

Chương Trình Giảng Dạy Kinh tế Fulbright

Học kỳ Thu năm 2015

Các Phương Pháp Phân Tích Định Lượng

GỢI Ý ĐÁP ÁN -- BÀI TẬP 4

PHÂN PHỐI MẪU VÀ ƯỚC LƯỢNG THỐNG KÊ

Ngày Phát: Thứ ba 27/10/2015

Ngày Nộp: Thứ ba 03/11/2015

Bản in nộp trước 8h20 tại Hộp nộp bài tập trong phòng Lab.

Bản điện tử nộp trước 8h20 tại địa chỉ: <http://www.fetp.edu.vn/vn/tai-nguyen/hoc-vien-hien-tai/>

Câu 1: (20 điểm)

Để triển khai chương trình tín dụng vi mô cho các hộ nghèo tại một trong những xã nghèo nhất tỉnh Lai Châu với khoảng 500 hộ nghèo đang sinh sống, 40 trong tổng số 500 hộ nghèo đã được lấy mẫu ngẫu nhiên và khảo sát về mong muốn tham gia chương trình tín dụng vi mô. Kết quả cho thấy có 25/40 hộ mong muốn tham gia chương trình tín dụng vi mô và 12/40 được đánh giá là có khả năng trả nợ. Dựa vào thông tin trên, anh/chị hãy tính:

- a. Số hộ mong muốn tham gia chương trình trên thực tế khi chương trình được triển khai với độ tin cậy 95%?

$N = 500$ (hộ)

$n = 40$ (hộ).

$n/N = 8\% > 5\% \Rightarrow$ Tổng thể là hữu hạn. Do vậy lưu ý phải nhân hệ số điều chỉnh khi tính độ lệch chuẩn của phân phối mẫu của tỷ lệ mẫu.

$$\hat{p} = \frac{25}{40} = 0,625 = 62,5\%$$

$$\hat{q} = 1 - \hat{p} = 1 - 62,5\% = 37,5\%$$

$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p} * \hat{q}}{n}} * \sqrt{\frac{N - n}{N - 1}} = \sqrt{\frac{62,5\% * 37,5\%}{40}} * \sqrt{\frac{500 - 40}{500 - 1}} = 7,35\%$$

Với $\alpha = 5\% \Rightarrow Z_{\alpha/2} = 1,96$

Với độ tin cậy 95%, tỷ lệ hộ tham gia chương trình trên thực tế sẽ là:

$$p \in \hat{p} \pm Z_{\frac{\alpha}{2}} * \sigma_{\hat{p}}$$

Vậy $p \in (62,5\% \pm 1,96 * 7,35\%) \Leftrightarrow p \in [48,1\% ; 76,9\%]$

Vậy số hộ tham gia chương trình trên thực tế là $[48,1\% * 500 ; 76,9\% * 500] \approx [240 ; 385]$ (người)

- b. Số hộ có khả năng trả nợ nếu chương trình cho 250 hộ vay vốn và kết quả đánh giá của cán bộ tín dụng có độ tin cậy 90%?

$$N = 250 \text{ (hộ)}$$

$$n = 40 \text{ (hộ)}.$$

$n/N = 8\% > 5\% \Rightarrow$ Tổng thể là hữu hạn. Do vậy lưu ý phải nhân hệ số điều chỉnh khi tính độ lệch chuẩn của phân phối mẫu của tỷ lệ mẫu.

$$\hat{p} = \frac{12}{40} = 0,3 = 30\%$$

$$\hat{q} = 1 - \hat{p} = 1 - 30\% = 70\%$$

$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p} * \hat{q}}{n}} * \sqrt{\frac{N - n}{N - 1}} = \sqrt{\frac{30\% * 70\%}{40}} * \sqrt{\frac{250 - 40}{250 - 1}} = 6,65\%$$

$$\text{Với } \alpha = 10\% \Rightarrow Z_{\alpha/2} = 1,645$$

Với độ tin cậy 90%, tỷ lệ hộ tham gia chương trình trên thực tế sẽ là:

$$p \in \hat{p} \pm Z_{\frac{\alpha}{2}} * \sigma_{\hat{p}}$$

$$\text{Vậy } p \in (30\% \pm 1,645 * 6,65\%) \Leftrightarrow p \in [19,1\% ; 40,9\%]$$

