Welcome to Statistics

Golder College Prep
2015-2016

Dear Learner,

Welcome to senior year and statistics. This is unlike any math course you have ever taken. Statistics is important to nearly every aspect in life: sports, entertainment, engineering, art, political science, finance, banking, education, culture, and on and on.

You might be wondering, "What is statistics?" To put it simply, statistics is the study of populations through analysis of data and data are numbers with a context.

Nothing will be taught in a vacuum. Everything we do has a purpose and a context. We will collect and describe data. We will study chance behavior and probability. We will make inference–make conclusions based on data.

I promise to be your guide as we explore statistics together. All I ask of you is to put forth strong and consistent effort and keep your mind open.

Enjoy your summer.

Sincerely,

Mr. Coleman

About the assignment:
The purpose of this assignment is to give you experience with the four main themes of statistics: exploring data, analyzing data, probability, and designing studies. It is divided into 5 sections. Answer the questions on your own paper.

Time frame:
Each section should take approximately 1 to 1.5 hours.

Due Date:
First day of class

For help on the assignment: I encourage you to reach out to each other as we often learn best from our peers. Use your resources: peers, parents, the Internet, past experiences, the textbook, etc.

For participation points: Use the class stream on Google Classroom to post questions and help each other. I will be monitoring the boards to track participation over the summer.

Register for the class at:
classroom.google.com

Class Codes:

AP Statistics: `ku2fjn8`

Statistics: `ow5kgv1`

Supplies for home:
- Textbook: The Practice of Statistics (TPS4e)

Supplies to bring to class:
- TI-83/84 Graphing Calculator
- Pen & Pencil
- Notebook

You can reach me by email this summer. Expect a reply within a few days. Also, reach out to your classmates in Google Classroom or any other means.

Contact Info:
bcoleman@noblenetwork.org

PART 1: Where do data come from?

Directions: If you drink Snapple, then you've probably noticed a "Real Fact" printed on each bottle cap. An example of one of these "Real Facts" is: *Licking a stamp burns 10 calories.*

Consider the following "Real Facts." For each one, (a) discuss whether you believe the claim and why, and (b) suggest how you might find data to support or oppose the "fact."

1. The average American walks 18,000 steps a day.
2. August has the highest percent of births.
3. The average person spends about 2 years on the phone in a lifetime.
4. Termites eat through wood twice as fast when listening to rock music.
5. 27% of all Americans have spent at least one night in jail.

Go to the Snapple Real Facts Page (http://www.snapple.com/real-facts/cap-view)

6. Find two Real Facts that are different from the ones above that involve a numerical result. Record each fact. Then tell how you think the data were produced and whether the results are logically sound.

PART 2: Talking about data

Statistics is the art and science of dealing with data. Data are made on **individuals** and organized in **variables**

Individuals are the objects described by a set of data. Individuals may be people, animals, or things.

A **variable** is any characteristic of an individual. A variable can take different values for different individuals. Variables can take two forms: **categorical** or **quantitative.**

A **categorical variable** places an individual into one of several groups or categories.

A **quantitative variable** takes numerical values we use mathematical operations on quantitative variables that can help describe the data set.

Directions: For the following data sets describe the (a) individuals, (b) all variables given for the individuals, and (c) whether each variable is categorical or quantitative.

Example: Here are a few lines from my AP Statistics grade book:

Name	Sex	Advisor	GPA	Period	Test Grade
Armstrong, M	F	Ceven	3.82	3	81
Chi, S	F	Mack	2.24	2	92
Johnson, T	M	French	2.86	4	87

Individuals: Students in Mr. Coleman's AP Statistics class
Variables: Name (categorical), Sex (categorical), Advisor (categorical), GPA (quantitative), period (categorical), test grade (quantitative)

1. The badminton team is planning to go to Grand Rapids, Michigan for a three-day tournament. The sponsor of the tournament provided a list of hotels to stay in the following table.

Hotel	Pool	Internet ($/day)	Distance to tournament	Exercise room?	Room rate ($/day)
Comfort Inn	Out	0	8.2 mi	Y	149
Fairfield Inn	In	0	8.3 mi	Y	119
Chase Suite Hotel	Out	15	1.5 mi	N	139
Marriot	In	9.95	0.0 mi	Y	145

   a. What individuals does this data set describe?
   b. Clearly identify each of the variables. Which are categorical and which are quantitative?
   c. If you were a member of the badminton team, which hotel would you recommend? Why?

2. Popular magazines rank colleges and universities on their "academic quality." Describe four variables that you would like to see measures for each college if you were choosing where to apply. Identify each as categorical or quantitative. Give reasons for your choices.

3. According to the "Real Facts" page at www.snapple.com, Americans on average eat 18 acres of pizza every day.
   a. How do you think the data were produced?
   b. Do you think the data results support Snapple's claim? Why or why not?

PART 3: How do we get data?

There are several methods from which we get data. Sometimes all an observer needs to do is watch.

An **observational study** observes individuals and measures variables of interest but does not attempt to influence the responses. The purpose of an observational study is to describe a group or situation.

