

Chương Trình Giảng Dạy Kinh tế Fulbright

Học kỳ Thu năm 2014

Các Phương Pháp Phân Tích Định Lượng

ĐÁP ÁN BÀI TẬP 1

THỐNG KÊ MÔ TẢ

Ngày Phát: Thứ ba 30/09/2014

Ngày Nộp: Thứ ba 07/10/2014

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

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

Bài 1: (25 điểm)

Dữ liệu thống kê một số chỉ số vĩ mô của Việt Nam và Hoa Kỳ trong 10 năm trở lại đây được trình bày như bảng dưới đây:

Năm	Tỷ giá VND/USD	Lạm phát VND	Lạm phát USD
2004	15,778	9.82%	3.21%
2005	15,915	8.82%	3.68%
2006	16,051	6.70%	2.20%
2007	16,030	12.63%	4.08%
2008	17,486	19.89%	1.46%
2009	18,495	6.58%	1.91%
2010	19,500	11.75%	1.70%
2011	21,036	18.13%	3.07%
2012	20,855	6.81%	1.78%
2013	21,125	6.03%	1.24%

Nguồn: IMF (2014) và VCB (2014)

- a. Dựa vào định nghĩa và công thức, Anh/chị hãy tính các giá trị trung bình, trung vị, nhỏ nhất, lớn nhất, khoảng biến thiên, phương sai, độ lệch chuẩn và hệ số biến thiên của lạm phát VND và lạm phát Hoa Kỳ. (8đ)

- Giá trị trung bình

$$\overline{\Pi}_{VND} = \frac{\sum_{i=1}^n \Pi_{VND}(i)}{N} = \frac{(9.82\% + 8.82\% + \dots + 6.81\% + 6.03\%)}{10} = 10.72\%$$

$$\overline{\Pi}_{USD} = \frac{\sum_{i=1}^n \Pi_{USD}(i)}{N} = \frac{(3.21\% + 3.68\% + \dots + 1.78\% + 1.24\%)}{10} = 2.43\%$$

- Giá trị trung vị: để tính giá trị trung vị, hãy sắp xếp các giá trị trong mỗi biến theo thứ tự tăng dần hoặc giảm dần. Gọi n là số giá trị có trong biến X bất kỳ. Nếu n là số chẵn thì giá trị trung vị của biến X được chọn là giá trị trung bình của giá trị thứ $(n/2)-1$ và giá trị thứ $(n/2)+1$ của biến X . Cụ thể trong trường hợp này là giá trị thứ 5 và thứ 6 như trong bảng sau:

Bảng sắp xếp lạm phát VND và USD theo thứ tự tăng dần:

STT	Lạm phát VND tăng dần	Lạm phát USD tăng dần
1	6.03%	1.24%
2	6.58%	1.46%
3	6.70%	1.70%
4	6.81%	1.78%
5	8.82%	1.91%
6	9.82%	2.20%
7	11.75%	3.07%
8	12.63%	3.21%
9	18.13%	3.68%
10	19.89%	4.08%

Như vậy áp dụng định nghĩa trên ta có:

$$\text{Median}\Pi_{\text{VND}} = \frac{\Pi_{\text{VND}(5)} + \Pi_{\text{VND}(6)}}{2} = \frac{8.82\% + 9.82\%}{2} = 9.32\%$$

$$\text{Median}\Pi_{\text{USD}} = \frac{\Pi_{\text{USD}(5)} + \Pi_{\text{USD}(6)}}{2} = \frac{1.91\% + 2.20\%}{2} = 2.06\%$$

- Min, Max, Range: dựa vào bảng sắp xếp thứ tự lạm phát tăng dần ở trên, ta có:

$\text{Min}\Pi_{\text{VND}} = 6.03\%$	$\text{Min}\Pi_{\text{USD}} = 1.24\%$
$\text{Max}\Pi_{\text{VND}} = 19.89\%$	$\text{Max}\Pi_{\text{USD}} = 4.08\%$
$\text{Range}\Pi_{\text{VND}} = 19.89\% - 6.03\% = 13.86\%$	$\text{Range}\Pi_{\text{USD}} = 4.08\% - 1.24\% = 2.84\%$

