1.1 What is Statistics?

_____ - study of how to collect, organize, analyze, and interpret numerical information from data.

_____ - consists of methods for organizing, displaying, and describing data by using tables, graphs, and summary measures.

Suppose we have information on the test scores of students enrolled in a statistics class. In statistical terminology, the whole set of numbers that represents the scores of students is called a _____, the name of each student is called an _____, and the score of each student is called an _____.

Example: The total number of home runs hit by Aaron Judge in last night's game.

_____ - consists of methods that use sample results to help make decisions or _____ about a population.

Example: The total number of home runs that will be hit by Aaron Judge by the end of this season.

For each of the following, identify as descriptive or inferential statistics:
1) The high temperature recorded over the last week in Central Park. _____
2) The forecast high temperature predicted for tomorrow. _____
3) The expected score on the SAT by a student taking it for a 3rd time. _____
4) The number of times Dwayne Wade shot free throws during last years NBA Finals._____
5) The number of home runs Albert Pujols will hit over his career. _____
6) Projected sales of the new Lexus model over the next five years. _____
7) The last year Jim Carrey won an Academy Award. _____
8) The number of season tickets the Mets are likely to sell in 2019. _____
9) The average time at which Mary left for school over the last month. _____
10) The final exam scores of students in this class on the Prob. & Stat. Final. _____

Populations & Samples:

_____ - people or objects included in the study.

_____ - all measurements or observations of interest.

_____- a part of the population.

_____- numerical measure that describes an aspect of a population.

_____ - numerical measure that describes an aspect of a sample.

** When referring to a population or sample, be sure to give the *quantity* being measured or the *quality* being observed.

_____ - characteristic of the individual to be measured or observed.

Four types of variables:

a) _____ - has a numerical measurement for which operations such as addition or averaging makes sense

b) _____ - describes an individual by placing the individual in a category or group, such as male or female.

c) _____- variable whose values are countable. A discrete variable
can assume only certain values with no intermediate values.

d) _____- can assume any numerical value over a certain interval or intervals.

For each of the following examples state the individual, variable (qualitative or quantitative), population, and sample (if mentioned). Then, note if the parameter or statistic is used:

1) The department of tropical agriculture is doing a study of pineapples in an experimental field. In this case the data under consideration are the individual weights of all pineapples in the field.

Individual - _____

Variable - _____

Population - _____

Sample - _____

Parameter or Statistic - _____

2) Television station QUE wants to know the proportion of TV owners in Virginia who watch the station's new program at least once a week. The station asked a random group of 1000 TV owners if Virginia if they watch the program at least once a week.

Individual - _____

Variable - _____

Population - _____

Sample - _____

Parameter or Statistic - _____

3) The department of education is conducting a study which is looking at seniors in high school. They would like to know how
many high school seniors in this country have scored above a 1300
on the SAT.

Individual - _____

Variable - _____

Population - _____

Sample - _____

Parameter or Statistic - _____

4) Airlines are trying to figure out how much additional money people while on the flight (ex. movies, radio, drinks, etc…). The airline companies have taken a survey of the last 500 customers they have served.

Individual - _____

Variable - _____

Population - _____

Sample - _____

Parameter or Statistic - _____

5) Major League Baseball is conducting a study on how many players in the league sign autographs before each game. They have asked each player to give the total amount of time each game spent to signing autographs.

Individual - _____

Variable - _____

Population - _____

Sample - _____

Parameter or Statistic - _____

1.1 Homework:

1) Marketing: Fast Food *USA Today* reported that 44.9% of those surveyed (1261 adults) ate in fast-food restaurants from one to three times each week.
(a) Identify the variable.

(b) Is the variable quantitative or qualitative?

(c) What is the implied population?

2) Advertising: Auto Mileage What is the average miles per gallon (mpg) for all new cars? Using *Consumer Reports,* a random sample of 35 new cars gave an average of 21.1 mpg.
(a) Identify the variable.

(b) Is the variable quantitative or qualitative?

(c) What is the implied population?

3) _Ecology: Wetlands_ Government agencies carefully monitor water quality and its effect on wetlands. Of particular concern is the concentration of nitrogen in water draining from fertilized lands. Too much nitrogen can kill fish and wildlife. Twenty-eight samples of water were taken at random from a lake. The nitrogen concentration (milligrams of nitrogen per liter of water) was determined for each sample.
(a) Identify the variable.

