Eighth Edition
Elementary Statistics
A Step by Step Approach
Bluman
Mc Graw Hill
Chapter 1

The Nature of Probability and Statistics

Objectives

After completing this chapter, you should be able to

1. Demonstrate knowledge of statistical terms.
2. Differentiate between the two branches of statistics.
3. Identify types of data.
4. Identify the measurement level for each variable.
5. Identify the four basic sampling techniques.
6. Explain the difference between an observational and an experimental study.
7. Explain how statistics can be used and misused.
8. Explain the importance of computers and calculators in statistics.

Outline

Introduction

1-1 Descriptive and Inferential Statistics
1-2 Variables and Types of Data
1-3 Data Collection and Sampling Techniques
1-4 Observational and Experimental Studies
1-5 Uses and Misuses of Statistics
1-6 Computers and Calculators

Summary
Introduction

- **Statistics** is the science of conducting studies to collect, organize, summarize, analyze, and draw conclusions from data.
A variable is a characteristic or attribute that can assume different values.

Variables whose values are determined by chance are called random variables.

The values that a variable can assume are called data.

A population consists of all subjects (human or otherwise) that are studied.

A sample is a subset of the population.
E.g. A data set in table form (columns and rows):

<table>
<thead>
<tr>
<th>id</th>
<th>gender</th>
<th>age</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>male</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>male</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>female</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>male</td>
<td>23</td>
</tr>
<tr>
<td>5</td>
<td>female</td>
<td>30</td>
</tr>
</tbody>
</table>
Population and Sample

Population

Sample

Use parameters to summarize features

Use statistics to summarize features

Inference on the population from the sample
Census versus Sample

- **Census**
  Collection of data from *every* member of a population

- **Sample**
  *Subcollection* of members selected from a population
1-1 Descriptive and Inferential Statistics

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

- **Inferential statistics** consists of generalizing from samples to populations, performing estimations and hypothesis tests, determining relationships among variables, and making predictions.
1-2 Variables and Types of Data

Data

- Qualitative
  - Categorical

- Quantitative
  - Numerical, Can be ranked
    - Discrete
      - Countable
      - 5, 29, 8000, etc.
    - Continuous
      - Measurable
      - 2.59, 312.1, etc.
1-2 Variables and Types of Data

Another way to classify data is to use **levels of measurement**. Four of these levels are discussed in the following slides.
1-2 Variables and Types of Data

**Levels of Measurement**

1. **Nominal** – categorical (names)
2. **Ordinal** – nominal, plus can be ranked (order)
3. **Interval** – ordinal, plus intervals are consistent (differences but no natural zero starting point)
4. **Ratio** – interval, plus ratios are consistent, true zero (difference and a natural zero starting point)

**NOTE:** To remember these four terms, think of the French word **NOIR**. Noir in French is the name of the color black. But the letters of the word represent the first letter of the above four terms. It will help you remember them.
# Variables and Types of Data

Determine the measurement level.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Nominal</th>
<th>Ordinal</th>
<th>Interval</th>
<th>Ratio</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hair Color</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
<td>Nominal</td>
</tr>
<tr>
<td>Zip Code</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
<td>Nominal</td>
</tr>
<tr>
<td>Letter Grade</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td>Ordinal</td>
</tr>
<tr>
<td>ACT Score</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Interval</td>
</tr>
<tr>
<td>Height</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Ratio</td>
</tr>
<tr>
<td>Age</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Ratio</td>
</tr>
<tr>
<td>Temperature (F)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Interval</td>
</tr>
</tbody>
</table>
1-3 Data Collection and Sampling Techniques

Some Sampling Techniques

- **Probability sampling (random sampling):**
  - **Random** – random number generator
  - **Systematic** – every $k^{th}$ subject
  - **Stratified** – divide population into group called “strata”
  - **Cluster** – use intact groups

- **Non-probability sampling:** cannot be used to infer from the sample to the general population.
  - **Convenient** – mall surveys

Bluman, Chapter 1, 07/2011
Q: Select random sample of 15 subjects out of 85 subjects:
   Skip the number > 85 or already been selected.
A: 12, 27, 75, 62, 57, 13, 31, 06, 16, 49, 46, 71, 53, 41, 02
Systematic Samples

• Decide on Sample Size: \( n \)
• Divide Frame of \( N \) individuals into Groups of \( k \) Individuals: \( k = \frac{N}{n} \)
• Randomly Select One Individual from the 1\text{st} Group
• Select Every \( k\)-th Individual Thereafter

\[
\begin{align*}
N &= 64 \\
n &= 8 \\
k &= 8
\end{align*}
\]
Stratified Samples

- Population Divided into 2 or More Groups According to Some Common Characteristic
- Simple Random Sample Selected from Each Group
- The Two or More Samples are Combined into One

E.g. gender difference
Cluster Samples

- Population Divided into Several “Clusters,” Each Representative of the Population
- A Random Sampling of Clusters is Taken
- All Items in the Selected Clusters are Studied
1-4 Observational and Experimental Studies

- In an observational study, the researcher merely observes and tries to draw conclusions based on the observations.

- The researcher manipulates the independent (explanatory) variable and tries to determine how the manipulation influences the dependent (outcome) variable in an experimental study.

- A confounding variable influences the dependent variable but cannot be separated from the independent variable.
1-5 Uses and Misuses of Statistics

- **Suspect Samples**
  - Is the sample large enough?
  - How was the sample selected?
  - Is the sample representative of the population?

- **Ambiguous Averages**
  - What particular measure of average was used and why?
1-5 Uses and Misuses of Statistics

- **Changing the Subject**
  - Are different values used to represent the same data?

- **Detached Statistics**
  - One third fewer calories than what?

- **Implied Connections**
  - Studies *suggest* that *some* people *may* understand what this statement means.
1-5 Uses and Misuses of Statistics

- **Misleading Graphs**
  - Are the scales for the x-axis and y-axis appropriate for the data?

- **Faulty Survey Questions**
  - Do you feel that statistics teachers should be paid higher salaries?
  - Do you favor increasing tuition so that colleges can pay statistics teachers higher salaries?
1-6 Computers and Calculators

- **TI-83/84**
- **Microsoft Excel**
- **Microsoft Excel with MegaStat**
- **Minitab**
- **SAS (Statistical Analysis Software)**
- **SPSS (Statistical Package for the Social Sciences)**

Bluman, Chapter 1, 07/2011
Highlights

- Statistical studies use random variables and values of the variable are called data.
- Statistics can be divided into two main branches, descriptive and inferential (based on probability theory) statistics.
- A very important requirement of a statistical study is to define the population and select a random sample.
- Various type of data and measurement levels.