Section 6.1

1)

a) Yes. Explanatory variable is the number of semesters. Response variable is the number of credits earned.

b) No. These variables have nothing to do with one another.

c) Yes. Explanatory variable is the number of years employed. Response variable is the annual salary.

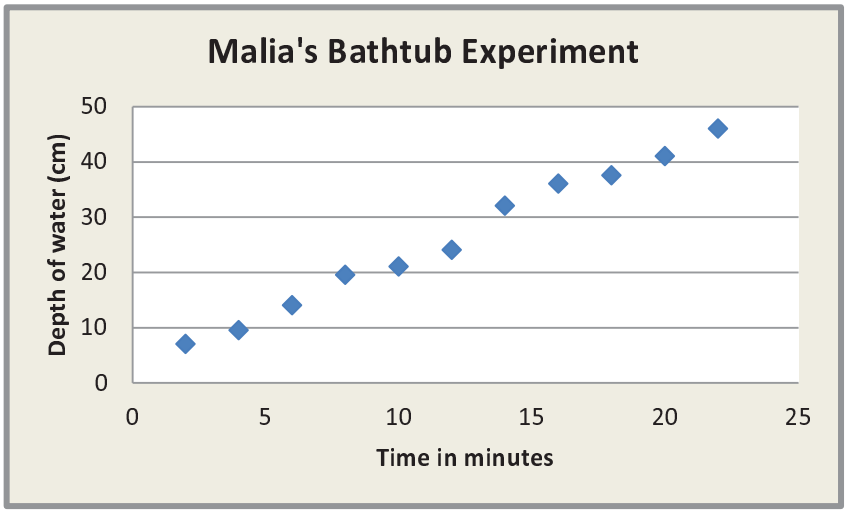
d) Yes. Explanatory variable is the number of months having owned the IPod. Response variable is the number of songs.

2) *The relationship between the amount of time studying by a student and his or her score on the mid-term exam is strong and roughly linear. The association is positive and has no outliers. As the number of hours studying increases, so does the score on the test.*

3) Malia's Bathtub Experiment

a) Explanatory variable is the time (in minutes) that the water has been running. Response variable is the depth (in centimeters) of the water.

b)



c*) The relationship between the number of minutes that the water has been running and the depth (in centimeters) of the water in the bathtub is extremely linear and very strong. There are no outliers and the association is clearly positive. As the time increases, so does the depth of the water.*

4) Peanut Butter

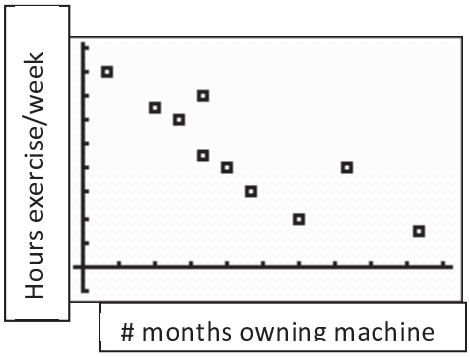
a) Explanatory variable is the quality rating. The response variable is the price.

b) *The relationship between the quality rating (on a scale of 0 to 100) and the price per ounce for these various brands of peanut butter is very weak. There is a slight positive trend, showing that peanut butter brands with higher quality ratings tend to have higher prices. However, the relationship has no clear form and a few possible outliers.*

5) Mr. Exercise

a) Explanatory variable is the number of months having owned the machine. Response variable is the number of hours of exercise per week.

b)



c) *As the number of months that a customer has owned an exercise machine increases, the number of hours that they exercise per week decreases. The relationship is moderately strong and shows a negative, linear trend. There are a couple possible outliers, such as the person who has owned their machine for 11 months and the one who has owned their machine for 14 months.*

6) Nevada Temperatures & Elevations

a) Explanatory variable is the elevation. Response variable is the mean annual Celsius temperature.

b*) For these locations in Nevada, the relationship between elevation and mean temperature is moderately strong and fairly linear. The association is negative, with a couple possible outliers. The graph shows that as the elevation for these locations in Nevada increases, the mean annual temperature decreases.*

