

Chương Trình Giảng Dạy Kinh tế Fulbright

Học kỳ Thu năm 2014

Các Phương Pháp Phân Tích Định Lượng

BÀI TẬP 3

THỐNG KÊ MÔ TẢ

Ngày Phát: Thứ ba 14/10/2014

Ngày Nộp: Thứ ba 21/10/2014

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

Bản điện tử gửi lên <http://www.fetp.edu.vn/vn/tai-nguyen/giang-vien-nhan-vien/>

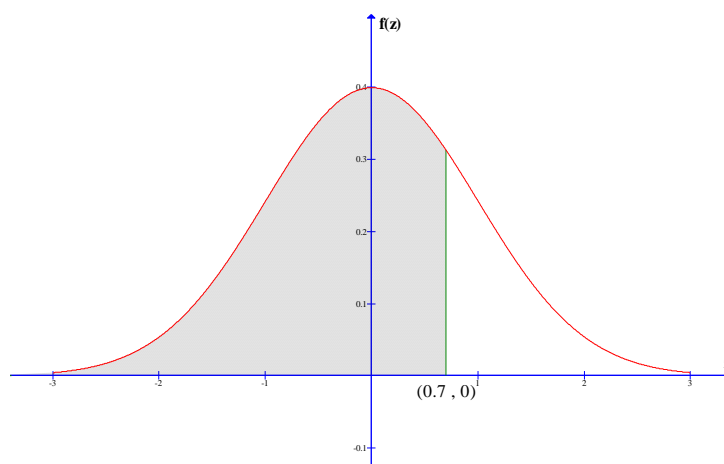
Câu 1: (25 điểm)

Cho Z là một biến ngẫu nhiên tuân theo phân phối chuẩn có $\mu = 0$ và $\sigma = 1$. Phương trình hàm mật độ xác suất của phân phối chuẩn như sau:

$$y = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

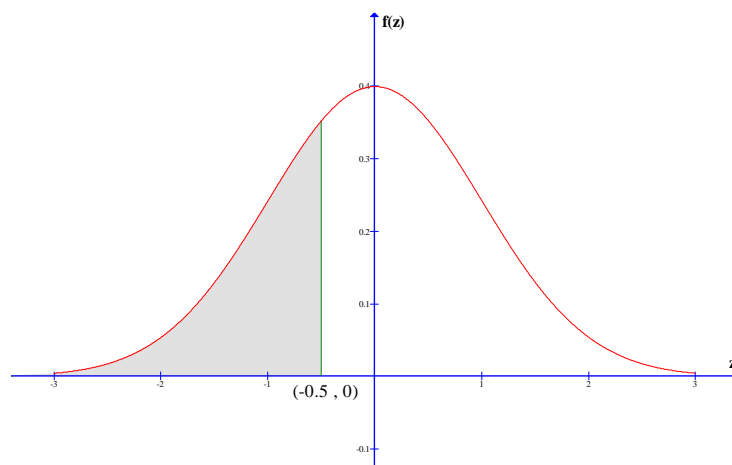
a. *Bằng hình ảnh minh họa, anh chị hãy thể hiện diện tích bên dưới đường cong chuẩn hóa và trên trục hoành? Tính diện tích tương ứng cho từng phần?*

- Bên trái điểm $Z = 0.7$



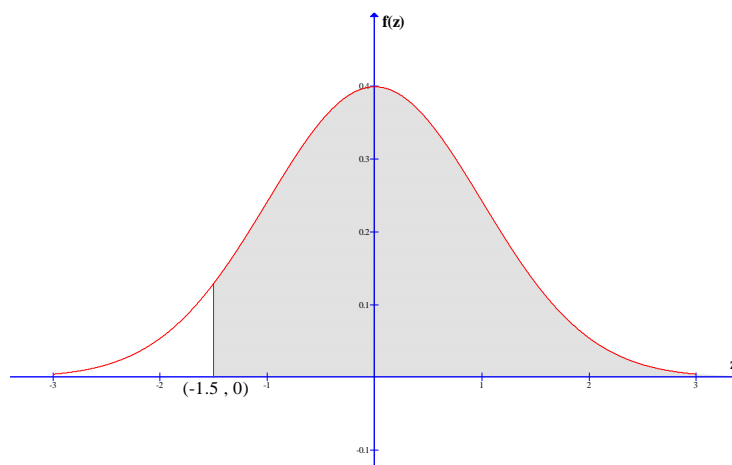
$$S = P(Z < 0.7) = \text{Normsdist}(0.7) = 0.758$$

