Chapter 3

Numerical Summaries of Center and Variation

Table Number:_____ Group Name: _____

Group Members:_____ _____ _____

What is Typical?

Part I: Making Predictions

For each of the following variables measured on the *Student Survey* (your section), make a prediction for a **typical value** *for all students enrolled in your statistics class this term*. A typical value is a single number that summarizes the class data for each variable.

1. Write that prediction in the *First Prediction* column.

Attribute from Student Survey	First Prediction	Revised Prediction from DotPlot	Statistics from *StatCrunch*	
			Mean	Median
Age				
College credits registered for this semester				
Number of states visited				
Height (in inches)				
Number of siblings				
Hours a week spent studying				
Hours a week spent working at a paying job				

Open MyMathLab and click on the *StatCrunch* link in the menu on the left. Click on the link "Stat Crunch Website,"

View the data sets from your textbook in StatCrunch.

Visit the StatCrunch website to perform complex analyses with the StatCrunch statistical software, search shared data sets, take online surveys, and more.

then "**My Groups.**" Click on our group, then under the "Preview Data," click on "Class data."

2. Now use *StatCrunch* to create dot plots of each variable to see if your original predictions seem reasonable. Based on the dotplots, make revised predictions for the typical value for each of the variables.

3. To make a dot plot in *StatCrunch*, click on the **Graph** button at the top of the spreadsheet, choose Dotplot.
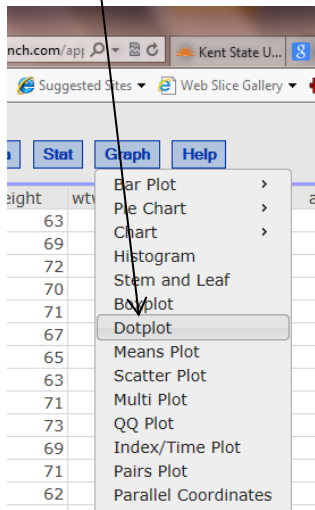
4. Choose the column of interest, label the axes, then click on "Compute!" at the bottom of the page.
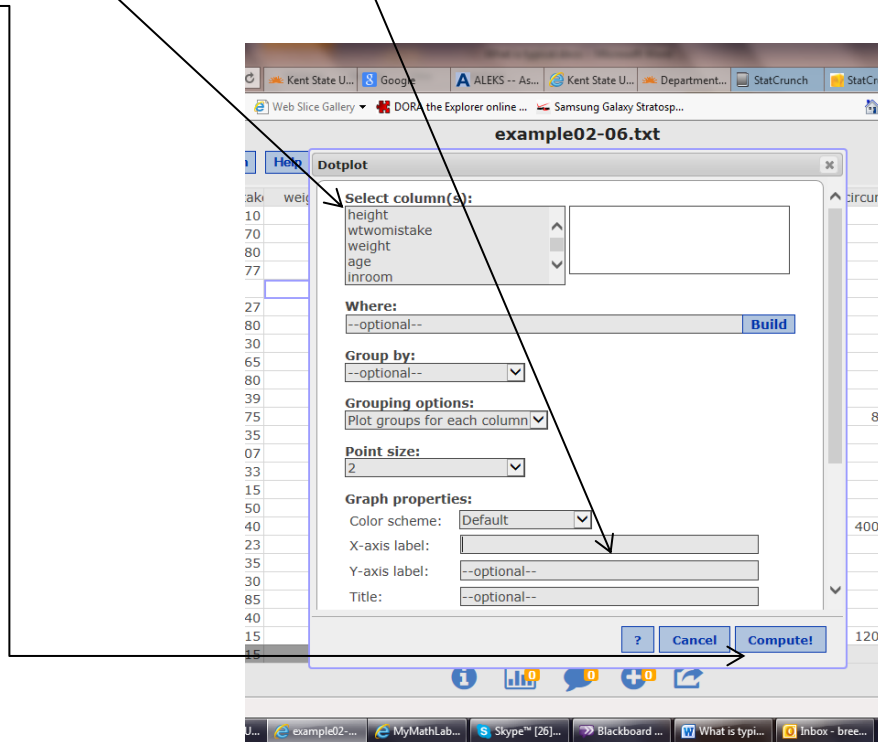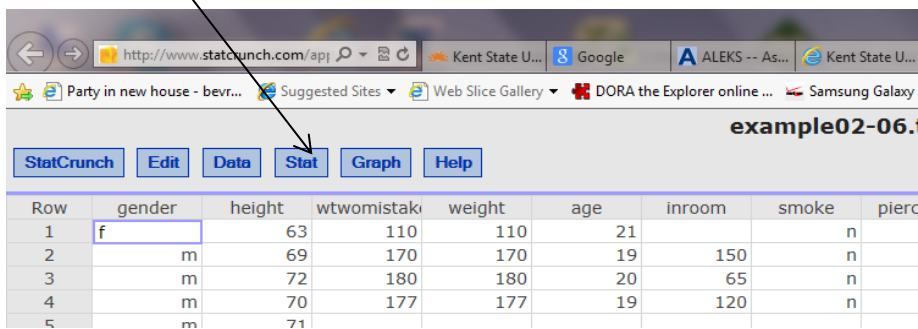
5. Write these new predictions in the *Revised Prediction* column in the table.
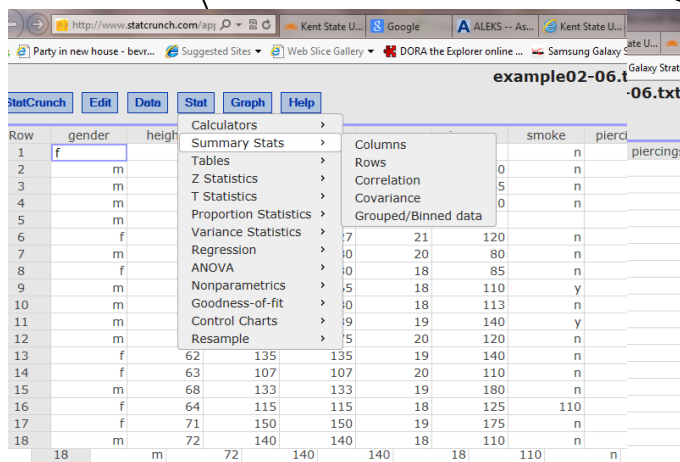
.