7) (4/52)(3/51) = 1/221 = 0.0045

8) (13/52)(13/52) =1 /16 = 0.0625

9)

a) (7/41)(11/40) = 77/1640 = 0.0470

b) (27/41)(26/40) = 351/820 = 0.4280

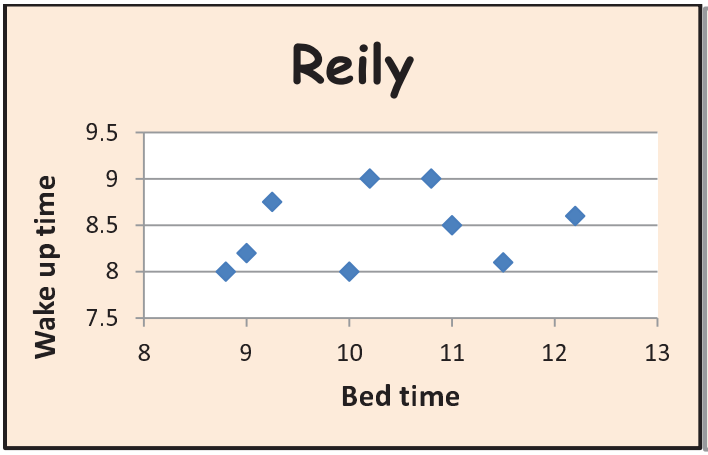
Section 6.2

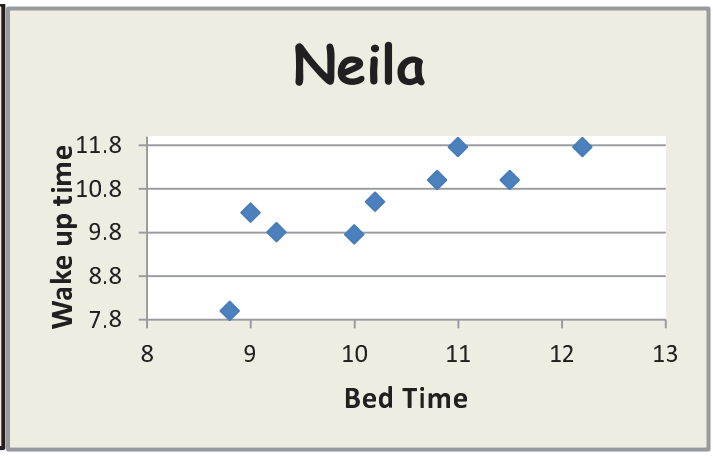
1) The direction and the strength of the linear relationship between two numerical variables.

2) $r = \sqrt{r^2} = \sqrt{0.805} = 0.8972$

The correlation will be positive because you can see in the graph that the association is positive. So r = +0.8972.

3) a) For each graph the explanatory variable is the bed time and wake up time is response.





b) C. Reily's correlation will be close to zero.

c) B. Neila's correlation will likely be positive, but not close to 1.

4) Answers will vary.

a) No. Ice cream sales and drowning deaths will both increase in the summer months. So this is an example of common response.

b) No. The number of pirates and the amount of global warming have nothing to do with one another. This is most likely a coincidence.

c) No. More severe fires require more fire fighters and do more damage. So, this is an example of common response.

d) No. The fact that each player has selected his or her own stick, and its amount of flex, is a lurking variable. Perhaps the players who are high-scorers are more likely to select a more flexible stick. This is an example of a confounding.

5) *No. Maybe the relationship is reversed. Perhaps men who abuse alcohol are more likely to get divorced. Or, perhaps these men were under major stress due to some other issues in their lives, which then caused them to abuse alcohol and to fight with their wives.*

6) *No. These people wanted to lose weight, so it is likely that they also exercised and ate more healthy food.*

7) Graph #1 = E; Graph #2 = C; Graph #3 = B; Graph #4 = A; Graph #5 = D

8) Answers will vary. A scatterplot that resembles a parabola is a possible example.

9) Beach Visitors

a) Explanatory is average daily temperature and response variable is number of beach visitors.

b) Correlation will be positive and close to 1. Approximately r = +0.88.

c) *The relationship between the average daily temperature and the number of beach visitors is strong and positive. The trend is moderately linear with one outlier at 87° and about 125 visitors. As the average daily temperature increases, the number of visitors to the beach also increases.*

10) Experimental probability of getting tails is: $P(Tails) = \frac{34}{93} \approx 0.3656 \approx 36.56\%$

11) 0.258 (20) = 5.16. So, I would expect Stephanie to get approximately 5. 16 hits out of her next 20 times at bat.

