

## CÁC PHƯƠNG PHÁP ĐỊNH LƯỢNG 1

### LỜI GIẢI GỢI Ý BÀI TẬP 2

Ngày phát: 09/11/2023

Hạn nộp: 8h20, 14/11/2023

**Bài làm được yêu cầu chỉ nộp bản điện tử trên Microsoft Teams**

-----

**Câu 1 (20đ):** Thu nhập hàng tháng của của một cặp vợ chồng được thể hiện bằng hai biến ngẫu nhiên  $X$  và  $Y$  có hàm phân phối như bảng dưới đây:

$f(x; y)$		$y$	
		30	40
$x$	10	0.20	0.15
	20	0.05	0.30
	30	0.20	0.10

Với  $X$ : thu nhập chồng,  $Y$ : thu nhập vợ. Đơn vị: triệu VNĐ.

Hãy tính các hàm và giá trị sau đây:

- a. Hàm phân phối xác suất biên và hàm phân phối xác suất có điều kiện cho từng thành viên.

**Hàm phân phối xác suất biên:**

$f(x; y)$		$y$		
		30	40	
$x$	10	0.20	0.15	<b>0.35</b>
	20	0.05	0.30	<b>0.35</b>
	30	0.20	0.10	<b>0.30</b>
		<b>0.45</b>	<b>0.55</b>	

**Hàm phân phối xác suất có điều kiện thu nhập của chồng khi biết thu nhập của vợ:**

$$P(X = 10|Y = 30) = \frac{P(X = 10; Y = 30)}{P(Y = 30)} = \frac{0.20}{0.45} = \frac{4}{9} = 0.4444$$

$$P(X = 20|Y = 30) = \frac{P(X = 20; Y = 30)}{P(Y = 30)} = \frac{0.05}{0.45} = \frac{1}{9} = 0.1111$$

$$P(X = 30|Y = 30) = \frac{P(X = 30; Y = 30)}{P(Y = 30)} = \frac{0.20}{0.45} = \frac{4}{9} = 0.4444$$

$$P(X = 10|Y = 40) = \frac{P(X = 10; Y = 40)}{P(Y = 40)} = \frac{0.15}{0.55} = \frac{3}{11} = 0.273$$

$$P(X = 20|Y = 40) = \frac{P(X = 20; Y = 40)}{P(Y = 40)} = \frac{0.30}{0.55} = \frac{6}{11} = 0.545$$

$$P(X = 30|Y = 40) = \frac{P(X = 30; Y = 40)}{P(Y = 40)} = \frac{0.10}{0.55} = \frac{2}{11} = 0.182$$

	X=10	X=20	X=30
$P(X Y = 30)$	0.444	0.111	0.444
$P(X Y = 40)$	0.273	0.545	0.182

Hàm phân phối xác suất có điều kiện thu nhập của vợ khi biết thu nhập của chồng:

$$P(Y = 30|X = 10) = \frac{P(X = 10; Y = 30)}{P(X = 10)} = \frac{0.20}{0.35} = \frac{4}{7} = 0.571$$

$$P(Y = 40|X = 10) = \frac{P(X = 10; Y = 40)}{P(X = 10)} = \frac{0.15}{0.35} = \frac{3}{7} = 0.429$$

$$P(Y = 30|X = 20) = \frac{P(X = 20; Y = 30)}{P(X = 20)} = \frac{0.05}{0.35} = \frac{1}{7} = 0.143$$

$$P(Y = 40|X = 20) = \frac{P(X = 20; Y = 40)}{P(X = 20)} = \frac{0.30}{0.35} = \frac{6}{7} = 0.857$$

$$P(Y = 30|X = 30) = \frac{P(X = 30; Y = 30)}{P(X = 30)} = \frac{0.20}{0.30} = \frac{2}{3} = 0.667$$

$$P(Y = 40|X = 30) = \frac{P(X = 30; Y = 40)}{P(X = 30)} = \frac{0.10}{0.30} = \frac{1}{3} = 0.333$$

	Y=30	Y=40
$P(Y X = 10)$	0.571	0.429
$P(Y X = 20)$	0.143	0.857
$P(Y X = 30)$	0.667	0.333

- b. Hàm giá trị kỳ vọng (trung bình kỳ vọng) có điều kiện của vợ (Y) có phụ thuộc vào thu nhập hằng tháng của chồng (X) hay không?

