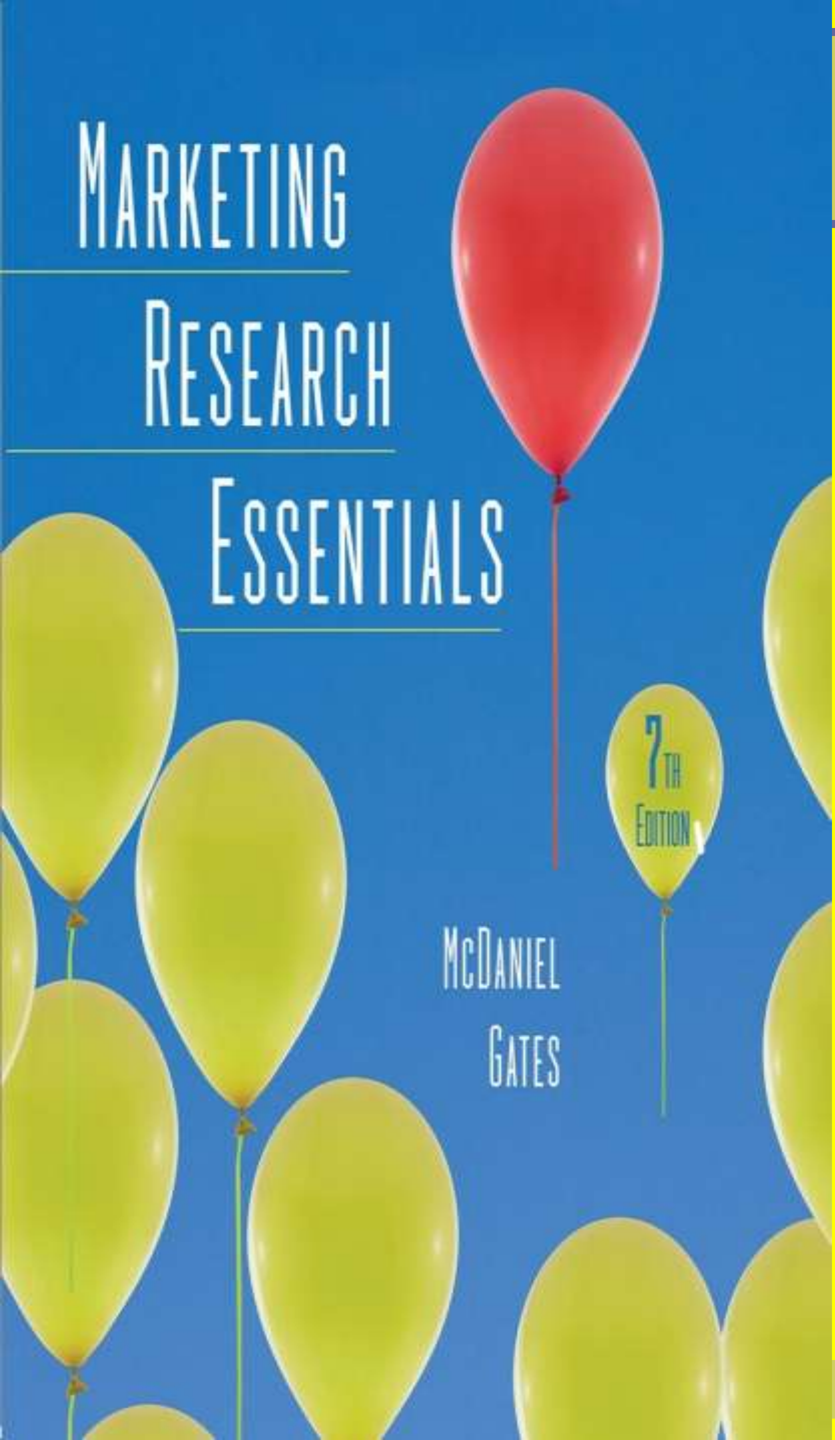


Chapter Fourteen

Bivariate Correlation and Regression

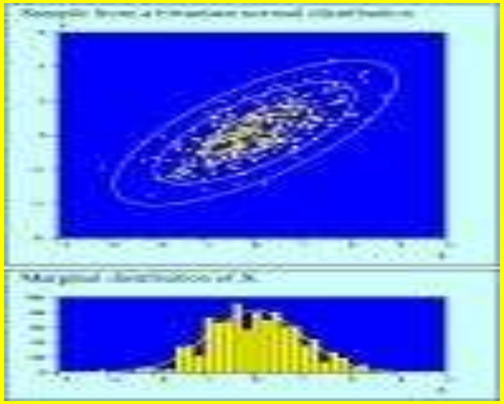
Chapter Fourteen Objectives

• To comprehend the nature of correlation analysis.
• To understand bivariate regression analysis.
• To become aware of the coefficient of determination of R^2 .
• To understand Spearman rank-order correlation.

Bivariate Analysis of Association

Bivariate Techniques:

- Statistical methods of analyzing the relationship between variables.