- Phương sai (Var), độ lệch chuẩn (SD)

$$\text{Var}_{\Pi_{\text{VND}}} = \frac{\sum_{i=1}^n (\Pi_{\text{VND}(i)} - \overline{\Pi_{\text{VND}}})^2}{n-1} = \frac{(6.03\% - 10.72\%)^2}{9} + \dots + \frac{(19.89\% - 10.72\%)^2}{9} = 0.0024$$

$$\text{SD}_{\Pi_{\text{VND}}} = \sqrt{\text{Var}_{\Pi_{\text{VND}}}} = \sqrt{0.0024} = 4.93\%$$

$$\text{Var}_{\Pi_{\text{USD}}} = \frac{\sum_{i=1}^n (\Pi_{\text{USD}(i)} - \overline{\Pi_{\text{USD}}})^2}{n-1} = \frac{(1.24\% - 2.43\%)^2}{9} + \dots + \frac{(4.08\% - 2.43\%)^2}{9} = 0.0001$$

$$\text{SD}_{\Pi_{\text{USD}}} = \sqrt{\text{Var}_{\Pi_{\text{USD}}}} = \sqrt{0.0001} = 1\%$$

- Hệ số biến thiên (CV)

$$CV_{\Pi_{VND}} = \frac{SD_{\Pi_{VND}}}{\Pi_{VND}} = \frac{4.93\%}{10.72\%} = 0.46$$

$$CV_{\Pi_{USD}} = \frac{SD_{\Pi_{USD}}}{\Pi_{USD}} = \frac{1.00\%}{2.43\%} = 0.41$$

- b. Dựa vào các hàm trong Excel, anh/chị hãy tính toán lại các giá trị đã tính ở câu a.**

(8đ)

Tên hàm	Chỉ tiêu	Lạm phát VND	Lạm phát USD
Average	Trung bình	10.72%	2.43%
Median	Trung vị	9.32%	2.06%
Min	Nhỏ nhất	6.03%	1.24%
Max	Lớn nhất	19.89%	4.08%
Range	Khoảng biến thiên	13.86%	2.84%
Var	Phương sai	0.24%	0.01%
Stdev	Độ lệch chuẩn	4.93%	1.00%
CV=Stdev*100/Average	Hệ số biến thiên	46.00	40.96

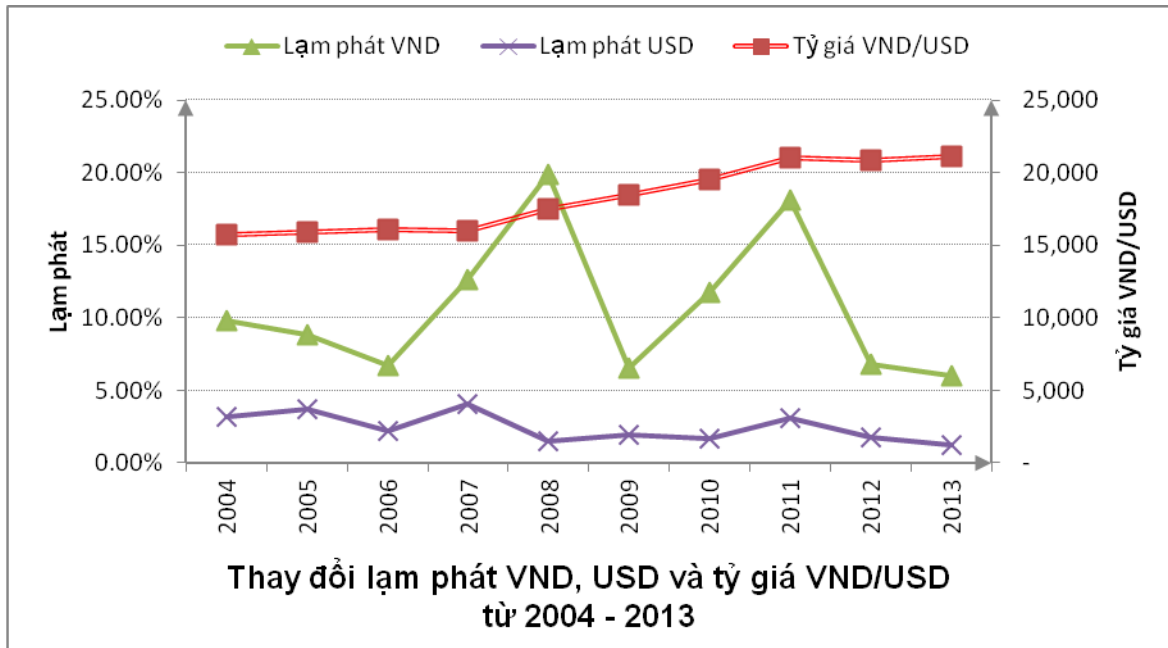
- c. Dựa vào các giá trị đã tính toán ở trên, anh/chị hãy cho biết nhận xét của mình về lạm phát của 2 nước? (3đ)**