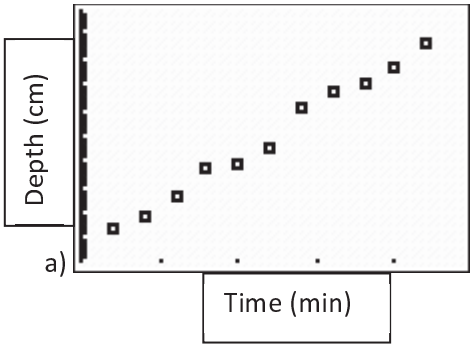
12) Experimental probability of getting a Yahtzee: $P(Yahtzee) = \frac{3}{79} \approx 0.03797 \approx 3.797\%$

13) Theoretical probability of getting a Yahtzee: $P(Yahtzee) = \frac{6}{6^5} \approx 0.00077 \approx 0.077\%$

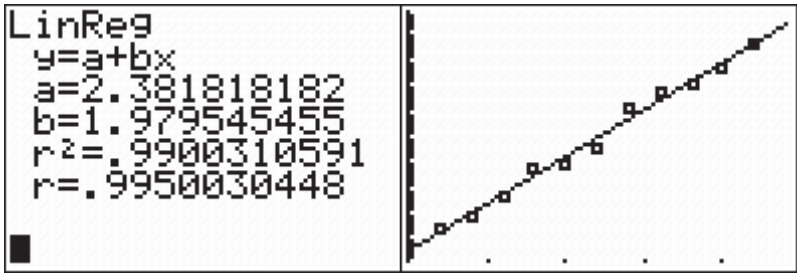
Section 6.3

1) Malia's Bathtub Experiment

a)



b)



$\hat{y} = 2.3818 + 1.9795x$

x = number of minutes water has been running

$\hat{y}$ = predicted depth (in centimeters) of the water

c) r = 0.9950. This tells us that the relationship is positive and very strong.

d) Slope is 1.9795. For each increase of one minute that the water runs, the depth of the water is predicted to increase by 1.9795 centimeters.

e)

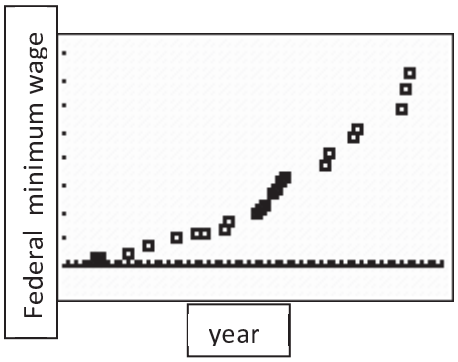
$\hat{y} = 2.3818 + 1.9795(17) = 36.0333\ cm$

$\hat{y} = 2.3818 + 1.9795(60) = 121.1518\ cm$

f) The first answer is reasonable, because it is within our actual data (interpolation). However, the answer for 60 minutes is way beyond the data (extrapolation) and is not reasonable at all. 121.1518 centimeters is much taller than the average bathtub, the water would have started running over and flooding long before this.

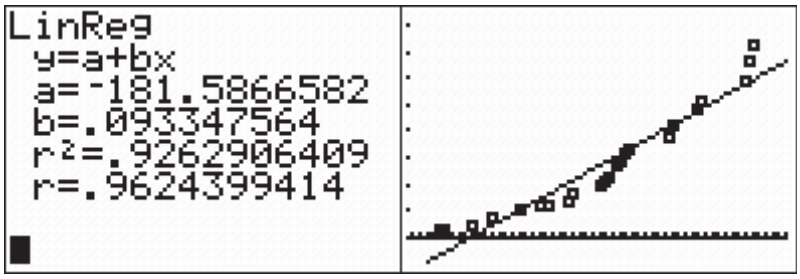
2) U.S. Minimum Wage

a)



b) *The relationship between the years from 1938 to 2009 and the Federal Minimum Wage is extremely strong and positive. The graph shows a clearly curved relationship between minimum wage and year. This non-linear relationship has no obvious outliers. As the years have increased, so has the Federal Minimum Wage in the United States.*

c)



$\hat{y} = -181.5867 + 0.0933x$

X = the year

$\hat{y}$ = the predicted Federal Minimum Wage in the U.S.