- Bên trái điểm $Z = -0.5$



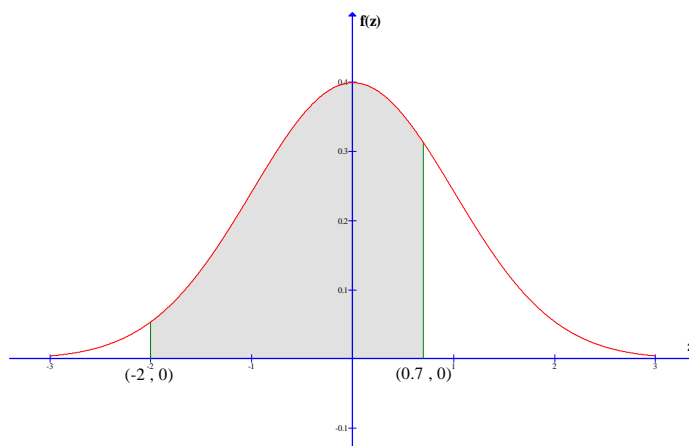
$$S = P(Z < -0.5) = \text{Normsdist}(-0.5) = 0.309$$

- Bên phải điểm $Z = -1.5$



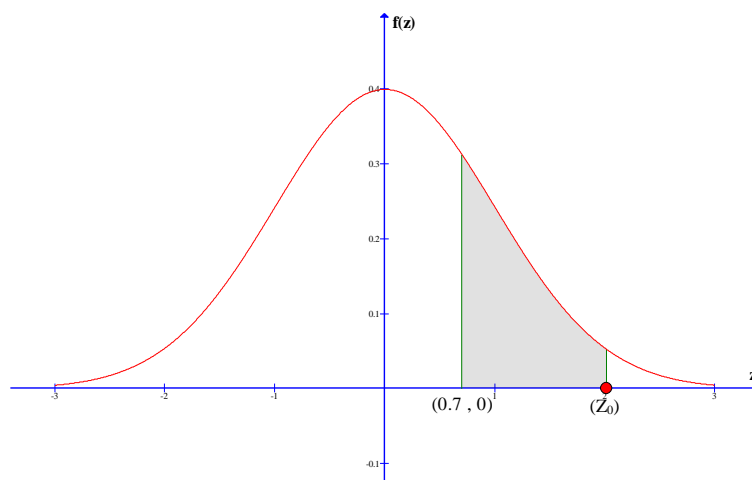
$$S = 1 - P(Z < -1.5) = 1 - \text{Normsdist}(-1.5) = 0.933$$

- Trong khoảng $-2 < Z < 0.7$



$$S = P(Z < 0.7) - P(Z < -2) = \text{Normsdist}(0.7) - \text{Normsdist}(-2) = 0.735$$

b. Tìm Z_0 sao cho $P(0.7 < Z < Z_0) = 22\%$



$$P(0.7 < Z < Z_0) = P(Z < Z_0) - P(Z < 0.7) = P(Z < Z_0) - 0.758 = 22\% = 0.22$$

$$\Rightarrow P(Z < Z_0) = 0.978$$

- Khi dùng hàm Normsinv (0.978) ta có $Z_0 = 2.014$
- Hoặc tra bảng Z với $P(0 < Z < Z_0) = 0.978 - 0.5 = 0.498 \Rightarrow Z_0 = 2.014$

Câu 2: (25 điểm)

Dữ liệu VIC_Stock.xls ghi nhận lịch sử giá cổ phiếu của Tập đoàn Vin-Group (Mã – VIC) từ 19/09/2007 đến 08/10/2014 gồm 1,755 quan sát.

a. Tính giá trị trung bình, trung vị và độ lệch chuẩn cho giá cổ phiếu VIC? Anh chị hãy cho biết phân phối giá cổ phiếu của VIC có bị lệch xiên không? Lệch bên nào, tại sao?