(b) Is the variable quantitative or qualitative?

(c) What is the implied population?

4) _Archaeology: Ireland_ The archaeological site of Tara is more than 4000 years old. Tradition states that Tara was the seat of the high kings of Ireland. Because of its archaeological importance, Tara has received extensive study. Suppose an archaeologist wants to estimate the density of ferromagnetic artifacts in the Tara region. For this purpose, a random sample of 55 plots, each of size 100 square meters, is used. The number of ferromagnetic artifacts for each plot is determined.
(a) Identify the variable.

(b) Is the variable quantitative or qualitative?

(c) What is the implied population?

1.2 Statistical Studies

Basic Statistical Terms:

_____ - specific subject or object about which the information is collected.

_____ - characteristic under study that assumes different values for elements.

_____ - value of a variable for an element.

_____- collection of observations on one or more variables.

Examples: Look at the table below and point out the data set, elements, variables, and observations.

2004 Profits of Seven U.S. Companies

Company	2004 Profits (millions of dollars)
Wal-Mart Stores	10,267
Exxon	25,330
General Electric	16,593
Citigroup	17,046
Home Depot	5,001
Pfizer	11,361
Target	3,198

Number of Mosquito Bites Reported on a Day in June

City	Number of Bites
Center City	47
Elm Grove	32
Franklin Square	51
Bay City	44
Oakdale	12
Sand Point	32

Oil Reserve Totals in Six Countries

Country	Oil Reserves
Saudi Arabia	261.7
Iraq	112.0
Kuwait	97.7
Iran	94.4
United Arab Emirates	80.3
Venezuela	64.0

Basic Guidelines for Planning a Statistical Study:

1) _____ the individuals or objects of interest.

2) Specify the _____ as well as protocols for taking measurements or making observations.

3) Determine if you will use an entire population or a representative _____. If using a sample, decide on a viable sampling method.

4) In your data collection plan, address issues of ethics, subject confidentiality, and privacy. If you are collecting data at a business, store, college, or other institution, be sure to be courteous and to obtain _____ as necessary.

5) Collect the data.

6) Use appropriate _____ statistics methods and make decisions using appropriate inferential statistics methods.

7) Note any concerns you might have about your data collection methods and list any recommendations for future studies.

Statistical Study Methods

_____ – gaining an accurate picture of the population without disturbing it or disturbing it as little as possible.

_____ – imposing some treatment on units or subjects in order to observe a given response.

_____ – numerical facsimile of real-world phenomena. - "dry lab" (studying real world problems but not going forth with it)

_____ – measurements from the entire population are used.

Which techniques (sample, experiment, simulation, or census) for gathering data do you think might be the most appropriate for the following studies? Why? *There can be more than one correct answer

1) Study the effect of stopping the cooling process of a nuclear reactor.

2) Study the amount of time college students taking a full course load spend watching television.

3) Study on the effects of a new drug on children with ADHD.

4) Study on the effects of freezing temperatures on human beings.

5) Study on the amount of money spent by American on home improvements.

6) Study the effects of a calcium supplement given to young girls.

7) Study of the credit hours of each student enrolled at the University of North Carolina at the end of the semester.

8) Study on the heights of skyscrapers in New York City.

9) Study which explains the effects of dehydration on athletes.

10) Study on students at St. Francis who drive cars.

_____ – asking questions – used after you have decided on sampling or experimentation

Some Problems of Surveys:
1) _____ – individuals either cannot be contacted or refuse to participate.
2) _____ of response – respondents may lie intentionally or inadvertently.
3) _____ – Respondents may lie intentionally or inadvertently.
4) _____ – The question may be worded in such a way as to elicit a specific response. The order of the questions might lead to biased responses.
5) _____ – Words such as "often," "seldom," and "occasionally" mean different things to different people.
6) Interviewer _____ – Factors such as tone of voice, body language, dress, gender, authority, and ethnicity of the interviewer might influence responses.
7) _____ – Individuals with strong feelings about a subject are more likely than others to respond. Such a study is interesting but not reflective of the population.

Look at these examples and make comments about the usefulness of the data collected:
1) A uniformed law officer interviews a group of 20 college freshmen. She asks each one his or her name and then if he or she has used an illegal drug in the last month.