Part II: Test Your Conjectures

Use *StatCrunch* to find the mean and median for each of these variables. (Follow the directions below.)

1. Click the Stat button.
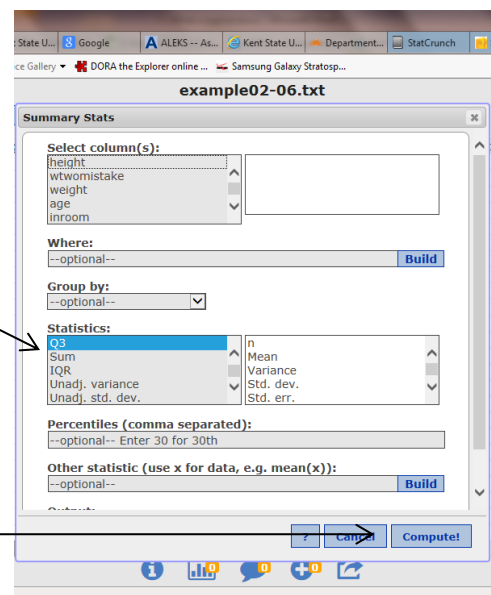
2. Click on "Summary Stats," then "Columns." Choose the columns of interest. You can select more   than one by holding down the Ctrl key on your keyboard.

3. Click on the stats you want. You can choose more than one by holding down the Ctrl key on your keyboard
   4. Click "Compute."

6. How close were your *revised predictions* to the "typical" values produced in *StatCrunch*? For which attributes were your predictions most accurate?

7. What was most surprising to you? Why?

8. In general, were your *revised predictions* closer to the means or medians?

Things to Consider

- How close were your predicted typical values?

- Which measure of center were your guesses closer to, the mean or median?

- What information do means and medians tell us about a distribution?

- How do we decide whether to use the mean or median to summarize a data set?

- In statistics, what do we mean by what is typical?

Reference

Adapted from AIMS http://www.tc.umn.edu/~aims/aimstopics.htm

Table Number:_____   Group Name: _____

Group Members:_____   _____   _____

What does a *Mean* Mean?

Things to Think About: Discuss with your group:

- Suppose the average age (mean) for students in the class is 21 years. Then what do we know about the distribution of students' ages for this class (e.g., Are they all about 21 years old?)? Explain.

- Where does that number (21) come from? Explain.

Dot Plots

In your small group, open the applet "Mean and Median" located in the Chapter 3 folder in Blackboard and type in "17" for lower limit and "25" for upper limit. Click "Update."   Click above the axis and find a point as close to 21 as possible.   If you put the cursor on your dot, you can move it from left to right.   Then form dot plots according to the directions below. Each dot plot will be created so that 10 dots average to 21 years.

1. Of course, the easiest way to do this is to stack them all above the value of 21. If we do this first, what does the plot look like in terms of shape, center, and spread?

2. Move one dot to 24 years. How can you move *one or more* of the other dots so that you keep the mean at 21 years? Is there more than one way to do this? Give at least three different ways. Explain in detail.

3. Now, clear the graph by clicking "Trash" and start with the original distribution with all dots on (or near) 21. Next, move one dot to 17 years. Describe what else you had to do to keep the mean at 21 years.

