CHAPTER 6 SOLUTIONS

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 seem to have nothing to do with one another.

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

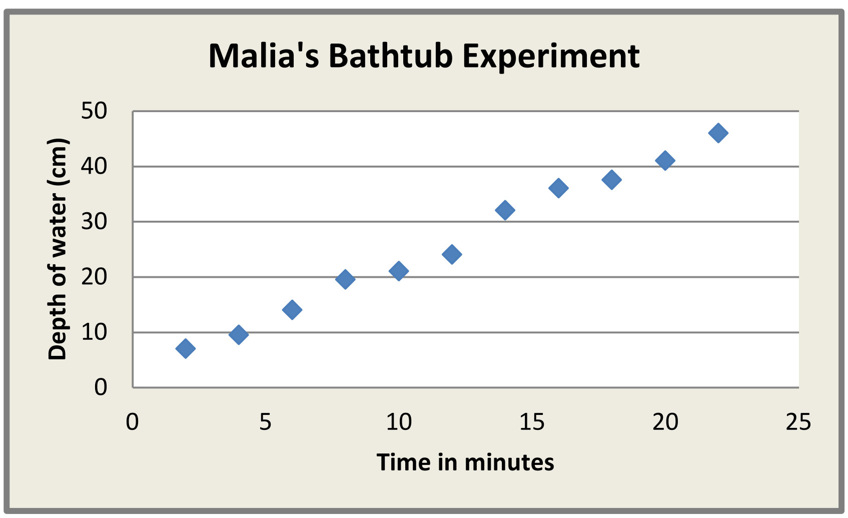
d) Yes. Explanatory variable is the number of months having owned the cell phone. Response variable is the number of applications downloaded.

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 it has no clear outliers. As the number of hours studying increased, so did the score on the test.*

3)

a) The explanatory variable is the time (in minutes) that the water has been running. The 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 linear and very strong. There are no outliers and the association is clearly positive. As the time increased, so did the depth of the water.*

4)

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 weak. There appears to be a positive association, showing that peanut butter brands with higher quality ratings tend to have higher prices. The relationship has no clear form and there are a few possible outliers.*

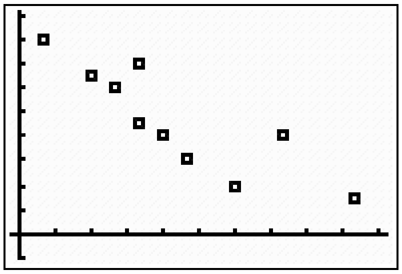
5)

a) The explanatory variable is the number of months having owned the machine and the response variable is the number of hours of exercise per week.

b)

Keep Working Out!

Hours of Exercise per Week



Months Owning Machine

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) The explanatory variable is the elevation. The 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 no clear outliers. The graph shows that as the elevation for these locations in Nevada increases, the mean annual temperature decreases.*

Review Exercises

7) $$\frac{4}{52}\cdot\frac{3}{51}=\frac{12}{2652}=\frac{1}{221}=0.0045$$

8) $$\frac{13}{52}\cdot\frac{13}{52}=\frac{169}{2704}=\frac{1}{16}=0.0625$$

9)

