NAME: _____ DATE: _____

Correlation and Regression Applet Exercises

This interactive correlation and regression applet at the **www.whfreeman.com/yates2e** Web site shows how the correlation r and the least-squares regression line respond to changes in a scatterplot. No amount of talking or writing can show so clearly how these statistical measures behave. The following exercises point to some important facts, but we hope that you will experiment a bit to gain some feeling for correlation and regression. To erase an entire scatterplot and start over, click on the trash can.

App3.1 Match the correlation. You are going to make scatterplots with 10 points that have correlation close to 0.7. The lesson is that many patterns can have the same correlation. Always plot your data before you trust a correlation.

(a) Stop after adding the first two points. What is the value of the correlation? Why does it have this value?

(b) Make a lower left to upper right pattern of 10 points with correlation about $r = 0.7$. (You can drag points up or down to adjust r after you have 10 points.) Make a rough sketch of your scatterplot.

(c) Make another scatterplot with 9 points in a vertical stack at the left of the plot. Add one point far to the right and move it until the correlation is close to 0.7. Make a rough sketch of your scatterplot.

(d) Make yet another scatterplot with 10 points in a curved pattern that starts at the lower left, rises to the right, then falls again at the far right. Adjust the points up or down until you have a quite smooth curve with correlation close to 0.7. Make a rough sketch of this scatterplot also.

App3.2 Is regression useful? In the previous exercise, you created three scatterplots having correlation about $r = 0.7$ between the horizontal variable x and the vertical variable y. Correlation $r = 0.7$ is considered reasonably strong in many areas of scientific work. Because there is a reasonably strong correlation, we might use a regression line to predict y from x. In which of your three scatterplots does it make sense to use a straight line for prediction?

App3.3 Influence on correlation. Click on the scatterplot to create a group of 10 points in the lower left corner of the scatterplot with a strong straight-line pattern (correlation about 0.9).

(a) Add one point at the upper right that is in line with the first 10. How does the correlation change?

(b) Drag this last point down until it is opposite the group of 10 points. How small can you make the correlation? Can you make the correlation negative? You see that a single outlier can greatly strengthen or weaken a correlation. Always plot your data to check for outlying points.

App3.4 Influence in regression. As in the previous exercise, create a group of 10 points in the lower left corner of the scatterplot with a strong straight-line pattern (correlation at least 0.9). Click the "Show least-squares line" box to display the regression line.

(a) Add one point at the upper right that is far from the other 10 points but exactly on the regression line. Why does this outlier have no effect on the line even though it changes the correlation?

(b) Now drag this last point down until it is opposite the group of 10 points. You see that one end of the least-squares line chases this single point, while the other end remains near the middle of the original group of 10. What about the last point makes it so influential?

App3.5 Guessing a regression line. Click on the scatterplot to create a group of 15 to 20 points from lower left to upper right with a clear positive straight-line pattern (correlation around 0.7). Click the "Draw line" button and use the mouse (right-click and drag) to draw a line through the middle of the cloud of points from lower left to upper right. Note the "thermometer" above the plot. The red portion is the sum of the squared vertical distances from the points in the plot to the least-squares line. The green portion is the "extra" sum of squares for your line – it shows by how much your line misses the smallest possible sum of squares.

(a) You drew a line by eye through the middle of the pattern. Yet the right-hand part of the bar is solid green. What does that tell you?

(b) Now click the "Show least-squares line" box. Is the slope of the least-squares line smaller (the new line is less steep) or larger (line is steeper) than that of your line? If you repeat this exercise several times, you will consistently get the same result. The least-squares line minimizes the *vertical* distances of the points from the line. It is *not* the line through the "middle" of the cloud of points. This is one reason why it is hard to draw a good regression line by eye.

App3.6 A special point. With the last scatterplot and least-squares line still on the graph, click in the box "Show mean X and mean Y lines." The vertical blue line has equation $x = \bar{x}$, and the horizontal line has equation $y = \bar{y}$. What are the coordinates of the intersection of these two lines? What can you say about this intersection point relative to the least-squares line?

Click on one of the scatterplot points, and drag it all over the screen. What happens to the intersection point as the least squares line is pulled one way and then the other? Write a sentence that describes the relationship between the point and the least-squares line.