

Chương Trình Giảng Dạy Kinh tế Fulbright

Học kỳ Thu năm 2011

Các Phương Pháp Phân Tích Định Lượng

Gợi ý lời giải Bài tập 3

PHÂN PHỐI XÁC SUẤT

Ngày Phát: Thứ Hai, 17/10/2010

Ngày Nộp: 8:20 sáng, Thứ Hai, 24/10/2010

Bài 1: (20 điểm)

Tìm diện tích của đường cong chuẩn chuẩn hóa

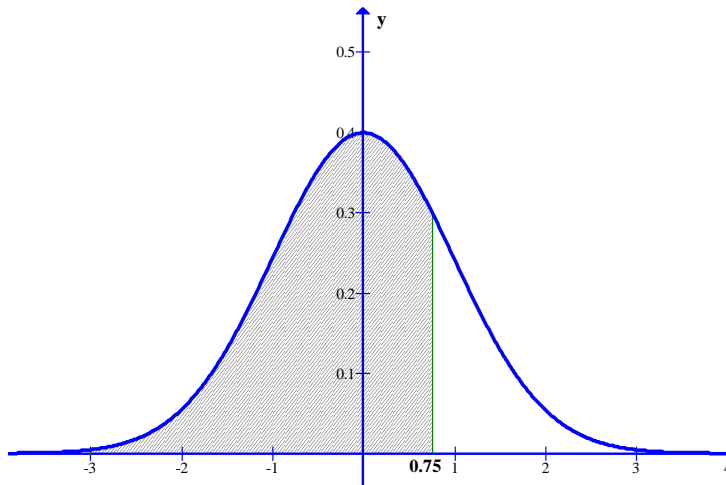
a. Bên trái điểm $Z = 0.75$

$$S = P(Z < 0.75) = \text{NORMSDIST}(0.75) = 0.7734$$

Hoặc tra bảng phân phối Z , ta có thể tính như sau:

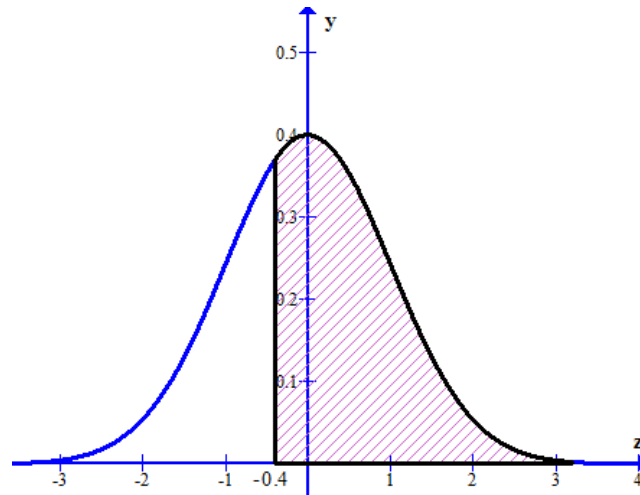
$$S = P(Z < 0.75) = P(Z < 0) + P(0 < Z < 0.75) = 0.5 + 0.2734 = 0.7734.$$

(Hình vẽ trong câu này chỉ để minh họa, không nhất thiết phải trình bày)



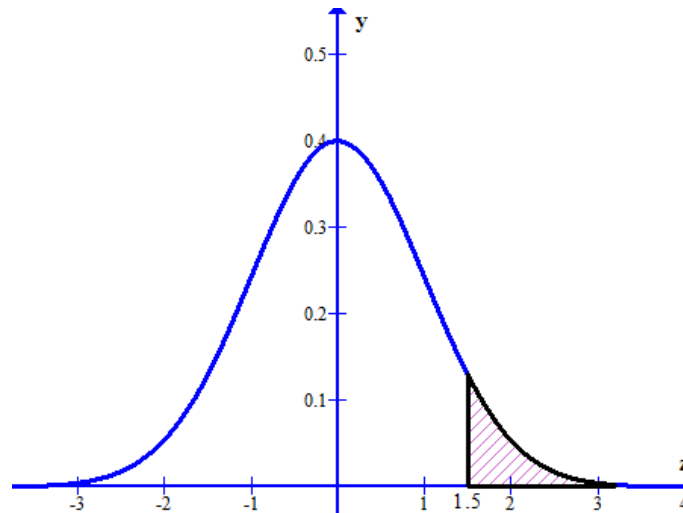
b. Bên phải điểm $Z = -0.40$

$$S = P(Z > -0.40) = 1 - P(Z < -0.40) = 1 - \text{NORMSDIST}(-0.4) = 1 - 0.3446 = 0.6554$$