4. Next, move all the dots so that none of the ages are 21 years, yet the mean age is still 21 years. Describe what you did.

Part II

Draw graphs of *two* different dot plots below such that:

- Each graph has 10 values (10 dots).

- The mean of the values in each dot plot is 21.

- One dot plot has a *little spread* and the other dot plot has *a lot of spread*.

18   20   22   24

18   20   22   24

Things to Consider: Discuss with your group

- What strategies did you use to keep a mean of 21 in each plot?

- Was it important to pay attention to the distance of each dot from the mean value of 21? How did this help you to keep the mean of 21 years?

Part III

We will use the term "**deviation**" to represent the distance of each data value from the mean. This can also be viewed as the number of units to the left or right of the mean for each data value.

1. Look at your last dot plot, and write down the deviations for each data point on the graph. List the deviations neatly, separated by a comma.

2. If you change one value so it has a deviation of -3, what do you have to do one or more values to keep the mean at 21 years?

Reference
Adapted from AIMS http://www.tc.umn.edu/~aims/aimstopics.htm

Table Number:_____   Group Name: _____

Group Members:_____   _____   _____

Standard deviation – The Basics

1. Find the Deviations. Given $\bar{x} = 2.8$

	Deviations from $\bar{x}$	Deviations SQUARED
0		
5		
6		
0		
2		
4		

2. The top speeds (in mph) for a sample of five new automobile brands are listed below. Calculate the standard deviation of the speeds.

	Deviations from $\bar{x}$	Deviations SQUARED
160		
125		
190		
185		
105		

3. The book cost (in dollars) for one semester's books are given below for a sample of five college students. Calculate the sample standard deviation of the book costs.

	Deviations from $\bar{x}$	Deviations SQUARED
200		
130		
400		
500		
345		

4. Each of the following distributions have a mean of 60.   Without calculating the standard deviation, decide which of the following data sets has the largest standard deviation. Which has the smallest?   Explain to your group members WHY you think so. Then check with your graphing calculator.

a)   60 60 60 60 60 60

b)   10 20 25 95 100 110

c)   50 50 55 65 70 70

The following list shows the age at appointment of U.S. Supreme Court Chief Justices appointed since 1900. Use the data to answer the question.

Last Name	Age
White	65
Taft	63
Hughes	67
Stone	68
Vinson	56

Last Name	Age
Warren	62
Burger	61
Rehnquist	61
Roberts	50

5. The US Supreme Court Chief Justice data was used to create the following output in an Excel spreadsheet. Choose the statement that best summarizes the variability of the dataset.

Column1	
Standard Error	1.864217971
Standard Deviation	5.592653912
Sample Variance	31.27777778
Kurtosis	1.177934931
Skewness	-1.064306225
Range	18
Minimum	50
Maximum	68
Sum	553
Count	9

   a. The ages of most of the US Supreme Court Chief Justices are between 50 and 68 years.
   b. The age of most of the US Supreme Court Chief Justices since 1900 are within 5.6 of the mean age.
   c. The age of most of the US Supreme Court Chief Justices since 1900 are within 31.3 years of the mean age.
   d. None of these.

6. The top nine scores on the organic chemistry midterm are as follows: 74, 39, 80, 57, 43, 69, 21, 34, 69.
   Christine is currently taking college astronomy. The instructor often gives quizzes and the following are the top seven scores on the last quiz: 44   20   37   28   19   52   55.
   Which set of scores is shows more variation and why?

7. Below is the standard deviation for extreme 10K finish times for a randomly selected group of women and men. Which statement best summarizes the meaning of standard deviation?

   Women: $s = 0.16$ Men: $s = 0.25$
   e. The distribution of women's finish times is less varied than the distribution of men's finish times.
   f. The distribution of men's finish times is less varied than the distribution of women's finish times.
   g. On average, men's finish times will be 0.25 hours faster than the overall average finish time.
   h. On average, women's finish times will be 0.16 hours less than men's finish times.

8. Which of the following measurements is likely to have the **most** variation?
   i. The volume of individual pop cans measured in fluid ounces from a randomly selected twenty-four pack.
   j. The individual weights in ounces of tennis balls in a randomly selected can of tennis balls.
   k. The individual weights in ounces of potatoes in a randomly selected crate of potatoes.

Table Number:_____   Group Name: _____

Group Members:_____   _____   _____

What makes the standard deviation larger or smaller?