Giá trị kỳ vọng có điều kiện của chồng:

$$E(X|Y = 30) = \frac{4}{9} \times 10 + \frac{1}{9} \times 20 + \frac{4}{9} \times 30 = 20$$

$$E(X|Y = 40) = \frac{3}{11} \times 10 + \frac{6}{11} \times 20 + \frac{2}{11} \times 30 = 19.09$$

Giá trị kỳ vọng có điều kiện của vợ:

$$E(Y|X = 10) = \frac{4}{7} \times 30 + \frac{3}{7} \times 40 = 34.29$$

$$E(Y|X = 20) = \frac{1}{7} \times 30 + \frac{6}{7} \times 40 = 38.57$$

$$E(Y|X = 30) = \frac{2}{3} \times 30 + \frac{1}{3} \times 40 = 33.33$$

Từ kết quả trên, giá trị kỳ vọng (trung bình kỳ vọng) có điều kiện của vợ (Y) có phụ thuộc vào thu nhập hằng tháng của chồng (X)

- c. Giá trị kỳ vọng không có điều kiện và phương sai cho thu nhập vợ, chồng ( $E(X), E(Y), \sigma_X, \sigma_Y$ ).

Giá trị kỳ vọng không có điều kiện cho thu nhập của chồng (X):

$$E(X) = \sum_{i=1}^n x_i P_i = 0.35 \times 10 + 0.35 \times 20 + 0.30 \times 30 = 19.5$$

Giá trị kỳ vọng không có điều kiện cho thu nhập của vợ (Y):

$$E(Y) = \sum_{i=1}^n y_i P_i = 0.45 \times 30 + 0.55 \times 40 = 35.5$$

Phương sai thu nhập của chồng (X):

$$\begin{aligned} \text{Var}(X) &= \sum_{i=0}^n (x - E(X))^2 P(X = x_i) \\ &= (10 - 19.5)^2 \times 0.35 + (20 - 19.5)^2 \times 0.35 + (30 - 19.5)^2 \times 0.30 = 64.75 \end{aligned}$$

$$\text{Độ lệch chuẩn } \sigma_X = \sqrt{\text{Var}(X)} = \sqrt{64.75} = 8.047$$

Phương sai thu nhập của vợ (Y):

$$\begin{aligned} \text{Var}(Y) &= \sum_{i=0}^3 (y_i - E(Y))^2 P(Y = y_i) = (30 - 35.5)^2 \times 0.45 + (40 - 35.5)^2 \times 0.55 \\ &= 24.75 \end{aligned}$$

$$\text{Độ lệch chuẩn } \sigma_Y = \sqrt{\text{Var}(Y)} = \sqrt{24.75} = 4.975$$

d. Hiệp phương sai và hệ số tương quan.

Hiệp phương sai:

$$\begin{aligned} \text{cov}(X, Y) &= 0.20 \times (10 - 19.5) \times (30 - 35.5) + 0.15 \times (10 - 19.5) \times (40 - 35.5) \\ &\quad + 0.05 \times (20 - 19.5) \times (30 - 35.5) \\ &\quad + 0.30 \times (20 - 19.5) \times (40 - 35.5) \\ &\quad + 0.20 \times (30 - 19.5) \times (30 - 35.5) \\ &\quad + 0.10 \times (30 - 19.5) \times (40 - 35.5) = -2.25 \end{aligned}$$

Hệ số tương quan:

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{-2.25}{8.047 \times 4.975} = -0.0562$$

e (\*). Theo bảng phân phối này, thu nhập của hai vợ chồng có phải là hai biến ngẫu nhiên độc lập không?

$f(x; y)$		$y$		
		30	40	
$x$	10	0.20	0.15	<b>0.35</b>
	20	0.05	0.30	<b>0.35</b>
	30	0.20	0.10	<b>0.30</b>
		<b>0.45</b>	<b>0.55</b>	

Ta có:  $f_X(x) f_Y(y) = 0.35 \times 0.45 = 0.1575$

Vì  $f_X(x) f_Y(y) \neq f(x; y)$  nên hai biến ngẫu nhiên này không độc lập

**Câu 2 (20đ):** Một công ty khảo sát thu thập được số lần sử dụng website chính phủ trong một năm từ 20 người ngẫu nhiên như sau:

15, 11, 13, 12, 9, 11, 13, 14, 10, 11, 12, 8, 13, 11, 12, 12, 11, 14, 9, 10

Hãy trả lời những câu hỏi sau đây:

a. Lập bảng giá trị tần suất cho mẫu này.