Sử dụng công cụ Descriptive Statistic ta có thống kê mô tả giá cổ phiếu lịch sử của VIC như sau:

Giá đóng cửa - VIC	
Mean	85.088
Standard Error	0.635
Median	78.500
Mode	78.000
Standard Deviation	26.592
Sample Variance	707.116
Kurtosis	1.828
Skewness	1.193
Range	159.700
Minimum	33.300
Maximum	193.000
Sum	149328.600
Count	1755.000

Trung bình (Mean)	85.088 nghìn VND/cp
Trung vị (Median)	78.500 nghìn VND/cp
Độ lệch chuẩn (SD)	26.592 nghìn VND/cp

- Kết quả cho thấy Median = 78.50 (nghìn đồng/cp) nhỏ hơn Mean = 85.08 (nghìn đồng/cp), do đó phân phối giá cổ phiếu của VIC bị lệch xiên và lệch xiên bên phải.

- Nhìn vào hệ số Skewness = 1.19 > 0, cũng cho kết quả tương tự như trên.

b. Lập bảng phân phối xác suất cho giá cổ phiếu VIC phân theo khoảng giá như sau:

$\mu-3\sigma \leftrightarrow \mu-2\sigma$	$\mu-2\sigma \leftrightarrow \mu-\sigma$	$\mu-\sigma \leftrightarrow \mu$	$\mu \leftrightarrow \mu+\sigma$	$\mu+\sigma \leftrightarrow \mu+2\sigma$	$\mu+2\sigma \leftrightarrow \mu+3\sigma$
---	--	----------------------------------	----------------------------------	--	---

Anh chị hãy tính xác suất để giá cổ phiếu của VIC nằm ngoài vùng $\mu \pm 3\sigma$?

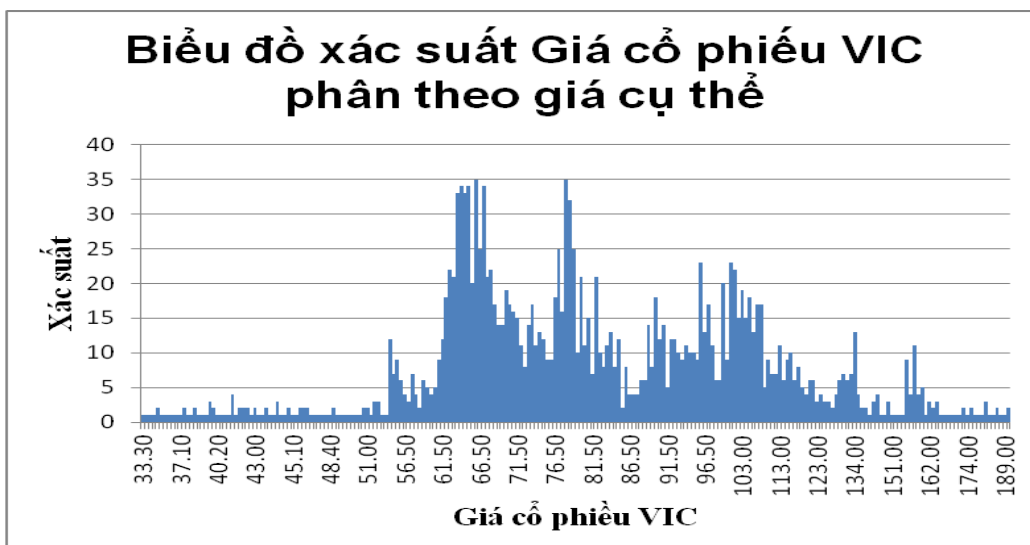
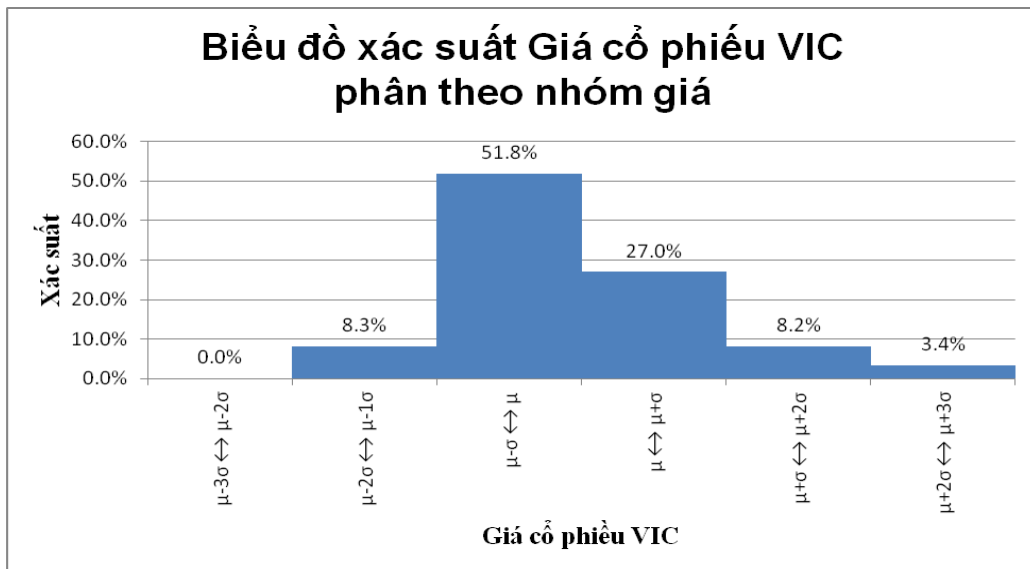
Sử dụng hàm *countif* trong Excel, ta có được bản phân phối xác suất giá cổ phiếu của VIC:

Phân nhóm	$\mu-3\sigma \leftrightarrow \mu-2\sigma$	$\mu-2\sigma \leftrightarrow \mu-\sigma$	$\mu-\sigma \leftrightarrow \mu$	$\mu \leftrightarrow \mu+\sigma$	$\mu+\sigma \leftrightarrow \mu+2\sigma$	$\mu+2\sigma \leftrightarrow \mu+3\sigma$
Số quan sát	0	145	909	473	144	59
Xác suất	0.0%	8.3%	51.8%	27.0%	8.2%	3.4%
Ký hiệu	p1	p2	p3	p4	p5	p6

Xác suất để giá cổ phiếu VIC nằm ngoài vùng $\mu \pm 3\sigma$ là:

$$P(\overline{\mu \pm 3\sigma}) = 1 - \text{Sum}(p1:p6) = 1 - 98.6\% = 1.4\%$$

c. Vẽ đồ thị phân phối xác suất của giá cổ phiếu VIC?



d. **Tính xác suất giá cổ phiếu VIC nằm trong khoảng $\mu \pm \sigma$; $\mu \pm 2\sigma$ và $\mu \pm 3\sigma$?**

$$P(\mu - \sigma < X < \mu + \sigma) = p3 + p4 = 51.8\% + 27\% = 78.8\%$$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = p2 + p3 + p4 + p5 = 8.3\% + 51.8\% + 27\% + 8.2\% = 95.2\%$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = \text{sum}(p1 : p6) = 98.6\%$$

Dựa vào các kết quả ở các câu trước, theo anh/chị, chúng ta có thể kết luận rằng giá cổ phiếu VIC có tuân theo phân phối xác suất chuẩn hay không? Chúng ta có thể kết luận rằng giá cổ phiếu VIC có không tuân theo phân phối xác suất chuẩn hay không? Vì sao?

Việc kiểm định một biến có thể tuân theo phân phối xác suất chuẩn không thể chỉ dựa vào các kết quả ở trên mà phải nhờ công cụ phần mềm khác hỗ trợ và kiểm định giả thuyết.

