Chương Trình Giảng Dạy Kinh tế Fulbright

Học kỳ Thu năm 2013

Các Phương Pháp Phân Tích Định Lượng

GỢI Ý GIẢI BÀI TẬP 4

LẤY MẪU VÀ ƯỚC LƯỢNG

Ngày Phát: Thứ Ba 22/10/2012
Ngày Nộp: Thứ Ba 29/10/2012

Bài 1: (20 điểm)

Cho 6 sản phẩm điện tử có tuổi thọ lần lượt là 2, 4, 6, 6, 7, 8 (đơn vị thời gian). Bây giờ chọn mẫu ngẫu nhiên 4 trong 6 sản phẩm nói trên và tìm hiểu về tuổi thọ trung của 4 sản phẩm đó. Để đáp ứng mục tiêu này hãy trả lời các câu hỏi sau đây:

a. *Nếu coi 6 sản phẩm trên là một tổng thể, hãy tính các tham số thống kê: μ_X và σ_X.*

b. *Tiến hành chọn mẫu ngẫu nhiên như đã nói ở trên. Tìm phân phối xác suất của trung bình mẫu $\overline{X}$.*

c. *Tính các trị thống kê: $\mu_{\overline{X}}$ và $\sigma_{\overline{X}}$ trực tiếp từ phân phối xác suất của $\overline{X}$. Nêu ý nghĩa của các trị số này.*

d. *Chỉ ra và giải thích các mối quan hệ giữa trị thống kê và tham số thống kê nói trên.*

Đáp án:

Gợi ý cách làm:

Gọi 6 sản phẩm điện tử trên lần lượt là x_1; x_2; x_3; x_4; x_5; x_6

Chọn ngẫu nhiên 4 trong 6 sản phẩm

a.

Coi 6 sản phẩm trên là một tổng thể thì:

$$\mu_x = \frac{\sum x}{N} = \frac{2+4+6+6+7+8}{6} = \frac{33}{6} = 5{,}5 \text{ (đơn vị thời gian).}$$

$$\sigma_x^2 = \frac{\sum (x_i - \mu)^2}{N} = \frac{(2-5{,}5)^2 + (4-5{,}5)^2 + (6-5{,}5)^2 + (6-5{,}5)^2 + (7-5{,}5)^2 + (8-5{,}5)^2}{6}$$

=> $\sigma_x^2 = 3{,}916666667$ (đơn vị thời gian)^2

=> $\sigma_x = 1{,}979057015$ (đơn vị thời gian)

b.

Với tổng thể 6 sản phẩm trên, ta có tất cả $C_6^4 = 15$ cách chọn mẫu và tính ra được trung bình mẫu $\overline{X}$ theo bảng như sau:

Đơn vị: thời gian

Mẫu thứ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
						2	2	2	2	2	2	2	2	2	2
		4	4	4	4					4	4	4	4	4	4
	6		6	6	6		6	6	6				6	6	6
	6	6		6	6	6		6	6		6	6			6
	7	7	7		7	7	7		7	7		7		7	
	8	8	8	8		8	8	8		8	8		8		
$\overline{X}$	6,75	6,25	6,25	6	5,75	5,75	5,75	5,5	5,25	5,25	5	4,75	5	4,75	4,5

Vậy phân phối xác suất trung bình của mẫu $\overline{X}$ là:

$\overline{X}$	4,5	4,75	5	5,25	5,5	5,75	6	6,25	6,75
Tần suất p($\overline{X}$)	1/15	2/15	2/15	2/15	1/15	3/15	1/15	2/15	1/15

c.

Áp dụng công thức tính trung bình mẫu:

$\mu_{\overline{X}} = \Sigma \overline{X}.\ p(\overline{X}) = 5{,}5$ (đơn vị thời gian)

Áp dụng công thức tính độ lệch chuẩn của $\overline{X}$:

$\sigma^2_{\overline{x}} = \sum (x - \overline{X})^2 p(\overline{X}) = 0{,}312916389$ (đơn vị thời gian)^2

=> $\sigma_{\overline{X}} = 0{,}625832779$ (đơn vị thời gian)

Ý nghĩa của các trị số này:

- $\mu_{\overline{X}}$ cho biết giá trị trung bình của các trung bình mẫu đã chọn ở trên, giá trị này cũng bằng với kỳ vọng của tổng thể, $\mu_{\overline{X}}$ là ước lượng không chệch của trung bình tổng thể.

- $\sigma_{\overline{X}}$ thể hiện sự biến thiên hay mức độ phân tán của các trung bình mẫu xung quanh giá trị kỳ vọng của chúng, có thể nhận thấy các mẫu có mức độ phân tán nhỏ hơn so với tổng thể.