2) Jessica saw some data that show that cities with more low-income housing have more homeless people. Does building low-income housing cause homelessness?

3) A survey about food in the student cafeteria was conducted by having forms available for customers to pick up at the cash register. A drop out box for completed forms was available outside the cafeteria.

4) Extensive studies on coronary problems were conducted using men over age 50 as the subjects.

5)A survey is conducted in which Americans are asked how much money they put into automobiles each year. This survey was completed by 100 New Yorkers.

6) Fast-Food companies were compared with Health-Food companies to look at the overall nutritional value of the food. To conduct this study the researchers compared McDonald's with Cherry Valley.

7) As technology is advancing, so are education techniques. Students with autism can benefit from a curriculum which includes technology. To study this, researchers incorporated technology into a classroom at PS 145.

Levels of Measurement:

1) _____ – lowest level, "in name only"- consist of names only or qualities with no implied criteria by which the data can be identified as greater than or less than other data items.
Example: Mets, Yankees, and Jets are the names of three professional teams from the population of all professional teams in New York.

2) _____ – data may be arranged in some order, but actual differences between data values either cannot be determined or are meaningless.
Example: In CHSAA Football, St. Francis Prep ranks 1st, St. Anthony's 2nd, Mount St. Michael 3rd, and Holy Cross 99th.

3) _____ – has same qualities as ordinal level, but it has the additional property that meaningful differences between data values can be computed.
Example: Body temperature (in degrees Celsius) of trout swimming in the Yellowstone River.

4) _____ – highest level, similar to the interval level, but includes an inherent zero as a starting point for all measurements.
Example: Length of trout swimming in the Yellowstone River.

For each of the following examples, indicate the corresponding level of measurement:

1) Length of time it takes to run a mile. _____
2) The name of your older sibling. _____
3) Age of people in this class. _____
4) The years that the Tarheels have won the national championship. _____
5) The food in the cafeteria can be rated: poor, acceptable, good. _____
6) Hyundai produces the Sonata, Elantra, and Azera. _____
7) In a high school graduating class, Joe ranks 45th and Jen ranks 12th. _____
8) Amount of times that Democrats have won the presidential election. _____
9) Length of trout swimming in the Yellowstone River. _____
10) Music, Math, and Science are three subjects that Sean goes to summer school for. _____

_____– data collected on different elements at the same point in time or for the same period of time.

Example:

Talk Show	Average Daily Viewers (millions)
The Oprah Winfrey Show	8.6
Dr. Phil	6.5
Live With Regis and Kelly	4.7
Maury	4.1
Jerry Springer	3.3
Montel Williams	3.2
Ellen	2.2

_____– data collected on the same element for the same variable at different points in time or for different periods of time.

Example:

Recorded Car Accidents in Queens

Year	Number of Collissions
1990	1,990
1995	2,775
2000	6,323
2002	6,556

1.2 Homework:

1) Categorize these measurements associated with student life according to level: nominal, ordinal, interval, or ratio.
(a) Length of time to complete an exam

(b) Time of first class

(c) Major field of study

(d) Course evaluation scale: poor, acceptable, good

(e) Score on last exam (based on 100 possible points)

(f) Age of student

2) Categorize these measurements associated with a robotics company according to level: nominal, ordinal, interval, or ratio.
(a) Salesperson's performance: below average, average, above average

(b) Price of company's stock

(c) Names of new products

(d) Temperature (°F) in CEO's private office

(e) Gross income for each of the past 5 years

3) Categorize these measurements associated with fishing according to level: nominal, ordinal, interval, or ratio.

(a) Species of fish caught: perch, bass, pike, trout

(b) Cost of rod and reel

(c) Time of return home

(d) Guidebook rating of fishing area: poor, fair, good

(e) Number of fish caught

(f) Temperature of water

4) You are interested in the weights of backpacks students carry to class and decide to conduct a study using the backpacks carried by 30 students.

(a) Give some instructions for weighing the backpacks. Include unit of measure, accuracy of measure, and type of scale.

(b) Do you think each student asked will allow you to weigh his or her backpack?

(c) Do you think telling students ahead of time that you are going to weigh their backpacks will make a difference in the weights?