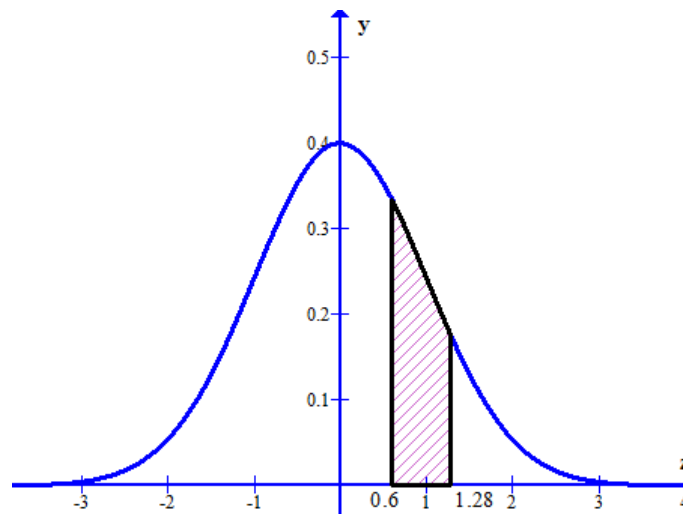
c. Bên phải điểm $Z = 1.50$

$$S = P(Z > 1.5) = 1 - P(Z < 1.5) = 1 - \text{NORMSDIST}(1.5) = 1 - 0.9332 = 0.0668$$



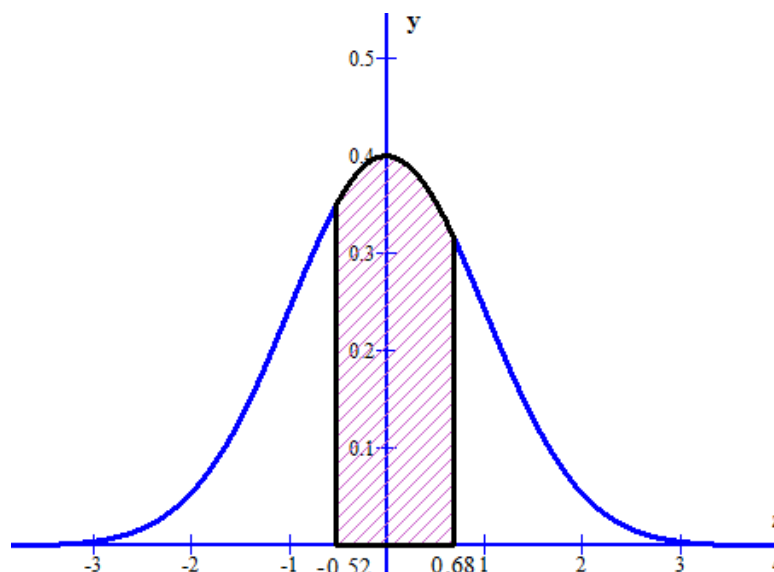
d. Giữa $Z = 0.60$ và $Z = 1.28$

$$S = P(0.60 < Z < 1.28) = P(Z < 1.28) - P(Z < 0.60) = \text{NORMSDIST}(1.28) - \text{NORMSDIST}(0.6) = 0.8997 - 0.7257 = 0.174$$



e. Giữa $Z = -0.52$ và $Z = 0.68$

$$S = P(-0.52 < Z < 0.68) = P(Z < 0.68) - P(Z < -0.52) = \text{NORMSDIST}(0.68) - \text{NORMSDIST}(-0.52) = 0.7517 - 0.3015 = 0.4502$$



Bài 2: (20 điểm)

Biến ngẫu nhiên X tuân theo phân phối chuẩn với $\mu = 1200$ và $\sigma = 12$.

a. Tìm xác suất để giá trị của X lớn hơn 2 lần độ lệch chuẩn so với trị trung bình.

$$P(X \geq \mu + 2\sigma) = P(Z \geq 2) = 1 - \text{NORMSDIST}(2) = 1 - 0.9773 = 0.0227$$

b. Tìm xác suất để giá trị của X lớn hơn hoặc bằng 3 lần độ lệch chuẩn so với trị trung bình.

$$P(X \geq \mu + 3\sigma) = P(Z \geq 3) = 1 - \text{NORMSDIST}(3) = 1 - 0.9987 = 0.0013$$

c. Tìm xác suất để giá trị của X nằm trong khoảng lớn hơn và nhỏ hơn một lần độ lệch chuẩn so với trị trung bình.