Part I

Before we begin the attached worksheet, we are going to think about the meaning of typical, or standard, deviation from the mean. First, examine the following dot plot which has the mean marked in the plot. Think about how large the deviations would be for each data point (dot).
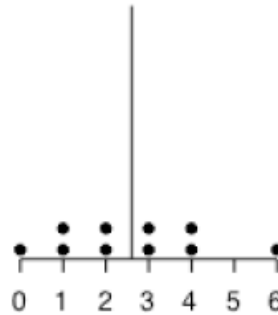
Next we will draw in each deviation from the mean.
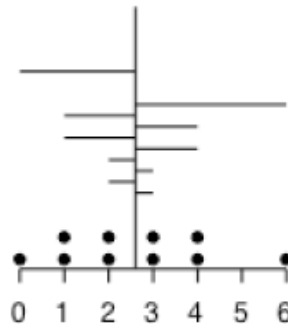
Now, think about the average size (length) of all of those deviations, and use this as an estimate for the size of the Standard Deviation. Don't worry about whether the deviation is to the left or right of the mean. Just consider all of the lengths. Draw the length of the standard deviation below.

Based on the scale in the graph, estimate a numerical value for the length of the line you drew above.

Repeat the process with this dot plot to help you draw and estimate the length of the standard deviation.
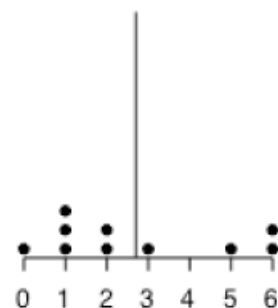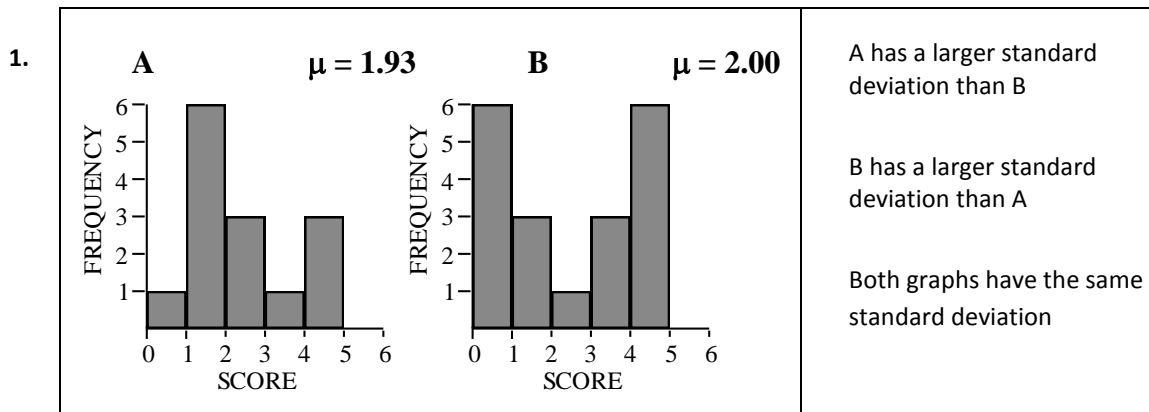
Now, try to draw and estimate the length of the standard deviation with the following histogram. The mean of the scores is 2.5. (Hint: Sketch in the appropriate number of dots in each bar of the histogram to make sure you have the appropriate number of deviations.)
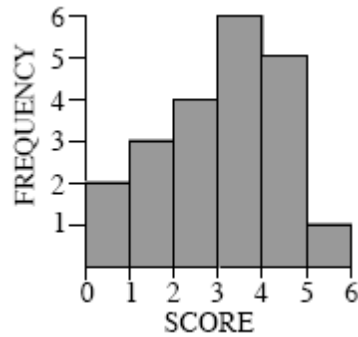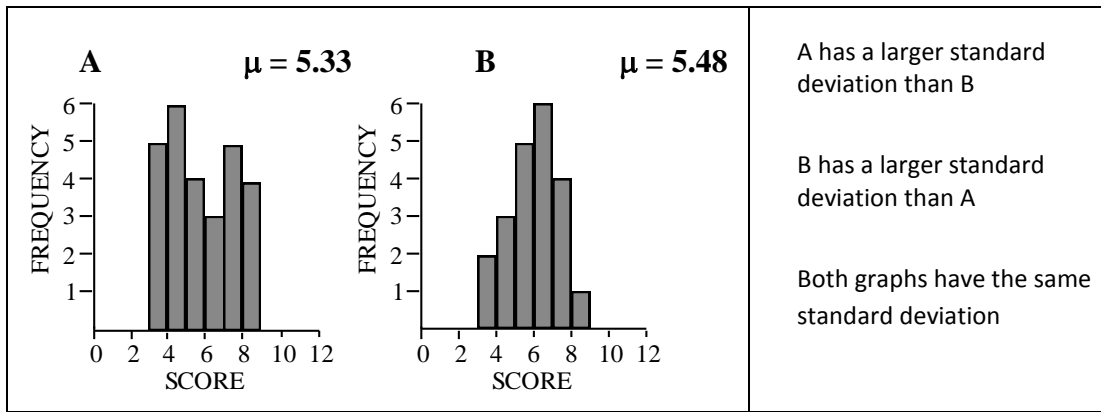

Comparing Standard Deviations

Below, you will find five pairs of graphs. The mean for each graph μ is given just above each histogram. For each pair of graphs presented,

- Indicate which one of the graphs has a larger standard deviation or if the two graphs have the same standard deviation.