- Đây là các trị thống kê của mẫu, có thể sử dụng các trị thống kê này để ước lượng các tham số của tổng thể.

d.

Chỉ ra và giải thích các mối quan hệ giữa trị thống kê và tham số thống kê nói trên

- Kỳ vọng của phân phối trung bình mẫu bằng kỳ vọng của tổng thể vì kỳ vọng của trung bình mẫu chính là trung bình của các số trung bình của mẫu. Kỳ vọng của phân phối trung bình mẫu là ước lượng không chệch của trung bình tổng thể.

- Độ lệch chuẩn của phân phối trung bình mẫu nhỏ hơn độ lệch chuẩn của tổng thể, vì các trung bình mẫu dao động nhỏ hơn so với các giá trị trong tổng thể. Do tổng thể là hữu hạn nên có mối liên hệ sau:

$\sigma_{\overline{X}} = \dfrac{\sigma_X}{\sqrt{n}}.\sqrt{\dfrac{N-n}{N-1}} = \dfrac{1{,}979057}{\sqrt{4}}.\sqrt{\dfrac{6-4}{6-1}} = 0{,}625832779$ (đơn vị thời gian)

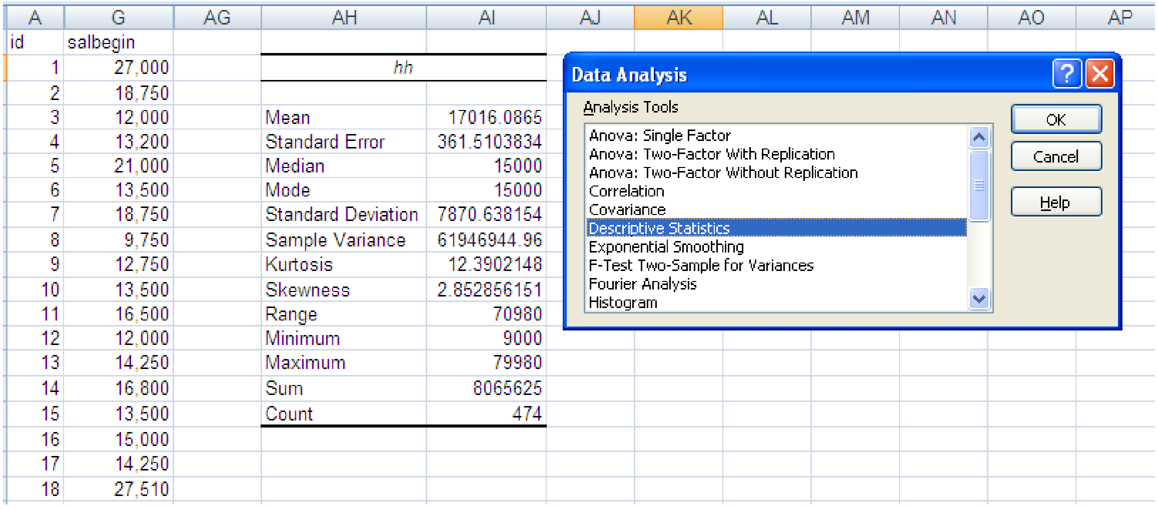
<u>Bài 2</u>: (25 điểm)

*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. Coi đây là một tổng thể.*

a. Tính toán các giá trị thống kê mô tả của tập dữ liệu Salbegin.

Trong Ms Excel 2007, ta sử dụng chức năng Descriptive Statistic để tính toán các giá trị thống kê mô tả của biến Salbegin như sau:

Chọn **Data -> Data Analysis -> Descriptive Statistic** và chọn như hình dưới:



Ta được kết quả như sau:

Salbegin (USD)		
Mean	17,016.08650	USD
Standard Error	361.51038	USD
Median	15,000.00000	USD
Mode	15,000.00000	USD
Standard Deviation	7,870.63815	USD
Sample Variance	61,946,944.96000	USD^2
Kurtosis	12.39021	
Skewness	2.85286	
Range	70,980.00000	USD
Minimum	9,000.00000	USD
Maximum	79,980.00000	USD
Sum	8,065,625.00000	USD
Count	474	USD

Lưu ý: Với các tính này, Excel coi tập dữ liệu là một mẫu thay vì tổng thể, do đó các giá trị Phương sai và Độ lệch chuẩn được tính theo công thức của mẫu.

