          | Proportion      Std. Err.      [95% Conf. Interval]
-----+-----
nhakiencok~n |
```

```
0 | .8576812 .0043321 .8491893 .8661732
1 | .1423188 .0043321 .1338268 .1508107
```

. proportion comaytinh [pw=wt9]

Proportion estimation Number of obs = 9189

```
-----
      | Proportion  Std. Err.  [95% Conf. Interval]
-----+-----
comaytinh |
      0 | .8849155  .0039906   .877093   .892738
      1 | .1150845  .0039906   .107262   .122907
```
