

Dr. Saeed A. Dobbah Alghamdi

Department of Statistics

Faculty of Sciences

Building 90, Office 26F41

King Abdulaziz University

<http://saalghamdy.kau.edu.sa>



Dr. Saeed's Website

Statistics for Business & Economics

David R. Anderson
Dennis J. Sweeney
Thomas A. Williams

Copyright © South-Western Cengage Learning All rights reserved.

Data and Statistics

Chapter 1



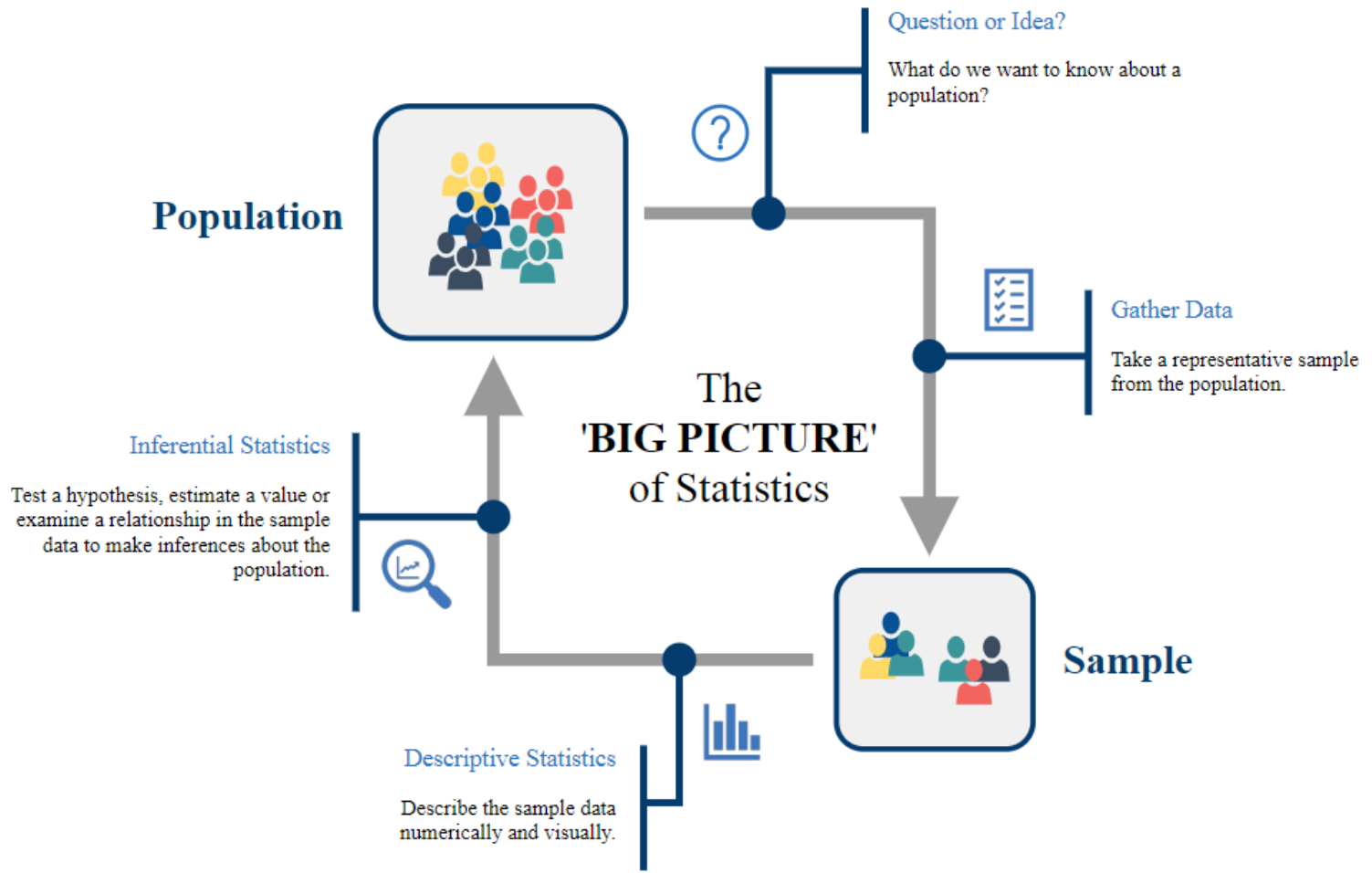


Learning Objectives

- LO1** Understand the importance of statistics in the business world
- LO2** Know data and sources of data
- LO3** Know the differences between descriptive and inferential statistics.
- LO4** Understand the differences between a sample and a population.
- LO5** Explain the difference between categorical and quantitative variables.
- LO6** Recognize the levels of measurement in data.
- LO7** Data sources.

Introduction

The 'Big Picture' of statistics



Applications in Business and Economics

■ Accounting

Public accounting firms use statistical sampling procedures when conducting audits for their clients.

■ Finance

Financial advisors use a variety of statistical information, including price-earnings ratios and dividend yields, to guide their investment recommendations.

■ Marketing

Electronic point-of-sale scanners at retail checkout counters are being used to collect data for a variety of marketing research applications.

■ Production

A variety of statistical quality control charts are used to monitor the output of a production process.

■ Economics