Kết quả cho thấy lạm phát VND hằng năm đều cao hơn lạm phát USD (lạm phát nhỏ nhất của VND (6.03%) vẫn lớn hơn lạm phát cao nhất của USD (4.08%); theo đó, lạm phát trung bình VND và lạm phát USD lần lượt là 10.72% và 2.43%.

Kết quả cũng cho thấy độ lệch chuẩn của lạm phát VND là 4.93% cao hơn rất nhiều so với độ lệch chuẩn của lạm phát USD (tương tự như khoảng biến thiên). Tuy nhiên, điều này không phản ánh chính xác lạm phát của quốc gia nào có độ biến thiên nhiều hơn do lạm phát trung bình của 2 nước khác nhau. Để biết được lạm phát VND hay USD có độ biến thiên nhiều hơn, ta xem xét hệ số biến thiên (CV) giữa 2 quốc gia. Kết quả cho thấy hệ số biến thiên của lạm phát VND (0.46) lớn hơn độ biến thiên của lạm phát USD (0.41). Do đó, ta có thể kết luận, lạm phát VND có độ biến thiên cao hơn so với lạm phát USD.

- d. Vẽ biểu đồ dạng line (line chart) biểu diễn sự biến thiên lạm phát VND, USD và tỷ giá VND/USD trong cùng 1 đồ thị. Nhìn vào đồ thị, anh/chị có nhận xét gì về xu**

hướng của tỷ giá VND/USD trong 10 năm trở lại đây? Giải thích ngắn gọn xu hướng trên dựa vào các thông tin đã cung cấp trong câu 1. (6đ)



Kết quả đồ thị cho thấy sự biến thiên lạm phát VND và USD khác nhau, tuy nhiên tỷ giá VND/USD trong 10 năm trở lại đây (2004 – 2013) có xu hướng tăng dần. Điều này là do lạm phát VND luôn cao hơn lạm phát USD, đồng tiền VND ngày càng mất giá so với đồng tiền USD. Do đó, tỷ giá VND/USD sẽ có khuynh hướng tăng dần để phản ánh sự mất giá của đồng tiền VND so với đồng tiền USD (nếu không có can thiệp từ chính sách của Chính phủ).

Câu 2. (25đ)

Tập tin BabyMilk.xlsx ghi nhận dữ liệu về một số mặt hàng sữa bột dùng cho trẻ em trong độ tuổi từ 0 – 36 tháng tuổi đang bán tại thị trường Việt Nam. Số liệu ghi nhận từ website <http://sua.vn> (cập nhật ngày 22/09/2014)

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

- a. Hãy phân biệt các biến trên: định tính hay định lượng, loại thang đo của từng biến? (5đ)**

STT	Tên biến	Loại biến	Thang đo
1	Quy cách đóng gói	Định tính	Thứ bậc, định danh
2	Công ty	Định tính	Định danh

