Multivariate Statistical Analysis

- Statistical methods that allow the simultaneous investigation of more than two variables
A Classification of Selected Multivariate Methods

- **All multivariate methods**
  - Are some of the variables dependent on others?
    - **Yes**
      - Dependence methods
    - **No**
      - Interdependence methods

Dependence Methods

- A category of multivariate statistical techniques; dependence methods explain or predict a dependent variable(s) on the basis of two or more independent variables.
Dependence Methods

How many variables are dependent

One dependent variable
Several dependent variables
Multiple independent and dependent variables

One dependent variable

Metric
Multiple regression analysis

Nonmetric
Multiple discriminant analysis
Dependence Methods

How many variables are dependent

Several dependent variables

Metric

Multivariate analysis of variance

Nonmetric

Conjoint analysis

Dependence Methods

How many variables are dependent

Multiple independent and dependent variables

Metric or nonmetric

Canonical correlation analysis
Interdependence Methods

- A category of multivariate statistical techniques; interdependence methods give meaning to a set of variables or seek to group things together.
Interdependence methods

Are inputs metric?

Metric

Factor analysis  Cluster analysis  Metric multidimensional scaling

Interdependence methods

Are inputs metric?

Nonmetric

Nonmetric
Multiple Regression

• An extension of bivariate regression
• Allows for the simultaneous investigation
  – two or more independent variables
  – a single interval-scaled dependent variable

Multiple Regression Equation

\[ Y = a + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_n X_n \]
Multiple Regression Analysis

\[ Y = a + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_n X_n \]

Coefficients of Partial Regression

\( \beta_1 \)

Independent variables correlated with one another

The % of the variance in the dependent variable that is explained by a single independent variable, holding other independent variables constant
Coefficient of Multiple Determination

- $R^2$
- The % of the variance in the dependent variable that is explained by the variation in the independent variables.

Statistical Results of a Multiple Regression

- $Y = 102.18 + 0.387X_1 + 115.2X_2 + 6.73X_3$
- Coefficient of multiple determination ($R^2$) .845
- $F$-value 14.6
F-Test

\[ F = \frac{(SSr)/k}{(SSe)/(n - k - 1)} \]

Degrees of Freedom (d.f.) are Calculated as Follows:

- d.f. for the numerator = k
- for the denominator = n - k - 1
Degrees of Freedom

- $k =$ number of independent variables
- $n =$ number of observations or respondents

F-test

$$ F = \frac{(SSr)/k}{(SSe)/(n-k-1)} $$

where

- $k =$ number of independent variables
- $n =$ number of observations
Multiple Discriminant Analysis

- A statistical technique for predicting the probability of objects belonging in two or more mutually exclusive categories (dependent variable) based on several independent variables

\[ Z_i = b_1X_{1i} + b_2X_{2i} + \ldots + b_nX_{ni} \]

- where
- \( Z_i \) = ith applicant’s discriminant score
- \( b_n \) = discriminant coefficient for the nth variable
- \( X_{ni} \) = applicant’s value on the nth independent variable
Discriminant Analysis

\[ Z_i = b_1 X_{1i} + b_2 X_{2i} + \ldots + b_n X_{ni} \]

\( X_{ji} \) = applicant’s value on the jth independent variable

\( b_j \) = discriminant coefficient for the \( j^{th} \) variable

\( Z_i \) = \( i^{th} \) applicant’s discriminant score
Canonical Correlation

- Two or more criterion variables (dependent variables)
- Multiple predictor variables (independent variables)
- An extension of multiple regression
- Linear association between two sets of variables

\[ Z = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n \]
\[ W = b_1 Y_1 + b_2 Y_2 + \ldots + b_n Y_n \]
Factor Analysis

• Summarize the information in a large number of variables
• Into a smaller number of factors
• Several factor-analytical techniques

Factor Analysis

• A type of analysis used to discern the underlying dimensions or regularity in phenomena. Its general purpose is to summarize the information contained in a large number of variables into a smaller number of factors.
Cluster Analysis

- A body of techniques with the purpose of classifying individuals or objects into a small number of mutually exclusive groups, ensuring that there will be as much likeness within groups and as much difference among groups as possible.
Multidimensional Scaling

- A statistical technique that measures objects in multidimensional space on the basis of respondents’ judgments of the similarity of objects

Multivariate Analysis of Variance (MANOVA)

- A statistical technique that provides a simultaneous significance test of mean difference between groups for two or more dependent variables