Giá trị	8	9	10	11	12	13	14	15
Tần suất	1	2	2	5	4	3	2	1
Tỷ lệ	0.05	0.1	0.1	0.25	0.2	0.15	0.1	0.05

b. Tính giá trị trung bình, phương sai và độ lệch chuẩn cho mẫu quan sát này.

Giá trị trung bình của mẫu:

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{8 + 9 \times 2 + 10 \times 2 + 11 \times 5 + 12 \times 4 + 13 \times 3 + 14 \times 2 + 15}{20} = 11.55$$

Phương sai của mẫu:

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} = \frac{(8 - 11.55)^2 + 2 \times (9 - 11.55)^2 + \dots + (15 - 11.55)^2}{20} = 3.313$$

Độ lệch chuẩn:  $\hat{\sigma} = \sqrt{s^2} = \sqrt{3.313} = 1.82$

c. Tính giá trị trung vị, yếu vị (mode) và độ trải (range) của mẫu này.

Giá trị trung vị là 11.5 (chia đôi tập dữ liệu: 4 giá trị nhỏ hơn và 4 giá trị lớn hơn)

Giá trị yếu vị là 11 vì có tần suất xuất hiện trong mẫu lớn nhất (5 lần)

Độ trải (khoảng cách từ giá trị lớn nhất đến giá trị nhỏ nhất) là 7

d. Tính điểm tứ phân vị thứ nhất và thứ ba của mẫu này (the first and third quartiles).

Sắp xếp các giá trị trong mẫu theo thứ tự tăng dần ta có:

8, 9, 9, 10, 10, 11, 11, 11, 11, 11, 12, 12, 12, 12, 13, 13, 13, 14, 14, 15

Vị trí của tứ phân vị thứ nhất là:

$$Q_1 = 25 \times \frac{n + 1}{100} = 25 \times \frac{20 + 1}{100} = 5.25$$

Vị trí thứ 5 và 6 trong mẫu có giá trị lần lượt là 10 và 11

→ Giá trị tứ phân vị thứ nhất là  $Q_1 = 10 + 0.25 \times (11 - 10) = 10.25$

Vị trí của tứ phân vị thứ ba là:

$$Q_3 = 75 \times \frac{n + 1}{100} = 75 \times \frac{20 + 1}{100} = 15.75$$

Vị trí thứ 15 và 16 trong mẫu là giá trị 13 và 13

→ Giá trị tứ phân vị thứ nhất là  $Q_3 = 13 + 0.75 \times (13 - 13) = 13$

**Câu 3 (20đ):** Cho bảng thu thập doanh thu (R) và chi phí quảng cáo (C) của một mẫu rất nhỏ gồm 6 doanh nghiệp. Hãy tính giá trị hiệp phương sai và hệ số tương quan mẫu. Hệ số này có thay đổi hay không nếu ta đổi đơn vị doanh thu về triệu VNĐ?

R (tỷ VNĐ)	2	3	3	4	5	6
C (triệu VNĐ)	25	60	35	50	50	80

Giá trị trung bình mẫu của doanh thu:

$$\bar{R} = \frac{2 + 3 + 3 + 4 + 5 + 6}{6} = 3.833 \text{ (tỷ VNĐ)}$$

Giá trị trung bình mẫu của chi phí quảng cáo:

$$\bar{C} = \frac{25 + 60 + 35 + 50 + 50 + 80}{6} = 50 \text{ (triệu VNĐ)}$$

Phương sai và độ lệch chuẩn mẫu của doanh thu:

$$S_R^2 = \frac{\sum_{i=1}^6 (R_i - \bar{R})^2}{n - 1} = \frac{(2 - 3.833)^2 + (3 - 3.833)^2 + (3 - 3.833)^2 + (4 - 3.833)^2 + (5 - 3.833)^2 + (6 - 3.833)^2}{5} = 2.167$$

$$\hat{\sigma}_R = S_R = \sqrt{2.167} = 1.472$$

$$S_C^2 = \frac{\sum_{i=1}^6 (C_i - \bar{C})^2}{n - 1} = \frac{(25 - 50)^2 + (60 - 50)^2 + (35 - 50)^2 + (50 - 50)^2 + (50 - 50)^2 + (80 - 50)^2}{5} = 370$$

$$\hat{\sigma}_C = S_C = \sqrt{370} = 19.235$$

Hiệp phương sai mẫu:

$$\widehat{cov}(R, C) = \frac{\sum_{i=1}^n (R_i - \bar{R})(C_i - \bar{C})}{n - 1} = \frac{(2 - 3.833) * (25 - 50) + (3 - 3.833) * (60 - 50) + (3 - 3.833) * (35 - 50) + (4 - 3.833) * (50 - 50) + (5 - 3.833) * (50 - 50) + (6 - 3.833) * (80 - 50)}{6 - 1} = 23$$