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = P(-1 \leq Z \leq 1) = 0.8413 - 0.1587 = 0.6826$$

d. Tìm xác suất để cho X nằm trong khoảng 1050 đến 1350

$$P(1050 \leq X \leq 1350) = P\left(\frac{1050 - 1200}{12} \leq Z \leq \frac{1350 - 1200}{12}\right) = P(-12.5 \leq Z \leq 12.5)$$

Do $P(-3 < Z < 3) \approx 1$ và $(-3; 3) \subset (-12.5; 12.5)$ nên $P(-12.5 \leq Z \leq 12.5) \approx 1$

e. Tìm xác suất để X bằng 1050

$P(X = 1050) = 0$ do X là biến ngẫu nhiên liên tục nên xác suất tại một điểm bất kỳ sẽ bằng 0.

Bài 3: (20 điểm)

Người ta quan sát sự tiêu thụ số Tấn nguyên liệu hóa màu tại một xí nghiệp mỹ phẩm, tổng số ngày quan sát là 150 ngày. Kết quả được trình bày trong bảng sau:

Số tấn nguyên liệu X	0	1	2	3	4	5
Số ngày có giá trị X	5	15	30	50	30	20

X là một đại lượng ngẫu nhiên rời rạc. Bảng phân phối xác suất của X như sau

X	0	1	2	3	4	5
P(X=xi)(xấp xỉ)	0.0333	0.100	0.200	0.3333	0.200	0.1333

- a. Hãy xác định xác suất để số tấn nguyên liệu X lớn hơn hoặc bằng 3

$$P(X \geq 3) = P(X=3) + P(X=4) + P(X=5) = 0.3333 + 0.200 + 0.1333 = 0.6666$$

- b. Hãy xác định xác suất để số tấn nguyên liệu X lớn hơn 2

$$P(X > 2) = P(X \geq 3) = 0.6666$$

- c. Hãy xác định xác suất để số tấn nguyên liệu X lớn hơn hoặc bằng 1

$$P(X \geq 1) = 1 - P(X=0) = 1 - 0.0333 = 0.9667$$

- d. Hãy xác định giá trị kỳ vọng của số tấn nguyên liệu X

$$E(X) = \sum_{i=1}^n x_i p_i = 0 * 0.0333 + 1 * 0.10 + 2 * 0.20 + 3 * 0.3333 + 4 * 0.20 + 5 * 0.1333 \approx 2.97 \text{ (tấn)}$$

- e. Hãy xác định phương sai và độ lệch chuẩn của số tấn nguyên liệu X

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

Nếu xem dữ liệu là của tổng thể, có thể tính phương sai như sau:

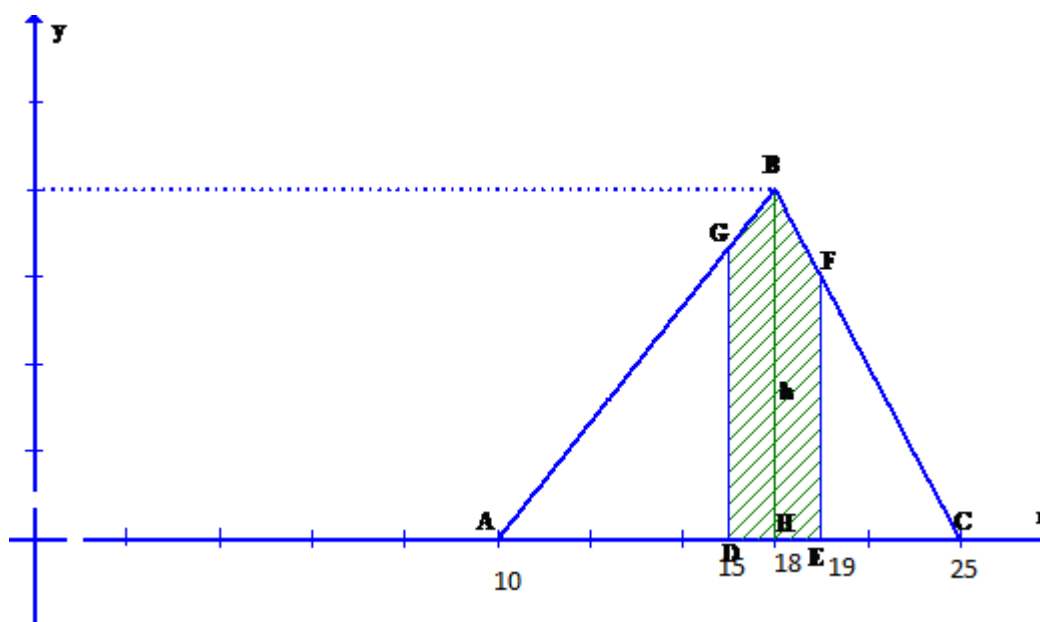
$$\text{Ta có } E(X^2) = 0^2 * 0.0333 + 1^2 * 0.10 + 2^2 * 0.20 + 3^2 * 0.3333 + 4^2 * 0.20 + 5^2 * 0.1333 = 10.422$$