3	Tên sản phẩm	Định tính	Định danh
4	Lứa tuổi	Định tính	Thứ bậc, định danh
5	Giá bán	Định lượng	Tỷ lệ

b. Theo anh chị, công ty nào có số lượng sản phẩm sữa cho lứa tuổi từ 0 – 6 tháng tuổi nhiều nhất và ít nhất? Số sản phẩm sữa cho lứa tuổi nào nhiều nhất trong số các lứa tuổi ở trên? (8đ)

Sử dụng công cụ Pivot Table của Excel ta có bảng thống kê sau như sau:

Count of Tên sản phẩm Công ty	Lứa tuổi			Tổng
	0-6 tháng	12-36 tháng	6-12 tháng	
Abbott	7	5	5	17
Dumex	1		1	2
Friso	6	3	3	12
Nestle	3	3	3	9
NutiFood	2	3	2	7
Vinamilk	3	6	3	12
Tổng	22	20	17	59

Kết quả cho thấy công ty có số lượng sản phẩm sữa cho lứa tuổi từ 0-6 tháng tuổi nhiều nhất là Abbott với 7 sản phẩm, công ty có số lượng sản phẩm sữa cho lứa tuổi từ 0-6 tháng tuổi ít nhất là Dumex với 1 sản phẩm. Số sản phẩm sữa nhiều nhất là dành cho lứa tuổi từ 0 – 6 tháng tuổi với 22 sản phẩm.

c. Anh/chị chuẩn bị đến chơi một gia đình người thân, gia đình này có một bé trai 18 tháng tuổi, anh/chị muốn mua sữa làm quà. Với số tiền dự định mua quà là 300,000 VND, anh chị có bao nhiêu lựa chọn sản phẩm (giả định rằng ngoài các sản phẩm đã được liệt kê trong file Babymilk.xlsx, anh/chị không còn lựa chọn nào khác). (6đ)

Tạo một biến mới có tên “Mua quà” (bất kỳ tên gì) với các giá trị trong biến thỏa mãn 2 điều kiện:

- Giá bán nhỏ hơn hoặc bằng 300,000 VND
- Sản phẩm thuộc lứa tuổi 12 – 36 tháng tuổi

Để tìm các giá trị cho biến “Mua quà” vừa tạo ra, ta dùng hàm “IF AND” như sau:

$IF(AND(Giá bán_i = < 300000, Lứa tuổi_i = “12-36 tháng”), Giá bán_i, “ ”)$

Kết quả những dòng nào thỏa 2 điều kiện trên sẽ bằng chính giá bán sản phẩm dòng đó và dòng nào không thỏa điều kiện trên sẽ nhận giá trị ô trống.

Dùng hàm “Count” để đếm số giá trị trong biến “Mua quà”.

Kết quả có 13 giá trị tương ứng với 13 sự lựa chọn để anh/chị ra quyết định mua sản phẩm nào phù hợp với 2 tiêu chí trên.

Lưu ý: Anh/chị có thể dùng công cụ khác của Excel để tính toán như filter, sum, count, countif, ...

d. Tính các giá trị trung bình, trung vị, nhỏ nhất, lớn nhất, khoảng biến thiên, số yếu vị của giá sữa phân theo nhóm đóng gói 400g. (6đ)

Dùng lệnh Filter để sắp xếp giá trị biến “Quy cách đóng gói” thành nhóm 400g và 900g.

Dùng một số hàm thống kê, ta có một số kết quả thống kê cho nhóm 400g như sau:

<i>Hàm</i>	<i>Chỉ tiêu</i>	Giá trị (VND/sản phẩm)
Average	Trung bình	191,893
Median	Trung vị	194,500
Min	Nhỏ nhất	72,000
Max	Lớn nhất	285,000
Max - Min	Khoảng biến thiên	213,000
Mode*	Yếu vị	128,000

