1 Statistics

1.1 What Is Statistics?

What is Statistics?

Statistics has become the universal language of the sciences. As potential users of statistics, we need to master both the “science” and the “art” of using statistical methodology correctly.

Careful use of statistical methods will enable us to obtain accurate information from data. These methods include
(1) carefully defining the situation,

(2) gathering data,

(3) accurately summarizing the data, and

(4) deriving and communicating meaningful conclusions.

What is Statistics?

Statistics involves information, numbers, and visual graphics to summarize this information, and its interpretation.

The field of statistics can be roughly subdivided into two areas: descriptive statistics and inferential statistics.

Descriptive statistics is what most people think of when they hear the word *statistics*. It includes the collection, presentation, and description of sample data.

What is Statistics?

The term **inferential statistics** refers to the technique of interpreting the values resulting from the descriptive techniques and making decisions and drawing conclusions about the population.

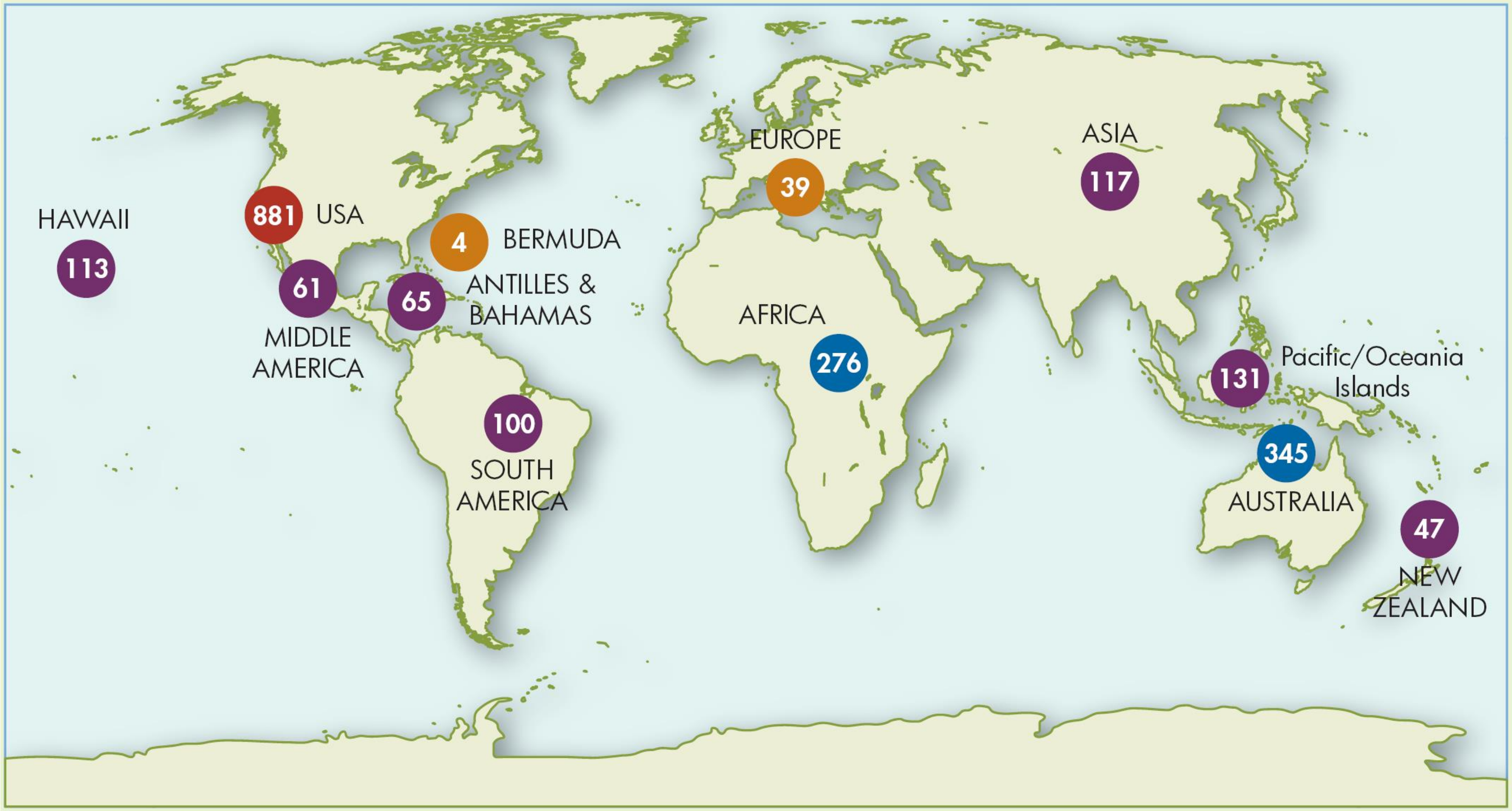
Statistics The science of collecting, describing, and interpreting data.