Tuy nhiên, khi xem xét một số đặc tính (đây chỉ là đặc tính mô tả không phải đặc tính để kiểm định có phải phân phối chuẩn hay không) của phân phối xác suất chuẩn như:

- *Hàm mật độ giá cổ phiếu VIC phải đối xứng qua giá trị trung bình, hay nói cách khác 03 giá trị trung bình, giá trị trung vị và giá trị yếu vị phải bằng nhau.*
- *Xác suất 95,5% các quan sát nằm trong khoảng biến thiên từ $\mu \pm 2\sigma$*
- *Xác suất 99.7% các quan sát nằm trong khoảng biến thiên từ $\mu \pm 3\sigma$*

....

Trong khi đó, kết quả các tính toán ở trên cho thấy mặc dù xác suất giá cổ phiếu VIC biến thiên trong khoảng $\mu \pm 2\sigma$ và $\mu \pm 3\sigma$ lần lượt là 95.2% và 98.6% (xấp xỉ so với 95.5% và 99.9% của phân phối chuẩn) nhưng điều đó không đảm bảo phân phối xác suất giá cổ phiếu của VIC tuân theo phân phối chuẩn bởi các kết quả ở trên đã chỉ ra: **phân phối giá cổ phiếu VIC bị lệch xiên bên phải; 3 giá trị Mean, Mode và Median không trùng nhau. Đồ thị phân phối xác suất giá của cổ phiếu VIC cũng cho thấy tính chất không đối xứng,...** Như vậy, chúng ta hoàn toàn có cơ sở để khẳng định rằng phân phối giá cổ phiếu của VIC không tuân theo quy luật phân phối xác suất chuẩn.

Câu 3: (25 điểm)

X là tổng mức đầu tư (TMĐT) một công trình giao thông. Đơn vị tư vấn đã lập khái toán sơ bộ tổng mức đầu tư của dự án là 10 triệu USD. Công ty A là Chủ đầu tư dự án giao thông trên. Sau khi xem xét chi tiết hồ sơ khái toán của công ty tư vấn, công ty A nhận thấy TMĐT như trên là

hợp lý. Tuy nhiên, vì là công ty có nhiều năm kinh nghiệm trong lĩnh vực thi công, công ty A cho rằng TMĐT thực tế khi thi công của dự án sẽ thay đổi tuân theo quy luật tam giác bất cân xứng với đỉnh tam giác là 10 triệu USD. Nghĩa là khả năng TMĐT của dự án bằng 10 triệu USD là cao nhất nhưng TMĐT có thể biến thiên trong khoảng sau: trong trường hợp xấu nhất, TMĐT của dự án có thể tăng lên khoảng 50% so với dự báo của tư vấn; nếu thi công thuận lợi và quản lý tốt nhất, TMĐT của dự án có thể giảm khoảng 30% so với dự báo của tư vấn. Giả định nhận xét trên của công ty A là chính xác.

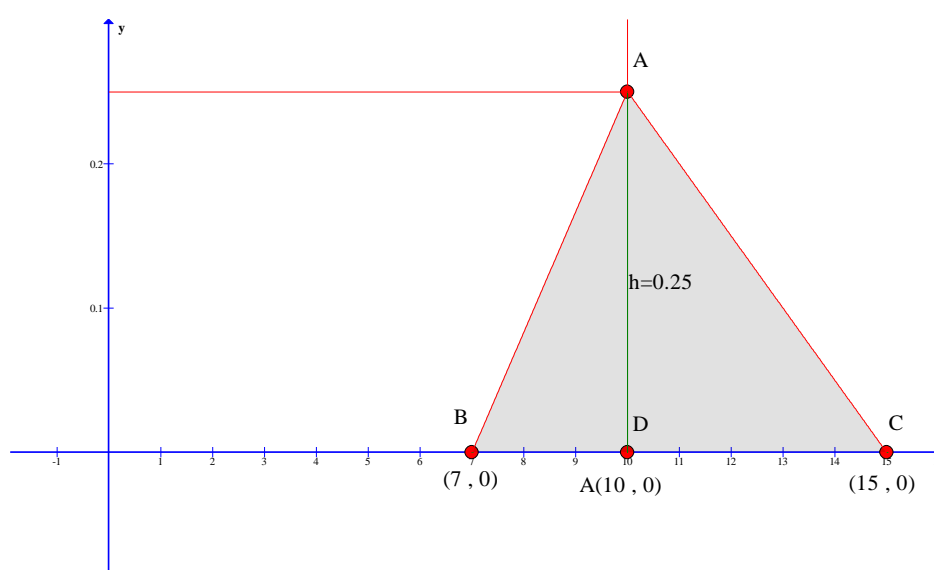
Ta có:

