201110 Chapter 3

Correlation

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

What do we study in this unit ?

- 1. We also study the correlation which measures the strength and direction of the linear relationship between two quantitative variables.

Introduction

- This chapter will discuss the methods of displaying and describing the relationship between two quantitative variables.
- The data used to study the relationship between two variables are bivariate data

The Bivariate data

- The bivariate data obtained by measuring both variables on the same individual unit. (X,Y)

  For example :

  The data of midterm score (X) and final score for a sample of students (Y).

  These data will help us study the association between two variables.

Correlation Coefficient

- In this section , we turn to a numerical measure of the strength of the linear relationship ,called the sample correlation coefficient , 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

Correlation Coefficient

- In this section , we turn to a numerical measure of the strength of the linear relationship ,called the <u>sample correlation coefficient , denoted by</u> r
- <u>Definition</u>

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

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 positive
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

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

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

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

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 :

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

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

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

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

Coefficient of determination(r2)

The coefficient of determination is the proportion of the variation in dependent variable this is described by the independent variable

we have

$$r = 0.8261$$

so that

$$r^2 = 0.6824$$

The independent variable (x) can describe the variation in dependent variable (y) = 68.24%