Applied Example 4 – *Shark Attack*

Consider the **International Shark Attack File** (ISAF), which is a compilation of all known shark attacks that is administered by the American Elasmobranch Society and the Florida Museum of Natural History and is shown in the chart and graph below.

Territory	Total Attacks	Fatal Attacks	Last Fatality
USA (w/out Hawaii)	881	38	2005
Australia	345	135	2006
Africa	276	70	2004
Asia	117	55	2000
Pacific/Oceania Islands (w/out Hawaii)	131	50	2007
Hawaii	113	15	2004
South America	100	23	2006
Antilles & Bahamas	65	19	1972
Middle America	61	31	1997
New Zealand	47	9	1968
Europe	39	19	1984
Bermuda	4	0	
Unspecified	20	6	1965
WORLD	2,199	470	2007

Applied Example 4 – *Shark Attack*



Source: http://www.flmnh.ufl.edu/fish/sharks/statistics/GAttack/World.htm

Applied Example 4 – *Shark Attack*

Common sense? Using common sense while reviewing the graph, one would certainly stay away from the United States if he or she enjoys the ocean. The United States has two-fifths of the world's shark attacks! U.S. waters must be full of sharks and the sharks must be mad!

Common sense—remember? What is going on with this graph? Is it a bit misleading? What else could be influencing the statistics shown here? First, one must take into consideration how much coastline of a country or continent comes in contact with an ocean.

Applied Example 4 – *Shark Attack*

Secondly, who is keeping track of these attacks? Notice the source of the map and chart, the Florida Museum of Natural History, a museum in the United States.

Apparently, the United States is trying to keep track of unprovoked shark attacks. What else is different about the United States as compared to the other areas?

Is the ocean a recreational area in the other places? What is the economy of these other areas and/or who is keeping track of their shark attacks?

What is Statistics?

Remember to consider the source when reading a statistical report. Be sure you are looking at the complete picture.

The following are a few examples of how and where statistics are used:

- In education, descriptive statistics are frequently used to describe test results.

- In science, the data resulting from experiments must be collected and analyzed.

What is Statistics?

- In government, many kinds of statistical data are collected all the time. In fact, the U.S. government is probably the world's greatest collector of statistical data.

To further continue our study of statistics, we need to "talk the talk." Statistics has its own jargon, terms beyond *descriptive statistics* and *inferential statistics,* that needs to be defined and illustrated. The concept of a population is the most fundamental idea in statistics.

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

What is Statistics?

The population is the complete collection of individuals or objects that are of interest to the sample collector.

The population of concern must be carefully defined and is considered fully defined only when its membership list of elements is specified.

The set of "all students who have ever attended a U.S. college" is an example of a well-defined population.

Typically, we think of a population as a collection of people. However, in statistics the population could be a collection of animals, manufactured objects, whatever.

What is Statistics?

For example, the set of all redwood trees in California could be a population.

There are two kinds of populations: finite and infinite. When the membership of a population can be (or could be) physically listed, the population is said to be **finite**.

When the membership is unlimited, the population is **infinite**. The books in your college library form a finite population; the OPAC (Online Public Access Catalog, the computerized card catalog) lists the exact membership.

What is Statistics?

All the registered voters in the United States form a very large finite population; if necessary, a composite of all voter lists from all voting precincts across the United States could be compiled.