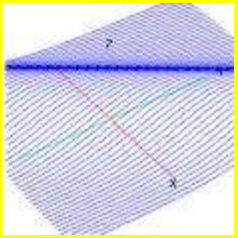
Independent Variable:

- Variable believed to affect the value of the dependent variable.

Bivariate Analysis of Association

Dependent Variable:

- Variable expected to be explained or caused by the independent variable.



Bivariate Regression Analysis:

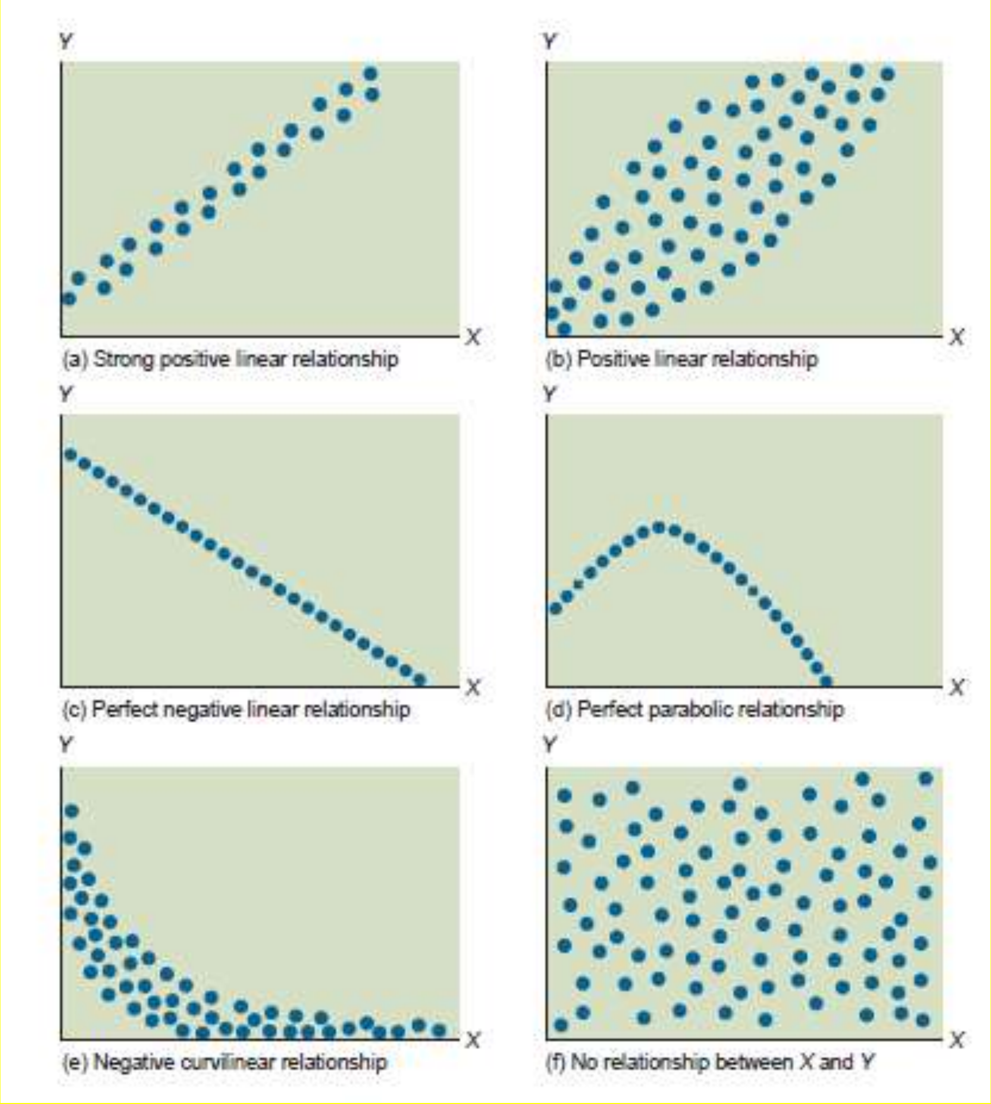
- The analysis of the strength of the linear relationship between variables when one is considered the independent variable and the other is the dependent variable.

Scatter Diagram

Scatter Diagram:

Graphic plot of the data with dependent variable on the Y (vertical) axis and the independent variable on the X (horizontal) axis. Shows the nature of the relationship between the two variables, linear or nonlinear.

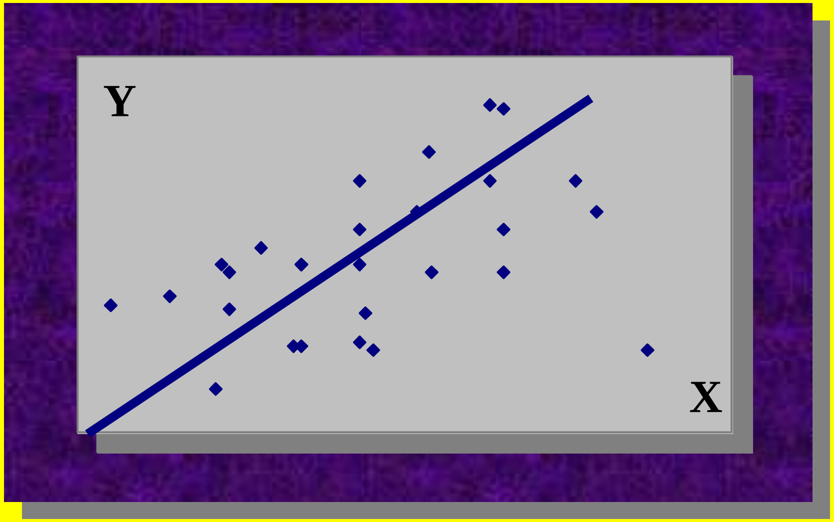
Types of Relationships Found in Scatter Diagrams