d) r = 0.9624. While this is a very high correlation, it is clear by looking at the graph that a curved model would be a much more appropriate fit for this data.

e)

$\hat{y} = -181.5867 + 0.0933(2012) \approx \$6.13$ *(with calculator it is $$6.23)*

No, this is not an accurate prediction. We know that the minimum wage is higher than this. The actual data points in the graph are curving up at a much higher rate than the line that we fit to the data is. So, using the LSRL equation to make predictions is not appropriate, especially prior to 1950 and after 2007.

f)

$\hat{y} = -181.5867 + 0.0933(1968) \approx \$2.03$ *(with calculator it is $2.12)*

The actual minimum wage in 1968 was $1.60. So, our prediction was fairly close, but too high. Our prediction was 43 cents above the actual minimum wage.

3) Father & Son IQ's

a) Explanatory variable is father's IQ and response variable is son's IQ.

b) Slope is 0.9. For each increase of 1 point in the father's IQ, there is a predicted increase of 0.9 points in the son's IQ.

c) y-intercept is 12. For a father with an IQ of zero, the son's IQ is predicted to be 12 points. This is clearly ridiculous and is an example of extrapolation.

d) The interpretation of the slope is reasonable based on the equation, but the interpretation of the y-intercept is nonsense. There is no person with an IQ anywhere near zero (or 12). This is extrapolation.

e) $\hat{y} = 12 + 0.9(120) = 120\ IQ\ points$

$\hat{y} = 12 + 0.9(140) = 138\ IQ\ points$

f) If our actual data ranges from 108 to 145, then using 170 in our equation would be an example of extrapolation. The results would not be trustworthy because we would have no idea how the data behaves beyond 145 points.

4) (0.79)(0.79)(0.79) = 0.79^3 = 0.4930 = 49.3% chance

5) (0.79)(0.79)(0.21)*(3 orders) = 0.3932 = 39.32% chance

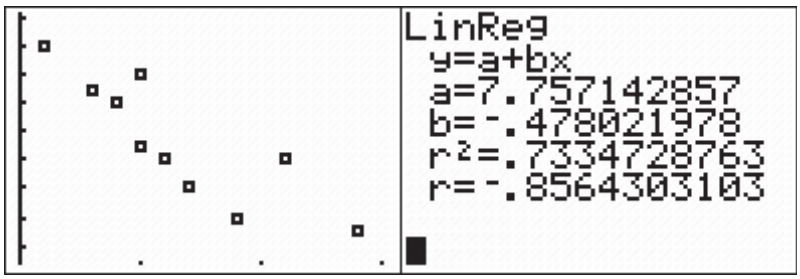
6) 1 – P(miss none) = 1 – [((0.79)(0.79)(0.79)(0.79)] = 1 – 0.79^4 = 0.6105 = 61.05% chance

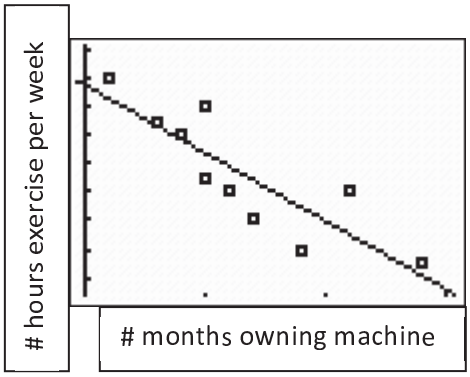
7) C is not appropriate because 00 to 79 is actually 80 numbers.

Section 6.4

1) Mr. Exercise

a)





$\hat{y} = 7.757 - 0.478x$

x = # months owning machine

$\hat{y}$ **= predicted # hours of exercise per week**

b) The slope is – 0.478. For each increase of one month a person owns their exercise machine, there is a predicted decrease of 0.478 hours that they exercise per week.

c) The y-intercept is 7.757. A person who has owned his or her exercise machine for zero months is predicted to exercise 7.757 hours per week.

d) The correlation is r = – 0.8564. This tells us that the relationship between the number of months these people have owned their exercise machines and the number of hours they exercise each week is negative and moderately strong.

e)

$\hat{y} = 7.757 - 0.478(12) = 2.021$. After 12 months, it is predicted that a person will exercise 2.021 hours per week.