On the other hand, the population of all people who might use aspirin and the population of all 40-watt light bulbs to be produced by General Electric are infinite.

Large populations are difficult to study; therefore, it is customary to select a *sample* and study the data in that sample.

What is Statistics?

Sample A subset of a population.

A sample consists of the individuals, objects, or measurements selected from the population by the sample collector.

Variable (or response variable) A characteristic of interest about each individual element of a population or sample.

A student's age at entrance into college, the color of the student's hair, the student's height, and the student's weight are four variables.

What is Statistics?

Data value 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.

For example, Bill Jones entered college at age “23,” his hair is “brown,” he is “71 inches” tall, and he weighs “183 pounds.” These four data values are the values for the four variables as applied to Bill Jones.

What is Statistics?

Data The set of values collected from the variable from each of the elements that belong to the sample. Once all the data are collected, it is common practice to refer to the set of data as the sample.

The set of 25 heights collected from 25 students is an example of a set of data.

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

An experiment includes the activities for both selecting the elements and obtaining the data values.

What is Statistics?

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

The “average” age at time of admission for all students who have ever attended our college and the “proportion” of students who were older than 21 years of age when they entered college are examples of two population parameters.

A parameter is a value that describes the entire population. Often a Greek letter is used to symbolize the name of a parameter.

What is Statistics?

For every parameter there is a *corresponding sample statistic.* The statistic describes the sample the same way the parameter describes the population.

Statistic A numerical value summarizing the sample data.

The “average” height, found by using the set of 25 heights, is an example of a sample statistic. A statistic is a value that describes a sample. Most sample statistics are found with the aid of formulas and are typically assigned symbolic names that are letters of the English alphabet (for example, $\bar{x}$, *s*, and *r*).

Example 5 – *Applying the Basic Terms*

A statistics student is interested in finding out something about the average dollar value of cars owned by the faculty members of our college.

Each of the eight terms just described can be identified in this situation.

1. The *population* is the collection of all cars owned by all faculty members at our college.

2. A *sample* is any subset of that population. For example, the cars owned by members of the mathematics department is a sample.

Example 5 – *Applying the Basic Terms*

3. The *variable* is the “dollar value” of each individual car.

4. One *data value* is the dollar value of a particular car. Mr. Jones’s car, for example, is valued at $9400.

5. The *data* are the set of values that correspond to the sample obtained (9400; 8700; 15,950; . . .).

6. The *experiment* consists of the methods used to select the cars that form the sample and to determine the value of each car in the sample. It could be carried out by questioning each member of the mathematics department, or in other ways.

Example 5 – *Applying the Basic Terms*

7. The *parameter* about which we are seeking information is the “average” value of all cars in the population.

8. The *statistic* that will be found is the “average” value of the cars in the sample.

What is Statistics?

Note

If a second sample were to be taken, it would result in a different set of people being selected—say, the English department—and therefore a different value would be anticipated for the statistic “average value.” The average value for “all faculty-owned cars” would not change, however.

There are basically two kinds of variables:
(1) variables that result in *qualitative* information and

(2) variables that result in *quantitative* information.

What is Statistics?

Qualitative, or attribute, or categorical, variable
A variable that describes or categorizes an element of a population.

Quantitative, or numerical, variable A variable that quantifies an element of a population.

A sample of four hair-salon customers was surveyed for their “hair color,” “hometown,” and “level of satisfaction” with the results of their salon treatment.

What is Statistics?

All three variables are examples of qualitative (attribute) variables because they describe some characteristic of the person, and all people with the same attribute belong to the same category.

The data collected were {blonde, brown, black, brown}, {Brighton, Columbus, Albany, Jacksonville}, and {very satisfied, satisfied, somewhat satisfied, not satisfied}.

The “total cost” of textbooks purchased by each student for this semester’s classes is an example of a quantitative (numerical) variable.