Vậy số hộ tham gia chương trình trên thực tế là $[19,1\% * 250 ; 40,9\% * 250] \approx [47 ; 102]$ (người)

Câu 2: (30 điểm)

Một dự án bệnh viện tại nông thôn đang chuẩn bị được đầu tư. Để đánh giá nhu cầu khám chữa bệnh của người dân trong địa bàn (khoảng 10000 dân). Kết quả khảo sát về nhu cầu khám bệnh sử dụng BHYT thông qua việc ghi nhận khả năng khám BHYT thay vì khám dịch vụ của hầu hết người dân trong địa bàn cho thấy có 17% người dân chọn hình thức khám BHYT.

Bệnh viện dự báo sẽ có khoảng 100 lượt khách/ngày đến khám tại bệnh viện. Đồng thời, năng lực khám bệnh BHYT của bệnh viện đang được thiết kế tối đa là 20 lượt khách/ngày.

- a. Anh/chị hãy xác định mô hình phân phối mẫu của tỷ lệ người khám BHYT tại bệnh viện trên khi đi vào hoạt động nếu dự báo và năng lực thiết kế không thay đổi so với thực tế?

$$N = 10.000 \text{ (người)}$$

$$n = 100 \text{ (hộ)}.$$

$n/N = 1\% < 5\% \Rightarrow$ Do vậy lưu ý không phải nhân hệ số điều chỉnh khi tính độ lệch chuẩn của phân phối mẫu của tỷ lệ mẫu. Phân phối mẫu của tỷ lệ mẫu sẽ có phân phối xấp xỉ chuẩn tắc.

$$\hat{p} = 17\%$$

$$\hat{q} = 1 - \hat{p} = 1 - 17\% = 83\%$$

$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p} * \hat{q}}{n}} = \sqrt{\frac{17\% * 83\%}{100}} = 3,76\%$$

$$\text{Vậy } \bar{p} \sim N(17\% ; 3,76\%^2)$$

- b. Với tỷ lệ khảo sát 17% người dân chọn hình thức khám BHYT, bệnh viện dự kiến bố trí nhân sự khám BHYT phù hợp với số lượt khách từ 15 – 20 lượt khám BHYT/ngày. Tính xác suất để số lượt khách khám BHYT nằm trong khoảng 15 – 20 lượt khám BHYT/ngày.

$$P(15 < X < 20) = P(15\% < p < 20\%) = \frac{15\% - 17\%}{3,76\%} < Z < \frac{20\% - 17\%}{3,76\%} = 49,05\%$$

- c. Tính xác suất bệnh viện không đáp ứng nhu cầu khám BHYT? Xác suất để số lượt khám BHYT thấp hơn số lượt khách dự kiến?

- Bệnh viện không đáp ứng nhu cầu khi $X > 20$ lượt khám BHYT hay $p > 20\%$

$$p(X > 20) = P(p > 20\%) = p\left(Z > \frac{20\% - 17\%}{3,76\%}\right) = 21,22\%$$

- d. Tính khoảng biến thiên cho số lượt khám BHYT tại bệnh viện nếu giả định số lượt khách đến khám ổn định ở mức 100 lượt khách/ngày với độ tin cậy 95%?

Với $\alpha = 5\% \Rightarrow Z_{\alpha/2} = 1,96$

$$p \in \hat{p} \pm Z_{\alpha/2} * \sigma_{\hat{p}} \leftrightarrow p \in 17\% \pm 1,96 * 3,76\% \leftrightarrow p \in [9,64\% ; 24,36\%]$$

Vậy số lượt khác đến khám BHYT hằng ngày là khoảng 9 đến 25 người.