$\hat{y} = 7.757 - 0.478(5) = 5.367$. After 5 months, it is predicted that a person will exercise 5.367 hours per week.

f) $9 = 7.757 - 0.478(x)$

$1.243 = -0.478(x)$

$-2.6004 = x$ They have owned it a negative number of months? Ridiculous! According to our equation, this is impossible. No one in our sample exercises more than 8 hours per week. This is an extrapolation.

2) Absences & Final Exam Scores

a) Explanatory is number of absences and response is grade on final exam.

b) *The relationship between the number of absences a student in this Philosophy 103 class has and her or his grade on the final exam is strong and fairly linear. The association is clearly negative and has no obvious outliers. Students with higher number of absences tend to have lower scores on the final exam in this philosophy course.*

c)

$\hat{y} = 91.704 - 1.654(25) = 50.354$. Jeremy is predicted to have approximately 50 points on the final exam.

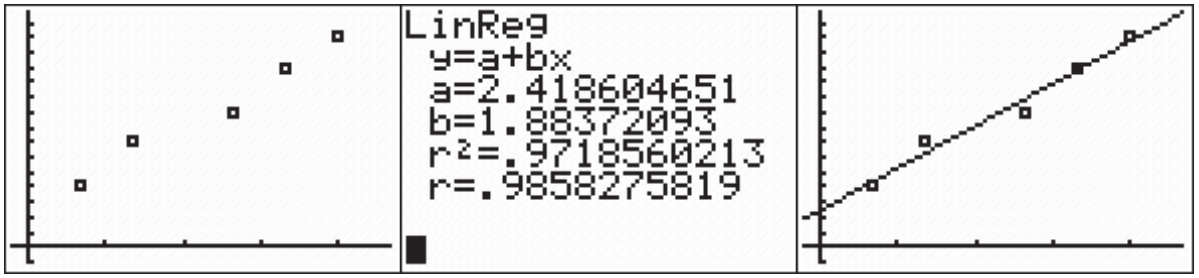
$\hat{y} = 91.704 - 1.654(43) = 20.582$. Lucy is predicted to have approximately 21 points on the final exam.

d) Correlation is r = – 0.9345. This tells us that the relationship between number of absences and grade on the final is negative and strong.

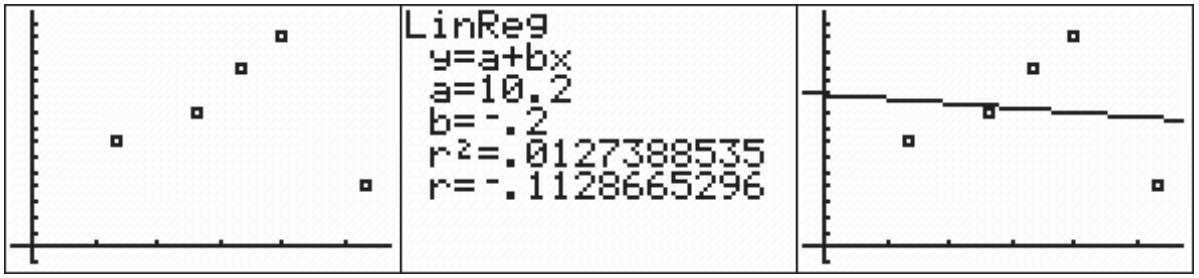
e) – 1.654 is the slope. For each increase of one absence, there is a predicted decrease of 1.654 points on the grade on the final exam.

3) Grade Level & Reading Level

a)



b)



c) There are many changes visible. First of all, the second set of data (b) contains a very obvious outlier. Second of all, the equation changed a ton (both slope and y-intercept). Third, the correlation changed from nearly perfect (r = 0.9858) to almost zero (r = –0.1129). Lastly, the LSRL equation changed from having a positive direction to having a negative direction. Apparently, outliers can really affect the LSRL equations and the correlations of bivariate data!

4) Taco Bell Burritos

a – f) Answers to these will vary, because the students select the data they wish to analyze.