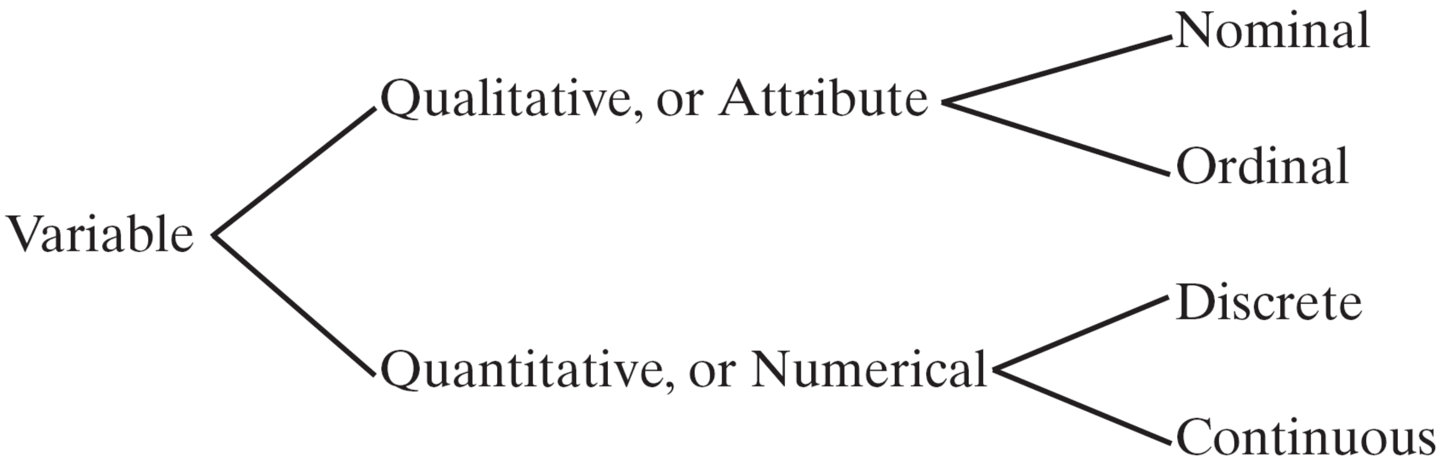
What is Statistics?

A sample resulted in the following data: \$238.87, \$94.57, \$139.24. [To find the "average cost," simply add the three numbers and divide by 3:
(238.87 + 94.57 + 139.24)/3 = \$157.56.]

Note
Arithmetic operations, such as addition and averaging, are meaningful for data that result from a quantitative variable.

What is Statistics?

Each of these types of variables (qualitative and quantitative) can be further subdivided as illustrated in the following diagram.



What is Statistics?

Qualitative variables may be characterized as nominal or ordinal.

Nominal variable A qualitative variable that characterizes (or describes, or names) an element of a population. Not only are arithmetic operations not meaningful for data that result from a nominal variable, but an order cannot be assigned to the categories.

What is Statistics?

In the survey of four hair-salon customers, two of the variables, “hair color” and “hometown,” are examples of nominal variables because both name some characteristic of the person and it would be meaningless to find the sample average by adding and dividing by 4.

For example, (blonde + brown + black + brown)/4 is undefined. Furthermore, color of hair and hometown do not have an order to their categories.

Ordinal variable A qualitative variable that incorporates an ordered position, or ranking.

What is Statistics?

In the survey of four hair-salon customers, the variable "level of satisfaction" is an example of an ordinal variable because it does incorporate an ordered ranking: "Very satisfied" ranks ahead of "satisfied," which ranks ahead of "somewhat satisfied."

Another illustration of an ordinal variable is the ranking of five landscape pictures according to someone's preference: first choice, second choice, and so on.

Quantitative or numerical variables can also be subdivided into two classifications: *discrete* variables and *continuous* variables.

What is Statistics?

Discrete variable A quantitative variable that can assume a countable number of values. Intuitively, the discrete variable can assume any 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, the continuous variable can assume any value along a line interval, including every possible value between any two values.

What is Statistics?

In many cases, the two types of variables can be distinguished by deciding whether the variables are related to a count or a measurement.

The variable “number of courses for which you are currently registered” is an example of a discrete variable; the values of the variable may be found by counting the courses. (When we count, fractional values cannot occur; thus, gaps can occur between the values.)

