What is Statistics?
Table of Contents

LEARNING OBJECTIVES ......................................................................................................................... 3
WHAT IS STATISTICS? ............................................................................................................................... 3
  STATISTICS DEFINED ............................................................................................................................ 3
  BRANCHES OF STATISTICS ................................................................................................................ 3
  DESCRIPTIVE STATISTICS .................................................................................................................. 4
  INFERENTIAL STATISTICS ................................................................................................................... 4
  INFERENTIAL STATISTICS IN USE ....................................................................................................... 4
  UNDERSTANDING DATA .................................................................................................................... 4
  DATA DEFINED ................................................................................................................................... 5
  POPULATION DEFINED ....................................................................................................................... 5
  WHAT IS A CENSUS? ............................................................................................................................. 5
  WHAT IS A SAMPLE? ............................................................................................................................... 6
SAMPLING METHODS ............................................................................................................................... 6
  RANDOM SAMPLING ............................................................................................................................ 7
  SYSTEMATIC SAMPLING ..................................................................................................................... 7
  CLUSTER SAMPLING ............................................................................................................................. 7
  STRATIFIED SAMPLING ......................................................................................................................... 8
  SAMPLING WITH REPLACEMENT ..................................................................................................... 8
  SAMPLING WITHOUT REPLACEMENT ............................................................................................. 8
TYPES OF VARIABLES ............................................................................................................................. 8
  UNDERSTANDING VARIABLES ......................................................................................................... 9
  QUALITATIVE DATA ............................................................................................................................. 9
  QUANTITATIVE DATA ............................................................................................................................ 9
DISCRETE AND CONTINUOUS .................................................................................................................... 10
  DISCRETE VARIABLE .......................................................................................................................... 10
  CONTINUOUS VARIABLES ............................................................................................................... 10
  WHAT IS AN EXPERIMENT? ................................................................................................................. 11

© 2019 by OpusWorks. All rights reserved. August, 2019

Terms of Use
This guide can only be used by those with a paid license to the corresponding course in the e-Learning curriculum produced and distributed by OpusWorks. No part of this Student Guide may be altered, reproduced, stored, or transmitted in any form by any means without the prior written permission of OpusWorks.

Trademarks
All terms mentioned in this guide that are known to be trademarks or service marks have been appropriately capitalized.

Comments
Please address any questions or comments to your distributor or to OpusWorks at info@OpusWorks.com.
Learning Objectives

Upon completion of this course, student will be able to:

• Introduced to some basic terminology used in Statistics
• Identify the different kinds of data and ways to gather or collect the data
• Begin learning ways to organize information into a manageable form for the purpose of making informed decisions

What is Statistics?

Are men paid more than women for doing the same job when all other factors such as educational level, experience, and age are considered?

What percent of all American adults read news on the Internet at least once a week?

These questions and others can be answered by using statistics.

We use statistics every day in quality control, inventory control, forecasting, market surveys, and in the creation of the Consumer Price Index.

Whether we like it or not, statistics is involved in our everyday life.

Statistics Defined

Statistics is the science of collecting, organizing, analyzing, and interpreting information. Statistics consists of methods and procedures to reduce a lot of information into a more manageable form.

Statistics has its own terminology and vocabulary. Some example terms are "population," "parameter," "random sampling," and "time-series data."

Branches of Statistics

The use of statistics will help you make informed decisions. Information is collected, organized, summarized, and analyzed in order to draw conclusions. Conclusions are fact-based and data-driven.

With this in mind, the application of statistics can be divided into two areas or branches -- "descriptive statistics" and "inferential statistics." We will start with descriptive statistics.
**Descriptive Statistics**

Descriptive Statistics is the use of graphical and numerical methods to organize, summarize and present data in a meaningful way for the purpose of making more informed decisions.

It involves using charts, tables, and graphs to summarize "raw data" for the purpose of extracting useful information.

**Inferential Statistics**

Inferential Statistics uses information gathered from sample data to make estimates, predictions, or other generalizations about a larger set of data called a "population."

Inferential statistics involves making an inference about the value of a population parameter on the basis of a sample statistic.

**Inferential Statistics in Use**

Using random sampling, a survey indicates that the woman is the primary investor in 56% of married couples. This statistic, 56%, will be used to make a prediction or inference of the population parameter, the percent of married couples where the woman is the primary investor.