- Explain why. (Hint: Try to identify the characteristics of the graphs that make the standard deviation larger or smaller.)
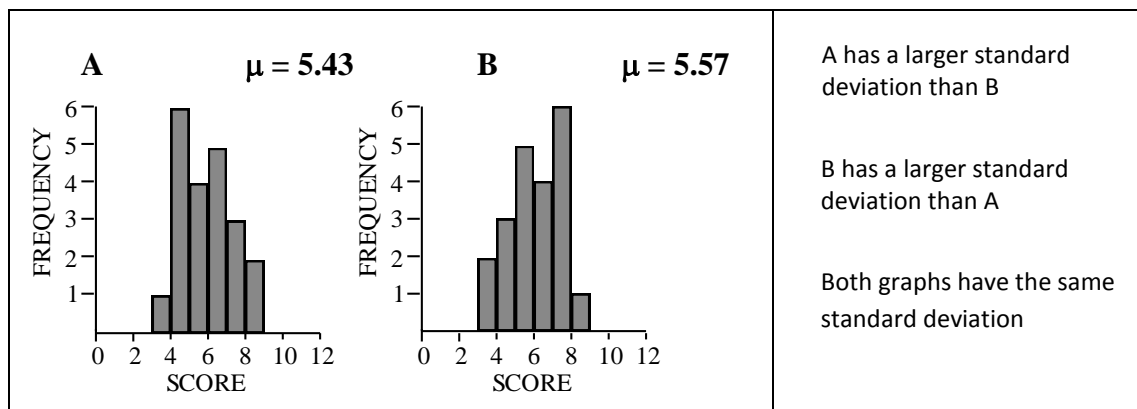
1.

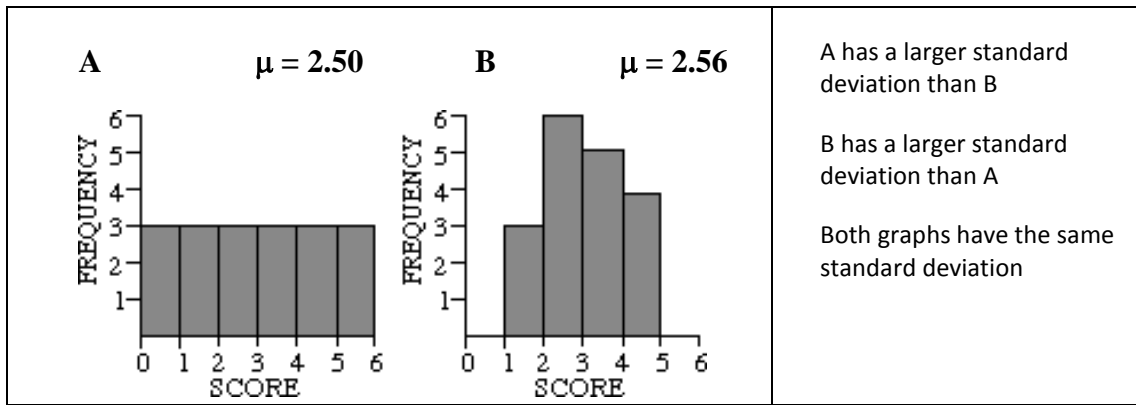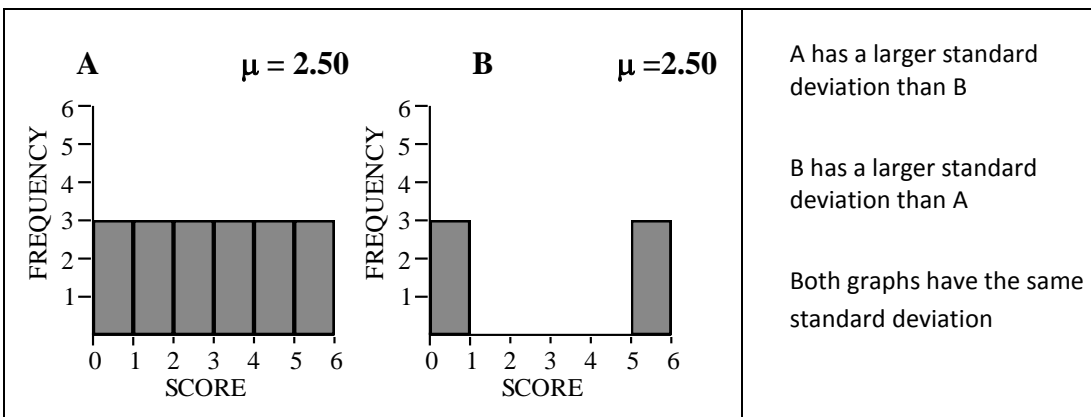A has a larger standard deviation than B

B has a larger standard deviation than A

Both graphs have the same standard deviation

Explain.