Nếu coi tập dữ liệu là tổng thể thì:

σ^2 = 61816255,2 USD^2 (sử dụng lệch *Varp(salbegin)* trong Excel)

=> σ = 7862,33141 USD

b. *Chọn ngẫu nhiên một mẫu có cỡ bất kỳ từ tổng thể nêu trên, hãy dự đoán một cách định tính:*

- *Giá trị kỳ vọng của mẫu này.*
- *Độ phân tán của giá trị trung bình trong mẫu này xung quanh giá trị trung bình của tổng thể. Nêu rõ các giả định cần thiết để dự đoán của bạn có độ tin cậy cao.*

Tự chọn ngẫu nhiên một mẫu với cỡ mẫu bất kì, giả sử cỡ mẫu lớn hơn 30. Theo Định lý giới hạn Trung tâm, nếu các mẫu ngẫu nhiên gồm n quan sát được rút ra từ một tổng thể không chuẩn tắc với trung bình **μ** và độ lệch chuẩn **σ**, thì khi **n** lớn, phân phối chọn mẫu của trung bình mẫu $\overline{X}$ được phân phối xấp xỉ chuẩn tắc với $\mu_{\bar{x}} = \mu$ và $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

Vậy có thể dự đoán, giá trị kỳ vọng của mẫu bằng với giá trị kỳ vọng của tổng thể.

Độ phân tán của giá trị trung bình trong mẫu xung quanh giá trị trung bình của tổng thể (hay nói cách khác là độ lệch chuẩn của phân phối trung bình mẫu) nhỏ hơn độ lệch chuẩn của tổng thể.

Giả sử chọn mẫu bất kỳ với cỡ mẫu n = 50 từ tổng thể trên, ta có:

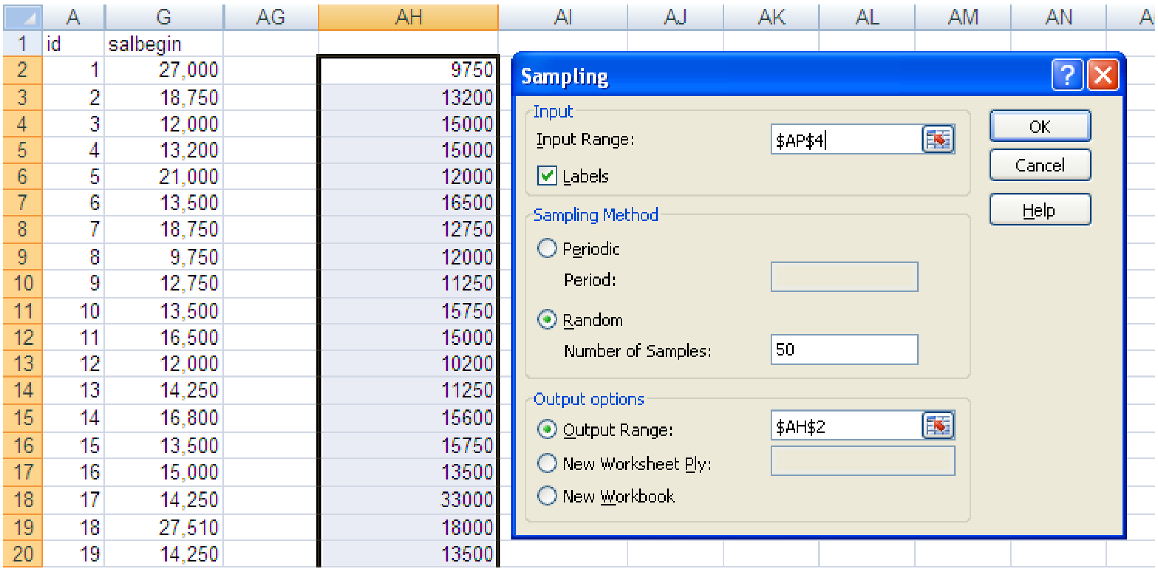
$$\mu_{\bar{x}} = \mu = 17016{,}0865 \text{ USD}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{7862{,}33141}{\sqrt{50}} = 1111.901571 \text{ USD}$$

Để có thể dự đoán các giá trị như trên, cần giả định rằng việc chọn mẫu là ngẫu nhiên, kích thước mẫu **n** là lớn (n>=30).

c. *Hãy chọn ngẫu nhiên một mẫu có cỡ n = 50 từ tập dữ liệu nói trên, chỉ ra các giá trị:*

Ta tiếp tục chọn ngẫu nhiên một mẫu có cỡ n = 50 như sau:



- *Kỳ vọng của mẫu.*

Sử dụng lệnh Mean, ta tính được kỳ vọng của mẫu là: 15.609 USD

- *Khoảng cách từ kỳ vọng của mẫu tới kỳ vọng của tổng thể:*

Khoảng cách (hay chênh lệch) giữa kỳ vọng của mẫu tới kỳ vọng của tổng thể là:

15609 – 17.016 = -1407 (USD)[1]

- *So sánh với dự đoán ở phần (b) và giải thích các quan sát thu được.*

So sánh với kết quả ở câu (b) ta thấy trung bình của mẫu vừa chọn là 15.609 có chênh lệch không lớn với kỳ vọng của trung bình mẫu là 17.016.

