Your Name: _____

Section:

36-201 INTRODUCTION TO STATISTICAL REASONING Computer Lab #2 Comparing Features of Distributions

Objectives:

- 1. To read data from a Minitab data file into Minitab.
- 2. To develop expertise in comparing features of distributions using histograms and boxplots.
- 3. To investigate the features of distributions of random samples taken from the same population.

Getting Started

Instead of typing data into Minitab by hand, we will often make data available to you in a file. Some of these files will be in a special Minitab format. To access these files you will need to copy them using the file server. The following are the steps you need to follow:

COPYING FILES

- 1. In the following order: From the 🖒 menu in the upper left hand corner of the titlebar, choose Chooser. In the dialog box that appears *click* Appleshare. In the bottom left box called "AppleTalk Zones" *select* BH. In the top right box where it says "Select a file server" select HSSHE-LIOS. *Click* Ok.
- 2. In the dialog box that appears login using "36201" as your Name and "36201" as your password. Click **Ok**. In the dialog box that appears highlight "Class". (Do not put a check next to "Class.") Click **Ok**.
- 3. A "Class" icon will appear to the right. Double click on the icon. Then double click on 36201. You will see many files in a folder.
- 4. Copy all of the files that you will need to the student folder. Do this by dragging the files to the student folder. For today, the data files you need are *cars.mtw* and *graduation.mtw*.
- 5. Close all windows.
- 6. **IMPORTANT:** Drag the *class* icon into the trash. This is necessary to let others in the lab get access to the data.

To start Minitab, either double-click on the cars.mtw icon, or

- 1. *Double-click* the hard drive icon.
- 2. *Double-click* the Applications icon.
- 3. *Double-click* the Minitab 10.5 Xtra Power folder.
- 4. Double-click Minitab 10.5 Xtra Power.

Part I. Case Identification

In this section you will learn about some features of Minitab that will help you in exploring a data set and understanding better the features of the data. Recall that when you describe the features of a distribution you should always comment on: CENTRAL LOCATION, SPREAD, SHAPE, and UNUSUAL OBSERVATIONS.

If you didn't already open cars.mtw, you will now open it as follows.

- From the File menu, choose Open Worksheet.
- *Click* the **Desktop** button on the right side of the dialog box.
- Select the file called "cars.mtw" and then *click* **Open**.

These data are measurements collected by *Consumer Reports* on 38 cars made during the 1978-1979 model year. Today we will only be interested in the four variables COUNTRY - where the car is made, CAR - which refers to the model type, WEIGHT - which is the car weight in thousand pounds, and MPG - miles per gallon.

Scroll through the window to look at the different countries and different make of cars in the datafile. *You should always become familiar with your data before you start performing a data analysis.*

Question #1 What type of variables are WEIGHT, CAR, COUNTRY, MPG? (quantitative or categorical?)

Make a **Histogram** of the weight variable. To get a different impression of the shape of the distribution of weight, it will be convenient to change the scale of the histogram. Recall that to do this, in the "Histogram" dialog box, *click* **Options** and then *click* the button next to "Mid/cutpoint positions". Type 2:5/.3 and *click* **OK** twice. Move the histogram window so that you can see it and the Worksheet window.

Question #2 Using the histogram describe the features of the distribution of car weights. Sketch the histogram in the space below.

Question #3 Give a simple practical explanation as to why the histogram has the notable feature you described in question #2.

Minitab has a very helpful sorting feature. From the **Manip** menu, choose **Sort**. *Click* the box under "Sort column(s):". *Drag* down the list of columns and variables in the list box on the left of the dialog box. Make sure all of them are highlighted. *Click* **Select**. *Click* the box under "Store sorted column(s) in: ". Once again, *drag* down the list of columns and variables in the list box on the left and then *click* **Select**. Then, *click* the box next to the first instance of "Sort by column:". Type *Weight*. *Click* the check box to the left of "Descending". *Click* **OK**. Now the cars are sorted from heaviest to lightest.

A good answer to question #3 would have been that the values around the first mode in the histogram are the smaller, hence lighter, cars and that the values around the second mode are the bigger, heavier cars.

Question #4 Which countries correspond to the weights of cars located around the second mode and higher? How many cars from each of these countries are there in this upper tail of the distribution of weight?

You now want to use Minitab to calculate the mean, standard deviation, and sample size for WEIGHT. From the **Stat** menu, choose the **Basic Statistics** sub-menu, and then choose **Descriptive Statistics**. Type *Weight* in the box under "Variables:". Click **Okay**. The Session window will provide you with descriptive statistics, such as, the sample size, mean, median, standard deviation (StDev), minimum, and maximum for each variable requested, along with some other descriptive statistics you will learn about later.

Question #5 What is the value of the mean, median and standard deviation of WEIGHT? How many observations are there in the data set?

Question #6 What do you think is the relationship between WEIGHT and MPG (miles per gallon)? (This is a thought question and you shouldn't use Minitab to answer it.)

Leave the window with the histogram for WEIGHT open. Make a histogram of MPG. To clear the settings from the previous histogram, hold the *apple-key* down and simultaneously press *R*. Note the features of the distribution of MPG. You will now resort the data, but this time according to MPG. Follow the same steps from before, only choose to sort by column MPG and DO NOT choose "Descending".

4 Question #7 Compare the values of MPG and WEIGHT. Is this result consistent with what you predicted in your answer to question #6? Explain.