Yếu vị của TMĐT là khả năng xảy ra cao nhất = 10 triệu USD

Khoảng biến thiên của TMĐT nằm trong khoảng từ:

$$[10*(1-30%) , 10*(1+50%)] = [7 , 15] \text{ triệu USD}$$

TMĐT tuân theo quy luật tam giác bất cân xứng, ta có kiểu hình phân phối xác suất của TMĐT chính là tam giá ABC như hình sau:



TMĐT tuân theo quy luật của hàm mật độ xác suất của một biến ngẫu nhiên liên tục, do đó ta có S_{ABC} bằng 1.

$$\Rightarrow \text{Chiều cao } h = S_{ABC} * 2 / BC = 1 * 2 / 8 = 0.25$$

a. Tìm hàm mật độ xác suất của X

Từ đồ thị trên viết được:

- Phương trình đường thẳng $y_1 = \frac{1}{12}X - \frac{7}{12}$ đi qua 2 điểm A và B

- Phương trình đường thẳng $y_2 = -\frac{1}{20}X + \frac{3}{4}$ đi qua 2 điểm A và C

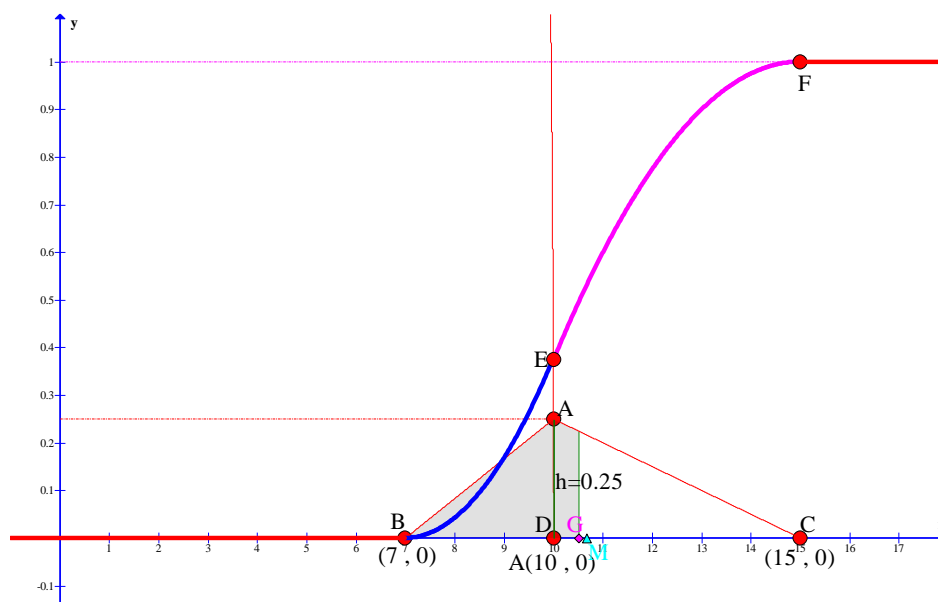
Như vậy, hàm mật độ xác suất của TMDT là:

$$f(x) = \begin{cases} \frac{1}{12}x - \frac{7}{12} & \forall x \in [7;10] \\ -\frac{1}{20}x + \frac{3}{4} & \forall x \in [10;15] \\ 0 & \forall x \notin [7;15] \end{cases}$$

b. Tìm hàm phân phối xác suất tích lũy của X?

$$F(X) = \begin{cases} 0 & \forall X < 7 \\ \frac{1}{24}X^2 - \frac{7}{12}X + \frac{49}{24} & \forall X \in [7,10] \\ -\frac{1}{40}X^2 + \frac{3}{4}X - 5 + 0.375 & \forall X \in [10,15] \\ 1 & \forall X > 15 \end{cases}$$

Hình ảnh minh họa hàm phân phối xác suất tích lũy trong khoảng [7 ; 15] triệu USD của X là kết hợp giữa đường parabol BE và đường parabol EF như hình dưới đây:



c. Xác định trung vị của X?