**Understanding Data**

Both branches of statistics require the use of data. "Data" is information collected and evaluated. We make conclusions from data.

Data can be collected from various sources. However, some sources of data are more reliable than others.
Data Defined

When working with data, we need to know the origin of the data and what the data represents.

In statistics, there are two important terms used to describe the collection of data elements: "population" and "sample."

Population Defined

A population is a collection of objects or individuals with at least one characteristic in common. For instance, a population may be the group of people who have credit cards. Or, a population may be the collection of people who have an account at a specific bank.

A population provides all outcomes, responses, measurements or counts that are of interest. Collecting data from every element in the population is called a "census."

What is a Census?

A census is a method of data collection from a population. It is a count or measure of an entire population. Every unit within the population must be counted or measured. It is often costly and difficult to perform.
The United States Government conducts a census every ten years in order to learn the demographic make up of the United States. Census data provides information such as the number of members in a household, number of years living at a residence, household income, and the median rent paid to lease an apartment.

**What is a Sample?**

Another method for collecting data from a population is sampling. A sample is a subset of a population.

Every element in a sample is an element of the population. It represents a portion of the population.

Measures from a census or population are known as population parameters.

Measures from samples are referred to as sample statistics. For example, when referring to a census from a population, the average income of everyone in the census would be a population parameter, and the average income of 40-49 year olds is a sample statistic.

There are various sampling methods that are used by statisticians to ensure that a sample is unbiased, accurate, precise, and representative of the population. Let's learn about some different sampling techniques.

**Sampling Methods**

Listed here are some methods used to collect data by sampling. Some methods are more reliable and effective than others depending on the population being studied. Statistical expertise should be consulted prior to developing any complex sampling plan.
Random Sampling

Random sampling is the process by which every member of the population has an equal chance of being selected for the sample. Samples need to be unbiased. Faulty data will result in faulty results.

If we want to collect information from a sample of 5 employees of a company, we could use an alphabetical listing of the company directory and generate a list of 5 unique random numbers. We would then locate the individuals on the list corresponding to the record number in the spreadsheet.

<table>
<thead>
<tr>
<th>Record #</th>
<th>Last Name</th>
<th>First Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Abbot</td>
<td>Donald</td>
</tr>
<tr>
<td>2</td>
<td>Alderman</td>
<td>Barbara</td>
</tr>
<tr>
<td>3</td>
<td>Baily</td>
<td>Rick</td>
</tr>
<tr>
<td>4</td>
<td>Blanks</td>
<td>James</td>
</tr>
<tr>
<td>5</td>
<td>Brewer</td>
<td>Joan</td>
</tr>
<tr>
<td>6</td>
<td>Brewer</td>
<td>Ron</td>
</tr>
<tr>
<td>7</td>
<td>Brown</td>
<td>Ed</td>
</tr>
<tr>
<td>8</td>
<td>Brown</td>
<td>Gary</td>
</tr>
<tr>
<td>9</td>
<td>Cox</td>
<td>Ann</td>
</tr>
<tr>
<td>10</td>
<td>Elliot</td>
<td>John</td>
</tr>
<tr>
<td>11</td>
<td>Hunter</td>
<td>Jerry</td>
</tr>
<tr>
<td>12</td>
<td>Jones</td>
<td>Scott</td>
</tr>
<tr>
<td>13</td>
<td>Martin</td>
<td>John</td>
</tr>
<tr>
<td>14</td>
<td>Parks</td>
<td>Lee</td>
</tr>
<tr>
<td>15</td>
<td>Ray</td>
<td>Ryan</td>
</tr>
<tr>
<td>16</td>
<td>Smith</td>
<td>Joe</td>
</tr>
<tr>
<td>17</td>
<td>Smith</td>
<td>Wanda</td>
</tr>
<tr>
<td>18</td>
<td>Williams</td>
<td>Jack</td>
</tr>
<tr>
<td>19</td>
<td>Williams</td>
<td>Maurice</td>
</tr>
<tr>
<td>20</td>
<td>Wilson</td>
<td>James</td>
</tr>
</tbody>
</table>

Systematic Sampling

Systematic sampling is the process by which the population is ordered in some way. Members are then selected at regular intervals. A quality control inspector may collect every 45th item, for example, as it comes from a production line.

