

Chương Trình Giảng Dạy Kinh tế Fulbright

Học kỳ Thu năm 2012

Các Phương Pháp Phân Tích Định Lượng

LỜI GIẢI ĐỀ NGHỊ BÀI TẬP 4

LẤY MẪU VÀ ƯỚC LƯỢNG

Ngày Phát: Thứ hai 22/10/2012

Ngày Nộp: Thứ hai 29/10/2012

Bản in nộp lúc **8h20 sáng**, tại Hộp nộp bài tập trong phòng Lab

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

Bài 1: (20 điểm)

Bạn hãy sử dụng file dữ liệu **Employee Data.xls** về 474 người lao động. Trong file này có cột Salbegin thể hiện mức lương khởi điểm trong năm tính bằng USD.

- Tính toán các trị thống kê mô tả của biến Salbegin (tạm gọi là tập dữ liệu Salbegin).
- Chọn ngẫu nhiên một mẫu có cỡ $n=50$ từ tập dữ liệu nêu trên, hãy dự đoán giá trị trung bình và độ lệch chuẩn của mẫu này. Nêu rõ các giả định cần thiết để dự đoán của bạn có độ tin cậy cao.
- Hãy chọn ngẫu nhiên một mẫu có cỡ $n=50$ từ tập dữ liệu nói trên, tính toán các giá trị: kỳ vọng và độ lệch chuẩn của mẫu. So sánh với giá trị dự đoán ở câu (b) và giải thích các quan sát thu được.
- Dùng lệnh Random trong Excel để chọn ra 20 mẫu ngẫu nhiên bất kỳ từ tập dữ liệu nói trên với cỡ mẫu tăng dần bắt đầu từ $n=30$. Với mỗi mẫu hãy tính giá trị trung bình và phương sai. Vẽ đồ thị của giá trị trung bình và phương sai theo sự tăng dần của n . Mô tả xu hướng thay đổi của các giá trị này theo n . Bạn rút ra được điều gì từ câu hỏi này?
- Giả sử rằng, số quan sát của biến Salbegin rất lớn (lớn hơn rất nhiều so với 474 quan sát) các giá trị trung bình và phương sai được cho là không đổi so với các giá trị trong câu (a). Với một mong muốn có độ lệch chuẩn của mẫu chọn từ tập dữ liệu này là 100 USD, thì cỡ mẫu cần thiết là bao nhiêu?

Đáp án bài 1: câu này chấm theo mẫu mà học viên tự chọn

Theo bảng của Excel

- Giá trị trung bình sẽ gần bằng trung bình tính được ở câu (a). Phương sai của trung bình mẫu nhỏ hơn Phương sai tính được ở câu (a) là n lần
- Giá trị trung bình sẽ hội tụ nhưng phương sai sẽ có khác biệt lớn so với giá trị dự đoán vì phân phối của tập dữ liệu này không hẳn là phân phối chuẩn.
- Giá trị trung bình sẽ hội tụ về giá trị tính được trong câu (a), độ lệch chuẩn có xu thế giảm dần theo cỡ mẫu (có thể có sự tăng lên nhưng không phải là xu thế).

<i>Salbegin [USD]</i>	
Mean	17016.09[USD]
Standard Error	361.5104[USD]
Median	15000[USD]
Mode	15000[USD]
Standard Deviation	7870.638[USD]
Sample Variance	61946945[USD] ²
Kurtosis	12.39021
Skewness	2.852856
Range	70980[USD]
Minimum	9000[USD]
Maximum	79980[USD]
Sum	8065625[USD]
Count	474

$$S = \frac{\sigma}{\sqrt{n}} = 100[\text{USD}] \Rightarrow n = \left(\frac{\sigma}{S}\right)^2 = \left(\frac{7870.638}{100}\right)^2 \approx 6195 \text{ mẫu}$$

Bài 2: (20 điểm)

Từ một mẫu tổng quát $W = (X_1, X_2)$ chúng ta xem xét hai ước lượng trung bình của tổng thể μ như sau:

$$\bar{X} = \frac{1}{2}X_1 + \frac{1}{2}X_2 \text{ và } \bar{X}' = \frac{1}{3}X_1 + \frac{2}{3}X_2$$

- Chứng tỏ rằng \bar{X} và \bar{X}' là các ước lượng không chệch của μ .
- Ước lượng nào hiệu quả hơn?