Mức chênh lệch -1407 cũng xoay quanh độ lệch chuẩn trung bình mẫu là 1111,901571.

[1] Cũng có thể tính khoảng cách như sau: $|15609 - 17.016| = 1407$ (USD)

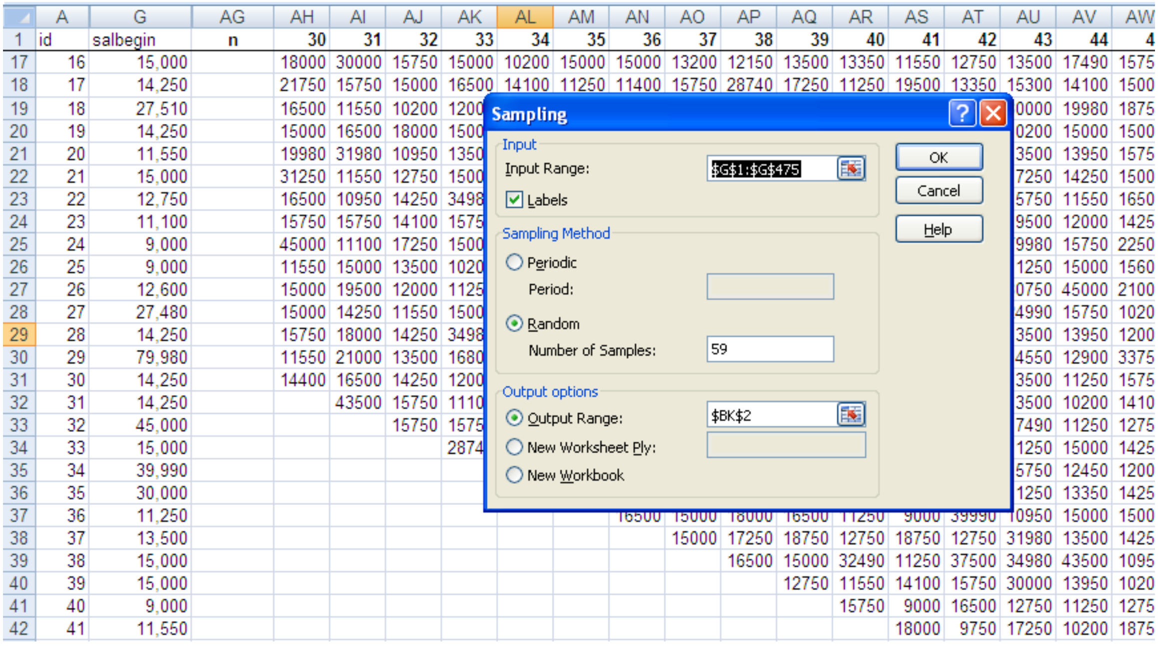
Giải thích:

- Với mẫu vừa chọn ở câu (c), giá trị trung bình của mẫu là 15.609 và là một đại diện trong phân phối trung bình mẫu mà ta đề cập tới ở câu (b), nên giá trị trung bình này (15.609) là một phần tử xoay quanh giá trị trung bình của phân phối mẫu (17.016,0865).
- Khoảng cách giữa kỳ vọng của mẫu tới kỳ vọng của tổng thể (cũng bằng kỳ vọng của phân phối trung bình mẫu) thể hiện mức độ phân tán của kỳ vọng của mẫu so với kỳ vọng của tổng thể, nên giá trị của nó cũng *xoay quanh* độ lệch chuẩn của phân phối trung bình mẫu.

d. *Dùng lệnh Random trong Excel để chọn ra 30 mẫu ngẫu nhiên bất kỳ từ tập dữ liệu trên với cỡ mẫu tăng dần bắt đầu từ n = 30.*

Để chọn ra 30 mẫu ngẫu nhiên bất kỳ từ tập dữ liệu trên với cỡ mẫu tăng dần bắt đầu từ n = 30, ta làm lần lượt 30 lần như sau:

Chọn **Data -> Data Analysis -> Sampling** và điền các thông số như sau:



Làm lần lượt như vậy 30 lần, mỗi lần thay đổi tăng số Number of Samples trong mục Random lên 1 đơn vị từ cỡ mẫu n = 30 cho đến khi n = 59, ta được 30 mẫu ngẫu nhiên bất kì thỏa mãn đề bài.

