

8: Regression and Correlation -- Answers

1.
 - a) 14.55 lbs.
 - b) 85.85 lbs. This is not reasonable (it is too high) since a child's weight gain does not remain constant
 - c) The y -intercept would change to 8.5 and the equation would be $y = 2.1x + 8.5$.
 - d) He gained 2 lbs each month.
2. The correct answer is (b). For (a), the y -intercept is obviously not 1.15; for (c), the slope is too large; for (d), the slope is obviously not negative.
3. They should look similar. The least-squares regression line is calculated by minimizing the sum of the squares of the vertical distances between points in the scatterplot and the line.
4. No, that would not minimize the sum of the squares of the vertical distances between points in the scatterplot and the line. We could just slide the line down and each of those vertical distances would be smaller.
5. Correlation measures the strength and direction of the linear relationship between two numerical variables.
6. (a) $r = 0$; (b) $r = -0.6$; (c) $r = 0.6$; (d) $r = 0.98$; (e) $r = 0.85$; (f) $r = -0.85$.
7.
 - a) No, an increase in age causes both height and math ability to increase.
 - b) No, cold weather causes people to stay inside and share germs and that causes colds.
 - c) No, drinking in a bar often happens while smoking and/or breathing in second-hand smoke. The smoke causes lung cancer.
8. A correlation of zero means there is no relationship between the variables what is being described is a negative relationship between coffee consumption and height.
9. The correlation would 1 since all the points would lie on the line $y = x - 3$, where x is height and y is arm span.
10.
 - a) Since the slope of the regression equation is negative, the correlation should be negative.
 - b) A correlation cannot be greater than 1.
11. As cars increase in age, their values decrease.
12.
 - a) $r \approx 0.91$. There is a high correlation between the heights and the arm spans.
 - b) $\hat{y} = 0.88x + 20.4$
13.
 - a) $\hat{y} = 2x + 3$; $r = 1$
 - b) Since the correlation equals 1, the regression equation fits the data perfectly.

14.

- a) As the number of assignments not turned in increases, the final grade decreases.
- b) For $x = 0$, $y = 3.01$. Thus, when there are no missing assignments, the average grade was 3.01.
- c) Every missing assignment lowered the average gpa by 0.248.
- d) 2.266
- e) The correlation was -0.473, which is not highly correlated, but there is some correlation.

15.

- a) The data closely follow to the linear equation $\hat{y} = 7.25x + 207$.
- b) 323 Million Metric Tons. Since the correlation was so high, we can be confident that our answer is quite accurate for the domain 1992 to 2001. Since the year 2008 is 7 years in the future of the given data, it may be accurate or things could change that would limit carbon emissions so they don't grow as much.