Đáp án bài 2

$$a. E(\bar{X}) = E\left(\frac{1}{2}X_1 + \frac{1}{2}X_2\right) = \frac{1}{2}E(X_1) + \frac{1}{2}E(X_2) = \frac{1}{2}\mu + \frac{1}{2}\mu = \mu$$

$$E(\bar{X}') = E\left(\frac{1}{3}X_1 + \frac{2}{3}X_2\right) = \frac{1}{3}E(X_1) + \frac{2}{3}E(X_2) = \frac{1}{3}\mu + \frac{2}{3}\mu = \mu$$

- Để đơn giản chúng ta giả sử rằng X_1, X_2 là độc lập thống kê (tuy nhiên nếu không có giả định này kết quả vẫn không thay đổi).

$$Var(\bar{X}) = Var\left(\frac{1}{2}X_1 + \frac{1}{2}X_2\right) = \frac{1}{4}Var(X_1) + \frac{1}{4}Var(X_2) = \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2 = \frac{1}{2}\sigma^2$$

$$Var(\bar{X}') = Var\left(\frac{1}{3}X_1 + \frac{2}{3}X_2\right) = \frac{1}{9}Var(X_1) + \frac{4}{9}Var(X_2) = \frac{1}{9}\sigma^2 + \frac{4}{9}\sigma^2 = \frac{5}{9}\sigma^2$$

Do $Var(\bar{X}) < Var(\bar{X}')$ nên ước lượng \bar{X} hiệu quả hơn

Bài 3: (20 điểm)

Tuổi thọ của một loại bóng đèn được biết theo quy luật chuẩn với độ lệch chuẩn 100 giờ. Chọn ngẫu nhiên 100 bóng để thử nghiệm, thấy mỗi bóng tuổi thọ trung bình là 1000 giờ.

- Tính tuổi thọ trung bình của bóng đèn xí nghiệp sản xuất với độ tin cậy 95%.
- Hãy ước lượng sai số chuẩn (độ lệch chuẩn) của phân phối mẫu.
- Với độ chính xác là 15 giờ. Hãy xác định độ tin cậy.
- Với độ chính xác là 25 giờ và độ tin cậy là 95% cần thử nghiệm bao nhiêu bóng?

Đáp án bài 3

- a. Áp dụng trường hợp: $n \geq 30, \sigma^2$ đã biết
Khoảng tin cậy 95% = 0.95 = 1 - $\alpha \Rightarrow \alpha = 0.05 \Rightarrow Z_{\alpha/2} = 1.96$.

$$n = 100, \sigma = 100, \bar{X} = 1000$$

$$b_1 = \bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 1000 - 1.96 \frac{100}{\sqrt{100}} = 980,4$$

$$b_2 = \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 1000 + 1.96 \frac{100}{\sqrt{100}} = 1019,6$$

Vậy với độ tin cậy là 95% thì tuổi thọ trung bình của bóng đèn mà xí nghiệp sản xuất ở vào khoảng (980,4 ; 1019,6) giờ.

- b. Phương sai của phân phối mẫu nhỏ hơn phương sai của tổng thể n lần. Do đó chúng ta có:

$$\sigma_{\bar{X}} = \sigma / \sqrt{n} = 100 / 10 = 10$$

c. $\varepsilon = 15, n = 100 \quad Z_{\alpha/2} = \frac{\varepsilon \sqrt{n}}{\sigma} = \frac{15 \sqrt{100}}{100} = 1,5 \Rightarrow \alpha / 2 = 0,0668$