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = 10.422 - 2.967^2 \approx 1.62 \text{ (tấn}^2\text{)}$$

$$\text{Độ lệch chuẩn của } X = 1.62^{(0.5)} \approx 1.273 \text{ (tấn)}$$

Bài 4: (20 điểm)

- a. X là một biến ngẫu nhiên tuân theo qui luật phân phối tam giác bất cân xứng trong khoảng từ 10 đến 25. Đỉnh của tam giác phân phối ở tại X = 18. Hãy tính xác suất để có được một giá trị của X nằm trong khoảng từ 15 đến 19. Vẽ hình minh họa, giải thích rõ ràng các tính toán và các giả định của Anh/Chị.



Do $P(10 \leq X \leq 25) = 1 \rightarrow S_{ABC} = h \times 15 / 2 = 1 \rightarrow h = 2/15 \approx 0.1333$

$P(15 \leq X \leq 19) = S_{DGBFE} = S_{ABC} - S_{ADG} - S_{CEF}$

Theo tính chất của tam giác đồng dạng, và định lý Ta-lét đã học ở bậc phổ thông, dễ dàng tính được độ dài của đoạn DG, và EF như sau:

$$\frac{GD}{BH} = \frac{AD}{AH} \Rightarrow GD = BH \frac{AD}{AH} = 0.1333 * \frac{15-10}{18-10} = 0.0833$$

$$\frac{FE}{BH} = \frac{CE}{CH} \Rightarrow FE = BH \frac{CE}{CH} = 0.1333 * \frac{25-19}{25-18} = 0.1143$$

Vậy $S_{DGBFE} = 1 - 0.5 * AD * GD - 0.5 * CE * FE$

$$= 1 - 0.5 * 5 * 0.0833 - 0.5 * 6 * 0.1143 = 0.4489$$

Hoặc bạn cũng có thể tính theo cách:

$$S_{DGFE} = S_{DGBH} + S_{BHEF}$$

Như vậy, $P(15 \leq X \leq 19) = 0.4489$

- b. Kết quả thi tuyển sinh của 20000 thí sinh vào một trường Đại Học tại Tp Hồ Chí Minh như sau: điểm trung bình trong kỳ thi là 18, với điểm nhỏ nhất là 9 và điểm cao nhất là 28. Giả sử rằng kết quả điểm thi trong kỳ thi có thể ước tính gần đúng theo phân phối chuẩn. Có bao nhiêu thí sinh đạt từ 22 đến 25 điểm?

Nếu chọn ngẫu nhiên một sinh viên từ 20000 sinh viên và quan sát điểm của thí sinh này là một phép thử, ta gọi X là đại lượng ngẫu nhiên cho biết điểm thi của thí sinh

Khoảng biến thiên của X là $Range = Max - Min = 28 - 9 = 19$ điểm

Vì X tuân theo phân phối chuẩn, nên từ Range, ta có thể ước lượng gần đúng độ lệch chuẩn theo công thức $s = Range / 6 = 19 / 6 \approx 3.17$ điểm

$$P(22 \leq X \leq 25) = P\left(\frac{22-18}{3.17} \leq Z \leq \frac{25-18}{3.17}\right) = P(1.26 \leq Z \leq 2.21) = 0.9864 - 0.8962 = 0.0902$$

Như vậy, số lượng thí sinh đạt từ 22 đến 25 điểm bằng $0.0902 * 20000$, và bằng 1804 thí sinh.