5) Calculator Output

a)

$\hat{y} = -2714.8589 + 35.0781x$

x = high temperature that day *(as forecast by meteorologists)*

$\hat{y}$ **= predicted number of people at the Swimtastic Pool & Water-Slides**

b) The slope is 35.0781. For each increase of one degree in high temperature, there is a predicted increase of 35.0781 people who will come to the Swimtastic Pool & Water-Slides that day.

c) The correlation is r = 0.7823. This tells us that the relationship between daily high temperature and number of people attending the Swimtastic Pool & Water-Slides is positive and moderately strong.

d)

$\hat{y} = -2714.8589 + 35.0781(91) = 477.2482$. We would predict about 477 swimmers on a 91° day.

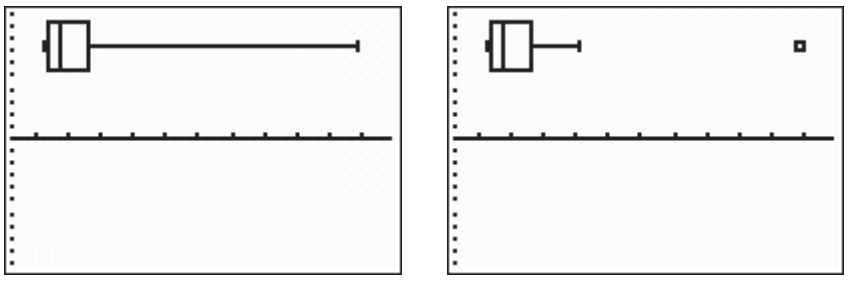
$\hat{y} = -2714.8589 + 35.0781(45) = -1136.3444$. We would predict about –1136 swimmers on a 45° day.

The first answer is reasonable. And it was within the 82-96° temperature range that the equation is based on. The second answer is completely ridiculous. This temperature is far from the actual data, so it is an extrapolation. This must be an outdoor waterpark!

6) Mean salary at Greezy's is $8.76. Standard deviation is $2.69.

7) Five-number-summary is {$7.25, $7.45, $7.80, $8.60, $16.90}

8) TI-84+ screenshots of box-plots (not modified & modified)



9) The median and IQR are much more appropriate here because of the strongly skewed right shape of the graph and the high outlier. The one extremely high salary would have a strong influence on both the mean and the standard deviation.

10) *The salaries at Greezy's Burger Boy range from $7.25 to $16.90 per hour. The distribution is heavily skewed to the right toward the one employee who makes $16.90 (an extreme outlier). The median salary is $7.80, with the majority of employees making less than $9.00 per hour.*

Section 6.5 Chapter 6 Review

1) False

2) True

3) True

4) False

5) r

6) Scatter-plot

7) Explanatory Variable

8) From – 1 to +1

9) The slope

10) Least Squares Regression Line

11) r = – 1 or r = +1

12) $r^2 = 0.7396$

13) r = 0.8775 or r = –0.8775

14) Extrapolation

15) Interpolation

16) Strength, Context, Outliers, Form, and Direction

17) A lurking variable – common response

18)

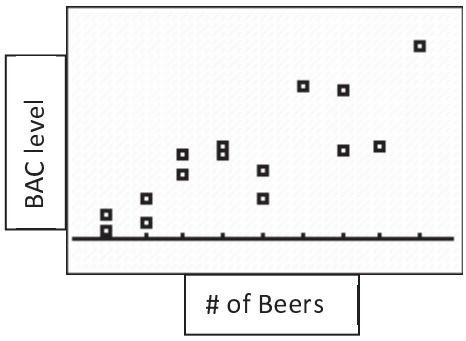
a) No. This is most likely a coincidence. Technology is advancing in many places and the number of cell phones is increasing because they are more affordable and more common. The number of starving children is due to some other factor, such as the population in poor countries increasing or famine, disease or natural disasters.

b) No. Perhaps these are both going up because the company has taken on a great more amount of work. This would make stress levels go up and (perhaps) encourage the company to give its employees raises. This could be common response. Or, maybe the relationship is reversed. And the owners see that the employees are stressed, so they are offering raises to try to help ease that stress.

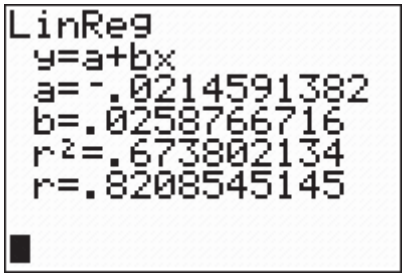
c) Not necessarily. Perhaps people who cannot sleep go smoke. Or, perhaps people who are depressed, stressed or ill have both of these issues – insomnia and cigarette addiction. The relationship is confounded, because we cannot see how one is related to the other.

