CORRELATION

Correlation refers to a process for establishing the relationships between two variables. You learned a way to get a general idea about whether or not two variables are related, is to plot them on a "scatter plot". While there are many measures of association for variables which are measured at the ordinal or higher level of measurement, correlation is the most commonly used approach.

Types of Correlation

- Positive Correlation – when the values of the two variables move in the same direction so that an increase/decrease in the value of one variable is followed by an increase/decrease in the value of the other variable.
- Negative Correlation – when the values of the two variables move in the opposite direction so that an increase/decrease in the value of one variable is followed by decrease/increase in the value of the other variable.
- No Correlation – when there is no linear dependence or no relation between the two variables.

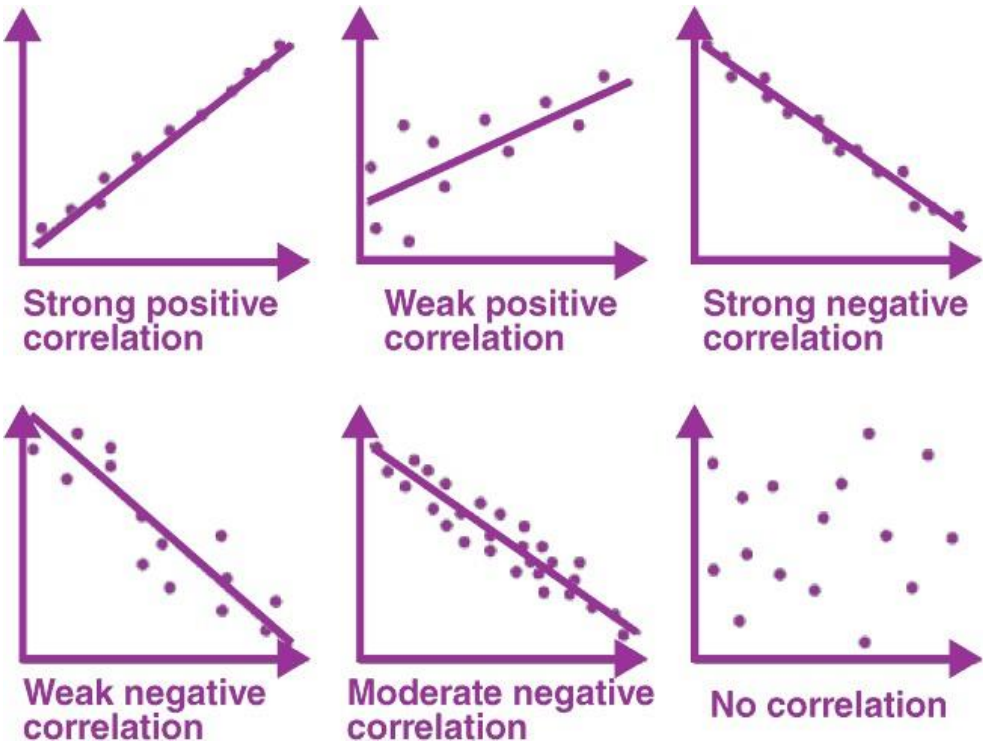


Table 6.2
THREE TYPES OF RELATIONSHIPS

A. POSITIVE RELATIONSHIP

FRIEND	SENT	RECEIVED
Doris	13	14
Steve	9	18
Mike	7	12
Andrea	5	10
John	1	6

B. NEGATIVE RELATIONSHIP

FRIEND	SENT	RECEIVED
Doris	13	6
Steve	9	10
Mike	7	14
Andrea	5	12
John	1	18

C. LITTLE OR NO RELATIONSHIP

FRIEND	SENT	RECEIVED
Doris	13	10
Steve	9	18
Mike	7	12
Andrea	5	6
John	1	14

SCATTERPLOTS

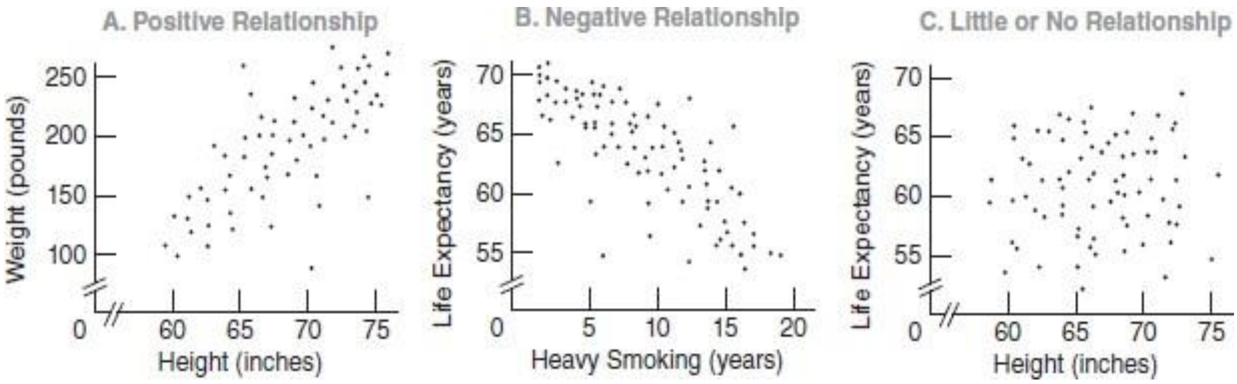
A scatter plot is a graph containing a cluster of dots that represents all pairs of scores. In other words Scatter plots are the graphs that present the relationship between two variables in a data-set. It represents data points on a two-dimensional plane or on a Cartesian system.

Construction of scatter plots

- The independent variable or attribute is plotted on the X-axis.
- The dependent variable is plotted on the Y-axis.
- Use each pair of scores to locate a dot within the scatter plot

Positive, Negative, or Little or No Relationship?

- The first step is to note the tilt or slope, if any, of a dot cluster. A dot cluster that has a slope from the lower left to the upper right, as in panel A of below figure reflects a positive relationship.
- A dot cluster that has a slope from the upper left to the lower right, as in panel B of below figure reflects a negative relationship.
- A dot cluster that lacks any apparent slope, as in panel C of below figure reflects little or no relationship.