Bài 5: (20 điểm)

Bạn hãy Sử dụng phần mềm Stata, mở file dữ liệu **hhexp08.dta** trong bộ dữ liệu VHLSS2008. Biến **foodnom** cho biết chi tiêu cho lương thực thực phẩm (theo giá hiện tại) của hộ trong 12 tháng qua – đơn vị tính là ngàn đồng, biến **urban08** (1 thành thị, 2 nông thôn) cho biết hộ được điều tra ở thành thị hay nông thôn. Biến **Reg8** cho biết hộ ở Vùng nào (có 8 vùng). Biến **pcexplrl** cho biết chi tiêu thực bình quân đầu người trong 12 tháng qua của hộ – đơn vị tính là ngàn đồng. Biến **hsize** cho biết tổng số người trong hộ.

a. Bạn hãy tính giá trị trung bình, độ lệch chuẩn, min, max, trung vị của biến **hsize**, **foodnom**, **pcexplrl**

Có nhiều cách tính, bạn có thể tính theo lệnh sau:

```
Tabstat hsize foodnom pcexplrl, stat (count min max mean sd median)
format(%10.2f)
```

stats	hsize	foodnom	pcexplrl
N	9189.00	9189.00	9189.00
min	1.00	956.75	490.36
max	15.00	107039.30	123343.63
mean	4.16	15966.50	7700.00
sd	1.66	9609.15	6077.58
p50	4.00	13978.05	6099.80

b. Trong ba biến **hsize**, **foodnom**, **pcexplrl**, theo bạn biến nào là biến rời rạc, biến nào là biến liên tục.

biến **hsize** là biến rời rạc; 2 biến còn lại là biến liên tục

c. Bạn hãy tạo bảng tần số cho biến **reg8**, cho biến **urban08**, cho biến **hsize** và cho biết những nhận định của mình từ kết quả tính toán

```
. tab reg8
```

reg8	Freq.	Percent	Cum.
1	1,944	21.16	21.16
2	1,317	14.33	35.49
3	429	4.67	40.16
4	1,014	11.03	51.19
5	852	9.27	60.46
6	582	6.33	66.80
7	1,188	12.93	79.73
8	1,863	20.27	100.00
Total	9,189	100.00	

. tab urban08

Khu vực	Freq.	Percent	Cum.
Urban	2,352	25.60	25.60
2	6,837	74.40	100.00
Total	9,189	100.00	

. tab hhsize

Tầng số ng-êi trong hế	Freq.	Percent	Cum.
1	383	4.17	4.17
2	1,004	10.93	15.09
3	1,568	17.06	32.16
4	2,864	31.17	63.33
5	1,769	19.25	82.58
6	901	9.81	92.38
7	398	4.33	96.71
8	169	1.84	98.55
9	74	0.81	99.36
10	27	0.29	99.65
11	19	0.21	99.86
12	6	0.07	99.92
13	3	0.03	99.96
14	3	0.03	99.99
15	1	0.01	100.00
Total	9,189	100.00	

Dựa vào các kết quả, bạn có thể mô tả một số thông tin mà bạn quan tâm.

d. Bạn hãy tính trung bình chi tiêu cho lương thực thực phẩm của hộ phân theo thành thị và nông thôn. Vẽ đồ thị thể hiện kết quả này.

Dưới đây là một gợi ý

```
. tab urban08, sum(foodnom)
```

Summary of nominal food expenditures			
Khu vực	Mean	Std. Dev.	Freq.
Urban	22034.872	13140.487	2352
2	13878.922	6905.9214	6837
Total	15966.505	9609.1549	9189

```
. table urban08, c(n foodnom mean foodnom) format (%10.2f)
```

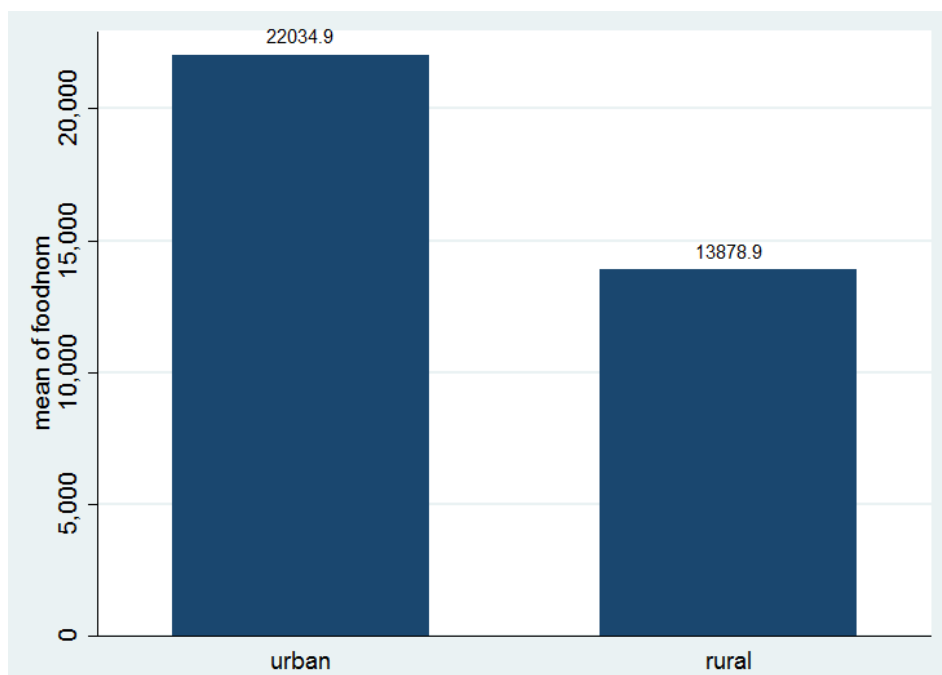
Khu vực	N(foodnom)	mean(foodnom)
Urban	2,352	22034.87
2	6,837	13878.92

```
gen urban=urban08
```