- *Với mỗi mẫu hãy lặp lại các yêu cầu trong phần (c).*

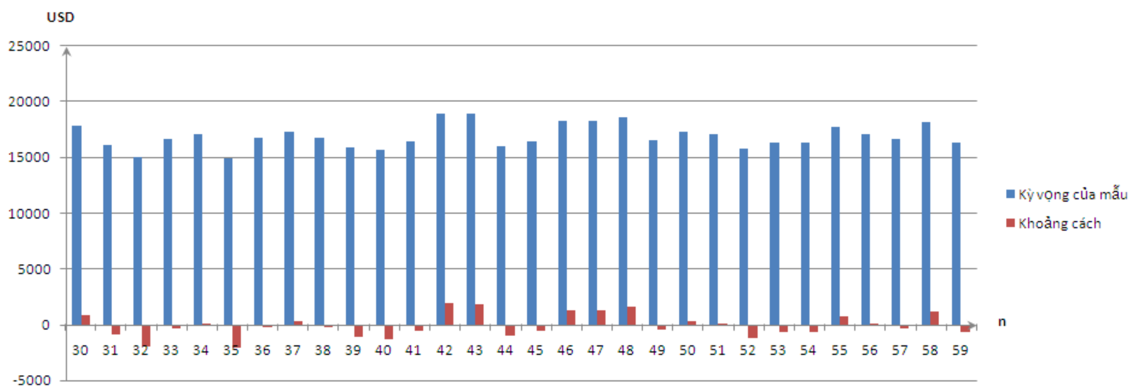
Với mỗi mẫu, tính kỳ vọng của mẫu; khoảng cách từ kỳ vọng của mẫu tới kỳ vọng của tổng thể như sau:

BK61 f_x =AVERAGE(BK2:BK60)

	A	G	AG	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG	BH	BI	BJ	BK
1	id	salbegin	n	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59
43	42	16,500		14250	12750	36750	9750	15750	32490	14250	11250	11225	15000	21480	18750	15750	13500	44100
44	43	14,250		13500	16500	11250	16500	13500	15750	15000	11225	23250	16500	27000	11250	32490	11250	19500
45	44	14,250		11250	13500	12000	16500	45000	15000	13050	14400	13200	11250	10950	18000	15000	15000	13200
46	45	13,500		15000	14250	28740	11250	29490	52500	33000	13500	15000	15000	34980	33750	14100	12750	15000
47	46	12,750			15750	13500	13500	11250	11250	13500	18000	16050	15750	12000	20550	15000	13500	10950
48	47	16,500				15750	20400	18000	13500	12750	15750	12750	13050	13500	10950	20250	18750	12600
49	48	14,100					34980	32490	12000	16500	13500	14250	10200	18000	32490	15750	18000	15000
50	49	16,500						13950	12750	18000	9750	11250	9750	13500	15000	9000	13500	27480
51	50	23,730							11250	10500	13950	13500	17250	15000	15750	10200	14250	15000
52	51	15,000								15750	16500	27480	35040	16500	30000	25500	21240	10200
53	52	15,000									9750	15000	15750	31980	15750	33000	10500	15000
54	53	26,250										27480	14250	15750	13500	12750	24990	11400
55	54	13,500											44100	33000	14250	23730	32490	13500
56	55	15,000												29490	20250	11250	12750	13350
57	56	13,500													27480	14250	15750	15750
58	57	15,750														13500	10200	15000
59	58	13,500															34980	12450
60	59	14,250																10200
61	60	15,000	Kỳ vọng của mẫu	16465	18294	18334	18631	16605	17331	17119	15829	16340	16320	17710	17108	16669	1816[illegible]	16379
62	61	9,750	Khoảng cách	-551	1278	1318	1615	-411	315.3	103.2	-1187	-676	-696	693.7	91.95	-347	1152	-637
63	62	21,750																
64	63	26,250					=AVERAGE(BK2:BK60)											
65	64	21,000																
66	65	14,550					=BK61-17016.0865											

- *Vẽ đồ thị của các giá trị vừa tìm được theo sự tăng dần của n.*

Đồ thị kỳ vọng của mẫu và Khoảng cách từ kỳ vọng của mẫu tới kỳ vọng của tổng thể



- *Mô tả xu thế thay đổi của các giá trị này theo n.*

- Khi n tăng lên, *Kỳ vọng của mẫu* vẫn biến động xung quanh giá trị kỳ vọng của tổng thể, *Khoảng cách từ kỳ vọng của mẫu tới kỳ vọng của tổng thể* cũng dao động xung quanh giá trị 0 và có vẻ như ngày càng dao động nhỏ hơn khi n càng tăng.