In this example, a company is interested in selecting 5 employees to serve on a committee. The employees names are arranged alphabetically, then the initial member is selected by choosing a random number for the first selection. The selection process continues by selecting at regular intervals, using a second random number, until the committee is chosen.

Cluster Sampling

Cluster sampling is when some of the clusters (group of elements) are selected, and then sampling from each of the selected clusters occurs. It is often used in populations where natural grouping of elements occurs.

Each cluster should be a small scale version of the total population.

Random sampling is used to select which clusters are sampled.

In single-stage cluster sampling, all the elements within a cluster are sampled. In two-stage cluster sampling, a random sample of elements is selected from each chosen cluster.
Stratified Sampling

Stratified sampling is used when subgroups within a population vary considerably. It is advantageous to sample each subgroup, or stratum, independently.

The process of grouping members of the population into relatively similar subgroups before sampling is called stratification. Random sampling is used to select elements from each stratum.

Stratified sampling is often used to reduce sampling error and concentrate on the important strata and ignore the irrelevant strata.

Sampling with Replacement

Sampling with Replacement is a method that selects an element from the population and replaces it prior to the selection of the next element.

In this method, the same element can be selected more than once.

Sampling without Replacement

In sampling without Replacement, an element is selected from the population and removed prior to the selection of the next element. In this method, the same element could only be selected once.

Types of Variables

Data is obtained by measuring or counting the values of one or more variables in the sample or population of interest. A variable may be classified as qualitative or quantitative.

Similarly, the data collected on a variable is called qualitative data or quantitative data.
Understanding Variables

In studying a population or a sample, we focus on one or more characteristics or properties of the units within the population.

A variable is any characteristic on which the elements of a population or sample differ from each other. For instance, we may be interested in age, gender, or years of education of a country’s workforce. Age, gender, and years of education would be defined as variables for this population.

Qualitative Data

Qualitative data cannot be measured on a numerical scale. Cereal brand, make of coffee, color of car, political party affiliation, and ZIP code are all examples of qualitative data.

Qualitative data is classified into categories and is also referred to as categorical data.

Often we assign arbitrary numerical values to qualitative data for ease of computer entry and analysis.

For instance, in assessing the quality of a product, we could assign a "0" for poor, a "1" for average, or a "2" for superior. The assigned numerical values are simply codes and cannot be added, subtracted, multiplied, or divided.

Quantitative Data

Quantitative variables provide numerical measures of the units of a population or sample. Quantitative data comes from measurements or counts.

Examples of quantitative data are weight, length, time, temperature, age, and scores on tests. Arithmetic operations such as addition and subtraction can be performed on quantitative data and provide meaningful results.
Discrete and Continuous

Quantitative variables can be divided into two types -- discrete variables and continuous variables.

Discrete variables produce discrete data which is countable.

Continuous variables produce continuous data which is measured.

Discrete Variable

Discrete variables are quantitative variables that produce a countable number of possible values.

The number of occupants in an elevator at 5 pm, the number of customer care letters sent per day, or the number of stocks whose share price increased on a given day are examples of discrete variables. In all three examples, the values are data that can be counted.

Continuous Variables

Continuous variables are quantitative variables that have an infinite number of possible values that are not countable.

Examples of continuous variables are the amount of water used per day from a water cooler, or the amount of time it takes to complete a customer service call.

These examples require a measurement. It is impossible to list all the possible values for a continuous variable.
What is an Experiment?

In business, we often define a problem, identify the variables, and discover we lack the necessary data to solve the problem. As we have seen previously, there are many sources of available data. In statistics, one important method often used to collect data is to run a statistically designed experiment.

A statistically designed experiment is when a treatment is applied to part of a population and responses are observed. The researcher exerts strict control over the entire experiment.

In some experiments, a second part of the population, called a control group, is used. The control group does not receive the treatment. Results from both groups are compared. Care must be taken in designing the experiment and the sampling methods used to ensure that both treatment and control groups are similar.

We often think of experiments as being run in a lab. But experiments are widely used in all facets of business to solve problems. For example, testing the effectiveness of a new prescription drug or changing the marketing strategy for one section of a territory will provide comparative information for effectiveness. Then the information provided will support a decision about a more effective strategy that a company should use.