a) $$\frac{7}{41}\cdot\frac{11}{40}=\frac{77}{1640}=0.0470$$

b) $$\frac{27}{41}\cdot\frac{26}{40}=\frac{702}{1640}=\frac{351}{820}=0.4280$$

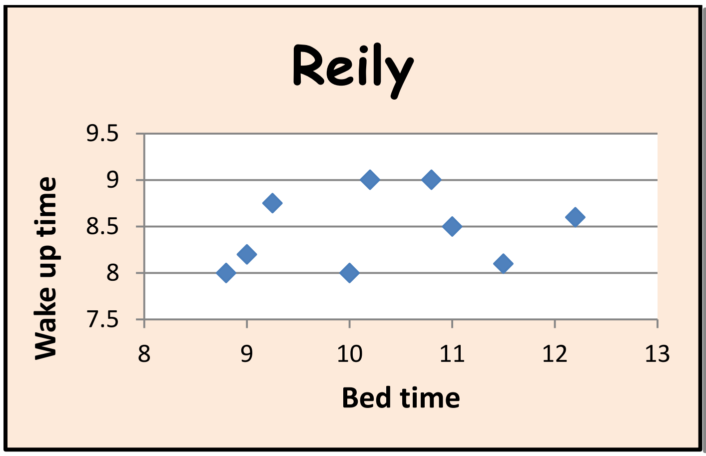
Section 6.2

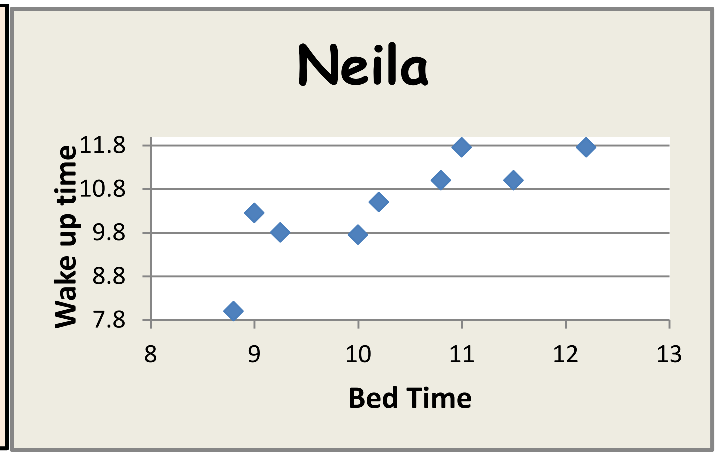
1) The correlation measures the **strength** and **direction** of a linear relationship between two variables.

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

The correlation is positive because the graph moves up from left to right. Therefore, 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 terribly close to 1.

4) Note that answers will vary on each question.

a) No. Ice cream sales and drowning deaths will both increase in the summer months when it is warmer outside. Warm weather tends to contribute to an increase in both ice cream sales and the number of people who will be swimming. This is an example of common response to the lurking variable of the time of year.

b) No. The number of pirates and the increase in 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 also will do more damage. This is an example of common response to the lurking variable of the size of the fire.

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. Maybe players who play defense (and don't score as much) tend to prefer sticks with more flex. This is an example of confounding.

5) No. It may be possible that getting divorced causes increased alcohol abuse but it seems just as likely that the the situation is actually reversed. Perhaps abuse of alcohol is a factor that contributes to the likelihood of getting divorced. Another possibility is that these men were under major stress due to some other issues in their lives which then caused them to abuse alcohol and to also strain their relationship with their wives.

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

7) a) The explanatory variable is average daily temperature and the response variable is number of beach visitors.

b) Correlation will be positive and close to 1. A value of about r = +0.8 or r = +0.9 would be reasonable.

c) The relationship between the average daily temperature and the number of beach visitors is strong and positive. The data is linear with one outlier at 87^{o} and about 125 visitors. As the average daily temperature increases, the number of visitors to the beach also increases.

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

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

Review Exercises

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

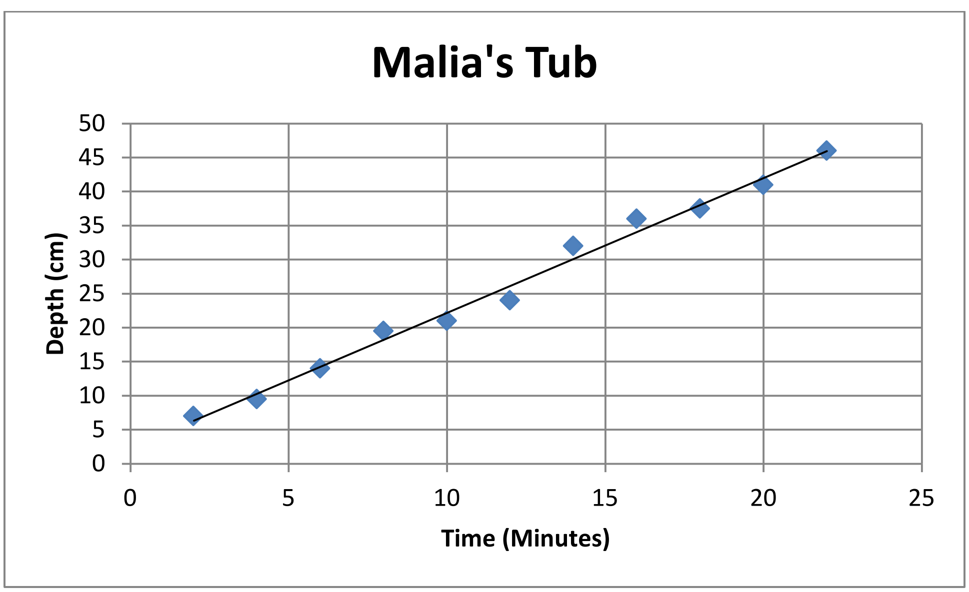
11) 0.258 (20) = 5.16. So, I would expect Stephanie to get about 5 or 6 hits out of her next 20 times at bat.