Economists use statistical information in making forecasts about the future of the economy or some aspect of it.

Basic Statistical Terminology

Statistics is the art and science of collecting, analyzing, presenting, and interpreting data.

In business, statistics is not just about numbers and mathematics but a tool that analyzes the available data and helps making informed and better business decisions.

EXAMPLE

The 2010 Census of Saudi Arabia reveals the following:

- Population is 27,136,977 persons.
- 69% of the population are Saudi citizens.
- Number of houses is 4,643,151.

Basic Statistical Terminology

In the stage of planning any statistical study, we need to determine the POPULATION of our study.

Population consists of all elements for the phenomenon under a study.

EXAMPLE

The 2010 Census of Saudi Arabia reveals the following:

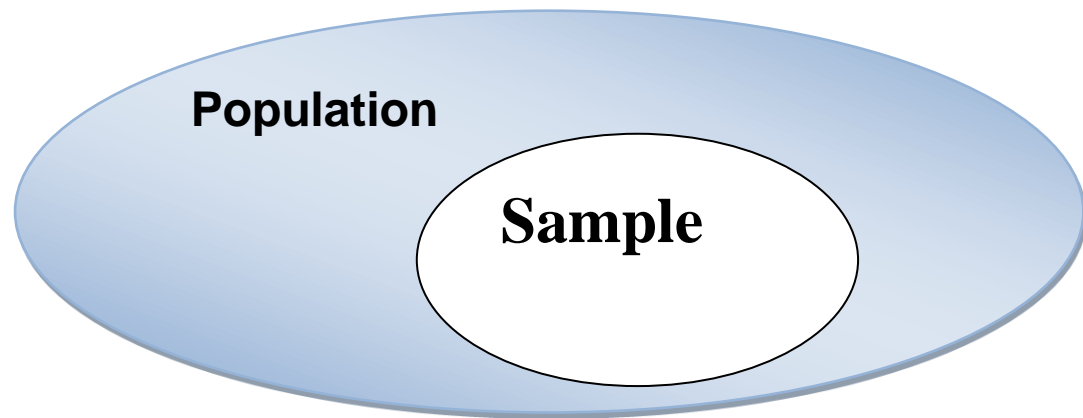
- Population is 27,136,977 persons.
- Number of houses is 4,643,151.

Census A survey to collect data on the entire population.

Basic Statistical Terminology

Sometimes it is not possible to collect the information from each element in the population because it is impractical or too expensive. Hence, we may select a representative portion from the entire population which is called **SAMPLE**.

Sample is a group of elements selected randomly from a population.



■EXAMPLE

A group of students is selected randomly from all students studying in King Abdulaziz University (KAU) for a research study.

Sample survey A survey to collect data on a sample.

