

THỐNG KÊ SUY DIỄN

Chương 4
TƯƠNG QUAN & HỒI QUY

- 1. Vector ngẫu nhiên
- 2. Hệ số tương quan
- 3. Hồi quy đơn
- 4. Máy tính Casio ES

1. Vector ngẫu nhiên

- Bộ gồm số nhiều m biến ngẫu nhiên được gọi là vector ngẫu nhiên m thành phần
 - Ví dụ:
 - X : biến ngẫu nhiên chiều cao học sinh THCS
 - Y : biến ngẫu nhiên cân nặng học sinh THCS
- bộ $(X ; Y)$ gọi là vector ngẫu nhiên 2 thành phần

2. Hệ số tương quan

- 2.1. Định nghĩa.**
- Là đại lượng đo lường mức độ *phụ thuộc tuyến tính* giữa hai biến ngẫu nhiên.
- Giả sử ta có mẫu ngẫu nhiên cỡ n về vector ngẫu nhiên (X, Y) là $(x_i, y_i); i = 1; 2;...; n$. Khi đó, hệ số tương quan mẫu r được tính theo công thức:

$$r = \frac{\overline{xy} - \overline{x}.\overline{y}}{\hat{s}_x.\hat{s}_y}; \overline{xy} = \frac{1}{n} \sum_{i=1}^n x_i.y_i.$$

2. Hệ số tương quan

2.2. Tính chất.

- $-1 \leq r \leq 1$.
- $r = 0 \rightarrow X$ và Y độc lập
- $r = \pm 1 \rightarrow X$ và Y phụ thuộc tuyệt đối
- $r > 0 \rightarrow X$ và Y có quan hệ thuận
- $r < 0 \rightarrow X$ và Y có quan hệ nghịch

2. Hệ số tương quan

VD 1. Kết quả đo lường độ cholesterol (Y) có trong máu của 10 đối tượng nam ở độ tuổi (X) như sau:

X	20	52	30	57	28	43	57	63	40	49
Y	1,9	4,0	2,6	4,5	2,9	3,8	4,1	4,6	3,2	4,0

Tính hệ số tương quan mẫu giữa X và Y .

3. Hồi quy đơn

3.1. Phân tích hồi quy.

- Là nghiên cứu sự phụ thuộc của một biến (biến phụ thuộc) dựa vào các biến độc lập khác (biến giải thích) với ý tưởng là ước lượng giá trị trung bình của biến phụ thuộc trên cơ sở các giá trị biết trước của các biến giải thích.
- Dự đoán mối quan hệ (hàm số) giữa biến phụ thuộc và các biến độc lập: $Y = f(X_1, X_2, ..., X_n)$

3. Hồi quy đơn

- 3.1. Phân tích hồi quy.
- Phạm vi môn học xét hồi quy tuyến tính và mối quan hệ biến phụ thuộc và một biến độc lập → *hồi quy đơn*: $Y = f(X) = a + bX$
- 3.2. Ước lượng hàm hồi quy mẫu.
- Từ dữ liệu mẫu, chúng ta ước lượng các hệ số của hàm hồi quy đơn tổng thể gọi là hàm hồi quy mẫu:

$$\boxed{Y = \hat{a} + \hat{b}X}$$

3. Hồi quy đơn

- 3.2. Ước lượng hàm hồi quy mẫu.
- Bằng phương pháp ước lượng bình phương độ lệch bé nhất, chúng ta có kết quả ước lượng sau:

$$\boxed{\hat{b} = \frac{\sum_{i=1}^n x_i y_i - \bar{n} \cdot \bar{x} \cdot \bar{y}}{\sum_{i=1}^n x_i^2 - \bar{n} \cdot \bar{x}^2}; \hat{a} = \bar{y} - \hat{b} \cdot \bar{x}}$$
- 3.3. Dự báo điểm.
- Với $X = x_0 \rightarrow$ dự báo $Y = y_0$ với $y_0 = \hat{a} + \hat{b}x_0$

Các ví dụ

- VD 2.** Đo chiều cao (X : m) và khối lượng (Y : kg) của 5 học sinh nam, ta có kết quả:
- | | | | | | |
|-----|------|------|------|------|------|
| X | 1,45 | 1,60 | 1,50 | 1,65 | 1,55 |
| Y | 50 | 55 | 45 | 60 | 55 |
- 1) Tìm hệ số tương quan r .
 - 2) Lập phương trình hồi quy tuyến tính của Y theo X .
 - 3) Dự đoán nếu một học sinh cao 1,62m thì nặng khoảng bao nhiêu kg?

Các ví dụ

VD 3. Số vốn đầu tư (X : triệu đồng) và lợi nhuận thu được (Y : triệu đồng) trong một đơn vị thời gian của 100 quan sát là:

$X \backslash Y$	0,3	0,7	1,0
1	20	10	
2		30	10
3		10	20

- 1) Lập phương trình hồi tuyến tính của X theo Y .
- 2) Dự đoán nếu muốn lợi nhuận thu được là 0,5 triệu đồng thì cần đầu tư bao nhiêu?

Các ví dụ

VD 4. Số thùng bia (Y : thùng) được bán ra phụ thuộc vào giá bán (X : triệu đồng/ thùng). Điều tra 100 đại lý về 1 loại bia trong một đơn vị thời gian có bảng số liệu:

$X \backslash Y$	100	110	120
0,150	5	15	30
0,160	10	25	
0,165	15		

- 1) Tính hệ số tương quan r .
- 2) Lập phương trình hồi tuyến tính của X theo Y .
- 3) Dự đoán nếu muốn bán được 115 thùng bia thì giá bán mỗi thùng cỡ bao nhiêu?

4. Máy tính Casio ES

- Mở màn số (đã học, nếu cần thiết)
- Mở chức năng: **mode** → **stat**(3) → **2**
- Nhập dữ liệu (như đã học), kết thúc: **AC**
- Xuất kết quả: **Shift** → **1** → **reg**(7) →
 $*(1): \hat{a}; *(2): \hat{b}; *(3): r$

.....Hết.....