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

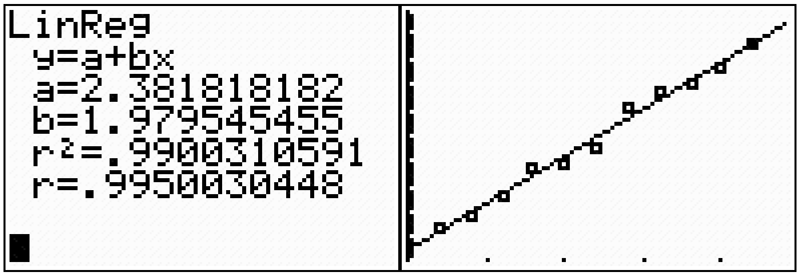
13) The theoretical probability of getting a Yahtzee on one roll is $\frac{1}{1296}$. Suppose you roll the 5 dice one at a time. The first die could land on anything. The chance that the second die matches the first die is $\frac{1}{6}$. The chance that the third die matches the first two is also $\frac{1}{6}$ and so on. This gives $1 \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{1296} \approx 0.00077 = 0.077\%$.

<u>**Section 6.3**</u>

1) a)



A screenshot from a TI-84 is shown below.



b) ŷ = 2.3818 + 1.9795x where ‘x’ is the number of minutes that the water has been running and ŷ is the predicted depth of the water in centimeters.

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) ŷ = 2.3818 + 1.9795(17) = 36.0333 cm which seems reasonable.

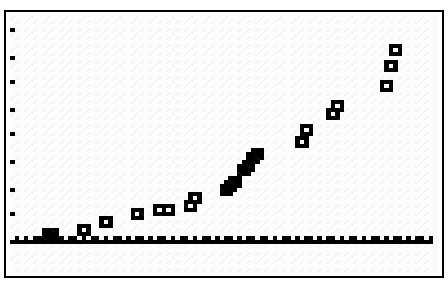
ŷ = 2.3818 + 1.9795(60) = 121.1518 cm which does not seem reasonable as this would make the tub overfill.

2)

a)

Fed.
Min.
Wage

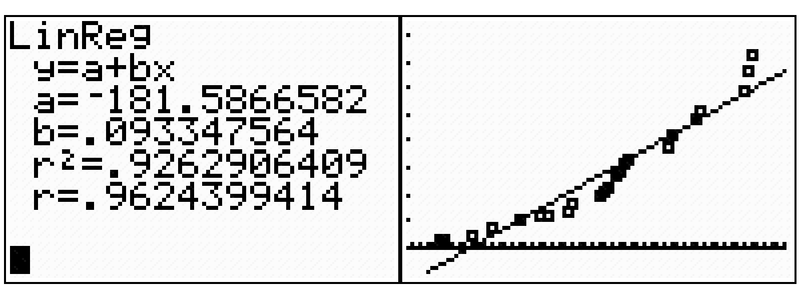
($/hr)



Year

b) The relationship between the years from 1938 to 2009 and the Federal Minimum Wage is 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) ŷ = -181.5867 + 0.0933x where ‘x’ is the year and ŷ is the predicted Federal Minimum Wage.



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

e) $\hat{y}$ = -181.5867 + 0.0933(2016)=$6.51 No, this is not an accurate prediction. We know that the minimum wage is higher than this. In addition, the actual data points in the graph are curving up at a much higher rate than our LSRL line is. Using the LSRL equation to make predictions is not appropriate here, especially prior to 1950 and after 2007. Finally, notice that 2016 is outside of the known data range so we should be very cautious about using this equation for this situation because it involves extrapolation.

f) $\hat{y}$ = -181.5867 + 0.0933(1968)=$2.03 The actual minimum wage in 1968 was $1.60. Our prediction was fairly close, but too high by $0.43.