2.

A $\mu = 5.33$

B $\mu = 5.48$

A has a larger standard deviation than B

B has a larger standard deviation than A

Both graphs have the same standard deviation

Explain.

3.

A $\mu = 5.43$

B $\mu = 5.57$

A has a larger standard deviation than B

B has a larger standard deviation than A

Both graphs have the same standard deviation

Explain.

4.

| A | $\mu = 2.50$ | B | $\mu = 2.56$ | A has a larger standard deviation than B |

B has a larger standard deviation than A

Both graphs have the same standard deviation

Explain.

5.

| A | $\mu = 2.50$ | B | $\mu = 2.50$ | A has a larger standard deviation than B |

B has a larger standard deviation than A

Both graphs have the same standard deviation

Explain.

Reference

delMas, R.C. (2001b). What makes the standard deviation larger or smaller? STAR library. Retrieved October 21, 2007, from http://www.causeweb.org/repository/StarLibrary/activities/delmas2001/

Table Number:_____   Group Name: _____

Group Members:_____   _____   _____

Standard Deviation – Applied Problems

1. The data below are the lifetimes (in hours) for 10 light bulbs from a new brand that our school is considering for use in the football stadium light fixtures:

   2009   2015   2002   1979   2032   1991   2016   2001   1900   2055

   a) What is the value of the standard deviation of the lifetimes for the 10 light bulbs from the new brand?   Show your calculations below.

   b) The standard deviation for the lifetimes of the bulbs from the brand currently in use is 40 hours. What does the standard deviation that you computed for the sample of light bulbs from the new brand tell you about how this brand might compare with the old brand?

   c) Replacing stadium light bulbs is difficult and requires special equipment. Because of this, rather than replace individual bulbs as they burn out, the school plans to replace *all* the stadium light bulbs as soon as one burns out. The mean lifetime is 2000 hours for both the current brand and the new brand under consideration, and the cost of the two brands is the same. Would you recommend that the school stay with the current brand or change to the new brand? Explain your reasoning.

2. A class of students took a test in Language Arts. The teacher determines that the mean grade on the exam is a 65%. She is concerned that this is very low, so she determines the standard deviation to see if it seems that most students scored close to the mean, or not. The teacher finds that the standard deviation is high. Interpret what might have caused this high standard deviation. Be detailed, but concise.

3. An employer wants to determine if the salaries in one department seem fair for all employees, or if there is a great disparity. He finds the average of the salaries in that department and then calculates the variance, and then the standard deviation. The employer finds that the standard deviation is slightly higher than he expected. Explain in some detail what might be causing the high standard deviation.

4. Consider the dotplots below showing the average temperature for the month of February 2006 for three cities. Which city would you expect to have the largest standard deviation in temperatures for the month? Explain how having this information might affect your choice of cities to which to move if given the opportunity. Be detailed, but concise.

February High Temperatures (°F)
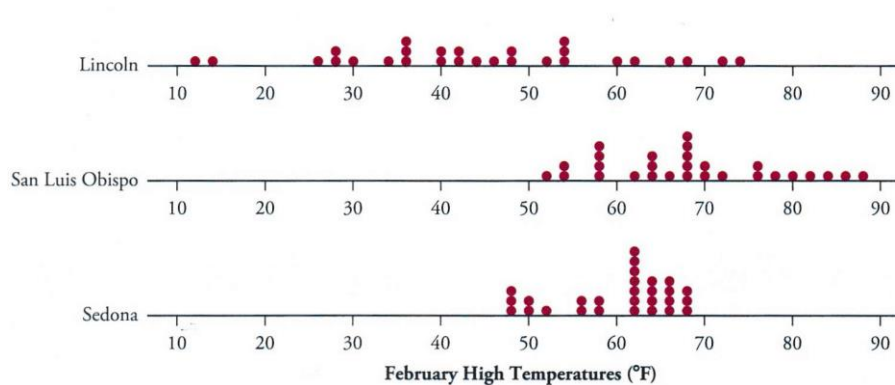
Are Female Hurricanes Deadlier than Male Hurricanes?