**Trong trường hợp này tập dữ liệu về giá sữa nhóm 400g có 03 giá trị yếu vị. Khi sử dụng hàm Mode của Excel (với tập dữ liệu lớn), trong trường hợp có nhiều giá trị yếu vị, excel sẽ chọn giá trị xuất hiện trước tiên trong số nhiều giá trị yếu vị để trả kết quả. Để xử lý và kiểm tra trường hợp này ta phải dùng công cụ PivotTable để đếm tần suất xuất hiện của từng giá trị, sau đó sắp xếp dãy tần suất đó theo thứ tự từ lớn đến nhỏ, chúng ta sẽ biết có bao nhiêu giá trị yếu vị.*

Câu 3 (25đ)

Dữ liệu từ tập tin WEO_Report_2014.xlsx ghi nhận ước tính GDP của 189 quốc gia và vùng lãnh thổ trên Thế giới (tính theo giá hiện hành và quy đổi sang cùng 1 đơn vị là USD). Số liệu ghi nhận từ Báo cáo kinh tế Thế giới, số tháng 04 năm 2014.

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

a. Anh chị hãy tính các giá trị GDP trung bình, trung vị. (4đ)

Sử dụng hàm Excel, ta có kết quả sau:

Hàm	Chỉ tiêu	Giá trị (tỷ USD)
Avarage	Trung bình	393.52
Median	Trung vị	30.80

b. Dựa vào giá trị tính toán ở câu a, theo anh chị phân phối GDP các nước có bị lệch xiên hay không? Nếu có, sử dụng hàm Excel để xác định độ lệch xiên và xác định lệch xiên bên trái hay bên phải? (6đ)

Ta thấy GDP trung bình của các nước (393.52 tỷ USD) lớn hơn so với giá trị GDP trung vị (30.80 tỷ USD). Như vậy, phân phối GDP các nước bị lệch xiên bên phải.

Kết quả tính toán lại bằng hàm “Skew” trong Excel cho thấy độ lệch xiên (Skewness) của GDP các nước là $8.34 > 0$. Do đó phân phối GDP các nước bị lệch xiên bên phải.

Ngoài ra, chúng ta còn nhận thấy giá trị trung vị rất nhỏ so với giá trị trung bình. Điều này là do giá trị trung bình thì phụ thuộc vào giá trị của từng quan sát trong tập dữ liệu quan sát, trong khi giá trị trung vị chỉ quan tâm đến thứ tự của các quan sát trong tập dữ liệu. Với việc giá trị trung vị nhỏ hơn rất nhiều so với giá trị trung bình cho thấy phần lớn các nước có GDP thấp hơn GDP trung bình, trong khi có rất ít các quốc gia có GDP lớn hơn giá trị trung bình nhưng lại tập trung phần lớn của cải của thế giới (28 quốc gia có GDP lớn hơn GDP trung bình).

c. Tìm giá trị chuẩn hóa Z cho giá trị quan sát lớn nhất và nhỏ nhất của GDP. Các giá trị này có lớn hay nhỏ bất thường không? (6đ)

Sử dụng công cụ Description Statistics ta có một số giá trị thống kê như sau:

Giá trị trung bình - \overline{GDP} (tỷ USD)	Mean	393.52
Độ lệch chuẩn - SD_{GDP} (tỷ USD)	Standard Deviation	1505.28
Giá trị lớn nhất - GDP_{Min} (tỷ USD)	Minimum	0.04
Giá trị nhỏ nhất - GDP_{Max} (tỷ USD)	Maximum	16,799.70

Như vậy:

- Giá trị chuẩn hóa Z cho giá trị quan sát lớn nhất và nhỏ nhất lần lượt là:

$$Z(GDP_{Max}) = \frac{GDP_{Max} - \overline{GDP}}{SD_{GDP}} = \frac{16,799 - 393.52}{1,505.28} = 10.90$$

$$Z(GDP_{Min}) = \frac{GDP_{Min} - \overline{GDP}}{SD_{GDP}} = \frac{0.04 - 393.52}{1,505.28} = -0.26$$

Kết quả cho thấy, với giá trị quan sát lớn nhất ($GDP_{USA} = 16,799$ tỷ USD) thì giá trị chuẩn hóa $Z = 10.90$ lớn hơn rất nhiều so với 3 (tương ứng giá trị trung bình + 3 độ lệch chuẩn). Do vậy đây là một giá trị bất thường. Hay nói cách khác, GDPUSA quá cao so với các nước trên TG.