3)

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

b) The 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) The y-intercept is 12. For a father with an IQ of zero, the son's IQ is predicted to be 12 points. It is clearly not realistic to be considering these points.

d) The interpretation of the slope is reasonable based on the equation, but the interpretation of the y-intercept is not reasonable. There is no person with an IQ anywhere near zero (or 12 for that matter). This is extrapolation because we would have certainly been outside any known data.

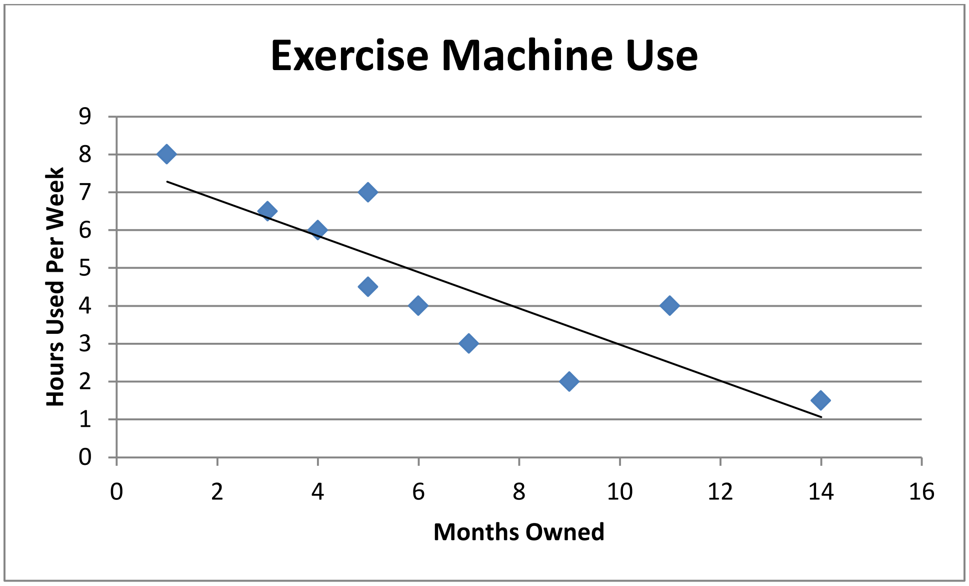
e) $\hat{y} = 12 + 0.9(120) = 120\ IQ\ points$

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

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)

a) $\hat{y} = 7.757 - 0.478x$ where 'x' is the months owned and $\hat{y}$ is the hours of exercise per week.



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

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

$1.243 = -0.478(x)$

$-2.6004 = x$

This suggests that they have owned their machine a negative number of months which is impossible! What you should notice is that 9 hours of exercise is outside the known data. We must be cautious when using extrapolation.

5)

a) The explanatory variable is the number of absences and response variable is the grade on the final exam.

b) The relationship between the number of times a student has been absent from this class and their grade on the final exam is strong and fairly linear. The association is clearly negative and with no obvious outliers. Students with higher number of absences tend to have lower scores on the final exam in this course.

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

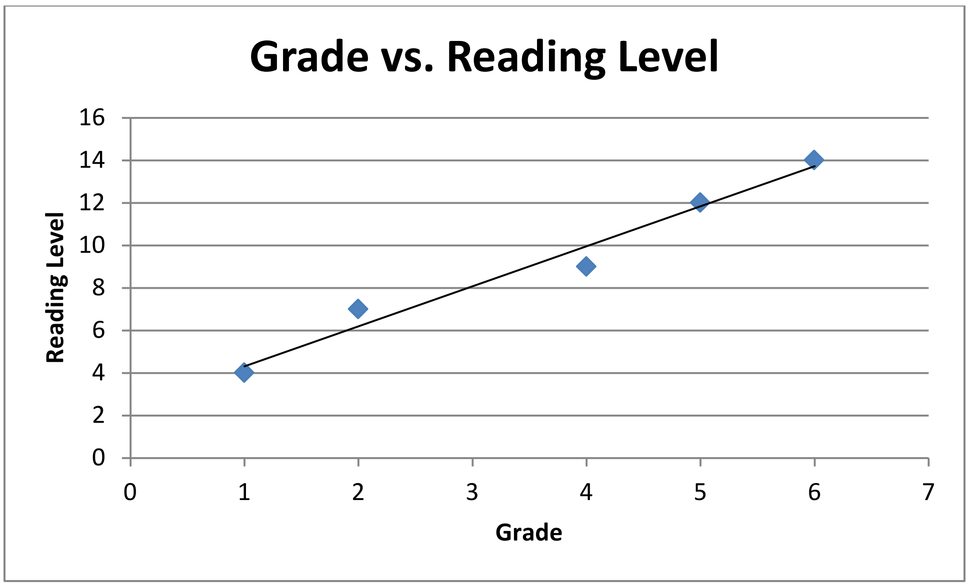
d) $\hat{y} = 91.704 - 1.654(43) = 20.582$. Lucy is predicted to have approximately 21 points on the final exam.