Đặt G ($G, 0$) là điểm trung vị của X .

Do G là trung vị của X , do vậy sẽ có 50% số quan sát của X nằm bên trái G (nhỏ hơn G) và 50% số quan sát của X nằm bên phải G (lớn hơn G). Do đó xác suất tích lũy tại điểm G hay $F(G)$ sẽ chính là tổng xác suất của 50% quan sát nằm bên trái G :

$$F(G) = \text{Tổng số quan sát tính đến quan sát thứ } G / \text{Tổng số quan sát của } X = 50\% = 0.5$$

$$\text{Vì } F(10) = S_{ABD} = ([10-7]*0.25)/2 = 0.375 \Rightarrow F(10) < F(G) \Rightarrow G > 10$$

Thay G vào phương trình hàm phân phối xác suất tích lũy trong khoảng từ $[10; 15]$ ta có:

$$-\frac{1}{40}G^2 + \frac{3}{4}G - 5 + 0.375 = 0.5$$

Giải phương trình ta có $G = 10.52$ triệu USD. Đây chính là điểm trung vị của X .

d. Xác định trung bình của X?

Gọi M ($M, 0$) là điểm trung bình của X , như vậy ta có:

$$M = E(X) = \int_7^{15} x * f(x) dx = \int_7^{10} x * f_1(x) dx + \int_{10}^{15} x * f_2(x) dx = \int_7^{10} x * y_1 dx + \int_{10}^{15} x * y_2 dx$$

$$\Leftrightarrow M = \int_7^{10} x * \left(\frac{1}{12}x - \frac{7}{12} \right) dx + \int_{10}^{15} x * \left(-\frac{1}{20}x + \frac{3}{4} \right) dx$$

$$\Leftrightarrow M = \int_7^{10} \left(\frac{1}{12}x^2 - \frac{7}{12}x \right) dx + \int_{10}^{15} \left(-\frac{1}{20}x^2 + \frac{3}{4}x \right) dx$$

$$\Leftrightarrow M = \left(\frac{1}{36}x^3 - \frac{7}{24}x^2 \right) \Big|_7^{10} + \left(-\frac{1}{60}x^3 + \frac{3}{8}x^2 \right) \Big|_{10}^{15}$$

$$\Leftrightarrow M = \left(\frac{1}{36}10^3 - \frac{7}{24}10^2 \right) - \left(\frac{1}{36}7^3 - \frac{7}{24}7^2 \right) + \left(-\frac{1}{60}15^3 + \frac{3}{8}15^2 \right) - \left(-\frac{1}{60}10^3 + \frac{3}{8}10^2 \right)$$

$$\Leftrightarrow M = 10.68$$

$$\Leftrightarrow M = \left(\frac{1}{36}10^3 - \frac{7}{24}10^2 \right) - \left(\frac{1}{36}7^3 - \frac{7}{24}7^2 \right) + \left(-\frac{1}{60}15^3 + \frac{3}{8}15^2 \right) - \left(-\frac{1}{60}10^3 + \frac{3}{8}10^2 \right)$$

$$\Leftrightarrow M = 10.68 \text{ triệu USD. Đây chính là điểm trung bình của } X$$

Câu 4: (25 điểm)