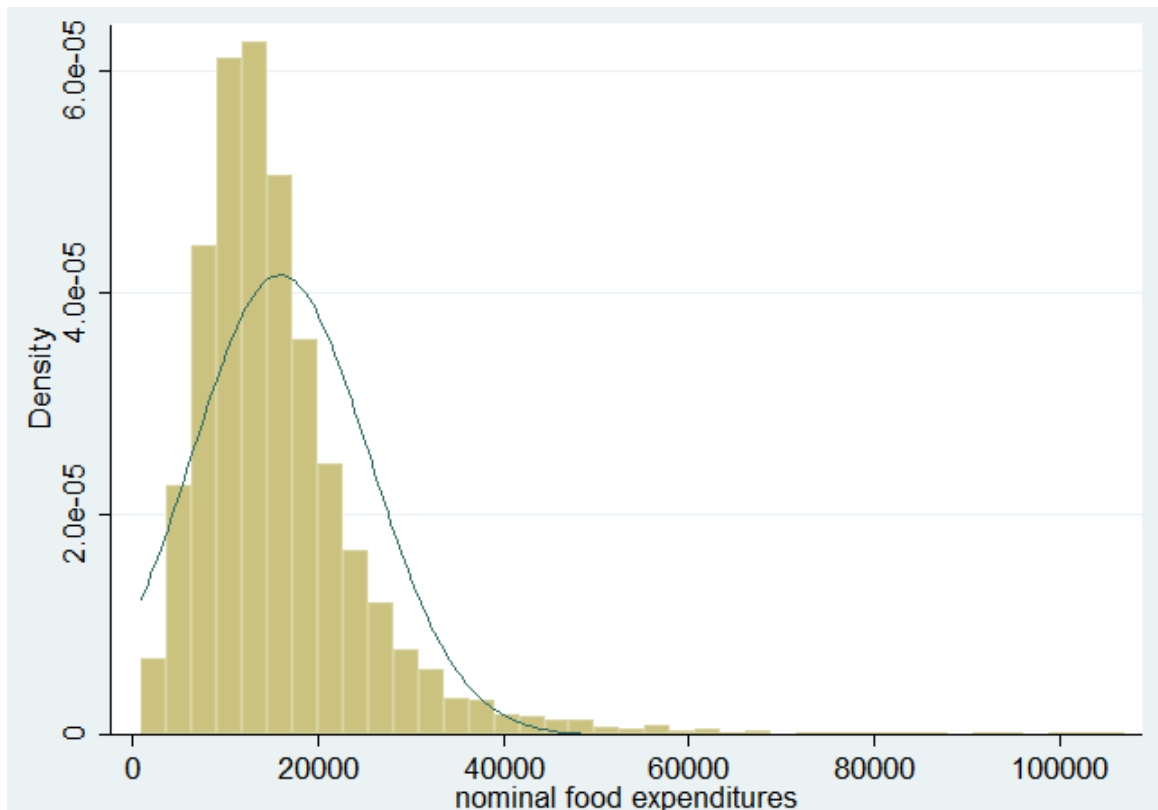
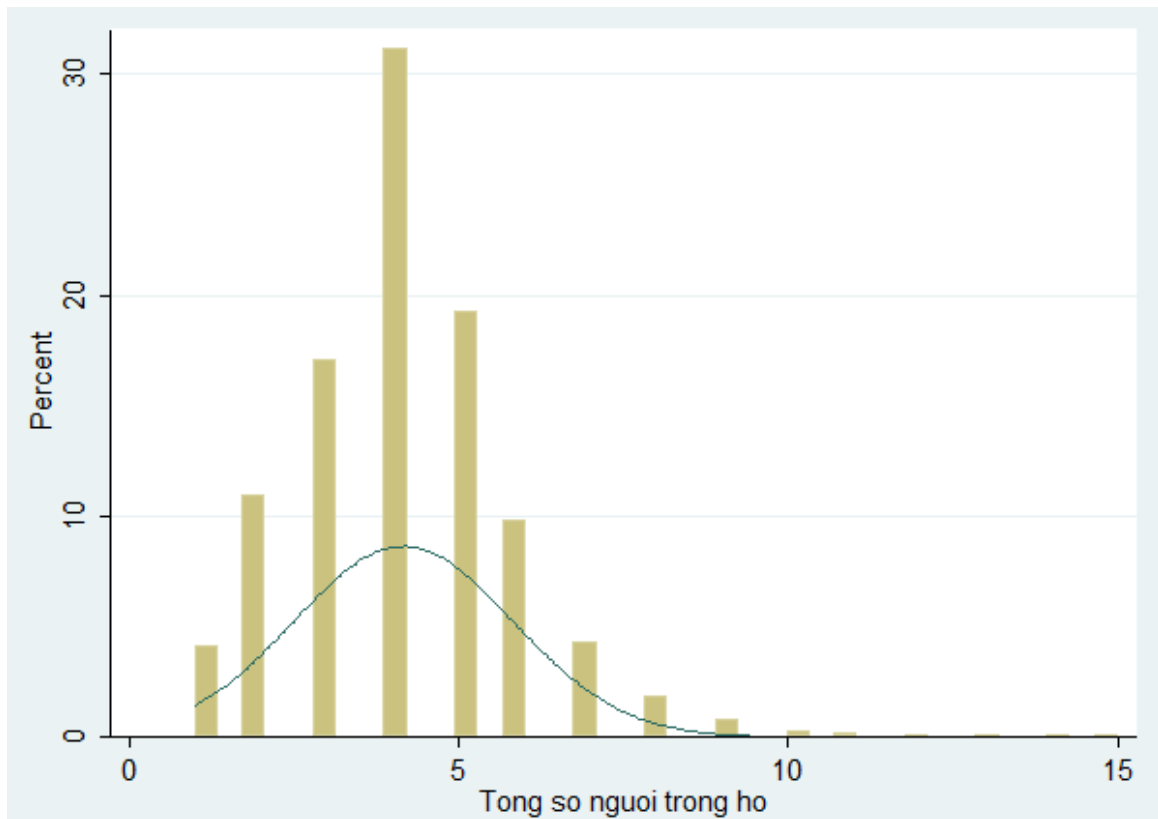
```
label define mkhuvuc 1 urban 2 rural
```