Background (Adapted from: "Female Hurricanes are Deadlier than Male Hurricanes, Study Says." by Holly Yan, CNN. June 3, 2014: http://www.cnn.com/2014/06/03/us/female-hurricanes-deadlier/)

Apparently sexism isn't just a social problem -- if you're in the path of a hurricane, gender bias might actually kill you. A study suggests people prepare differently for hurricanes depending on whether the storm has a male or female name. "Feminine-named hurricanes (vs. masculine-named hurricanes) cause significantly more deaths, apparently because they lead to a lower perceived risk and consequently less preparedness," a team of researchers wrote in the Proceedings of the National Academy of Sciences. In other words, a hurricane named "Priscilla" might not make people flee like a hurricane named "Bruno" would.

The study analyzed death rates from U.S. hurricanes from 1950 to 2012. "For severe storms, where taking protective action would have the greatest potential to save lives, the masculinity-femininity of a hurricane's name predicted its death toll," the study said. Hurricane Katrina in 2005, which left more than 1,800 people dead, was not included in the study because it was considered a statistical outlier. Neither was Hurricane Audrey in 1957, which killed 416 people. The study does note that both of those very deadly hurricanes had female names.

Why name hurricanes anyway? Giving hurricanes short, easy-to-remember names helps reduce confusion when two or more tropical storms are brewing at the same time, the National Hurricane Center said. For decades, all hurricanes were given female names in part because hurricanes were unpredictable, the study said, citing the "Encyclopedia of Hurricanes, Typhoons and Cyclones." "This practice came to an end in 1979 with increasing societal awareness of sexism, and an alternating male-female naming system was adopted," the report said.

Each year's list of hurricane names is alphabetical, alternating between male and female monikers. A U.N. World Meteorological Organization committee has already set up six years' worth of names. The lists repeat after each six-year cycle. "The only time that there is a change is if a storm is so deadly or costly that the future use of its name on a different storm would be inappropriate for obvious reasons of sensitivity," the National Hurricane Center said.

The hurricane data is given in an Excel Spreadsheet in the chapter 3 folder on Blackboard. It is also in our class data on Stat Crunch. It gives data that was used in the article *Female Hurricanes are Deadlier than Male Hurricanes* by Kiju Junga, Sharon Shavitta, Madhu Viswanathana, and Joseph M. Hil. In *Proceedings of the National Academy of Sciences of the United States of America*, May 2014.

*Note: hurricanes Katrina in 2005 (1833 deaths) and Audrey in 1957 (416 deaths) were removed from the data set.

Table Number:_____   Group Name: _____

Group Members:_____   _____   _____

Are Female Hurricanes Deadlier than Male Hurricanes?

1. Suggest a graph that might be used to compare the death totals for Female and Male named hurricanes. Explain why you chose the graph that you did.

2. Calculate the mean, standard deviation, and five-number summary of the death totals for Female and Male named hurricanes.

Gender	Mean	S.D.	Min	Q1	Median	Q3	Max
Female							
Male							

(a) Which measure, the mean or the median, do you think better represents a typical number of deaths from a hurricane? Why?

(b) Based upon the numerical calculations, do you think that the Female named hurricanes are more deadly? Why or why not?

3. For each of Female and Male named hurricanes, determine whether there are any outliers.

4. Construct comparative boxplots on Stat Crunch that display the distributions of the number of deaths for Female and Male named hurricanes. Copy and paste your boxplots below. Be sure to label your graphs' axes.

5. Thoroughly interpret the boxplots. Compare and contrast center and spread for the two distributions. Then, state your opinion on whether or not it seems that the Female named hurricanes are more severe.

6. How could the fact that all hurricanes had female names until 1979 bias the results from Question 5?

7. Now, consider only the *Female* named hurricanes. Earlier, it was noted that hurricanes Audrey and Katrina were omitted from the analysis. Add the death totals from these two hurricanes to your dataset and redo the summary calculations:

Katrina/Audrey Included	Mean	S.D.	Min	Q1	Median	Q3	Max
No							
Yes							

Which measure, the mean or the median, do you think better represents a typical number of deaths from a hurricane? Why?