Hệ số tương quan mẫu:

$$\hat{\rho}_{RC} = \frac{\widehat{cov}(R, C)}{S_R S_C} = \frac{23}{1.472 * 19.235} = 0.812$$

Học viên thực hiện tính toán tương tự khi đổi đơn vị doanh thu về triệu VNĐ

Giá trị trung bình mẫu của doanh thu:  $\bar{R} = 3833.33$

Phương sai và độ lệch chuẩn mẫu của doanh thu:  $S_R^2 = 2166666.68$

$$\hat{\sigma}_R = S_R = 1471.96$$

Hiệp phương sai mẫu:  $\widehat{cov}(R, C) = 23000$

Hệ số tương quan mẫu:

$$\hat{\rho}_{RC} = \frac{\widehat{cov}(R, C)}{S_R S_C} = \frac{23000}{1471.96 * 19.235} = 0.812$$

Như vậy, việc thay đổi đơn vị của doanh thu không làm thay đổi hệ số tương quan mẫu.

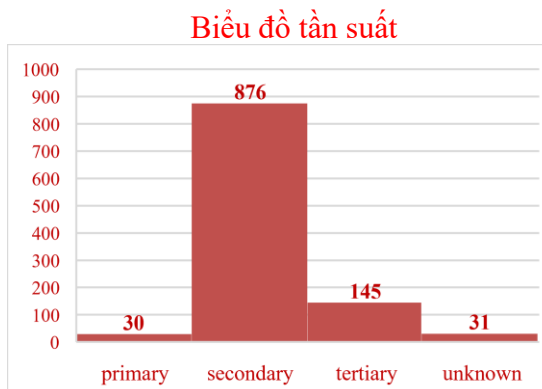
**Câu 4 (40đ):** File dữ liệu *Assignment 2.xlsx* cung cấp một mẫu bao gồm một số thông tin về khách hàng tại một ngân hàng ở Chile. Dữ liệu ở sheet *Data* và mô tả biến cung cấp ở sheet *Description*.

Hãy sử dụng Excel để hoàn thành các phần sau đây:

a. Tính tần suất và vẽ biểu đồ tần suất cho biến *education*.

Thực hiện thao tác trên Excel, ta có bảng tần suất cho biến *education*:

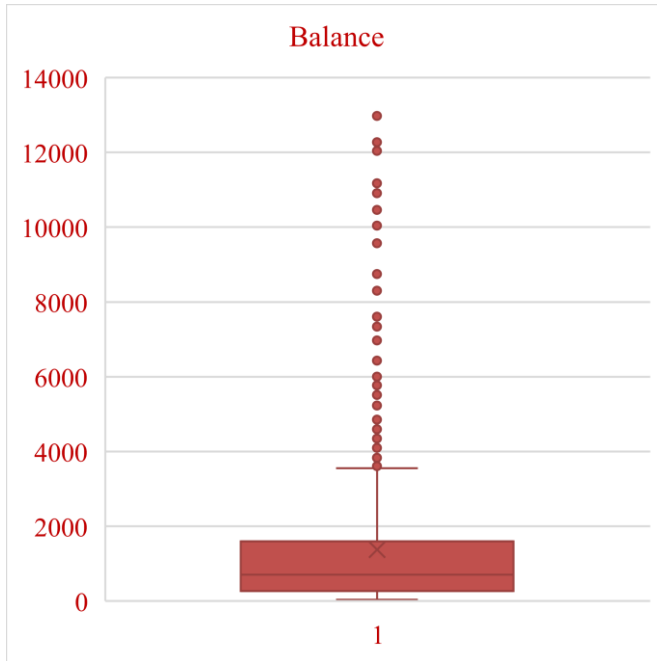
Giá trị	Tần suất
Primary	30
Secondary	876
Tertiary	145
Unknown	31