An **experiment** deliberately imposes some treatment on individuals in order to observe their responses. The purpose of an experiment is to study whether the treatment causes a change in the response.

You don't need to drink an entire pitcher of lemonade to know whether it's too sweet. A few sips can tell you what you want to know. This is the idea of **sampling**. If we want to know about a larger group of individuals we might take a smaller representative group within that group to make our judgments.

The **population** in a statistical study is the entire group of individuals about which we want to find information. The population is the group we want to study.

A **sample** is a part of the population from which we actually collect information, which is then used to draw conclusions about the whole.

A **census** is if we take a survey of an *entire* population

Short-Answer Question 1: What are the benefits and detriments to observational studies and experiments? What type of scenario would make a researcher prefer an observational study? When would a researcher prefer an experiment?

Short-Answer Question 2: Why would a researcher take a sample as opposed to a census? What benefits do samples provide? What are the limitations samples and what are the limitations to a census? Which is a more accurate way of describing a population?

Directions: In questions 1 & 2 identify the population and the sample as exactly as possible.

1. A furniture maker buys hardwood in large batches. The supplier is supposed to dry the wood before shipping. The furniture maker chooses five pieces of wood from each batch and tests their moisture content. If any piece exceeds 12% moisture content, the entire batch is sent back.

2. An insurance company wants to monitor the quality of its procedures for handling loss claims from its auto insurance policyholders. Each month the company selects a sample from all auto insurance claims filed that month to examine the accuracy and promptness from which they are handled.

3. Word-length can sometimes distinguish different types of writing. A student interested in this fact wants to study the lengths of words used by JK Rowling in her Harry Potter book series. The student opens a Harry Potter book at random and records the lengths of the first 50 words on the page.
    a. What is the population in this study? What is the sample?
    b. What variable does this student measure?

4. Go to the Gallup Poll Web site (gallup.com) and locate the results of a sample survey on a topic that interests you.
    a. What questions did the survey include?
    b. Who were the individuals in the sample?
    c. Summarize one or two of the important results from the survey. Be sure to cite source information for your article, including the title, date, and web address. *Note: this information is usually found in a PDF at the bottom of a survey of interest where you can click a link regarding survey methodology, question responses, and trends.*

5. To study the effect of living in public housing on family stability in low-income households, researchers obtain a list of all applicants for public housing in Chicago last year. Some applicants were accepted, while the housing authority turned others down. The researchers interview both groups and compare them. Is this an experiment or an observational study? Explain your answer.

6. An oldwives tale states that the color red makes bees angry. Here's a method I've designed to test that claim. I'll select half of my students (by drawing random names) to wear red clothes and the other half to wear white clothes. Then I'll turn a bunch of bees loose in the classroom and record how many times each student is stung. (Unrelated: bring a white and red t-shirt to school… uh… for science.)
    a. Is this an observational study or experiment? Why?
    b. What variables are recorded?
    c. If students wearing red clothes are stung much more often than students wearing white, can we conclude that the color red *causes* bees to sting more? Why or why not?
    d. Comment on any flaws you see with my methods.

PART 4: Critiquing Student Work

One of the best ways to learn is by critiquing and reflecting on the work of others. The following page shows the work of two students on a free-response question. The importance of the following exercise is not for you to understand the content asked in the question, but for you to get a feel of how to properly answer a complex statistics question. Take a look at the following student work.

1. What qualities of STUDENT A's work make you believe she was successful in answering the question?
2. What qualities of STUDENT B's work make you believe she was successful in answering the question?
3. What are the similarities in each student's work? Where do they differ?
4. If a rubric were created to grade the student responses what categories do you think would be included on the rubric?
5. Which student deserves the higher score? Explain your reasoning.

STUDENT A

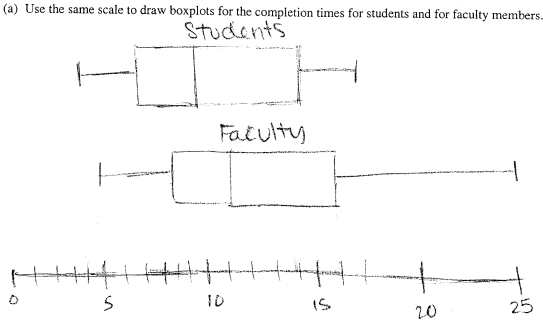
1 of 2

2 of 2

5. At a school field day, 50 students and 50 faculty members each completed an obstacle course. Descriptive statistics for the completion times (in minutes) for the two groups are shown below.

	Students	Faculty Members
Mean	9.90	12.09
Median	9.25	11.00
Minimum	3.75	4.50
Maximum	16.50	25.00
Lower quartile	6.75	8.75
Upper quartile	13.75	15.75

(a) Use the same scale to draw boxplots for the completion times for students and for faculty members.



Students
Min 3.75
Q_1 6.75
Med 9.25
Q_3 13.75
max 16.5

Faculty
min 4.5
Q_1 8.75
med 11
Q_3 15.75
max 25

GO ON TO THE NEXT PAGE.