Table Number:_____   Group Name: _____

Group Members:_____   _____   _____

Interpreting Boxplots

Part I

The following graph shows the distribution of ages for 72 recent Academy Award winners split up by gender (36 females and 36 males).
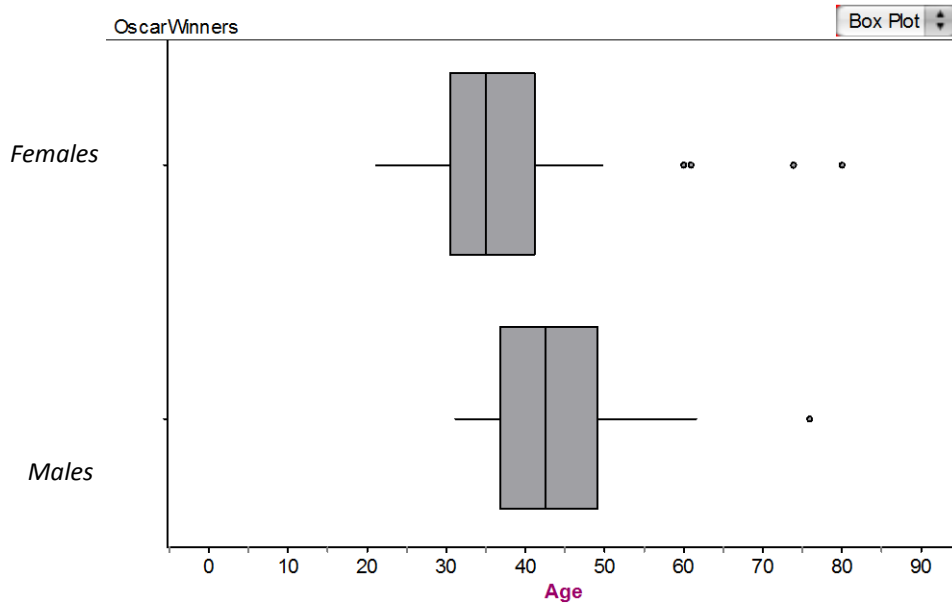
Use the graph to help answer the following questions.

a) Estimate the percentage of female Oscar winners who are younger than 40.

b) The oldest 50% of Oscar winners who are male are between which two ages?

c) What is the shape of the distribution of male Oscar winners? Explain.

d) Explain how to find the Inter-Quartile Range (IQR) for the female Oscar winners.

e)   Find the IQR for the female Oscar winners.

f)   What information does the IQR of the female Oscar winners offer us? Why would a statistician be more interested in the IQR than in the range?

g)   Compare the medians for male and female Oscar winners. What do you conclude about the ages of male and female Oscar winners? Explain.

h)   Compare the IQR for the male and female Oscar winners. What do you conclude about the ages of male and female Oscar winners now? Explain.

Part II

In the next problem, you will be given a descriptive scenario and a graph that shows two box plots. Use the graphs to make an informed comparison of the groups.

▪   Be sure to compare *shape*, *center* and *spread* of the distributions.
▪   Also, be sure that you are comparing the groups using the context of the data and not just comparing two (or more) numbers.

Stephen wants to investigate differences in spending habits of males and females. He compares the amounts spent per week on reading materials by males and females in a random sample of college students by generating the following plots.
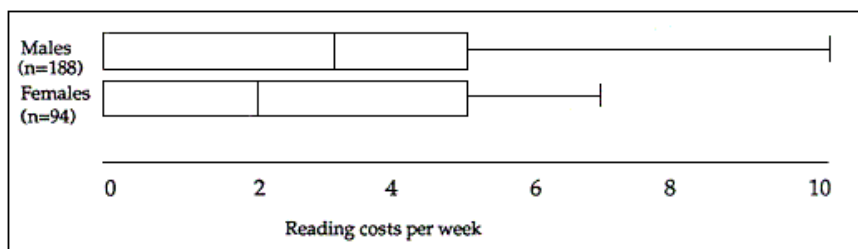
Help Stephen by comparing the two distributions in the space below.

Reference

Garfield, J., Zieffler, A., & Lane-Getaz, S. (2005). EPSY 3264 Course Packet, University of Minnesota, Minneapolis, MN.
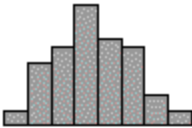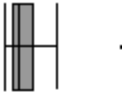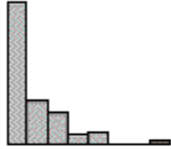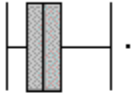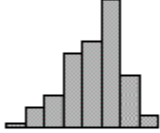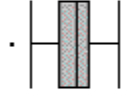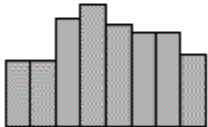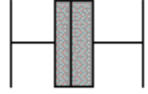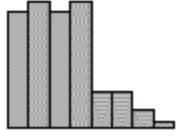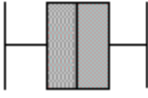
Table Number:_____ Group Name: _____

Group Members:_____ _____ _____

Matching Histograms to Boxplots

Consider all of the graphs below, the histograms and the boxplots, to be on the same scale. Match each histogram with its corresponding boxplot, by writing the letter of the box plot in the space provided.

Reference
Scheaffer, R.L., Watkins, A., Witmer, J., & Gnanadesikan, M. (2004a). *Activity-based statistics: Instructor resources* (2nd edition, Revised by Tim Erickson). Key College Publishing.