

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 1

THỐNG KÊ MÔ TẢ

Ngày Phát: Thứ hai 01/10/2012

Ngày Nộp: Thứ hai 8/10/2012

Bản in nộp tại Hộp nộp bài tập trong phòng Lab

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

Bài 1: (25 điểm)

Mẫu thống kê tỷ lệ sản phẩm lỗi trên dây chuyền may của 2 nhà máy X và Y trong 10 ngày liên tục được cho như sau:

Ngày	1	2	3	4	5	6	7	8	9	10
Nhà máy A	6%	1%	5%	8%	10%	12%	0%	4%	5%	11%
Nhà máy B	4%	5%	8%	10%	6%	8%	9%	13%	3%	5%

- Tính các giá trị trung bình, trung vị, khoảng cho tỷ lệ sản phẩm lỗi ở mỗi nhà máy
- Tìm phương sai, độ lệch chuẩn cho tỷ lệ sản phẩm lỗi ở mỗi nhà máy
- Từ các kết quả trên, theo Anh/Chị nhà máy nào vận hành tốt hơn và ổn định hơn, giải thích.

Đáp án Bài 1

- Sắp xếp các giá trị theo thứ tự tăng dần.

											Tổng	Trung bình
Nhà máy A	0%	1%	4%	5%	5%	6%	8%	10%	11%	12%	62%	6.2%
Nhà máy B	3%	4%	5%	5%	6%	8%	8%	9%	10%	13%	71%	7.1%

Từ bảng trên ta có các thông số cần tìm như sau:

	Nhà máy A	Nhà máy B
Trung bình	6.2%	7.1%
Trung vị	5.5%	7.0%
Khoảng	12.0%	10.0%

- Sử dụng bảng dưới đây để tính toán phương sai và độ lệch chuẩn:

Nhà máy A			Nhà máy B		
X_i	$(X_i - \text{tb})$	$(X_i - \text{tb})^2$	X_i	$(X_i - \text{tb})$	$(X_i - \text{tb})^2$
0	-6.20%	0.003844	3%	-4.10%	0.001681
1%	-5.20%	0.002704	4%	-3.10%	0.000961
4%	-2.20%	0.000484	5%	-2.10%	0.000441
5%	-1.20%	0.000144	5%	-2.10%	0.000441
5%	-1.20%	0.000144	6%	-1.10%	0.000121
6%	-0.20%	4E-06	8%	0.90%	8.1E-05
8%	1.80%	0.000324	8%	0.90%	8.1E-05
10%	3.80%	0.001444	9%	1.90%	0.000361
11%	4.80%	0.002304	10%	2.90%	0.000841
12%	5.80%	0.003364	13%	5.90%	0.003481
phương sai		0.00164			0.000943
độ lệch chuẩn		0.040497			0.030714

- c. Nhà máy A vận hành tốt hơn nhà máy B vì giá trị trung bình và trung vị trong số sản phẩm lỗi của nhà máy A thấp hơn các chỉ số này của nhà máy B.
 Tuy nhiên, nhà máy B sẽ ổn định hơn nhà máy A vì khoảng và hệ số biến thiên của nhà máy B thấp hơn của nhà máy A.

Bài 2: (25 điểm)

Khi nghiên cứu mức thu nhập của các hộ gia đình tại hai thành phố A và B, người ta thu được tập dữ liệu sau:

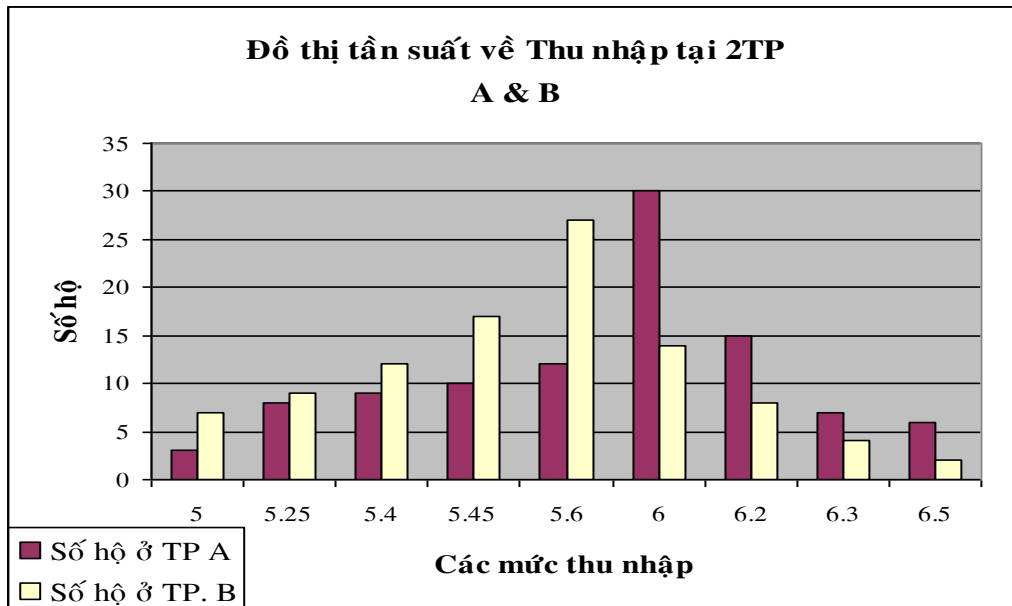
Thu nhập hàng tháng (ngàn đồng)	Số hộ ở TP. A	Số hộ ở TP. B
5.000	3	7
5.250	8	9
5.400	9	12
5.450	10	17
5.600	12	27
6.000	30	14
6.200	15	8
6.300	7	4
6.500	6	2
Cộng	100	100

Hãy:

- Vẽ đồ thị tần suất mô tả thu nhập của nhóm hộ gia đình tại hai thành phố A và B.
- Xác định thu nhập trung bình của hai nhóm hộ gia đình tại hai thành phố A và B. Xác định thêm số yếu vị, min, max cho từng nhóm và đưa ra nhận xét.

Đáp án Bài 2

- Vẽ đồ thị



b.

(Đơn vị: ngàn đồng)	Nhóm TP. A	Nhóm TP. B
Thu nhập trung bình	5.834	5.627
Thu nhập trung vị	5.600	5.600
Thu nhập yếu vị	6.000	5.600
Thu nhập max	6.500	6.500
Thu nhập min	5.000	5.000

Nhận xét: (10 điểm)

Hai nhóm có cùng các giá trị: thu nhập max, min và trung vị, nhưng nhóm TP.B có giá trị thu nhập trung bình và yếu vị nhỏ hơn nhóm TP. A; điều đó cho ta *cảm nhận ban đầu* là thu nhập hộ gia đình ở TP. B nhìn chung thấp hơn TP. A.

Bài 3: (25 điểm)

Bạn hãy sử dụng file dữ liệu **Employee Data.xls** về 474 người lao động. File này có các biến sau:

Id	Số thứ tự
Age	Tuổi
Gender	1: Nam, 2: Nữ
Jobcat (vị trí công việc)	1: Nhân viên, 2: Trưởng nhóm/Tổ trưởng/Quản đốc, 3: Quản lý
Educ	Số năm đi học
Prevexp	Số tháng đã đi làm
Salary	Tiền lương hiện tại trong một năm (USD)
Salbegin	Tiền lương khởi điểm trong một năm (USD)

- Phân biệt loại biến và thang đo cho 4 biến số Age, Gender, Jobcat và Salary
- Hãy phân nhóm người lao động theo vị trí công việc. Vẽ biểu đồ tần số theo vị trí công việc.
- Hãy phân nhóm người lao động theo giới tính. Vẽ biểu đồ tần số theo giới tính.
- Tìm các giá trị cực tiểu, cực đại, phân vị 25%, phân vị 50%, phân vị 75%, trung bình, trung vị và yếu vị của biến salary.

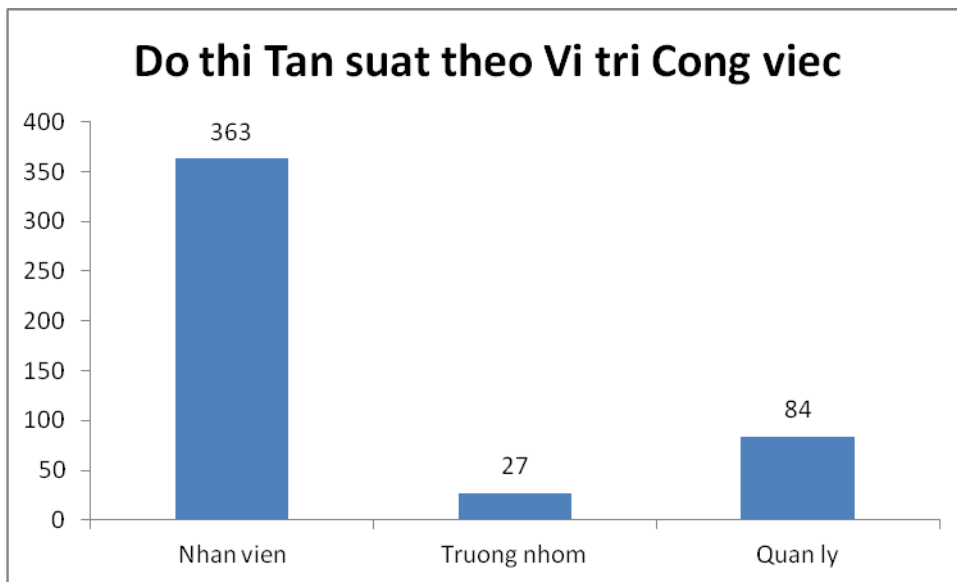
e. Có bao nhiêu người lao động ở vị trí quản lý và có số tuổi từ 40 đến 49