Vậy độ tin cậy $\gamma = 1 - 2\alpha/2 = 0,8664 = 86,64\%$

d. $\varepsilon = 25, \gamma = 95\%, \sigma = 100$

Do $\gamma = 95\%$ nên $Z_{\alpha/2} = 1.96$

$$n = \left(\frac{Z_{\alpha/2} \sigma}{\varepsilon} \right)^2 = \left(\frac{1,96 * 100}{25} \right)^2 = 61,47$$

Do cỡ mẫu là số nguyên nên chúng ta chọn $n = 62$

Bài 4: (20 điểm)

Năm 2012 là năm phát động phong trào “An toàn phòng chống cháy nổ”. Trước đó dựa trên báo cáo khi khảo sát 200 chung cư, khách sạn, nhà nghỉ mới xây dựng trong vòng 15 năm trở lại đây của Sở Xây dựng thì tỷ lệ không đảm bảo an toàn phòng chống cháy nổ là 30%. Để kiểm chứng

số liệu, UBND T.p Hồ Chí Minh đã chỉ đạo Sở Cảnh sát Phòng cháy Chữa cháy (PCCC) tiếp tục khảo sát và báo cáo kết quả. Sở này đã chọn ngẫu nhiên 250 chung cư để khảo sát.

- Hãy tính xác suất để tỷ lệ không đạt yêu cầu an toàn phòng chống cháy nổ trong mẫu điều tra của Sở Cảnh sát PCCC nằm trong khoảng từ 0,25 đến 0,35.
- Kết thúc cuộc điều tra, Sở này báo cáo có 32% trong mẫu khảo sát không đạt yêu cầu phòng chống cháy nổ. Hãy xác định khoảng tin cậy 95% của sự khác biệt về tỉ lệ không đạt yêu cầu an toàn phòng chống cháy nổ của hai Sở nêu trên?

Đáp án bài 4

- Ta có $n = 250$ và $p = 0,30$ gọi P_X là biến ngẫu nhiên chỉ tỉ lệ không đạt yêu cầu phòng chống cháy nổ ta có:

$$P(0,25 < \hat{P}_X < 0,35) = P\left(\frac{0,25 - p}{\sigma_{\hat{p}}} < \frac{\hat{P} - p}{\sigma_{\hat{p}}} < \frac{0,35 - p}{\sigma_{\hat{p}}}\right) = P\left(\frac{0,25 - p}{\sigma_{\hat{p}}} < Z < \frac{0,35 - p}{\sigma_{\hat{p}}}\right)$$

ng đó $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = 0,029$ nên

$$P(0,25 < \hat{P}_X < 0,35) = P\left(\frac{0,25 - 0,30}{0,029} < Z < \frac{0,35 - 0,30}{0,029}\right) = P(-1,72 < Z < 1,72)$$

Tra bảng ta có

$$P(0,25 < \hat{P}_X < 0,35) = F(1,72) - F(-1,72) = 0,9573 - (1 - 0,9573) = 0,9146$$

- Hãy xác định khoảng tin cậy 95% cho sự khác biệt về tỉ lệ không đạt yêu cầu an toàn phòng chống cháy nổ của hai Sở nêu trên .

Khoảng tin cậy 95% = 0.95 = 1 - α $\Rightarrow \alpha = 0.05 \Rightarrow z_{0,025} = 1.96$.

Sai số ước lượng:

$$z_{0,025} * \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = 1.96 * \sqrt{\frac{0,30*0,70}{200} + \frac{0,32*0,68}{250}} \approx 0.085892$$

Khoảng tin cậy 95% cho sự khác biệt về về tỉ lệ không đạt yêu cầu an toàn phòng chống cháy nổ của hai Sở nêu trên là:

$$\hat{p}_2 - \hat{p}_1 \pm 0.085892 = (0.32-0.30) \pm 0.085892,$$

hay $[-0.065892; 0.105892] = [-6,58\%; 10,58\%]$.

Bài 5: (20 điểm)

Bạn hãy sử dụng phần mềm Stata, mở file dữ liệu muc3b.dta trong bộ dữ liệu VHLSS2008. Giả sử rằng mẫu điều tra được chọn theo phương pháp ngẫu nhiên.