e) $R^2 = 0.8732$ so $r = \sqrt{0.8732} \approx 0.9345$ We note that the slope is negative so the correlation must be r = – 0.9345. This tells us that the relationship between number of absences and grade on the final is negative and strong.

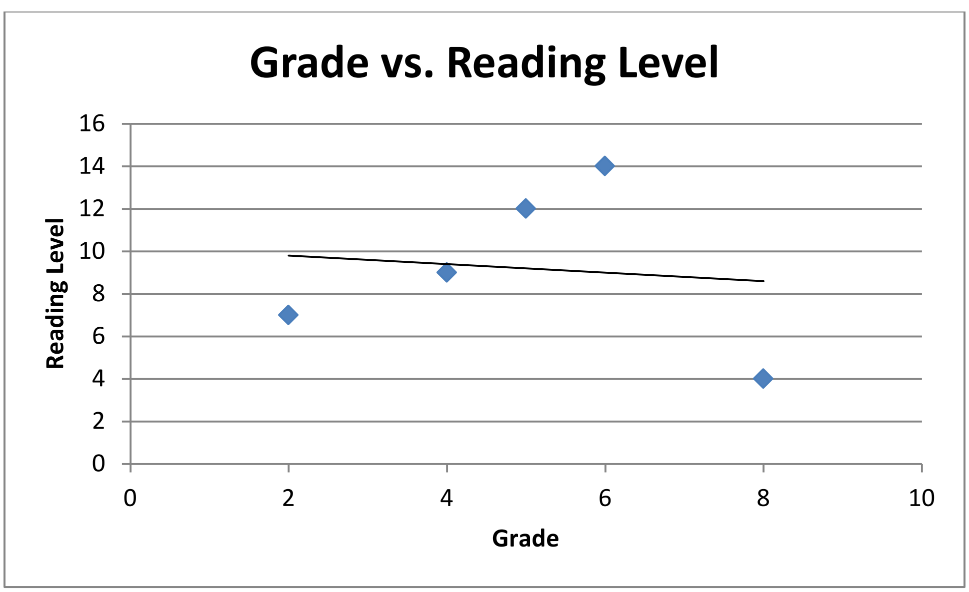
f) – 1.654 is the slope. For each increase of one absence, there is a predicted decrease of 1.654 points on the grade on the final exam.

6)

a) ŷ = 2.419 + 1.884x and r = +0.9858



b) ŷ = 10.2 - 0.2x and r = -0.1129



c) There are many visible changes. First of all, the second set of data (b) contains a very obvious outlier which is also an influential point. Secondly, the equation changed dramatically. 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. This shows that outliers can have a dramatic impact on the LSRL equations and the correlations of bivariate data!

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

8)

a) $\hat{y} = -2714.8589+35.0781x$ where 'x' is the temperature and $\hat{y}$ is the predicted number of people at the pool.

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. Also note that 91° was within the 82° to 96° temperature range that was used for the data. The second answer does not make any sense. Because this temperature is far from the known data, it is using extrapolation and we must be cautious about using results from extrapolation.