Với giá trị quan sát nhỏ nhất ($GDP_{TUVALU} = 0.04$ tỷ USD) thì giá trị chuẩn hóa $Z = -0.26$ nên vẫn nằm trong khoảng 3 độ lệch chuẩn so với trung bình. Do vậy, đây không phải là một giá trị bất thường. Hay nói cách khác, có rất nhiều các quốc gia cũng có GDP thấp như GDPTUVALU.

d. Anh chị hãy tính tổng GDP của 19 quốc gia có GDP thấp nhất? Tổng GDP của 19 quốc gia có GDP cao nhất? Tỷ lệ % so với tổng GDP của 189 quốc gia. Anh chị có nhận xét gì về kết quả này? (9đ)

Sử dụng kết hợp hàm Percentile trong Excel để tính số phân vị (p-10%) và (p-90%) để tính giá trị GDP mà tại đó có 10% số nước có GDP nhỏ hơn hoặc bằng (p-10%); giá trị GDP tại đó mà 10% số nước sẽ có GDP lớn hơn hoặc bằng (p-90%). Sau đó dùng hàm countif để tính tổng GDP của 19 quốc gia có GDP thấp nhất, tổng GDP của 19 quốc gia có GDP cao nhất.

Chỉ tiêu	Giá trị	Ghi chú
Tổng GDP các quốc gia	73,982.14	Tỷ USD
19 quốc gia có GDP thấp nhất:		(10% tổng số quốc gia)
<i>Tổng GDP</i>	<i>12.05</i>	Tỷ USD
<i>% / Tổng GDP các nước</i>	<i>0.016%</i>	
19 quốc gia có GDP cao nhất:		(10% tổng số quốc gia)
<i>Tổng GDP</i>	<i>58,507.33</i>	Tỷ USD
<i>% / Tổng GDP các nước</i>	<i>79.08%</i>	

Kết quả cho thấy 10% các nước có GDP thấp nhất chỉ chiếm 0.016% tổng GDP 188 nước được ghi nhận. Trong khi đó tỷ lệ này ở nhóm 10% các nước có GDP cao nhất chiếm đến 79.08% tổng GDP các nước. Điều này cho thấy phần lớn tài sản trên thế giới đang tập trung vào những cường quốc lớn như Mỹ, Trung Quốc, Nhật Bản, Đức, Pháp, Anh, Braxin, Nga...

Câu 4 (25đ)

Dữ liệu Employee Data.xls lưu trữ dữ liệu của 474 người lao động ở nhiều công ty khác nhau trong ngành Viễn Thông. Dữ liệu bao gồm các biến cụ thể như sau:

age	Tuổi
gender	Giới tính
jobcat	Vị trí công việc
educ	Số năm đi học
prevexp	Số tháng đã đi làm
salary	Lương hiện tại/năm (USD)
salbegin	Lương khởi điểm/năm (USD)

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

- a. Anh/chị hãy tính trung bình và trung vị, yếu vị, phương sai và độ lệch chuẩn cho biến biến educ và biến salary? Theo anh/chị 2 biến này có mối tương quan với nhau không? Nếu có, mối tương quan đó là như thế nào? (7đ)**