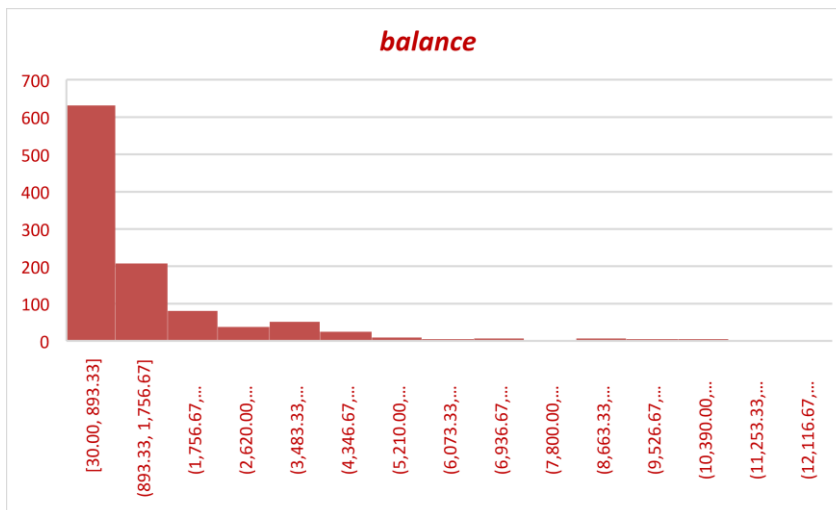
b. Tính giá trị trung bình mẫu, phương sai mẫu và độ lệch chuẩn mẫu cho biến *balance*.

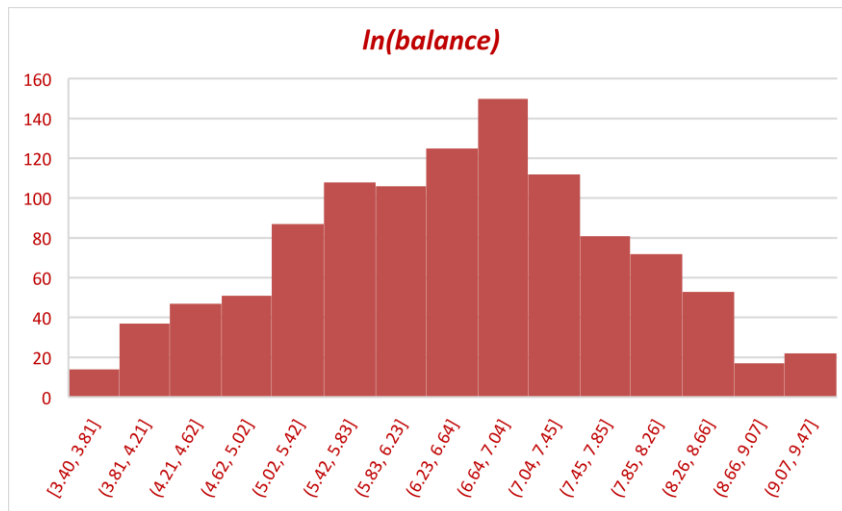
Giá trị	Hàm excel	Kết quả
Trung bình mẫu	=AVERAGE	1367.70
Phương sai mẫu	=VAR.S	3595485.88
Độ lệch chuẩn mẫu	=SQRT(VAR.S)	1896.18

c. Vẽ biểu đồ hộp (box plot) cho biến *balance*.



d (\*). Vẽ biểu đồ histogram cho biến *balance*. Sau đó tạo biến *ln\_balance* từ log tự nhiên (sử dụng hàm LN) sau đó vẽ histogram cho biến *ln\_balance* này. Theo bạn, phân phối tổng thể của biến *balance* sẽ theo dạng hàm phân phối nào đã được giới thiệu trong môn học?





Phân phối tổng thể của biến *balance* khi chuyển đổi thành dạng log tự nhiên sẽ tuân theo phân phối chuẩn, tuy nhiên phân phối của biến *balance* là phân phối loga chuẩn (log normal)

e. Tính hệ số tương quan mẫu cho hai biến *duration* và *balance*. Hai biến này có khả năng có mối quan hệ gì không?

Thực hiện tính hệ số tương quan mẫu trong Excel bằng hàm CORREL, ta có hệ số tương quan mẫu cho hai biến *duration* và *balance* là  $\hat{\rho} = -0.00184$ . Hai biến này có khả năng không có mối quan hệ vì hệ số tương quan gần như bằng 0.

---HẾT---