(b) Write a few sentences comparing the variability of the two distributions.

The students range is 12.75 min and the faculty's range is 20.5 min. This is probably due to the fact that students are more active and use to physical challenges in their youth than teachers and other faculty members.
It is interesting to note that the mean, median, minimum and Q_1 and Q_3 scores are quite close. the difference in range comes because one teacher took 25 min to finish the race.

(c) You have been asked to report on this event for the school newspaper. Write a few sentences describing student and faculty performances in this competition for the paper.

The students have a range of 12.75 min between the shortest & longest times while the teachers have a range of 20.5 min. While this seems like a large difference it may be noted that the means, medians and upper and lower quartile values are quite close. The large difference is because one teacher (who shall remain nameless) took a whopping 25 minutes to finish the obstacle course.

GO ON TO THE NEXT PAGE.

STUDENT B

1 of 2

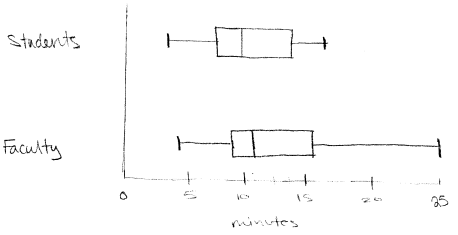
2 of 2

5. At a school field day, 50 students and 50 faculty members each completed an obstacle course. Descriptive statistics for the completion times (in minutes) for the two groups are shown below.

	X Students	Y Faculty Members	X-Y Dif
Mean	9.90	12.09	-2.19
Median	9.25 ✓	11.00 ✓	-1.75
Minimum	3.75 ✓	4.50 ✓	-0.75
Maximum	16.50 ✓	25.00 ✓	-8.5
Lower quartile	6.75 ✓	8.75 ✓	-2
Upper quartile	13.75 ✓	15.75 ✓	-2
Range	12.75	20.5	-7.75

IQR: 7 (both)

(a) Use the same scale to draw boxplots for the completion times for students and for faculty members.



GO ON TO THE NEXT PAGE.

(b) Write a few sentences comparing the variability of the two distributions.

The faculty had a much higher variability than the students. The ~~range~~ range of the faculty is 20.5 min, 7.75 min higher than the students. The interquartile range is the same, 7 min. The faculty's distribution is strongly skewed to the right, and the mean is 1.09 min higher than the median. The students' distribution is roughly symmetrical. The mean is 0.65 higher than the median, which suggests a slight skew to the right.

(c) You have been asked to report on this event for the school newspaper. Write a few sentences describing student and faculty performances in this competition for the paper.

In the student vs. faculty obstacle course, the students easily outperformed the faculty. On average, the students were 2.19 minutes faster than the faculty in completing the course. The fastest times and slowest times for the students, 3.75 min and 16.5 min, were both faster than the minimum and maximum times for the faculty, 4.5 min and 25 min. As a whole, the students were more consistent in their times, as shown by the range of 12.75 min, 7.75 min smaller than the faculty's range of 20.5 min!

GO ON TO THE NEXT PAGE.

PART 5: Parameters & Statistics and Means & Proportions

A **parameter** is a number that describes the population of interest. Parameters are fixed numbers and in practice we don't know its value. It's what we're trying to find when performing experiments and surveys.

A **statistic** is a number that describes a sample. We can know the value of a statistic, but depending on the sample it can vary. We use statistics to estimate parameters.

It's easy to remember which number describes each group: **P**arameters describe **p**opulations and **s**tatistics describe **s**amples.

The type of number a parameter or statistic can take is most often a **mean** or a **proportion**.

A **proportion** refers to the fraction of the total that possesses a certain attribute. For example: if 12 out of 30 students did their summer homework we would say the proportion who did their summer homework is $\frac{12}{30}$ or 0.40 or 40%

A **mean** is computed by adding all observations and dividing by the number of observations. For example: if the weights (in pounds) of 5 Golder teachers are: 155, 170, 135, 145, 190 we say the mean is 159 pounds.

Directions: Each scenario has **bold**faced numbers. Determine if the value is a parameter or a statistic AND if the value is a mean or a proportion.

1. On Tuesday, the bottles of Arizona iced tea filled in a plant were supposed to contain an average of **20** ounces of iced tea. Quality control inspectors sampled 50 bottles at random from the day's production. Those bottles contained an average of **19.6** ounces.

2. On a New York–Denver flight, **8**% of the 125 passengers were selected for random security screening prior to boarding. According to the TSA, **10**% of airline passengers are chosen for random screening.

3. A recent report in the journal *Nature* examined whether ducks keep an eye out for predators while they sleep. The researchers, from Indiana State, put four ducks in each of flour plastic boxes, which were arranged in a row. Ducks in the two end boxes slept with one eye open **31.8**% of the time, compared with only **12.4**% of the time for the ducks in the two center boxes.

4. **Review**: is the scenario in #3 above an observational study or an experiment? Explain in 1-2 sentences.