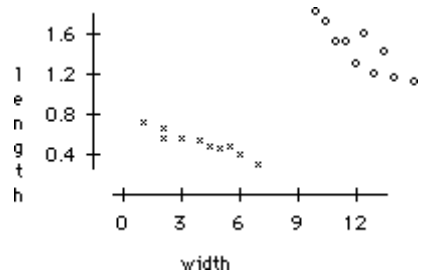
Ch3 Review

1. For children between the ages of 18 months and 29 months, there is approximately a linear relationship between height and age. The relationship can be represented by $\hat{y} = 64.93 + 0.63x$, where y represents height (in centimeters) and x represents age (in months). Joseph is 22.5 months old and is 80 centimeters tall. What is Joseph's residual?

` (a) 79.1 (b) -0.9 (c) 0.9 (d) 56.6 (e) 64.93

2. A study examined the relationship between the sepal length and sepal width for two varieties of an exotic tropical plant. Varieties A and B are represented by x's and o's, respectively, in the following scatterplot. Which of the following statements is FALSE?



   (a) Considering Variety A only, there is a negative correlation between sepal length and width.
   (b) Considering Variety B only, the least-squares regression line for predicting sepal length from sepal width has a negative slope.
   (c) Considering both varieties together, there is a positive correlation between length and width.
   (d) Considering each variety separately, there is a positive correlation between sepal length and width.
   (e) Considering both varieties together, the least-squares regression line for predicting sepal length from sepal width has a positive slope.

3. On May 11, 50 randomly selected subjects had their systolic blood pressure (SBP) recorded twice—the first time at about 9:00 a.m. and the second time at about 2:00 p.m. If one were to examine the relationship between the morning and afternoon readings, then one might expect the correlation to be
   (a) near zero, as morning and afternoon readings should be independent.
   (b) high and positive, as those with relatively high readings in the morning will tend to have relatively high readings in the afternoon.
   (c) high and negative, as those with relatively high readings in the morning will tend to have relatively low readings in the afternoon.
   (d) near zero, as correlation measures the strength of the linear association.
   (e) near zero, as blood pressure readings should follow approximately a Normal distribution.

4. Suppose we fit a least-squares regression line to a set of data. What is true if a plot of the residuals shows a curved pattern?
   (a) A straight line is not a good model for the data. (b) The correlation must be 0.
   (c) The correlation must be positive. (d) Outliers must be present.
   (e) The regression line might or might not be a good model for the data, depending on the direction of the curve.

One concern about the depletion of the ozone layer is that the increase in ultraviolet (UV) light will decrease crop yields. An experiment was conducted in a green house where soybean plants were exposed to varying levels of UV, measured in Dobson units. At the end of the experiment the yield (kg) was measured. A regression analysis was performed with the following results:

Parameter Estimates

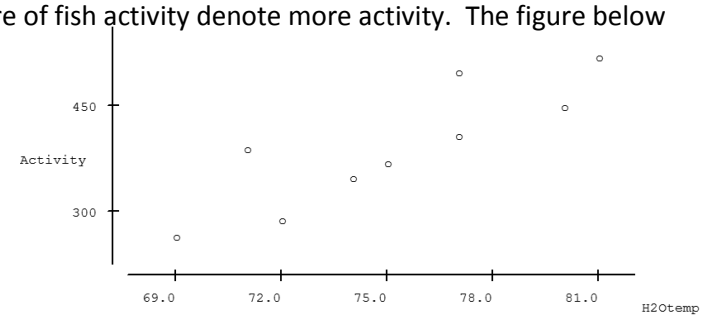
Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	3.9800118	0.053774	74.01	<.0001	3.8638398	4.0961838
uv	-0.046285	0.010741	* hidden *	0.0008	**** hidden ****	

5. The least-squares regression line is the line that
   (a) minimizes the sum of the squared differences between the actual UV values and the predicted UV values.
   (b) minimizes the sum of the squared residuals between the actual yield and the predicted yield.
   (c) minimizes the sum of the squared differences between the actual yield and the predicted UV.
   (d) minimizes the sum of the squared residuals between the actual UV reading and the predicted UV reading.
   (e) minimizes the total variation in the data.

6. Which of the following is correct?
   (a) If the UV reading increases by 1 Dobson unit, the yield is expected to increase by 0.0463 kg.
   (b) If the yield increases by 1 kg, the UV reading is expected to decline by 0.0463 Dobson units.
   (c) The estimated yield is 3.98 kg when the UV reading is 0 Dobson units.
   (d) The predicted yield is 4.3 kg when the UV reading is 20 Dobson units.
   (e) None of these

7. The following are resistant:
   (a) Least-squares regression line (b) Correlation coefficient
   (c) Both (a) and (b) (d) Neither (a) nor (b) (e) It depends

Joey read in his biology book that fish activity increases with water temperature, and he decided to investigate this issue by conducting an experiment. On nine successive days, he measures fish activity and water temperature (in degrees Fahrenheit) in his aquarium. Larger values of his measure of fish activity denote more activity. The figure below presents the scatterplot of his data.



8. Describe the scatterplot.
10. If temperature were measured in degrees Celsius instead of degrees Fahrenheit, how would the correlation change?
    Note that $F = \frac{9}{5}C + 32$.
11. Suppose a new point at (66, 500), that is, water temperature = 66°F and fish activity = 500, is added to the plot. What effect, if any, will this new point have on the correlation between fish activity and water temperature? Justify your answer.

12. Is cardiovascular fitness (as measured by time to exhaustion running on a treadmill) related to an athlete's performance in a 20-km ski cross country ski-race? The following data shows x = treadmill run time to exhaustion (in minutes) and y = 20km ski time (in minutes).

Person	Treadmill Time	Ski Time
1	7.7	71
2	8.4	71.4
3	8.7	65
4	9	68.7
5	9.6	64.4
6	9.6	69.4
7	10	63
8	10.2	64.6
9	10.4	66.9
10	11	62.6
11	11.7	61.7

a. Why should treadmill time be the explanatory variable? Explain.
b. Draw a scatterplot and interpret a scatterplot for the data.
c. Calculate the least squares line and graph it on the scatterplot.
d. Interpret the slope in the context of the problem.
e. Interpret the x- and y-intercepts in the context of the problem. Are these reasonable values in this context?
f. Find and interpret the value of the correlation coefficient.
g. If the time was measured in seconds, how would this value change?
h. If r is high, can we conclude that a change in treadmill time causes the ski time to change? Explain.
i. Calculate and interpret the residual for the first point in the data set.
j. Sketch the residual plot. What does it tell you?
k. Calculate and interpret the values of *s* and r^2 in the context of the problem.
l. If you were to use number of hours instead of number of minutes for the ski time, how would the values of r^2 change?
m. Predict the ski time for a runner who can last 8 minutes on the treadmill.
n. Suppose the observation (13, 59) was added to the data set. What effect will this have on the LSRL and the values of r, r^2, s?
o. Instead of (13, 59), suppose the new point was (13, 65). What effect will this have on the LSRL and the values of r, r^2, s?