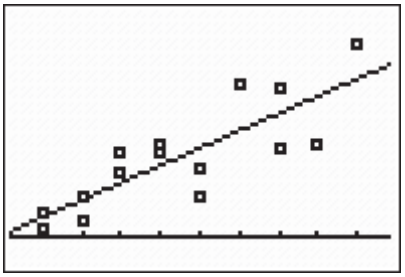
19) Number of beers and BAC level

a) Explanatory variable is the number of beers and response variable is the BAC level.



b)





$\hat{y} = -0.0215 + 0.0259x$

x = # of beers consumed

$\hat{y}$ **= predicted BAC level**

Correlation is r = 0.8209

c) The slope is 0.0259. For each increase of one beer consumed, there is a predicted increase of 0.0259 in the BAC level.

d) The y-intercept is –0.0215. A person who has consumed zero beers has a predicted BAC level of –0.0215. This is extrapolation and makes no sense.

e)

$\hat{y} = -0.0215 + 0.0259(6) = 0.1339.$ If a person drinks 6 beers during this time, they will have a predicted BAC level of 0.1339.

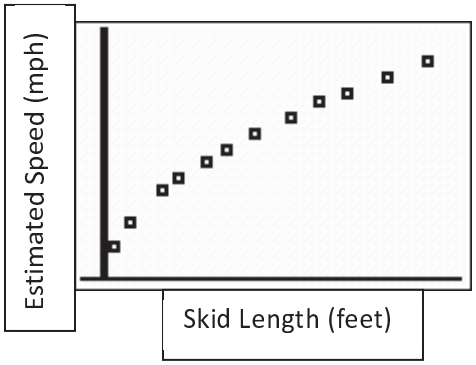
f)

$\hat{y} = -0.0215 + 0.0259(15) = 0.367.$ If a person drinks 15 beers during this time, they will have a predicted BAC level of 0.367.

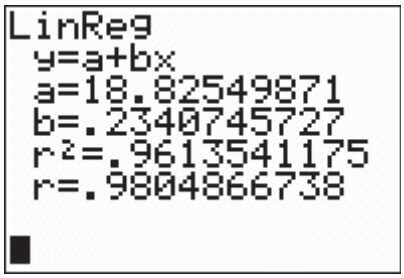
g) We can be fairly confident in the first answer because it is reasonable and it is within the range of our data. However, the second answer uses a number of beers way beyond that of our data, so any results are suspect. This is an extrapolation, so should not be trusted as an accurate prediction.

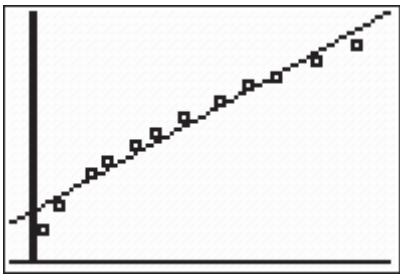
20) Speed of cars

a) Explanatory variable is length of skid mark (in feet) left by car and response is estimated speed (mph) of car at the time of the accident.



b)





$\hat{y} = 18.8255 + 0.2341x$

X = length of skid mark left by car (feet)

$\hat{y}$ = estimated speed of car (mph)

c) *The relationship between the length of the skid mark left by a car (in feet) and its estimated speed (mph) is extremely strong in a positive direction. However, the relationship is clearly not linear. The graph shows a very obvious curved pattern with no outliers. Longer skid marks are highly correlated with faster traveling vehicles.*

d) The correlation is r = 0.9805. Even though this is a very high correlation, it is obvious by looking at the graph that a curved model would fit this data much better than this line does.

e)

$\hat{y} = 18.8255 + 0.2341(157) = 55.5792$. A skid mark of 157 feet has an estimated vehicle speed of about 55.6 miles per hour.

$\hat{y} = 18.8255 + 0.2341(36) = 27.2531$. A skid mark of 36 feet has an estimated vehicle speed of about 27.3 miles per hour.

f) The actual data points on the right side of the graph seem to be increasing more slowly than that of the line, so predictions made with our line equation will most likely overestimate the actual speed of the vehicles. It appears that the line will be above the actual data values. Plus, we would be extrapolating.