- *Bạn rút ra được điều gì từ câu hỏi này?*

Như vậy, nếu chọn ngẫu nhiên cỡ mẫu càng lớn, thì sự biến động của của *Kỳ vọng của mẫu* và *Khoảng cách từ kỳ vọng của mẫu tới kỳ vọng của tổng thể* càng nhỏ, độ chính xác của trung bình mẫu so với kỳ vọng của tổng thể càng cao. Tuy nhiên đi kèm với đó là tốn thêm nhiều chi phí do cỡ mẫu ngày càng lớn.

e. *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 phần (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?*

Lúc này 474 quan sát có được là một mẫu của tổng thể. Ta có Trung bình = 17.016,0865 và Phương sai = 61.946.944,96

Với mong muốn có độ lệch chuẩn của mẫu chọn từ tập dữ liệu này là 100 USD, ta có:

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ = > n = $\frac{\sigma^2}{\sigma_{\bar{x}}^2}$ = $\frac{7870{,}6382^{\,2}}{100^2}$ = 6194,69

=>Cỡ mẫu cần thiết là n = 6195 (mẫu)

<u>Bài 3</u>: (20 điểm)

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

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ và $\bar{X}' = \sum_{i=1}^{n} X_i f_i$ với f_i là tần suất xuất hiện của X_i

Giả sử rằng $\bar{X}$ và $\bar{X}'$ là hai đại lượng ngẫu nhiên và độc lập với nhau.

a. Chứng tỏ rằng $\bar{X}$ và $\bar{X}'$ là các ước lượng không chệch của μ.

Giả sử rằng phương sai của X_1, X_2, …,X_n là như nhau (và bằng s)

- Ta có E($\bar{X}$) = E ($\frac{1}{n}\sum_{i=1}^{n} X_i$) = E ($\frac{1}{n}X_1 + \frac{1}{n}X_2 + \ldots + \frac{1}{n}X_n$)

= E($\frac{1}{n}X_1$) + E($\frac{1}{n}X_2$) + …+ E($\frac{1}{n}X_n$)

= $\frac{1}{n}$E(X_1) + $\frac{1}{n}$E(X_2) + …+ $\frac{1}{n}$E(X_n)

= $\frac{1}{n}$ **μ +…+** $\frac{1}{n}$ **μ = n.** $\frac{1}{n}$ **μ = μ**

Vậy $\bar{X}$ là ước lượng không chệnh của **μ**

- Tương tự, ta có E($\bar{X}'$) = E ($\sum_{i=1}^{n} X_i f_i$) = E ($X_1 . f_1 + X_2 . f_2 + \ldots + X_n . f_n$)

= E ($X_1 . f_1$) + E ($X_2 . f_2$) + … + E ($X_n . f_n$)

= f_1.E (X_1) + f_2.E (X_2) + … + f_n.E (X_n)

= f_1 . **μ** + f_2 . **μ** + … + f_n . **μ**

= μ ($f_1 + f_2 + \ldots + f_n$)

= μ

Vậy $\bar{X}'$ là ước lượng không chệnh của **μ**

b. Ước lượng nào hiệu quả hơn?

- Ta có Var($\bar{X}$) = Var ($\frac{1}{n}\sum_{i=1}^{n} X_i$) = Var ($\frac{1}{n}X_1 + \frac{1}{n}X_2 + \ldots + \frac{1}{n}X_n$)

= Var($\frac{1}{n}X_1$) + Var($\frac{1}{n}X_2$) + …+ Var($\frac{1}{n}X_n$)

= $\frac{1}{n^2}$ Var(X_1) + $\frac{1}{n^2}$ Var(X_2) + …+ $\frac{1}{n^2}$ Var(X_n)

= $\frac{1}{n^2}$ S^2 + $\frac{1}{n^2}$ S^2+ …+ $\frac{1}{n^2}$ S^2

=n. $\frac{1}{n^2}$ S^2

= $\frac{1}{n}$ S^2

- Ta có Var ($\overline{X}'$) = Var ($\sum_{i=1}^{n} X_i f_i$) = Var ($X_1 . f_1 + X_2 . f_2 + \ldots + X_n . f_n$)