Basic Statistical Terminology

To gain knowledge about any phenomenon, DATA are collected for VARIABLES that describe that phenomenon.

Data are the values that variables can assume to describe a phenomenon.

■EXAMPLE

The values that the variable weight can take when measuring the weights of students who are studying Statistics in KAU.

Variable is a characteristic or an attribute of an item or individual that can assume different values, e.g., flight classes, gender, color, students ID,

Cross-sectional data Data collected at the same or approximately the same point in time for two or categorical variables, e.g., flight classes and gender

Time series data Data collected over several time periods, e.g., weekly cars accidents

Basic Statistical Terminology

Data set All the data collected in a particular study.

Elements The entities on which data are collected.

Observation The set of measurements obtained for a particular element.

Data Sets

Company	Stock Exchange	Annual Sales(\$M)	Earn/Sh.(\$)
Dataram	AMEX	73.10	0.86
EnergySouth	OTC	74.00	1.67
Keystone	NYSE	365.70	0.86
LandCare	NYSE	111.40	0.33
Psychemedics	AMEX	17.60	0.13

The diagram includes the following labels and callouts:

- Variables:** Points to the column headers of the table.
- Elements:** Points to the list of company names in the first column.
- Observation:** Points to a single row of data.
- Data:** Points to the entire table.
- Datum:** Points to the specific value 111.40 in the table.

Variables and Types of Data

Qualitative or categorical variables data use labels or names to identify an attribute of each element. Categorical data use either the nominal or ordinal scale of measurement and may be nonnumeric or numeric, e.g., flight classes, gender, color, students ID,

Quantitative or scale variables data are numeric values that indicate how much or how many. Quantitative data use either the interval or ratio scale of measurement. Ordinary arithmetic operations are meaningful only if the data are quantitative. Therefore, statistical computations used for quantitative data are not always appropriate for categorical data, e.g., number of passengers, weight, age,

Variables and Types of Data

Qualitative or categorical variables data use labels or names to identify an attribute of each element. Categorical may be nonnumeric or numeric, e.g., flight classes, gender, color, students ID,

Quantitative or scale variables data are numeric values in nature that indicate how much or how many. Ordinary arithmetic operations are meaningful only if the data are quantitative, e.g., weight, age, number of students,

Quantitative or scale variables can also be classified into two groups:

Discrete—assume numerical values that can be counted, e.g., number of rooms in buildings, number of students in Stat 110, number of children in families,

Continuous—can assume all numerical values that can be measured and take values between any two specific values, e.g., heights, weights, temperature,

Variables and Types of Data

Qualitative or categorical variables can be classified into two groups according to measurements levels:

Nominal—classifies data into categories that can not be ordered or ranked. Thus, nominal data are labels or names used to identify an attribute of an element, e.g., gender, color, departments,

Ordinal—classifies data into categories that can be ordered or ranked, e.g., flight classes, ranking, grade letters,

Quantitative or scale variables can be classified into two groups according to measurements levels:

Interval—ranks data, and precise differences between units of measure do exist; however, there is no meaningful zero, e.g., temperature, dress size, IQ-score, ...

Ratio—the data show all the properties of interval data and the ratio of two values is meaningful; however, there is meaningful zero value, e.g., speed, distance, grade scores,

Data Sources

■ Existing Sources

- Data needed for a particular application might already exist within a firm. Detailed information is often kept on customers, suppliers, and employees for example.
- Substantial amounts of business and economic data are available from organizations that specialize in collecting and maintaining data.
- Government agencies are another important source of data.
- Data are also available from a variety of industry association and special interest organizations.

■ Internet

- The Internet has become an important source of data.
- Most government agencies, like the Bureau of the make their data available through a web site.
- More and more companies are creating web sites and providing public access to them.
- A number of companies now specialize in making information available over the Internet

Data Sources

■ Statistical Studies

- Statistical studies can be classified as either experimental or observational.
- In experimental studies the variables of interest are first identified. Then one or more factors are controlled so that data can be obtained about how the factors influence the variables.
- In observational (nonexperimental) studies no attempt is made to control or influence the variables of interest.
- A survey is perhaps the most common type of observational study.

