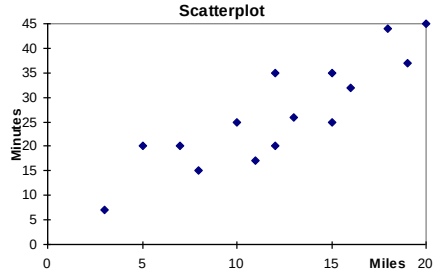
Chapter 1: Statistics



Chapter Goals

- Create an initial image of the field of statistics.
- Introduce several basic vocabulary words used in studying statistics: *population*, *variable*, *statistic*.
- Learn how to obtain sample data.

1.1 What is Statistics?

Statistics: The science of collecting, analyzing, presenting and interpreting data to make decisions.

Two areas of statistics:

Descriptive Statistics: collection, presentation, and description of sample data.

Inferential Statistics: Using sample info to make decisions and draw conclusions about populations.

Example: A recent study examined the math and verbal SAT scores of high school seniors across the country. Which of the following statements are descriptive in nature and which are inferential.

- The mean math SAT score was 492.
- The mean verbal SAT score was 475.
- Students in the Northeast scored higher in math but lower in verbal.
- 80% of all students taking the exam were headed for college.
- 32% of the students scored above 610 on the verbal SAT.
- The math SAT scores are higher than they were 10 years ago.

1.3 Population V. Sample

Population: A collection, or set, of individuals or objects or events whose properties are to be analyzed.

Two kinds of populations: *finite* or *infinite*.

Sample: A subset of the population.

Census: A survey that includes every member of the population.

Sample Survey : A survey that includes elements of a sample

Representative Sample: Represents the characteristics of the population as closely as possible. (proportional to the population)

Random Sample: Drawn in such a way that each element of the population has a chance of being selected.

1.4 Basic Terms

Element: A specific subject or object about which info is collected

Variable: A characteristic about each individual element of a population or sample.

Observation: The value of a variable for an element

Data Set : A collection of observations on one or more variables.

Experiment: A planned activity whose results yield a set of data.

Data (singular): The value of the variable associated with one element of a population or sample. This value may be a number, a word, or a symbol.

Data (plural): The set of values collected for the variable from each of the elements belonging to the sample.

Parameter: A numerical value summarizing all the data of an entire population.

Statistic: A numerical value summarizing the sample data.

***Example*: A college dean is interested in learning about the average age of faculty. Identify the basic terms in this situation.**

The variable is the age of all faculty members at the college.
A sample is any subset of that population. For example, we might select 10 faculty members and determine their age.
The observation is the "age" of each faculty member.
One element would be the age of a specific faculty member.
The data would be the set of values in the sample.
The experiment would be the method used to select the ages forming the sample and determining the actual age of each faculty member in the sample.
The parameter of interest is the "average" age of all faculty at the college.
The statistic is the "average" age for all faculty in the sample.

1.5 Types of variables:

Qualitative, or Attribute, or Categorical, Variable: A variable that categorizes or describes an element of a population.

***Note*: Arithmetic operations, such as addition and averaging, are *not* meaningful for data resulting from a qualitative variable.**

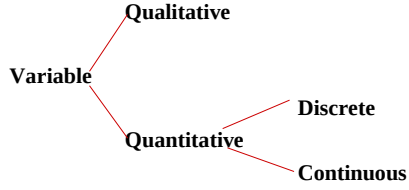
Quantitative, or Numerical, Variable: A variable that quantifies an element of a population.

***Note*: Arithmetic operations such as addition and averaging, are meaningful for data resulting from a quantitative variable.**

***Example*: Identify each of the following examples as attribute (qualitative) or numerical (quantitative) variables.**

1. **The residence hall for each student in a statistics class.**
2. **The amount of gasoline pumped by the next 10 customers at the local Unimart.**
3. **The amount of radon in the basement of each of 25 homes in a new development.**
4. **The color of the baseball cap worn by each of 20 students.**
5. **The length of time to complete a mathematics homework assignment.**
6. **The state in which each truck is registered when stopped and inspected at a weigh station.**

Quantitative variables may be further subdivided:



Discrete Variable: A quantitative variable that can assume a countable number of values. Intuitively, a discrete variable can assume values corresponding to isolated points along a line interval. That is, there is a gap between any two values.

Continuous Variable: A quantitative variable that can assume an uncountable number of values. Intuitively, a continuous variable can assume any value along a line interval, including every possible value between any two values.

Note:

1. In many cases, a discrete and continuous variable may be distinguished by determining whether the variables are related to a count or a measurement.
2. Discrete variables are usually associated with counting. If the variable cannot be further subdivided, it is a clue that you are probably dealing with a discrete variable.
3. Continuous variables are usually associated with measurements. The values of discrete variables are only limited by your ability to measure them.

Example: Identify each of the following as examples of *qualitative* or *quantitative* variables:

1. The temperature in Barrow, Alaska at 12:00 pm on any given day.
2. The make of automobile driven by each faculty member.
3. Whether or not a 6 volt lantern battery is defective.
4. The weight of a lead pencil.
5. The length of time billed for a long distance telephone call.
6. The brand of cereal children eat for breakfast.
7. The type of book taken out of the library by an adult.

Example: Identify each of the following as examples of variables that are discrete, continuous or neither:

1. The length of time until a pain reliever begins to work.
2. The number of chocolate chips in a cookie.
3. The number of colors used in a statistics textbook.
4. The brand of refrigerator in a home.
5. The overall satisfaction rating of a new car.
6. The number of files on a computer's hard disk.
7. The pH level of the water in a swimming pool.
8. The number of staples in a stapler.

Measure and Variability

- No matter what the response variable: there will always be variability in the data.
- One of the primary objectives of statistics: measuring and characterizing variability.

Example: A supplier fills cans of soda marked 12 ounces. How much soda does each can really contain?

- It is very *unlikely* any one can contains exactly 12 ounces of soda.
- There is variability in any process.
- Some cans contain a little more than 12 ounces, and some cans contain a little less.
- On the average, there are 12 ounces in each can.
- The supplier hopes there is little variability in the process, that most cans contain *close* to 12 ounces of soda.

Data Collection

- **First problem a statistician faces: how to obtain the data.**
- **It is important to obtain *good*, or *representative*, data.**
- **Inferences are made based on statistics obtained from the data.**
- **Inferences can only be as good as the data.**

Process of data collection:

1. **Define the objectives of the survey or experiment.**
   ***Example*: Estimate the average life of an electronic component.**
2. **Define the variable and population of interest.**
   ***Example*: Length of time for anesthesia to wear off after surgery.**
3. **Defining the data-collection and data-measuring schemes. This includes sampling procedures, sample size, and the data-measuring device (questionnaire, scale, ruler, etc.).**
4. **Determine the appropriate descriptive or inferential data-analysis techniques.**

1.5: Comparison of Probability and Statistics

Probability: Properties of the population are assumed known. Answer questions about the sample based on these properties.

Statistics: Use information in the sample to draw a conclusion about the population.

***Example*: A jar of M&M's contains 100 candy pieces, 15 are red. A handful of 10 is selected.**

Probability question: What is the probability that 3 of the 10 selected are red?

***Example*: A handful of 10 M&M's is selected from a jar containing 1000 candy pieces. Three M&M's in the handful are red.**

Statistics question: What is the proportion of red M&M's in the entire jar?