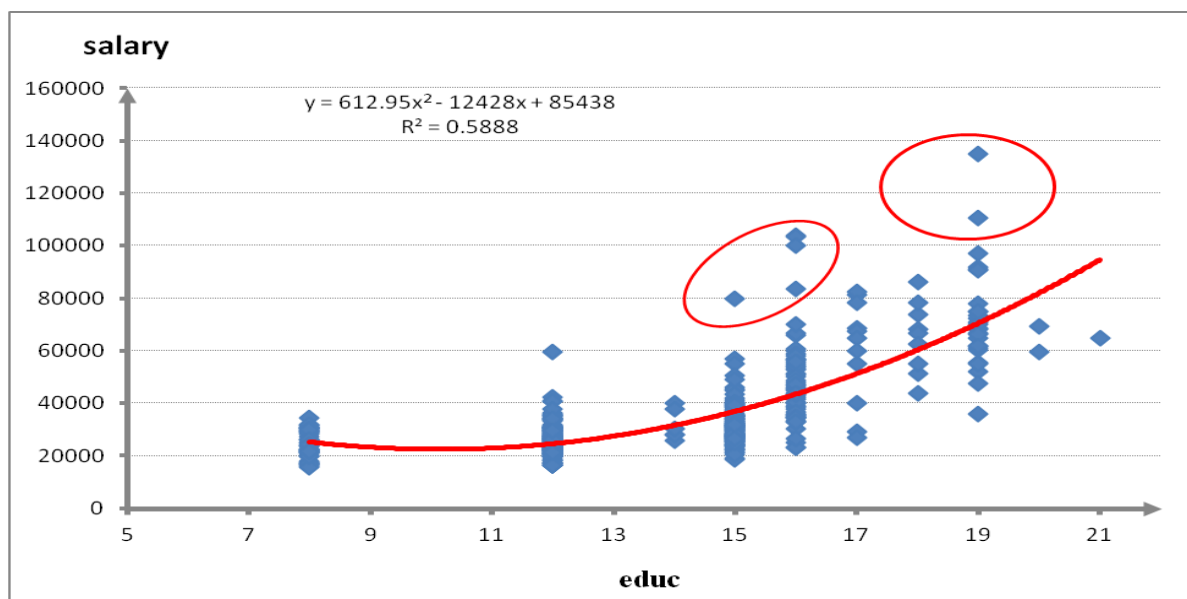
Sử dụng công cụ Descriptive Statistics trong Data\Data Analysis ta có kết quả một số chỉ tiêu thống kê mô tả như bảng sau:

Biến	Trung bình	Trung vị	Yếu vị	Độ lệch chuẩn	Phương sai
Variables	Average	Median	Mode	Standard Deviation	Sample Variance
educ	13.49	12.00	12.00	2.88	8.32
salary	34,420	28,875	30,750	17,076	291,587,214

Đây là dữ liệu về các lao động làm việc trong ngành viễn thông, do vậy, trình độ người lao động (trong đó có trình độ đào tạo hay học vấn) là một trong những tiêu chí quan trọng để đánh giá chất lượng lao động và ảnh hưởng trực tiếp đến tiền lương của người lao động. Mặc dù vậy, không phải lúc nào người có trình độ đào tạo cao cũng là người lao động giỏi, để trở thành lao động giỏi cần kết hợp nhiều yếu tố khác như kỹ năng, sự thích nghi,... Do vậy. Theo tôi, biến **educ** sẽ có tương quan với biến **salary** và là **tương quan thuận**.

- b. Vẽ biểu đồ phân tán (Scatter) thể hiện mối quan hệ giữa 2 biến educ và salary. Kết quả từ đồ thị có ủng hộ cho lập luận của anh/chị hay không? Hãy chỉ ra một vài điểm dẫn chứng. (6đ)**

Kết quả vẽ biểu đồ Scatter cho kết quả như sau:



Kết quả nhìn vào đường xu hướng cho thấy 02 biến **educ** và **salary** có mối quan hệ đồng biến, Tuy nhiên, bên cạnh đó vẫn xuất hiện một số điểm dị biệt cho thấy những người có số năm đi học thấp hơn nhưng lương lại cao hơn và ngược lại. Điều này sẽ làm giảm mức độ tương quan giữa 2 biến. Do vậy, mức độ tương quan giữa 2 biến chỉ ở mức độ vừa phải.