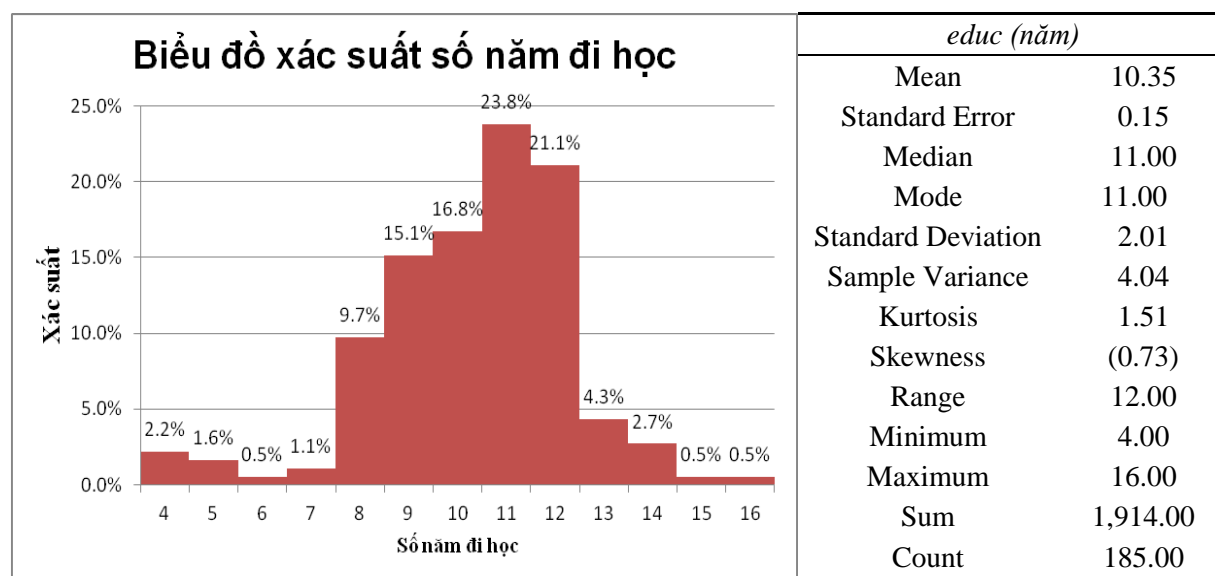
Tập tin Education.xls ghi nhận dữ liệu về trình độ học vấn (số năm đi học tính tại thời điểm khảo sát) của 185 người có hoàn cảnh đặc biệt được tham gia chương trình hỗ trợ việc làm tại Hoa Kỳ (năm 1974 và 1975).

Anh/chị hãy sử dụng tập tin trên và trả lời các câu hỏi sau:

a. Lập bảng phân phối xác suất về trình độ học vấn của tập hợp trên.

Số năm đi học	4	5	6	7	8	9	10	11	12	13	14	15	16	Tổng cộng
Quan sát	4	3	1	2	18	28	31	44	39	8	5	1	1	185
Xác suất	2.2%	1.6%	0.5%	1.1%	9.7%	15.1%	16.8%	23.8%	21.1%	4.3%	2.7%	0.5%	0.5%	100%

b. Vẽ biểu đồ phân phối xác suất trình độ học vấn của nhóm lao động trên.



Sử dụng công cụ Descriptive Statistics trong Excel ta có thống kê mô tả của biến educ như sau:

c. Xác suất để anh/chị chọn được một người có trình độ học vấn từ trung bình trở lên là bao nhiêu?

$$\begin{aligned}
 P(X > 10.35) &= P(X=11) + P(X=12) + P(X=13) + P(X=14) + P(X=15) + P(X=16) \\
 &= 23.8\% + 21.1\% + 4.3\% + 2.7\% + 0.5\% + 0.5\% = 53\%
 \end{aligned}$$

d. Xác suất để anh/chị chọn được một người có trình độ học vấn nhỏ hơn yếu vị?

$$P(X < 11) = 1 - P(X \geq 10) = 1 - P(X \geq 10.35) = 1 - 53\% = 47\%$$

- e. **Anh/chị hãy tính giá trị kỳ vọng trình độ học vấn của những lao động có hoàn cảnh đặc biệt trong khu vực khảo sát?**

$$E(X) = \sum_1^n X_i * p_i = 2.2\% * 4 + 1.6\% * 5 + \dots + 0.5\% * 8 + 0.5\% * 16$$

$$\Leftrightarrow E(X) = 10.35 \text{ (năm).}$$

Vậy giá trị kỳ vọng trình độ học vấn của những lao động có hoàn cảnh đặc biệt trong khu vực khảo sát là 10.35 (năm).