  = Var ($X_1 . f_1$) + Var ($X_2 . f_2$) + … + Var ($X_n . f_n$)

  = f_1^2 .Var (X_1) + f_2^2 .Var (X_2)+ … + f_n^2 .Var (X_n)

  = $f_1^2 . S^2 + f_2^2 . S^2 + \ldots + f_n^2 . S^2$

  = S^2 **(** $f_1^2 + f_2^2 + \ldots + f_n^2$)

- Ta sẽ chứng minh **(** $f_1^2 + f_2^2 + \ldots + f_n^2$) $\geq \frac{1}{n}$

Thật vậy:

<u>Cách 1:</u>

Ta có $\sum f_i^2 = \sum (\frac{k_i}{n})^2$ với k_i là tần suất xuất hiện của X_i.

=> $\sum f_i^2 = \frac{1}{n^2}\sum (k_i)^2$

Do $k_i \geq 1$ => $\sum (k_i)^2 \geq$ n

=> $\sum f_i^2 \geq \frac{1}{n^2}.n = \frac{1}{n}$

<u>Cách 2:</u>

$f_1^2 + f_2^2 + \ldots + f_n^2$ = 1 – (2 $f_1 f_2$ +2 $f_1 f_3$+…2 $f_1 f_n$ +…2 $f_{n-1} f_n$)

=> $f_1^2 + f_2^2 + \ldots + f_n^2 \geq$ 1- ($f_1^2 + f_2^2$)-($f_1^2 + f_3^2$)-…-($f_1^2 + f_n^2$) -…-($f_{n-1}^2 + f_n^2$)

=> f_1^2 + f_2^2 + … + f_n^2 ≥ 1- (n-1) f_1^2 + -(n-1) f_2^2 +… + (n-1) f_n^2

=>n(f_1^2 + f_2^2 + … + f_n^2) ≥ 1

=>(f_1^2 + f_2^2 + … + f_n^2) ≥ $\frac{1}{n}$

=> S^2 (f_1^2 + f_2^2 + … + f_n^2) ≥ $\frac{1}{n}$ S^2

=> Var ($\overline{X}'$) ≥ Var($\overline{X}$) =>Ước lượng $\overline{X}$ hiệu quả hơn ước lượng $\overline{X}'$.

<u>Bài 4</u>: (20 điểm)

Tuổi thọ của một loại bóng đèn được biết tuân theo quy luật chuẩn. Theo báo cáo của phòng kỹ thuật thì tuổi thọ trung bình của các bóng đèn này là 1012 giờ. Để đánh giá chất lượng một lô hàng mới xuất xưởng, người ta chọn ngẫu nhiên 1000 bóng để thử nghiệm, thấy có 70 bóng có tuổi thọ trung bình lớn 1015 giờ. Hãy trả lời các câu hỏi sau, nêu rõ cách tính toán và các lập luận.

Ta có:

μ= 1012 giờ

n = 1000 bóng

Chọn ngẫu nhiên 1000 bóng để thử nghiệm, thấy có 70 bóng có tuổi thọ trung bình lớn 1015 giờ => P(X>1015) = 7%

=> P ($\frac{X-\mu}{\sigma} > \frac{1015-1012}{\sigma}$) = 7%

=>P(Z> $\frac{3}{\sigma}$)=7%

=P(Z< $\frac{3}{\sigma}$)=1-7%= 93%

Đặt Z_0 = $\frac{3}{\sigma}$ =>P(Z< Z_0)=1-7%= 93%

Tra bảng phân phối chuẩn hóa hoặc dùng lệnh Normsinv trong Excel ta được
Z_0 = 1,475791028

=> $\frac{3}{\sigma}$ =1,475791028 => σ = 2,03280813 (giờ)

a. Tính tuổi thọ trung bình của bóng đèn xí nghiệp sản xuất với độ tin cậy 95%.

Biên sai số là: $Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

Với độ tin cậy là 95% => α = 1 -0,095 = 0,05

=> $Z_{\frac{\alpha}{2}}$ = 1,96

Như vậy biên sai số xấp xỉ bằng:

$1.96\dfrac{\sigma}{\sqrt{n}}$ =1,96 $\dfrac{2{,}03280813}{\sqrt{1000}}$ = 0,125994753 (giờ)

Vậy tuổi thọ trung bình của bóng đèn xí nghiệp sản xuất là 1012 giờ, với độ tin cậy là 95% thì biên sai số là 0,125994753 giờ.

Ta có khoảng tin cậy là [1011,874005; 1012,125995] giờ

b. Hãy ước lượng sai số chuẩn của phân phối mẫu.