- e. Để được cấp giấy phép đầu tư và tài trợ vốn từ các nhà tài trợ, dự án phải chứng minh có thể đảm bảo 90% nhu cầu khám BHYT của người dân trong tất cả các ngày. Theo anh/chị, năng lực khám bệnh BHYT của bệnh viện phải được thiết kế bao nhiêu lượt khách/ngày?

$$P(X < X_0) = P(p < p_0) = P(Z < Z_0) = 90\%$$

$$Z_0 = \text{Normsinv}(90\%) = 1,28$$

$$Z_0 = \frac{p_0 - 17\%}{3,76\%} \Rightarrow p_0 = 1,28 * 3,75\% + 17\% = 21,82\%$$

Vậy năng lực khám BHYT của BV phải được thiết kế cho khoảng 22 người/ngày nếu muốn đảm bảo 90% nhu cầu khám mỗi ngày.

Câu 3: (20 điểm)

Một hãng sản xuất sữa bột trẻ em công bố là cho hàm lượng protein trong 100g sữa bột là 13g và độ lệch chuẩn 0,3g. Trước khi đưa sản phẩm ra thị trường, công ty tiến hành tự đánh giá sự ổn định của dây chuyền sản xuất bằng cách cho sản xuất thử nghiệm 300 hộp. Sau đó 36 trong tổng số 300 hộp đã được lấy mẫu ngẫu nhiên.

- a. Tính xác suất để trung bình mẫu thu thập có hàm lượng protein trong 100g sữa bột nhỏ hơn 12g? Nếu số sản phẩm có trọng lượng dưới 12g từ 1 hộp trở xuống thì kết quả kiểm tra đạt yêu cầu và có thể sản xuất đồng loạt để tung ra thị trường. Kết luận của anh/chị trong trường hợp này?

$$N = 300 \text{ (hộp)}$$

$$n = 36 \text{ (hộp)}$$

$n/N > 5\% \Rightarrow$ có nhân hệ số điều chỉnh khi tính độ lệch chuẩn của trung bình mẫu.

$$\mu = \bar{x} = 13 \text{ (g)}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} * \sqrt{\frac{N-n}{N-1}} = \frac{0,3}{\sqrt{100}} * \sqrt{\frac{300-36}{300-1}} = 0,047 \text{ (g)}$$

$$P(\bar{x} < 12) = P\left(Z < \frac{12 - 13}{0,047}\right) = P(Z < -21,8) \approx 0$$

Lưu ý: sự khác biệt giữa phân phối của trung bình mẫu và phân phối của mẫu hay phân phối của tổng thể. Việc nhà sản xuất đưa ra ngưỡng 1 hộp có hàm lượng protein dưới 12g là do với 99% độ tin cậy hay áp dụng quy tắc 3 sigma thì hộp sữa có hàm lượng protein thấp nhất cũng phải là: $13 - 0,3*3 = 12,1$ (g). Như vậy việc 1 sản phẩm có hàm lượng protein dưới 12 g là rất khó xảy ra.

- b. Giả định sau kiểm tra nội bộ, sản phẩm đạt yêu cầu và được tung ra thị trường đồng loạt (số lượng rất lớn). Một đoàn kiểm tra chất lượng thị trường cũng tiến hành kiểm tra sản phẩm trên với 36 mẫu được thu thập ngẫu nhiên từ 36 cửa hàng trên địa bàn. Anh/chị hãy tính xác suất để trung bình mẫu thu thập có hàm lượng protein trong 100g sữa bột nhỏ hơn 12g? Nhận xét của anh/chị so với kết quả Câu a?

$$N = \text{rất lớn (hộp)}$$

$$n = 36 \text{ (hộp)}$$

$n/N < 5\% \Rightarrow$ không nhân hệ số điều chỉnh khi tính độ lệch chuẩn của trung bình mẫu.

$$\mu = \bar{x} = 13 \text{ (g)}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0,3}{\sqrt{100}} = 0,05 \text{ (g)}$$

$$P(\bar{x} < 12) = P\left(Z < \frac{12 - 13}{0,05}\right) = P(Z < -20) \approx 0$$

Trên thực tế, có thể suy luận nhanh xác suất ở câu a và câu b sẽ gần như bằng 0 vì xác suất để một hộp sữa bất kỳ có hàm lượng protein nhỏ hơn 12g là rất nhỏ vì với 99% độ tin cậy hay áp dụng quy tắc 3 sigma thì hộp sữa có hàm lượng protein thấp nhất cũng phải là:

$$13 - 0,3*3 = 12,1 \text{ (g)}$$

Do vậy, dù có lấy mẫu như thế nào thì khả năng trung bình của mẫu đó vượt qua ngưỡng 12,1g là rất thấp.