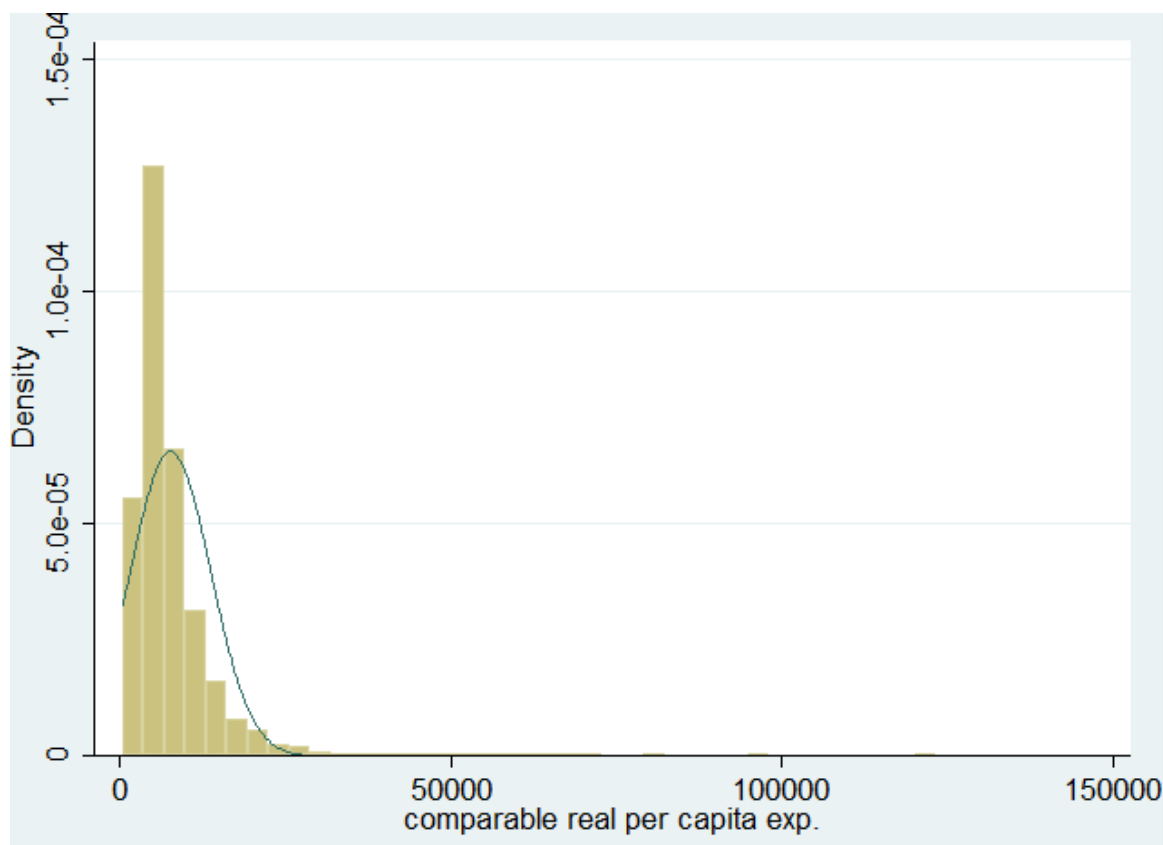
```
label val urban mkhuvuc
```

```
graph bar (mean) foodnom, over(urban) blabel(bar)
```