Sai số chuẩn (độ lệch chuẩn) của phân phối:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2{,}03280813}{\sqrt{1000}} = 0{,}064283037 \text{ (giờ)}$$

c. Với độ chính xác là 15 giờ. Hãy xác định độ tin cậy.

Độ chính xác là 15 giờ => $Z_{\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}} = 15$

=> $Z_{\frac{\alpha}{2}}\dfrac{2{,}03280813}{\sqrt{1000}}$ = 15 => $Z_{\frac{\alpha}{2}}$ = 233,34305

Dùng lệnh Normsdist trong excel => Độ tin cậy ≈ 100 %

d. Với độ chính xác là 25 giờ và độ tin cậy là 95% cần thử nghiệm bao nhiêu bóng?

Với độ chính xác là 25 giờ và độ tin cậy là 95%, ta có:

$Z_{\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}}$ = 25 => 1,96 . $\dfrac{2{,}03280813}{\sqrt{n}}$ = 25

=> n = 0.025399 (bóng)

Vậy chỉ cần phải thử nghiệm 1 bóng là đủ.

<u>Bài 5</u>: (15 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.

a. 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 20% đến 30%.

<u>Cách 1:</u>

Theo đầu bài ta có

n_0 = 200;

$\widehat{p}_0$ = 0,3

n_1 =250

Lúc này, coi như giá trị tỉ lệ $\widehat{p}_0$ = 0,3 là đã biết và xác định từ trước.

Gọi Px là biến ngẫu nhiên liên tục chỉ tỉ lệ không đạt yêu cầu phòng chống cháy nổ.

P(0,2 < $\hat{p}_x$ < 0,3) = P ($\frac{0,2-\hat{p}_0}{\sigma_{\hat{p}_x}} < \frac{\hat{p}_x-\hat{p}_0}{\sigma_{\hat{p}_x}} < \frac{0,3-\hat{p}_0}{\sigma_{\hat{p}_x}}$)

=P($\frac{0,2-\hat{p}_0}{\sigma_{\hat{p}_x}} < Z < \frac{0,3-\hat{p}_0}{\sigma_{\hat{p}_x}}$)

Với $\sigma_{\hat{p}_x} = \sqrt{\frac{\hat{p}_0(1-\hat{p}_0)}{n_1}} = \sqrt{\frac{0,3(1-0,3)}{250}}$ = 0,028982753

=> P(0,2 < $\hat{p}_x$ < 0,3) = P ($\frac{0,2-0,3}{0,028982753} < Z < \frac{0,3-0,3}{0,028982753}$)
= P(-3,450327797 <Z<0)
= P(Z<0) - P(Z<-3,450327797)
= 0,5 – 0,000279953
= 0,499720047 = 49,97%

Vậy, 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 20% đến 30% là 49,97%.

b. *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?*

Ta có: n_0 =200 ; $\hat{p}_0$ = 0,3
n_1 =250 ; $\hat{p}_1$ = 0,32

Ước lượng khoảng tin cậy (1 - α) 100% cho ($\hat{p}_0$ - $\hat{p}_1$)

$$(\hat{p}_0-\hat{p}_1) \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_0\hat{q}_0}{n_0}+\frac{\hat{p}_1\hat{q}_1}{n_1}}$$

Khoảng tin cậy là 95% nên $Z_{\frac{\alpha}{2}}$ =1,96

=>khoảng tin cậy là: $(\hat{p}_0-\hat{p}_1) \pm 1,96\sqrt{\frac{\hat{p}_0\hat{q}_0}{n_0}+\frac{\hat{p}_1\hat{q}_1}{n_1}}$

$$=(0,3-0,32) \pm 1,96\sqrt{\frac{0,3(1-0,3)}{200}+\frac{0,32(1-0,32)}{250}}$$

= - 0,02 $\pm$ 0,085
=> Khoảng tin cậy là **[-0,105 ; 0,065]**[2]
Vậy sự khác biệt về tỷ lệ không đạt yêu cầu an toàn phòng chống cháy nổ của 2 Sở trên nằm trong khoảng [-10,5%;6,5%] với độ tin cậy 95%.

[2] Chú ý: Cũng có thể xác định khoảng tin cậy bằng cách xác định khoảng tin cậy là

$(\hat{p}_1-\hat{p}_0) \pm 1,96\sqrt{\frac{\hat{p}_0\hat{q}_0}{n_0}+\frac{\hat{p}_1\hat{q}_1}{n_1}}$ và cho kết quả theo cách tương tự.