What is Statistics?

The variable “weight of books and supplies you are carrying as you attend class today” is an example of a continuous random variable; the values of the variable may be found by measuring the weight. (When we measure, any fractional value can occur; thus, every value along the number line is possible.)

When trying to determine whether a variable is discrete or continuous, remember to look at the variable and think about the values that might occur.

Do not look at only data values that have been recorded; they can be very misleading.

What is Statistics?

Consider the variable "judge's score" at a figure-skating competition. If we look at some scores that have previously occurred, 9.9, 9.5, 8.8, 10.0, and we see the presence of decimals, we might think that all fractions are possible and conclude that the variable is continuous.

This is not true, however. A score of 9.134 is impossible; thus, there are gaps between the possible values and the variable is discrete.

What is Statistics?

Note
Don’t let the appearance of the data fool you in regard to their type. Qualitative variables are not always easy to recognize; sometimes they appear as numbers. The sample of hair colors could be coded: 1 = black,
2 = blonde, 3 = brown.

The sample data would then appear as {2, 3, 1, 3}, but they are still nominal data. Calculating the “average hair color” [(2 + 3 + 1 + 3)/4 = 9/4 = 2.25] is still meaningless.

What is Statistics?

The hometowns could be identified using ZIP codes. The average of the ZIP codes doesn't make sense either; therefore, ZIP code numbers are nominal too.

Let's look at another example. Suppose that after surveying a parking lot, I summarized the sample data by reporting 5 red, 8 blue, 6 green, and 2 yellow cars.

You must look at each individual source to determine the kind of information being collected. One specific car was red; "red" is the data value from that one car, and red is an attribute. Thus, this collection (5 red, 8 blue, and so on) is a summary of nominal data.

What is Statistics?

Another example of information that is deceiving is an identification number. Flight #249 and Room #168 both appear to be numerical data.

However, the numeral 249 does not describe any property of the flight—late or on time, quality of snack served, number of passengers, or anything else about the flight.

The flight number only identifies a specific flight. Driver's license numbers, Social Security numbers, and bank account numbers are all identification numbers used in the nominal sense, not in the quantitative sense.

What is Statistics?

Remember to inspect the individual variable and one individual data value, and you should have little trouble distinguishing among the various types of variables.

Applied Example 6 – *The Big Paycheck*

World's Highest Paid Athletes




Forbes' list of the highest-paid athletes looks at earnings derived from salaries, bonuses, prize money, endorsements, and licensing income between June 2008 and June 2009 and does not deduct for taxes or agents' fees.

Here are the Top 5:

Rank	Athlete	Sport	Earnings
1	Tiger Woods	Golf	$110 million
2	Kobe Bryant	Basketball	$45 million
2	Michael Jordan	Basketball	$45 million
2	Kimi Raikkonen	Auto Racing	$45 million
5	David Beckham	Soccer	$42 million

Source: http://www.gettlisty.com/preview/highest-paid-athletes/

Applied Example 6 – *The Big Paycheck*

Let's face it; most of us dream about making this much in our whole lifetime.

If one is gainfully employed every year from 21 years old to 62 and makes $1 million each year, that would be $42 million for one's whole life.

Most of us cannot even wrap our heads around that concept. You probably are thinking, "Sign me up to be a superstar athlete!"

Applied Example 6 – *The Big Paycheck*

Let’s see how we can apply our new terminology to the “Big Paycheck.” First, the general population of interest would be professional athletes.

Furthermore, the information in the above table demonstrates the various types of variables. The athlete’s name is generally not considered to be a variable; it is for the purpose of identification only.

Applied Example 6 – *The Big Paycheck*

The other three kinds of information are variables:

1. Rank is qualitative and an ordinal variable since it incorporates the concept of ordered position.

2. Sport is qualitative and a nominal variable since it describes the athlete’s sport.

3. Earnings is quantitative and a continuous variable as it is a measure of the athlete’s income. Amounts of money are generally considered continuous since fractional parts of dollars are possible, even though the amount is generally rounded to the nearest dollar or penny.