Statistics Summer Work Packet

Name:

Dear Statistics Students,

I hope you are all enjoying your first few days of summer! Here is your summer work packet for your upcoming Statistics course. Included are problems that you should be familiar with (mean and median) as well as problems that may look new to you. These problems can be answered by using what you already know, information included in the packet, as well as the first chapter of the textbook, *The Practice of Statistics 5e* by Starnes, Tabor, Yates, and Moore (ISBN: 978-1464108730). You are also encouraged to look up any word or plot type you don't recognize. There are many online sources that cover these topics.

The goal of this packet is the review some old material, introduce you to some of the statistical capabilities of the graphing calculator, and teach you some of the basic vocabulary and concepts that we will encounter in the first chapter or two of Statistics. A TI-83 or TI-84 is required for this packet and the course. Make sure you have your own graphing calculator to use both at home and in class.

Your work on this packet will be worth four homework assignments and a quiz on this material will be given at the end of the first week of class. This packet is due on the first day of class. Show all of your work and staple any additional pages to the back of the packet. Make sure that you review this material at the end of the summer to keep the information fresh in your mind, even if you finish the packet on the first day of summer break.

Good luck and enjoy!

There are two types of variables that we will discuss in the first chapter of our textbook: categorical and quantitative.

A <u>*categorical variable*</u> *places an individual into one of several groups or categories.*

- *Examples of categorical variables: Favorite color, zip code, gender*

A <u>*quantitative variable*</u> *takes numerical values for which it makes sense to find an average.*

- *Examples of quantitative variables: age, height, grade on a test, number of states/countries visited*

1. Classify each variable below as either categorical or quantitative.

   a) Grade level (9^{th}, 12^{th}, etc.)

   b) Number of siblings

   c) Birth month

   d) Number of pets

   e) Age category (young, middle-aged, etc.)

   f) Favorite sport

   g) Years at key school

   h) Street Number

   i) How long you can hold your breath

   j) Favorite season

We use different types of plots and analysis when exploring trends and patterns in each type of variable.

For categorical variables, we can display our data in a bar graph or a pie graph.

For quantitative variables, we can display our data using a dotplot or histograph. We can also calculate the mean, median of a set of data, as well as a value called the range, which gives us a measure of how spread out the data is.

For quantitative data, we will also practice describing the distribution by discussing its shape, center, spread, and outliers.

2. On the next page is the distribution of bachelor's degrees awarded in 2010, according to a government survey. The middle row (Number of degrees) shows us the frequency of each degree type while the last column shows us the relative frequency of each degree type. Frequency is the count, or number in each category while the relative frequency is the percent or proportion in each category.

Major	Number of degrees	Percent of degrees
Business	358,293	21.7
Social sciences/ history	172,780	10.5
Health professions	129,634	7.9
Education	101,265	6.1
Psychology	97,216	5.9
Visual and performing arts	91,802	5.6
Biological and biomedical sciences	86,400	5.2
Communication and related programs	81,266	4.9
Engineering	72,654	4.4
English language and literature	53,231	3.2
Other	405,473	24.6

a) What type of variable is this?

b) Draw a pie graph of this data (by hand). Remember to label your graphs!

c) Draw a bar graph of this data (by hand). I recommend using the relative frequency on the graph. Remember to label your graphs!

d) Which graph was more difficult to draw (accurately)?

3. The U.S. Food and Drug administration limits the amount of caffeine in a 12-ounce can of carbonated beverage to 72 milligrams (mg). That translates to a maximum of 48mg of caffeine per 8-ounce serving. Data on the caffeine content of popular soft drinks (in mg per 8-ounces) are given below.

   15 15 16 20 23 23 23 24 24 25 25 26 26 27 27 27 28 28 28 28 29 29
   31 31 33 35 35 35 36 37 37 37 38 43 43 47 47

a) What type of variable is being displayed?

b) Calculate the mean of the data.

c) Calculate the median of the data.

A dotplot is a plot is a plot where each data value is shown as a dot above its location on a number line. If there is more than one data point for a given x value (there are two instances of 15 in the data above) you just stack the dots on top of each other. This plot allows us to see the shape of the data.

A histogram is similar to a dotplot but instead of plotting each data point at its exact value, we group nearby points together to get a better sense of the shape of the graph. There is no one correct way to group points, but usually you want 5 to 10 equally sized groups. For example, with the data from Problem 3, you could group all values between 15 and 19, all values between 20 and 24, all values between 25 and 29, and so on from there. The groups are referred to as "bins" and the length of these bins is called the "bin width". The "bin width" described above is 5. It is important to know how to do this by hand, but you can also do it using your calculator. You can find the instructions for constructing a histogram using your calculator online. There should be several resources, including videos, available to you.

4. Use the information above and the data from Problem 3 to create:
   a) The dotplot. Label your axes!

   b) A histogram of bin width 5. Label your axes!

   c) A histogram of bin width 8. Label your axes!

d) What is the difference between these three plots?

Once we have these plots, we can start to describe the data's distribution. To do this we discuss the shape, center and spread. We can even talk about outliers if we see them.

Shape. *To talk about this, we need to introduce some new vocabulary: symmetric, right-skewed, and left-skewed.*

- *A distribution is roughly symmetric if the right and left sides of the graph are approximately mirror images of each other.*
- *A distribution is skewed to the right or right-skewed if the right side of the graph (containing the half of the observations with the larger values) is much longer than the left side.*
- *A distribution is skewed to the left or left-skewed if the left side of the graph (containing the half of the observations with the smallest values) is much longer than the right side.*

We can also discuss how many peaks are in the data. Is there one (unimodal), two (bimodal), or many (multimodal)?

Center: Where is the center of your distribution? We typically use the mean or median to describe this quality.

Spread: How spread out is the data? You can use the range of the data (on what interval does the data lie?), the standard deviation or the interquartile range (IQR). We will discuss the last two more during the year so for now focus on the range. The range is the maximum value minus the minimum value. It should always be positive.

5. Use your data from problem 3 and your plots from problem 4 to describe the data's distribution.

Shape:

Center:

Spread:

6. For the following data, provide a plot and a description of the distribution. This may involve calculating values such as mean, median, and/or range.

   a) UK 14-year-old female height data (in centimeters): 160, 169, 152, 167, 164, 163, 160, 163, 169, 157,158, 153, 161, 165, 165, 159, 168, 153, 166, 158, 158, 166

b) Number of contacts that a sample of high school boys had in their cell phones: 124, 41, 29, 27, 44, 87, 85, 260, 290, 31, 168, 169, 167, 214, 135, 114, 105, 103, 96, 144

c) Number of contacts that a sample of high school girls had in their cell phones: 30, 83, 116, 22, 173, 155, 134, 180, 124, 33, 213, 218, 183, 110