Least-Square Estimation Procedure

Least-Square Estimation

- Used to fit data for X and Y not plotted;
- Enables estimation of non-plotted data points;
- Results in a straight line that fits the actual observations *(plotted dots)* better than any other line that could be fit to the observations.



Least-Square Estimation Procedure

Estimating the best line of fit:

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

Where:

Y = dependent variable

$\hat{a}$ = estimated Y intercept

$\hat{b}$ = estimated slope of the regression line

X = independent variable

e = error

Values for "a" and "b" can be calculated as follows:

$$\hat{b} = \frac{\Sigma X_i Y_i - n\bar{X}\bar{Y}}{\Sigma X_i^2 - n(\bar{X})^2}$$

$$\hat{a} = \bar{Y} - \hat{b}\bar{X}$$

Where:

$\bar{\mathbf{X}}$ = mean of value X

$\bar{\mathbf{Y}}$ = mean of value y

n = sample size

Measures of Association

Coefficient of Determination:

- Percentage of the total variation in the dependent variable explained by the manipulation of the independent variable(s).

$$R^2 = \frac{\text{Total Variation - Unexplained Variation}}{\text{Total Variation}}$$

$$R^2 = 1 - \frac{\sum_{I=1}^{n} (Y_i - \hat{Y}_i)^2}{\sum_{I=1}^{n} (Y_i - \bar{Y})^2}$$

Pearson Correlation:

- Analysis of the degree to which changes in one variable are associated with changes in another for use with metric data.

The Strength of Association – R^2 :

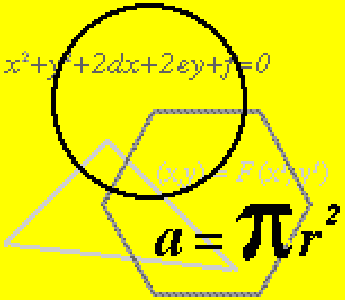
- The coefficient of determination: *the percentage of the total* … *dependent variable explained by the independent variable.*

$$R = + \text{ or } - \sqrt{R^2}$$

Sum of Squares

Total Variation: Sum of Squares (SST)

$$SST = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

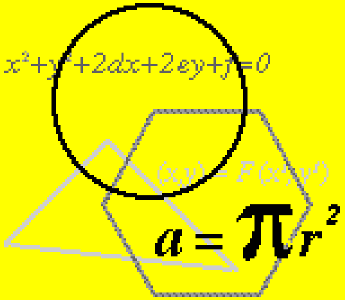


$$= \sum_{i=1}^{n} Y_i^{\,2} - \left(\frac{\sum_{i=1}^{n} Y_i^{\,2}}{n} \right)$$

Sum of Squares

Sum of Squares due to Regression (SSR)

$$SSR = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$$

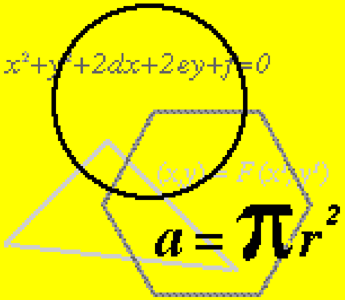


$$= a \sum_{i=1}^{n} Y_i + b \sum_{i=1}^{n} X_i Y_i - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}$$

Sum of Squares

Error Sums of Squares (SSE)

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y})^2$$



$$= \sum_{i=1}^{n} Y^2_i - a\sum_{i=1}^{n} Y_i - b\sum_{i=1}^{n} X_i Y_i$$

Hypotheses Concerning Overall Regression

Here we, as the researchers, are interested in hypotheses regarding the computed R^2 value for the problem. Is the amount of variance explained in the result (by our model) significantly greater than we would expect due to chance?

- ☐ Null hypothesis H_0: There is no linear relationship between *X* (average daily vehicular traffic) and *Y* (annual sales).
- ☐ Alternative hypothesis H_1: There is a linear relationship between *X* and *Y*.

Correlation

Assessing Measures of Association

Pearson Correlation:

Measure of Association using interval or ratio data.

Spearman Correlation:

Measure of Association using ordinal or rank order data.

Key Points:

Measures of Association:

- Do not mean there is a causal relationship between the relevant variables;
- Could simply represent coincidence between the relevant variables;
- Should be taken in context and with the timeliness of both data sets in mind;
- Can be used in conjunction with cross tabulations of the relevant data to add another perspective to the results.