- Bạn hãy ước lượng khoảng tin cậy 95% cho trung bình số lần khám chữa bệnh ngoại trú và trung bình chi phí khám chữa bệnh ngoại trú của người dân trong 12 tháng qua (với mốc thời gian là thời điểm trả lời bảng hỏi). (chú ý cần loại ra những giá trị bị lỗi)

- b. Bạn hãy ước lượng khoảng tin cậy 95% cho tỷ lệ người dân đến khám bệnh tại phòng khám đa khoa khu vực.
- c. Bạn hãy ước lượng khoảng tin cậy 95% cho tỷ lệ người dân đến cơ sở y tế để kiểm tra sức khỏe
- 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.

Đáp án bài 5

- a. Ước lượng khoảng tin cậy 95% cho trung bình số lần khám chữa bệnh và chi phí khám ngoại trú

. ci m3c10a

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]	
m3c10a	16959	2.567663	.0300173	2.508826	2.6265

. ci m3c10b

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]	
m3c10b	14803	491.3329	18.51754	455.0362	527.6296

- b. Ước lượng khoảng tin cậy 95% cho người dân đến khám tại phòng khám đa khoa khu vực

. tab m3c8b

M. CSYT	Freq.	Percent	Cum.
YT then b]n	137	0.81	0.81
Tr'm YT X/P	3,810	22.47	23.27
PK *a khoa KV	599	3.53	26.81
BV huyện/quEn	3,410	20.11	46.91
BV t@nh TP	2,627	15.49	62.40
BV TW	765	4.51	66.91
BV nhu n-íc kh,c	236	1.39	68.31
BV t- nh@n	409	2.41	70.72
BV kh,c	43	0.25	70.97
Ph&ng kh,m YT t- nh@n	3,041	17.93	88.90
Lang y	390	2.30	91.20
Dpch v@ y t@ c, th@	1,328	7.83	99.03
CSYT kh,c	164	0.97	100.00
Total	16,959	100.00	

. ci 16959 599

Variable	Obs	Mean	Std. Err.	— Binomial Exact — [95% Conf. Interval]	
	16959	.0353205	.0014174	.0325928	.038209

- c. Ước lượng khoảng tin cậy 95% cho người dân đến cơ sở y tế để kiểm tra sức khỏe.

. tab m3c9

9. Lý do *Ốn CSYT	Freq.	Percent	Cum.
Ti*m ph&ng	406	2.39	2.39
Kh,m thai, KHHG\$, sinh *f	491	2.90	5.29
KT s&oc kho&I	2,760	16.27	21.56
Ch+a b&nh	13,302	78.44	100.00
Total	16,959	100.00	

. ci 16959 2760

Variable	Obs	Mean	Std. Err.	— Binomial Exact — [95% Conf. Interval]	
	16959	.1627454	.0028345	.1572185	.1683883

d. Ước lượng các giá trị nói trên khi co kể đến trọng số

a. ci m3c10a [aw= wt9]

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]	
m3c10a	16959	2.65339	.0310671	2.592495	2.714284

. ci m3c10b [aw= wt9]

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]	
m3c10b	14803	523.7015	20.33493	483.8425	563.5605

b. gen DKV = (m3c8b==3)

. tab DKV

DKV	Freq.	Percent	Cum.
0	18,698	96.90	96.90
1	599	3.10	100.00
Total	19,297	100.00	

. proportion DKV [pw=wt9]

Proportion estimation Number of obs= 19297

	Proportion	Std. Err.	[95% Conf.	Interval]
DKV				
0	.9700864	.0012988	.9675406	.9726322
1	.0299136	.0012988	.0273678	.0324594

c. gen KSK = (m3c9==3)

. tab KSK

KSK	Freq.	Percent	Cum.
0	16,537	85.70	85.70
1	2,760	14.30	100.00
Total	19,297	100.00	

. proportion KSK [pw= wt9]

Proportion estimation Number of obs= 19297

	Proportion	Std. Err.	[95% Conf.	Interval]
KSK				
0	.8584721	.0027527	.8530765	.8638676
1	.1415279	.0027527	.1361324	.1469235