Name_____

Feeling Bloated for Statistics

Statistical objectives:

- Developing proficiency in using computer software (*Excel* and *Fathom*) to organize and analyze data.
- Becoming familiar with the visual characteristics of a scatter plot for two correlated variables.
- Realizing that correlation does not necessarily imply causation.
- Understanding what the correlation coefficient r tells about the strength of the linear relationship between two variables.
- Using the equation of the least-squares regression line to formulate and answer important questions arising from the data and its context.
- Interpreting the slope of the least-squares regression line.
- Realizing the danger of extrapolating too far from the given data.
- Recognizing how "influential" points affect the equation of the regression line.
- Thinking about how "influential" points may arise from the design of an experiment.
- Constructing residual plots using Fathom, and using them to identify "influential" points.
- Becoming familiar with the characteristics of the residual plots generated by two correlated variables.
- Understanding the difference between the correlation coefficient and the slope of the least squares regression line.
- Investigating real-world contexts in which correlation between two variables occurs.

1) Transfer the data for your creature's length and width from Excel to Fathom by following these steps:
   a. In your Excel document, highlight the data in the columns showing the length vs. width for your creature as it grew.
   b. Right click on the highlighted area and select "copy." There should be a dashed line moving around the data you wish to copy.
   c. Start Fathom. Drag a new collection (icon in upper left hand corner) onto the main screen.
   d. Right click on the new collection and select "paste cases."
   e. Right click on the new collection again, and select "inspect collection," and make sure that all cases have been transferred to the Fathom collection by scrolling through the data using the arrows at the bottom of the "inspect collection" box.

2) Use Fathom to construct a scatter plot, with length on the x-axis and width on the y-axis. Does it appear as if the two variables are correlated? If so, go on to #3. If not, use the data set from another student for this activity (pick a data set with two variables which do appear to be correlated).

3) Would it be reasonable to say that the length of the creature has an effect upon its width? Could length and width be correlated even if one of them does not cause the other? Why or why not?

4) How strong does the linear relationship appear to be between the two variables you have selected? If you were to estimate the value of r, what would it be?

5) Construct a "movable line" using Fathom. Do this by going to "graph" and selecting "movable line." Move the line until you think it fits the data well. Record the equation of your movable line here_____. Copy all of your Fathom output at this point into a MS Word Document by left clicking on your output, selecting "copy pictures" from the "edit menu," opening a MS Word Document, right clicking on the MS Word Document, and selecting "paste." Label the output "steps 1-4." Save the MS Word document as "_____lab1." (Fill in the blank with your last name).

6) What is the y-intercept of your movable line? Why is the y-intercept in the position that it is in? What useful information, if any, does the y-intercept give you?

7) Use Fathom to plot a regression line for your data. Do this by going to "graph" and selecting "least squares line." How close is the regression line to your movable line?

8) What is the y-intercept of the regression line? How did you find it? What is the slope of the regression line? How did you find it? How close were the slope and y-intercept of the movable line to the slope and y-intercept of the regression line? (After you answer these, go back to the "graph" menu and unselect your movable line).

9) What is the x-intercept of the regression line? What useful information, if any, does it give you?

10) What is the slope of the regression line? What useful information, if any, does it give you?

11) What is the difference between the slope of the least squares regression line and the correlation coefficient *r*?

12) Write a question whose answer would be given by the slope of your regression line.

13) Write a question that you could answer by using the equation for your regression line. Answer the question you have written, and explain how the regression line helped you find the answer.

14) Use Fathom to construct a residual plot for the data. Do this by left clicking on the graph, going to the "graph" menu, and selecting "make residual plot." What exactly is the residual plot showing you? What information does it give?

15) At this point, copy all of the Fathom output you have done in steps 6-14 into your MS Word Document "_____lab1." Label the output: "Steps 6-14."

16) Using your residual plot and/or scatter plot, identify 2 points lying farthest from the regression line. List the points here, and explain why you think they are so far from the regression line:

17) Delete the 2 data points which you found in #16. How does deleting these points affect the correlation between the two variables (i.e., what is the new value of *r*)? How does deleting the points effect the regression line equation? Record the new equation here:_____. How does deleting the points affect the graph? How does deleting the points affect the equation?

18) At this point, copy and paste the Fathom output from steps 16-17 into your MS Word Document "_____lab1." Label the output "Steps 16-17."

19) Go back and read the question you wrote for #13. Answer it again, this time using the "new" regression equation in #17. How do the two answers differ? Is one any "better" than the other? Why or why not?

Extension: Describe up to two different real-world situations where the correlation between two variables would be crucial to know. Tell what the application is, where or in what industry it would be important, how the correlation would be measured – include units used, and who would use this information. Site your sources – book, magazine, web page, etc.

"Feeling Bloated III" and the NCTM *Principles and Standards for School Mathematics*

Activity Objectives

- Developing proficiency in using computer software (*Excel* and *Fathom*) to organize and analyze data.
- Becoming familiar with the visual characteristics of a scatter plot for two correlated variables.
- Realizing that correlation does not necessarily imply causation.
- Understanding what the correlation coefficient *r* tells about the strength of the linear relationship between two variables.
- Using the equation of the least-squares regression line to formulate and answer important questions arising from the data and its context.
- Interpreting the slope of the least-squares regression line.
- Realizing the danger of extrapolating too far from the given data.
- Recognizing how "influential" points affect the equation of the regression line.
- Thinking about how "influential" points may arise from the design of an experiment.
- Constructing residual plots using Fathom, and using them to identify "influential" points.
- Becoming familiar with the characteristics of the residual plots generated by two correlated variables.
- Understanding the difference between the correlation coefficient and the slope of the least squares regression line.
- Investigating real-world contexts in which correlation between two variables occurs.

Principles addressed (source: http://standards.nctm.org/document/chapter2/index.htm)

Teaching principle: " Effective mathematics teaching requires understanding what students know and need to learn and then challenging and supporting them to learn it well." *As students work through "Feeling Bloated III," their responses to questions give the teacher an understanding of what they know about correlation. Their responses to the activity can help inform future instructional plans.*

Learning Principle: "Students must learn mathematics with understanding, actively building new knowledge from experience and prior knowledge." *Students use their background knowledge from Feeling Bloated II in order to begin to make sense of Feeling Bloated III. The context in which Feeling Bloated II took place helps them begin to understand the statistical concepts inherent in Feeling Bloated III.*

Assessment Principle: "Assessment should support the learning of important mathematics and furnish useful information to both teachers and students." *Feeling Bloated III incorporates a number of important concepts from statistics, as listed in the statistical objectives section at the beginning of the paper. As such, student replies to the parts of the activity give the teacher a well-rounded picture of how much students understand about the concept of correlation.*

Technology Principle: "Technology is essential in teaching and learning mathematics; it influences the mathematics that is taught and enhances students' learning." *In Feeling Bloated III, students learn the power of statistical software for exploratory data analysis.*

Alignment with Data Analysis and Probability Standard for Grades 9–12 (source: http://standards.nctm.org/document/chapter7/data.htm)

- Students "Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them" as they work through the extension.
- The standard recommends, "For bivariate measurement data, (students) be able to display a scatterplot, describe its shape, and determine regression coefficients, regression equations, and correlation coefficients using technological tools." These are all addressed in the activity.
- Students also "identify trends in bivariate data and find functions that model the data," as recommended by the standard.