e. Bạn hãy vẽ đồ thị thể hiện phân phối của biến hsize, foodnom, pcexp1rl và cho biết nhận xét của bạn. Hãy thực hiện những kiểm định cần thiết để đưa ra kết luận xem những biến ấy biến nào có phân phối chuẩn.





Nhìn vào đồ thị, ta có thể thấy rằng phân phối của các biến trên bị lệch phải.

Có nhiều cách khác nhau để kiểm định xem một biến có phân phối chuẩn hay không. Bạn có thể dùng lệnh `swilk`, hay lệnh `sktest`...

```
. swilk hhszpe pcexplrl foodnom
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
hhszpe	9189	0.97867	98.715	12.266	0.00000
pcexplrl	9189	0.70164	1380.727	19.312	0.00000
foodnom	9189	0.82811	795.472	17.839	0.00000

Kết quả kiểm định Shapiro-Wilk cho thấy P-value của mỗi biến đều < 0.05 , nên xét ở độ tin cậy 95%, các biến này đều không có phân phối chuẩn.

f. Bạn hãy tạo biến `lnpcexplrl` bằng `ln(pcexplrl)`, và kiểm định xem biến `lnpcexplrl` có phân phối chuẩn hay không?

```
. gen lnpcexplrl=ln(pcexplrl)
```

```
. swilk lnpcexplrl
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
lnpcexplr1	9189	0.99479	24.092	8.499	0.00000

. tabstat lnpcexplr1, stat (skewness, kurtosis)

variable	skewness	kurtosis
lnpcexplr1	.2837077	3.400326

. sktest lnpcexplr1

Skewness/Kurtosis tests for Normality

Variable	Obs	Pr (Skewness)	Pr (Kurtosis)	adj chi2 (2)	Prob>chi2
lnpcexplr1	9.2e+03	0.0000	0.0000	.	0.0000

Sau khi lấy ln, biến lnpcexplr1 cũng không có phân phối chuẩn.

g. Bạn hãy tạo bảng kết hợp giữa biến reg8 và biến urban08. Cho biết ở Vùng Đồng Bằng Sông Cửu Long, có bao nhiêu hộ được điều tra ở khu vực thành thị, bao nhiêu hộ ở khu vực nông thôn.

. tab reg8 urban08

reg8	Khu vực		Total
	Urban	2	
1	435	1,509	1,944
2	291	1,026	1,317
3	69	360	429
4	168	846	1,014
5	273	579	852
6	165	417	582
7	552	636	1,188
8	399	1,464	1,863
Total	2,352	6,837	9,189

Ở Đồng Bằng Sông Cửu Long, có 1863 hộ được điều tra, trong đó 399 hộ ở thành thị, và 1464 hộ ở nông thôn.