Đáp án Bài 3

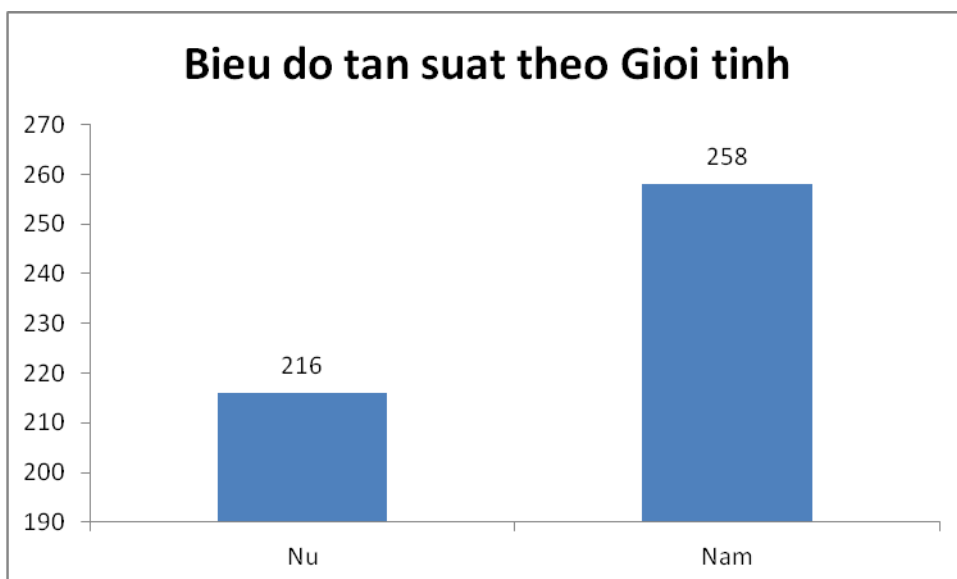
a. Phân biệt loại biến và thang đo

Biến	Phân loại	Thang đo
Age	Định lượng	Tỷ lệ
Gender	Định tính	Chi danh
Jobcat	Định tính	Thứ tự
Salary	Định lượng	Tỷ lệ

b. Nêu cách phân nhóm người lao động theo vị trí công việc, vẽ đồ thị



c. Nêu cách phân nhóm theo giới tính (2 điểm) và vẽ đồ thị (3 điểm)



d. Tìm các thông số theo yêu cầu

Trung bình	34,419.56
Trung vị	28,875.00
Yếu vị	30,750.00
Cực tiểu	15,750.00
Cực đại	135,000.00
Q ₁ (P25%)	24,000.00
Q ₂ (P25%)	28,875.00
Q ₃ (P25%)	36,937.50

e. Nêu rõ cách làm để có được số người ở vị trí quản lý có độ tuổi từ 40 đến 49: 13 người

Bảng Pivot table

Count of id	age					Grand Total
	20-29	30-39	40-49	50-59	60-69	
jobcat						
1	184	68	40	55	16	363
2		4	8	9	6	27
3	27	42	13	2		84
Grand Total	211	114	61	66	22	474

Bài 4: (25 điểm)

Tập tin Chi Phí Bảo Trì.xls ghi nhận dữ liệu về biến Y là chi phí bảo trì và biến X là số Km sử dụng của 57 xe ô tô.

