201110

Regression and Correlation
Part 2

Assoc.Prof Phisanu Chiawkhun
Statistics department
Faculty of Science
Chiangmai University

Correlation Coefficient

- In this section , we turn to a numerical measure of the strength of the linear relationship ,called the Pearson's correlation coefficient of , denoted by r
- Definition

  The sample correlation coefficient r measures the strength of the linear relationship between two quantitative variable .

Correlation Coefficient

The correlation coefficient describes the direction of linear association and indicates how closely the point in a scatter plot are to the least square regression line

Some of features of a Correlation Coefficient

1. Range : $-1 \leq r \leq 1$
2. Sign : The sign of the correlation coefficient indicates the direction of association negative or negative
3. Magnitude : The magnitude of the correlation coefficient indicates the strength of the linear association . If the data follow a straight line ,r=1 or r = -1 indicating a perfect linear association if r =0 there is no linear association

How to calculate r

- We can calculate the correlation coefficient by

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{[\sum (x-\bar{x})^2][\sum (y-\bar{y})^2]}}$$

$$r = \frac{n\sum xy - (\sum x)(\sum (y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

$$r = \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{[\sum x^2 - n\bar{x}^2][\sum y^2 - n\bar{y}^2]}}$$

Example 1

- Consider the heights (x) and weights (Y) of 10 basketball players. Find the correlation coefficient.

player	Weight , X (kg)	Height, Y (cm)
1	73	185
2	71	175
3	75	200
4	72	210
5	72	190
6	75	195
7	67	150
8	69	170
9	71	180
10	69	175

Solution

- For computing the correlation coefficient we calculate :

$$\sum x = 714 \quad \sum y = 1{,}830 \quad \sum xy = 130{,}990$$

$$\sum x^2 = 51{,}040 \quad \sum y^2 = 337{,}500 \quad \bar{x} = 71.4 \quad \bar{y} = 183$$

- $$r = \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{[\sum x^2 - n\bar{x}^2][\sum y^2 - n\bar{y}^2]}}$$

- $$r = \frac{130{,}990 - 10(71.4)(183)}{\sqrt{[51{,}040 - 10(71.4)^2][337{,}500 - 10(183)^2}} = 0.8261$$

Solution

we obtained the correlation coefficient :

$$r = 0.8261$$

This value of r is fairly close to 1, which indicates a fairly strong positive linear relationship between height and weight.

- ✓ There is a direct relationship between the r and b.
- ✓ r and b has the same sign.
- ✓ when r =0 , the slope b =0 and there is no linear relationship between x and y
- ✓ when r is positive , so is b, and there is a positive linear relationship between x and y
- ✓ when r is negative , so is b, and there is a negative linear relationship between x and y

Coefficient of Determination : R2

we obtained the correlation coefficient :

$$r = 0.8261$$

We can calculate the coefficient of determination by Calulating R squares

$$R2 = (0.8261)(0.8261) = 0.6824$$

Meaning

The variable X (weight) can describe the variation in variable Y (Height) about 68.24% .