For each statement below, explain why the statement <u>cannot</u> be true:

1) The correlation between hair color and GPA is -0.25.

2) The correlation between height and weight is 0.73 inches per pound.

3) Scientists have found a high correlation (r = 1.43) between lipid concentration in a blood sample and age of the blood sample taken.

Match each pair of variables to its most likely value of r.

4) Height and armspan in adult males   _____

             a.   -0.92

5) SAT score and height in inches of 16-year olds   _____

             b. +0.93

6) Number of hours a person has been up past their normal bedtime & _____
number of minutes it takes the person to do a crossword puzzle

             c. +0.71

7) Temperature outside and the number of minutes it takes an ice cube to _____
melt on the sidewalk

             d. Close to 0

             e. -1.00

8) The amount of time a person spends sleeping and amount of time awake _____
in a 24-hour period.

---

9) The correlation (r) between speed and gas mileage of a vehicle is close to zero, even though speed and gas mileage are definitely related. Explain how that is possible.

10) A couple of years ago a local newspaper published research results claiming a positive association between the number of years high school children had taken instrumental music lessons and their performances in school (as measured by GPA).

    a. What does "positive association" mean in this context?

    b. A group of parents then went to the School Board demanding more funding for music programs as a "for sure" way to improve student chances for academic success in high school. As a statistician, do you agree or disagree with their reasoning? Explain.

11) Researchers examining life in the Alps recorded the boiling point of water at a variety of air pressures and created the following linear regression.

	Coefficients	Standard Error	t Stat	P-value
Intercept	155.2965	0.927337	167.465	5.85E-26
Pressure (inches Hg)	1.901784	0.036756	51.74075	2.53E-18

R-sq = 99.4%   S = 0.44
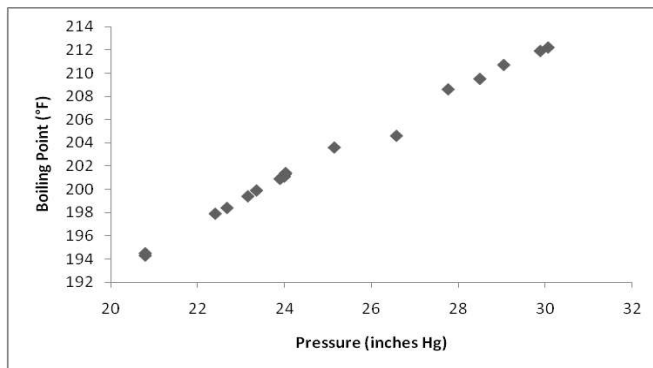
a) Write the equation of the least squares regression line. Be sure to define your variables.

b) Interpret the slope of the line.

c) Interpret the intercept of the line.

d) Find and interpret the correlation coefficient.

e) Interpret R^2.

f) Interpret S.

g) Use your model to predict the boiling point of water when the air pressure is 25 inches Hg.

12) If the point in the upper left corner of the scatterplot is removed, what will happen to the correlation (r) and the slope of the line of best fit (b)?


   a. They will not change.
   b. Both will increase.
   c. Both will decrease.
   d. r will increase and b will decrease.
   e. r will decrease and b will increase.

13) What is the name for the point removed in question #11 above?

14) A manufacturer creates a linear regression model of production (y) v. hours worked (x) for its workers. If a worker has a negative residual, what does that mean?

15) The residual plot for a linear model is shown. Which is true?
   a. The linear model is okay because approximately the same number of points are above the line as below it.
   b. The linear model is okay because the association between the two variables is fairly strong.
   c. The linear model is no good because the correlation is near 0
   d. The linear model is no good because some residuals are large.
   e. The linear model is no good because of the curve in the residuals.
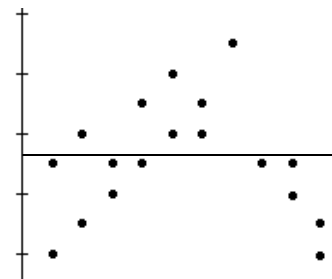
16) It's easy to measure the circumference of a tree's trunk, but not so easy to measure its height. Foresters developed a model for ponderosa pines that they use to predict the tree's height (in feet) from the circumference of its trunk (in inches): ln $\hat{h}$ = -1.2 + 1.4(ln C). A lumberjack finds a tree with a circumference of 60"; how tall does this model estimate the tree to be?

17) Choose the right word to fill in the blanks:

   a) If a model is _____ (linear/exponential), the same value is *added* to y each time x is increased by one unit.

   b) If a model is _____(linear/exponential) y is *multiplied* by the same value each time x is increased by one unit.

18) Students in a physics class are studying free-fall to determine the relationship between the distance an object has fallen and the amount of time since release. They record data from an experiment in which a rock is dropped from a platform. A scatterplot of distance (y) vs. time (x) showed that the data was not linear, so the students used the re-expression $\sqrt{\text{distance}}$ v. time. Examine the regression output and residuals below.

	Coefficients	Standard Error	t Stat	P-value
Intercept	0.111436	0.114482	0.973393	0.344006
Time (sec)	22.01768	0.203536	108.1758	1.43E-25

S = 0.199   R-sq = 99.8%
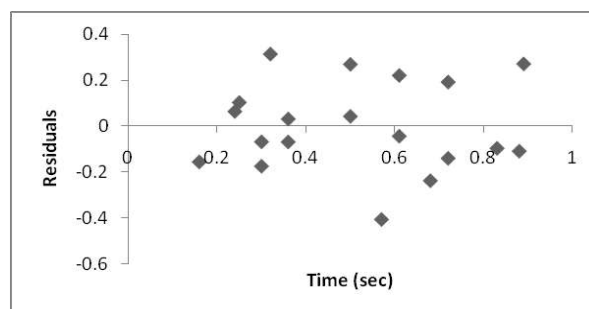
a) Does this model appear to be a good fit? Explain.

b) Write the equation of the least squares regression line. Be sure to define your variables.

c) Use the equation to predict the distance fallen after 2 seconds.

d) Even though this model was a good fit, you should not be too confident in your prediction for part (c). Explain why not.