<u>COMMENT.</u> It seems that the second mode in the histogram of weight corresponds to bigger, heavier cars mostly from the US that use up a lot of gas. The first mode corresponds to smaller cars which tend to get better gas mileage.

From the **Graph** menu, choose **Pie Chart**. *Click* the box next to "Chart data in". Type *Country*. *Click* **OK**. If you have difficulty reading the chart, *click* the Zoom box.

Question #8 From the pie chart which countries make (i) the most cars (ii) the second greatest number of cars, and (iii) the fewest cars?

We are finished with this data set. Move the cursor to the **File** menu and choose **Restart Minitab**. Choose **Don't Save** to the request to save the file. Choose **No** to the request to save changes to the Session window. You should now see blank Session and Worksheet windows.

Part II. A Data Analysis

From the **File** menu, choose **Open Worksheet**. *Click* "graduation.mtw" in the list of files. *Click* **Open**.

Background: These data are the percentages of freshman graduating from the 6 different colleges within a University in each of 8 years. From the **Graph** menu, choose **Boxplot**. Highlight the space below "Y" in the first row, then find the variable '% grad' in the box on the left. Double-click on it and it should appear in the selected box. Now *click* the box in row 1, column X if it is not already highlighted. Double-click the variable *School* in the left box.

Question #10: Although the variable is called SCHOOL it really refers to the colleges within the university. Is this a quantitative or categorical variable? Which college at the university traditionally has the highest % freshman graduation rate? Based on features of the distribution explain why.

4 Question #11: Compare the distribution of % freshman graduation rates for college B to college F. How are they similar and how are they different? Be specific in your answer referring to features in your boxplots (i.e., central location, spread, shape, outliers) and to approximate numerical summary measures such as the median and range, that you can estimate from the boxplot.

We are finished with this data set. From the **File** menu, choose **Restart Minitab**. Choose **Don't Save** to the request to save the file. Choose **No** to the request to save changes to the Session window. You should now see blank Session and Worksheet windows.

Part III. Samples from a Population

We have said that a characteristic of a variable is that it can take different values for different people or things. Therefore it follows that the distribution of a variable for one subset of people might have different features than the distribution of the same variable for another subset of people. For example, if we looked at the distribution of heights (the variable) for students in section A (i.e., a sample of students) it might have different features than the distribution of heights of students in section E (i.e., another sample of students). It might even have different features than the distribution of heights for all of the students in 36-201 combined (i.e., the population).

In this section you will explore how features of a distribution from a sample differ from one another and from the population. You will do this by creating 3 variables consisting of 25 values each that are random samples from a population distribution that is unimodal, symmetric, has a mean (and median) equal to 100 and a standard deviation equal to 15. Imagine that the variable is IQ score with a mean in the population of 100 and a standard deviation of 15 points. Each sample of size 25 could represent a different section of students in our class.

- 1. Go to the Calc menu, and from the Random Data sub-menu, choose Normal.
- 2. A dialog box will appear. In the first box after the word "Generate" type the number 25. Move the arrow pointer to the box under "Store in column(s):" and *click*. Type the column range *C1-C3*.
- 3. Click the box following "Mean". Delete the number that is there and replace it with 100. *Click* the box following "Standard Deviation". Delete the number that is there and replace it with 15.
- 4. Click **Ok**.

The three variables and their values will appear in the first three columns of the Worksheet window. Resize the window so that you can see all of the values.

Question #12 Reading down the first column, what is the largest value? What is the smallest value?

You now want to use Minitab to calculate the mean, median, quartiles, standard deviation, the minimum, and the maximum for each variable. From the **Stat** menu, from the **Basic Statistics** sub-menu, choose **Descriptive Statistics**. *Click* the box under "Variables:". In the list box on the left side of the dialog box highlight the variables C1, C2, and C3 by holding the mouse button down and *dragging* over the three variables. They should all be highlighted. *Click* **Select**. *Click* **OK**. The Session window will provide you with the mean, sample size, and standard deviation (StDev) for each variable.

Question #13: Using the information you just created fill in the following table for all three variables. *For our purposes you do not need to enter more than one decimal place.*

	<u>C1</u>	<u>C2</u>	<u>C3</u>
Population mean	100	100	100
Population standard deviation	15	15	15
Sample mean			
Sample size			
Sample standard deviation			
Sample median			

♣ Question #14 Did you expect that these 3 samples from the same population distribution would have means, medians, and standard deviations exactly equal to the population values? (This is like asking whether the features of the distribution of heights of the students in our class are exactly the same as the features of the distribution of heights of all students.) Why are they not equal?

Go to the **Manip** menu and *choose* **Stack**. In the first row under "Stack the following blocks" type C1; in the second row type C2; and in the third row type C3. In the row under "Store the results in" type C4. Click in the box preceding "Store the subscript in:" and type C5. (Remember each column must go on a separate line in the "Stack" dialog box.) Click **Ok.** From the **Graph** menu, choose **Boxplot**.

Question #16 Create a boxplot with column C4 as the Y variable and column C5 as the X. Make rough sketches of your boxplots below.

Question #17 Recall that these 3 variables were created from a symmetric distribution. Based on the boxplots are each of these 3 sample distributions symmetric? To answer this question, for each boxplot (i) note the location of the median relative to the quartiles and also (ii) judge whether the tails of each distribution are about the same length.

To quit Minitab, from the File menu, choose Quit. Do not save any files. Drag cars.mtw and graduation.mtw to the trash. Remember to delete any other files and folders that you might have created. LOG OUT LOG OUT LOG OUT LOG OUT Turn in your cover sheet.