- c. **Tính các giá trị đồng phương sai và hệ số tương quan giữa 2 biến educ và salary. Anh/chị có nhận xét gì về mối quan hệ của 2 biến này từ các hệ số tính được và so sánh nó với nhận xét ở các câu trên? (6đ)**

Kết quả tính toán bằng công thức và hàm trong Excel cho kết quả như sau:

Hàm	Chỉ tiêu	Giá trị	Ghi chú
Covar	Đồng phương sai tổng thể	32,471	(N) quan sát
Covar*N/(N-1)	Đồng phương sai mẫu	32,540	(N-1) quan sát
Correl	Hệ số tương quan	0.66	Cách 1
Sample Covar/(Sd_educ * Sd_salary)		0.66	Cách 2

Kết quả tính toán cho thấy đồng phương sai mẫu = 32,540 > 0, do đó 2 biến **educ** và **salary** có mối quan hệ tuyến tính đồng biến. Hệ số tương quan giữa 2 biến là 0.66 > 0, xét về giá trị tuyệt đối ở mức trung bình khá so với 1. Do vậy, dữ liệu ghi nhận cho thấy mối quan hệ tuyến tính thuận tương đối giữa 2 biến này như đã nhận định ở trên.

d. Căn cứ vào các biến có trong tập dữ liệu, theo anh chị, những biến nào có thể có liên quan đến biến salary? Giải thích ngắn gọn và tính hệ số tương quan giữa biến salary với biến mà anh/chị đề xuất. Kết luận của anh/chị nếu có là gì? (6đ)

Ngoài biến educ, theo tôi, tất cả các biến trên có thể có liên quan đến biến **salary** của một người lao động bao gồm các biến sau: **jobcat** (người càng có chức vụ, lương càng cao); **salbegin** (hầu hết quá trình tăng lương đều nằm trong quy chế chung, ngoại trừ một số trường hợp đặc biệt, do đó, lương khởi điểm đóng vai trò quan trọng, tạo ra sự khác biệt về lương giữa các lao động). Đối với biến **gender**, khó có khả năng tác động bởi sự bình đẳng giới ngày càng rõ nét. Biến **age** và biến **prevexp** cũng rất khó tạo ra sự khác biệt bởi mỗi công việc có một đặc thù riêng, có công việc cần kinh nghiệm nhưng cũng có công việc cần sự năng động và nét mới,...

Kết quả tính toán hệ số tương quan giữa biến salary và một số biến cho kết quả như sau:

Hệ số tương quan giữa biến salary và biến:

1	age	-0.15	Quan hệ nghịch biến, tương quan yếu
2	salbegin	0.88	Quan hệ đồng biến, tương quan mạnh
3	prevexp	-0.14	Quan hệ nghịch biến, tương quan yếu
4	educ	0.66	Quan hệ đồng biến, tương quan khá

Lưu ý: biến gender và biến jobcat là 2 biến định tính nên không thể tính hệ số tương quan với biến salary (hay hệ số tương quan tính được không có ý nghĩa về mặt thống kê).

Kết quả cho thấy, hai biến **salbegin** và **educ** là những biến có sự tương quan đồng biến ở mức độ khá cao với biến **salary** (trị tuyệt đối hệ số tương quan giữa 2 biến này với biến **salary** tiến về gần 1 hơn). Các biến còn lại là **age** và **prevexp** có tương quan nghịch biến nhưng mức độ tương quan ở mức trung bình và yếu. Điều này có thể giải thích như sau: cả 2 biến trên đều có đặc trưng cho kinh nghiệm làm việc, tuy nhiên, mức lương như đã thấy phụ thuộc lớn vào lương khởi điểm ($r = 0.88$), do vậy, khả năng những người có lương cao là những người mới vào công ty nên mức lương khởi điểm của họ theo mức lương thị trường hiện tại nên cao hơn những người đã ở công ty thời gian dài (lương khởi điểm thấp và mức độ tăng lương không đáng kể). Tuy nhiên, nhưng người có chức vụ lại thường là những người làm lâu năm, do vậy lại có lương cao hơn. Chính 2 điều mâu thuẫn này làm cho hệ số tương quan của 2 biến age và prevexp tiến gần về 0.

Lưu ý:

Các phần nhận xét và giải thích trong đáp án chỉ mang tính chất tương đối, các anh/chị có thể có các cách lập luận khác nhau từ kết quả tính toán, cơ sở lý thuyết và các nhận định chủ quan mà anh chị cho rằng phù hợp và logic.