Descriptive and Inferential Statistics

In any statistical study, you should start with the descriptive statistics and then the inferential statistics which is usually the goal of any study.

Descriptive statistics consists of the collection, organization, summarization, and presentation of data.

EXAMPLE

The 2010 Census of Saudi Arabia reveals the following:

- Population is 27,136,977 persons.
- 69% of the population are Saudi citizens.
- Number of houses is 4,643,151.

Inferential statistics consists of generalizing from samples to populations by performing estimations, hypothesis testing, determining relationships among variables, and making predictions.

EXAMPLE

- The estimated average height of KAU students is 1.75m
- There is a relationship between Math and Stat grades..
- The population of a country will be increased by 3 million into years.

Inferential Statistics

To estimate population parameters based on sample statistics

Sample
Statistics

Population
Parameters

- ◆ Sample mean: \bar{X} → ◆ Population mean: μ
- ◆ Sample variance: S^2 → ◆ Population variance: σ^2
- ◆ Sample proportion: \bar{P} → ◆ Population proportion: P

Data Mining

The subject of **data mining** deals with methods for developing useful decision-making information from large data bases. Using a combination of procedures from statistics, mathematics and computer science, analysts “mine the data” in the *warehouse* to convert it into useful information, hence the name *data mining*.

The term *data warehousing* is used to refer to the process of capturing, storing, and maintaining the data. Computing power and data collection tools have reached the point where it is now feasible to store and retrieve extremely large quantities of data in seconds. Analysis of the data in the warehouse may result in decisions that will lead to new strategies and higher profits for the organization.

The major applications of data mining have been made by companies with a strong consumer focus, such as retail businesses, financial organizations, and communication companies. Data mining has been successfully used to help retailers such as Amazon and Barnes & Noble determine one or more related products that customers who have already purchased a specific product are also likely to purchase. Then, when a customer logs on to the company’s website and purchases a product, the website uses pop-ups to alert the customer about additional products that the customer is likely to purchase

Summary

- **Statistics** The art and science of collecting, analyzing, presenting, and interpreting data.
- **Data** The facts and figures collected, analyzed, and summarized for presentation and interpretation.
- **Data set** All the data collected in a particular study.
- **Elements** The entities on which data are collected.
- **Variable** A characteristic of interest for the elements.
- **Observation** The set of measurements obtained for a particular element.
- **Nominal scale** Classifies data into categories that can not be ordered or ranked. Thus, nominal data are labels or names used to identify an attribute of an element,
- **Ordinal scale** Classifies data into categories that can be ordered of ranked

Summary

- **Interval scale** Ranks data, and precise differences between units of measure do exist; however, there is no meaningful zero
- **Ratio scale** The data show all the properties of interval data and the ratio of two values is meaningful; however, there is meaningful zero value.
- **Categorical data** Labels or names used to identify an attribute of each element. Categorical data use either the nominal or ordinal scale of measurement and may be nonnumeric or numeric.
- **Quantitative data** Numeric values in nature that indicate how much or how many of something. Quantitative data are obtained using either the interval or ratio scale of measurement.
- **Categorical variable** A variable with categorical data.
- **Quantitative variable** A variable with quantitative data.
- **Discrete variable** Assume numerical values that can be counted.
- **Continuous variable** can assume all numerical values that can be measured and take values between any two specific values.

Summary

- **Cross-sectional data** Data collected at the same or approximately the same point in time.
- **Time series data** Data collected over several time periods.
- **Descriptive statistics** Tabular, graphical, and numerical summaries of data.
- **Population** The set of all elements of interest in a particular study.
- **Sample** A subset of the population.
- **Census** A survey to collect data on the entire population.
- **Sample survey** A survey to collect data on a sample.
- **Statistical inference** The process of using data obtained from a sample to make estimates or test hypotheses about the characteristics of a population.
- **Data mining** The process of using procedures from statistics and computer science to extract useful information from extremely large databases.