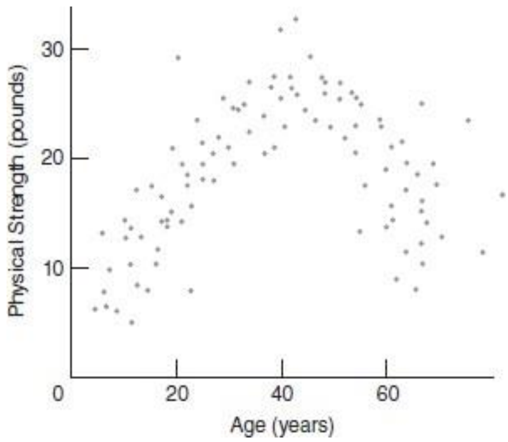


Perfect Relationship

A dot cluster that equals (rather than merely approximates) a straight line reflects a perfect relationship between two variables.

Curvilinear Relationship

The previous discussion assumes that a dot cluster approximates a straight line and, therefore, reflects a linear relationship. But this is not always the case. Sometimes a dot cluster approximates a bent or curved line, as in below figure, and therefore reflects a curvilinear relationship.



A CORRELATION COEFFICIENT FOR QUANTITATIVE DATA : *r*

The correlation coefficient, r, is a summary measure that describes the extent of the statistical relationship between two interval or ratio level variables.

Properties of r

- The correlation coefficient is scaled so that it is always between -1 and +1.
- When r is close to 0 this means that there is little relationship between the variables and the farther away from 0 r is, in either the positive or negative direction, the greater the relationship between the two variables.
- The sign of r indicates the type of linear relationship, whether positive or negative.
- The numerical value of r, without regard to sign, indicates the strength of the linear relationship.
- A number with a plus sign (or no sign) indicates a positive relationship, and a number with a minus sign indicates a negative relationship.

COMPUTATION FORMULA FOR *r*

Calculate a value for r by using the following computation formula:

CORRELATION COEFFICIENT (COMPUTATION FORMULA)

$$r = \frac{SP_{xy}}{\sqrt{SS_x SS_y}}$$

Where the two sum of squares terms in the denominator are defined as

$$SS_x = \Sigma\left(X - \overline{X}\right)^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{n}$$

$$SS_y = \Sigma\left(Y - \overline{Y}\right)^2 = \Sigma Y^2 - \frac{(\Sigma Y)^2}{n}$$

The sum of the products term in the numerator, SPxy, is defined in below formula

$$SP_{xy} = \Sigma\left(X - \overline{X}\right)\left(Y - \overline{Y}\right) = \Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{n}$$

Or the formula is written as

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

Where n = Number of Information

Σx = Total of the First Variable Value

Σy = Total of the Second Variable Value

Σxy = Sum of the Product of first & Second Value

Σx2 = Sum of the Squares of the First Value

Σy2 = Sum of the Squares of the Second Value