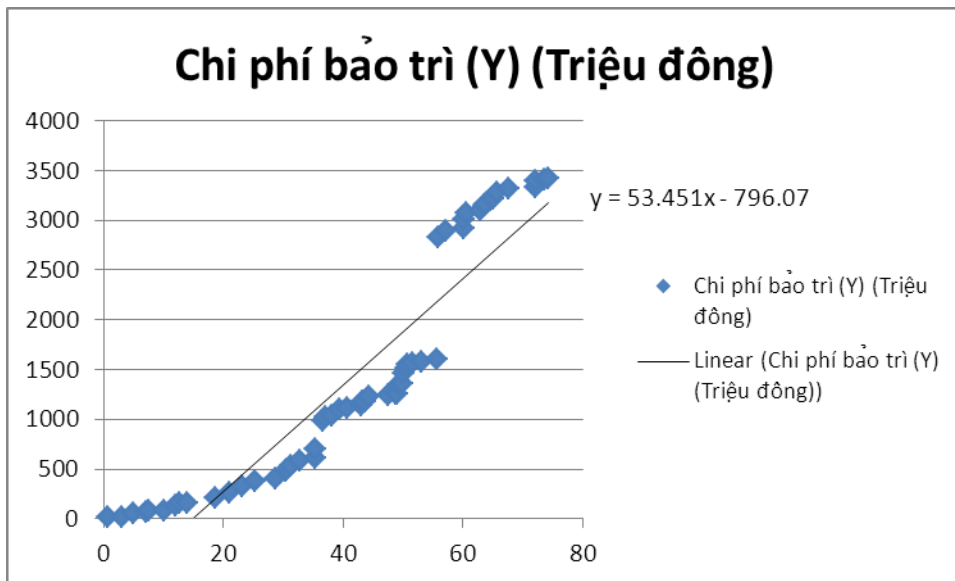
- Sử dụng Excel tính các giá trị thống kê để có thể mô tả 2 biến X và Y.
- So sánh các giá trị thống kê mô tả của biến X ở câu a, Anh/Chị có kết luận phân phối biến X bị lệch xiên hay không. Nếu có thì bị lệch về bên phải hay bên trái. Giải thích ngắn gọn câu trả lời của Anh/Chị.
- Không cần tính toán, theo nhận định của Anh/Chị 2 biến X, Y có mối quan hệ tuyến tính hay không. Nếu có thì mối quan hệ này đồng biến hay nghịch biến, mạnh hay yếu. Giải thích ngắn gọn về nhận định của Anh/Chị.
- Vẽ biểu đồ phân tán điểm thể hiện mối quan hệ giữa hai biến X và Y. Dựa trên đồ thị này, nhận xét mối quan hệ giữa 2 biến này. Nhận xét này có phù hợp với nhận định của Anh/Chị ở câu c hay không.
- Sử dụng Excel tính hệ số tương quan giữa 2 biến Y và X. Dựa trên hệ số tương quan này, nhận xét mối quan hệ giữa 2 biến này. So sánh nhận xét này với câu c và d.

Đáp án Bài 4

a. Tính toán các trị thống kê mô tả hai biến X và Y

	Chi phí bảo trì (Y) (ngàn km)	Số Km sử dụng (X) (triệu đồng)
Mean	1426.63	41.58
Standard Error	152.48	2.64
Median	1182	43.4
Mode	1182	35.3
Standard Deviation	1151.19	19.96
Sample Variance	1325229.49	398.22
Kurtosis	-1.04	-0.71
Skewness	0.59	-0.37
Range	3414	73.6
Minimum	11	0.8
Maximum	3425	74.4
Sum	81318	2370.3
Count	57	57

- b. Biến X có giá trị trung bình khác với giá trị trung vị nên phân phối bị lệch xiên. Cụ thể giá trị trung bình của X nhỏ hơn giá trị trung vị nên phân phối bị lệch xiên về bên trái.
- c. Hai biến X và Y có thể có quan hệ tuyến tính đồng biến với mối tương quan mạnh vì khi X tăng thì Y tăng, tuy nhiên tốc độ tăng của X nhanh hơn tốc độ tăng của Y.
- d. Theo đồ thị thì tương quan giữa hai biến X và Y là đồng biến. Cũng dựa trên đồ thị này, sự phụ thuộc giữa X và Y rất rõ ràng (xu hướng mạnh) do đó mối tương quan này khá mạnh. Các nhận định này hoàn toàn phù hợp với nhận định ở câu c.



e. Dùng phần mềm Excel ta tính được hệ số tương quan giữa X và Y là 0.92. Hệ số này mang dấu dương chứng tỏ hai biến này có quan hệ đồng biến, giá trị của chỉ số này rất gần 1 chứng tỏ hai biến có tương quan mạnh.