So với câu a, mặc dù cả 2 trường hợp đều cho xấp xỉ bằng 0, tuy nhiên, nếu nhìn vào 2 trường hợp ta sẽ thấy, khi cỡ mẫu là rất nhỏ so với tổng thể thì sai số ước lượng của trung bình mẫu sẽ tăng. Hay nói cách khác, để giảm sai số ước lượng ($\sigma_{\bar{x}}$) thì phải lấy mẫu đủ lớn. Vấn đề là đánh đổi chi phí và chất lượng thông tin thu thập.

- c. Tính toán khoảng ước lượng của trung bình mẫu trong trường hợp lấy mẫu như trường hợp ở câu b với 3 mức độ tin cậy khác nhau lần lượt là 90%; 95% và 99%. Nhận xét của anh/chị về xu hướng của khoảng ước lượng trong 3 trường hợp trên là gì? Giải thích?

$$\mu = \bar{x} = 13 \text{ (g)}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0,3}{\sqrt{100}} = 0,05 \text{ (g)}$$

Với $\alpha = 10\% \Rightarrow Z_{\alpha/2} = 1,64 \Rightarrow$ Khoảng tin cậy: $[13\% \pm 1,64 \cdot 0,05] = [12,92; 13,08\text{g}]$

Với $\alpha = 5\% \Rightarrow Z_{\alpha/2} = 1,96 \Rightarrow$ Khoảng tin cậy: $[13\% \pm 1,96 \cdot 0,05] = [12,90 ; 13,10\text{g}]$

Với $\alpha = 1\% \Rightarrow Z_{\alpha/2} = 2,57 \Rightarrow$ Khoảng tin cậy: $[13\% \pm 2,57 \cdot 0,05] = [12,87 ; 13,13\text{g}]$

Khi tăng độ tin cậy thì sai số ước lượng tăng, khoảng ước lượng của trung bình mẫu mở rộng và khả năng chứa thông tin về tổng thể dễ hơn nhưng chất lượng thông tin cung cấp sẽ kém hơn. Do vậy, lại thêm một quy luật đánh đổi là tăng độ tin cậy của ước lượng và giảm chất lượng thông tin cung cấp.

Câu 4: (30 điểm)

Để so sánh chất lượng đào tạo chuyên ngành kinh tế của 2 trường đại học A và B. Một nghiên cứu nhỏ đã được thực hiện với thiết kế như sau. Chọn những sinh viên đến từ trường Đại học A với điều kiện sinh viên này có điểm thi đầu vào Đại học từ 20 – 24 điểm và tốt nghiệp Đại học loại Khá. Cách chọn nhóm sinh viên của trường B cũng tương tự. Kết quả chỉ có 10 sinh viên trường Đại học A và 12 sinh viên trường Đại học B thỏa mãn các điều kiện trên và đồng ý tham gia nghiên cứu. Giả định cả 22 sinh viên ở trên đều có thể xin được việc làm đúng chuyên ngành ngay sau khi ra trường và kết quả thu nhập bình quân/tháng trong năm đầu tiên đi làm được tổng hợp như bảng sau:

Đại học A			Đại học B	
STT	Lương TB tháng		STT	Lương TB tháng
1	17		1	15
2	15		2	12
3	10		3	8
4	15		4	12
5	18		5	14
6	18		6	14
7	13		7	15
8	20		8	9
9	10		9	11
10	15		10	13
			11	12
			12	13

Dựa vào các thông tin ở trên anh/chị hãy thực hiện các yêu cầu sau:

- a. Anh/chị hãy tính các đặc trưng thống kê cần thiết để kết luận về sự biến thiên của mức lương trong 2 nhóm trên?

- Các trị thống kê cần thiết là các trị thống kê được tô đỏ trong bảng kết quả sau đây:

Đại học A		Đại học B	
Mean	15.10	Mean	12.33
Standard Error	1.058825	Standard Error	0.631656
Median	15	Median	12.5
Mode	15	Mode	12
Standard Deviation	3.3483	Standard Deviation	2.188122
Sample Variance	11.21111	Sample Variance	4.787879
Kurtosis	-0.68672	Kurtosis	0.065038
Skewness	-0.37473	Skewness	-0.7613
Range	10	Range	7
Minimum	10	Minimum	8
Maximum	20	Maximum	15
Sum	151	Sum	148
Count	10	Count	12
CV-A	0.22	CV-B	0.18

Kết quả trên cho thấy hệ số biến thiên lương của những người tốt nghiệp Đại học A có sự biến thiên lớn so với trường B do CV-A lớn hơn CV-B

- b. Ước lượng khoảng tin cậy 95% cho mức lương trung bình của các sinh viên chuyên ngành kinh tế thỏa mãn các yêu cầu về điểm thi ĐH đầu vào và tốt nghiệp loại khá ở mỗi trường?

Cỡ mẫu nhỏ $\Rightarrow$ dùng t thay vì z		$t_{\alpha/2} = t_{2.5\%} = 2.26$	
Đại học A		Đại học B	
$\bar{X}_A =$	15.10	$\bar{X}_B =$	12.33
$Min_{\mu} = \bar{x} - t_{\alpha/2} * \frac{s}{\sqrt{n}} =$	12.70	$Min_{\mu} = \bar{x} - t_{\alpha/2} * \frac{s}{\sqrt{n}} =$	10.90
$Min_{\mu} = \bar{x} + t_{\alpha/2} * \frac{s}{\sqrt{n}} =$	17.50	$Min_{\mu} = \bar{x} + t_{\alpha/2} * \frac{s}{\sqrt{n}} =$	13.76
Vậy khoảng ước lượng 95% của Đại học A là: [12.7 - 17.5] (triệu đồng/tháng)		Vậy khoảng ước lượng 95% của Đại học B là: [12.90 - 13.76] (triệu đồng/tháng)	

- c. Ước lượng khoảng tin cậy cho sự khác biệt về trung bình tiền lương của sinh viên giữa 2 trường với mức ý nghĩa $\alpha = 10\%$?

Gọi d là sự khác biệt về trung bình tiền lương giữa 2 trường ($d = A - B$) = 15,10 – 12,33 = 2,77

Với $\alpha = 10\% \Rightarrow t_{\frac{\alpha}{2}, df=20} = 1,725$

Giả định phương sai của 2 tổng thể là bằng nhau:

$$s^2 = \frac{(n_A - 1) * s_A^2 + (n_B - 1) * s_B^2}{n_A + n_B - 2} = \frac{(10 - 1) * 3,34^2 + (10 - 1) * 2,18^2}{10 + 10 - 2} = 7,67$$

$$\Rightarrow s = 2,77 \text{ (ĐVT)}$$

Khoảng ước lượng sự khác biệt trung bình tiền lương là:

$$d \in \left[d \pm t_{\frac{\alpha}{2}} * s \right] = [0,72 ; 4,83] \quad \left(\frac{\text{triệu đồng}}{\text{tháng}} \right)$$

d. Những kết luận của anh/chị rút ra từ các kết quả câu a, b, c ở trên?

- Lương trung bình sinh viên tốt nghiệp trường A cao hơn trường B nhưng sự biến thiên mức lương của sinh viên tốt nghiệp trường A lại lớn hơn. Do vậy, còn tùy vào năng lực của từng sinh viên và tuy môi trường làm việc.
- Khoảng ước lượng sự khác biệt tiền lương trung bình giữa sinh viên tốt nghiệp ở 2 trường thấp cũng là $0,73 > 0$. Do vậy, với độ tin cậy 90% có thể khẳng định lương trung bình của sinh viên trường A là cao hơn so với trường B.

---HẾT---