Review Exercises

9) $(0.79)(0.79)(0.79) = 0.79^3 = 0.4930 = 49.3\%$ chance

10) Marco could make two out of three free-throws by making either the first two free-throws, the first and third free-throws, or second and third free-throws. Notice that this means that Marco will make two of three shots and there are 3 ways for this to happen. $(0.79)(0.79)(0.21)(3) = 0.3932 = 39.32\%$

11) $1 – P(\text{miss none}) = 1 – [(0.79)(0.79)(0.79)(0.79)] = 1 – 0.79^4 = 0.6105 = 61.05\%$

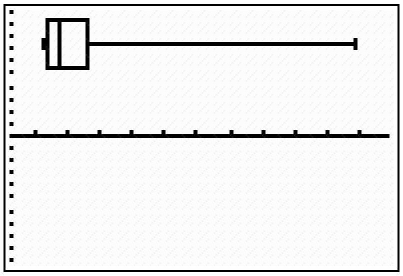
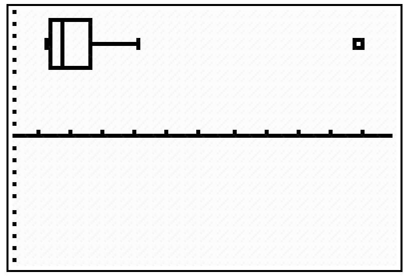
12) C is not appropriate because 00 to 79 is actually 80 numbers.

13)

a) The mean salary at Greezy's is \$10.76. The standard deviation is \$2.69.

b) The five-number-summary is {\$9.25, \$9.45, \$9.80, \$10.60, \$18.90}

c) TI-84+ screenshots of box-plots are shown below. (Not modified & modified)

d) 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.

e) The salaries at Greezy's Burger Boy range from \$9.25 to \$18.90 per hour. The distribution is heavily skewed to the right toward the one employee who makes \$18.90 (an extreme outlier). The median salary is \$9.80, with the majority of employees making less than \$11.00 per hour.

Section 6.4 Chapter 6 Review

1) False

2) True

3) True

4) False

5) r

6) Scatterplot

7) Explanatory Variable

8) From – 1 to +1

9) The Slope

10) Least Squares Regression Line

11) – 1 or +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)

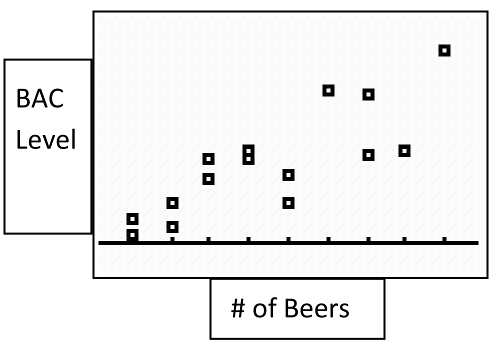
a) No. This is most likely a coincidence. Technology is advancing in many places and the number of cell phones is increasing because they are more affordable and more common. The number of starving children is due to some other factor, such as the population in poor countries increasing, famine, disease, 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. It could also possibly be that the relationship is reversed. In other words, the owners see that the employees are stressed, so they decided to offer raises to try to help ease that stress.

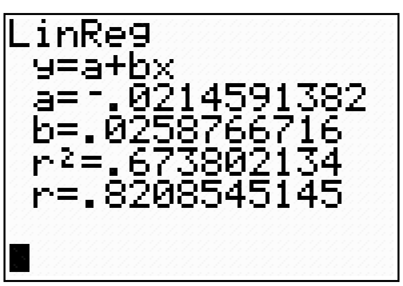
c) Not necessarily. It certainly seems that some people who cannot sleep might smoke in an effort to help themselves relax. Maybe however, there are people who are depressed, stressed, or ill and they happen to have both of these issues. In other words, maybe insomnia and cigarette addiction are responses to other existing issues. Maybe excessive cigarette smoking keeps people from sleeping well. This relationship is confounded, because we cannot see for sure how one issue is related to the other.

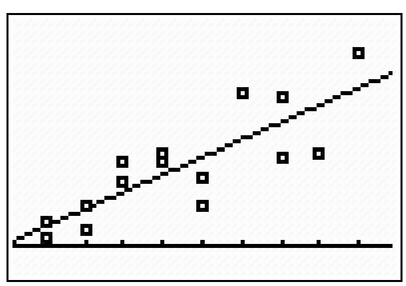
19) Number of beers and BAC level

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



b) $\hat{y}$ = -0.0215+0.0259x where ‘x’ stands for the # of beers and $\hat{y}$ represents the BAC level. 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 when considering the context of the problem. Zero beers should equate to a BAC level of 0.

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 well beyond that of our data, so any results are suspect. This is an extrapolation so we should not trust this as an accurate prediction.

h) 0.122 = -0.0215 + 0.0259x

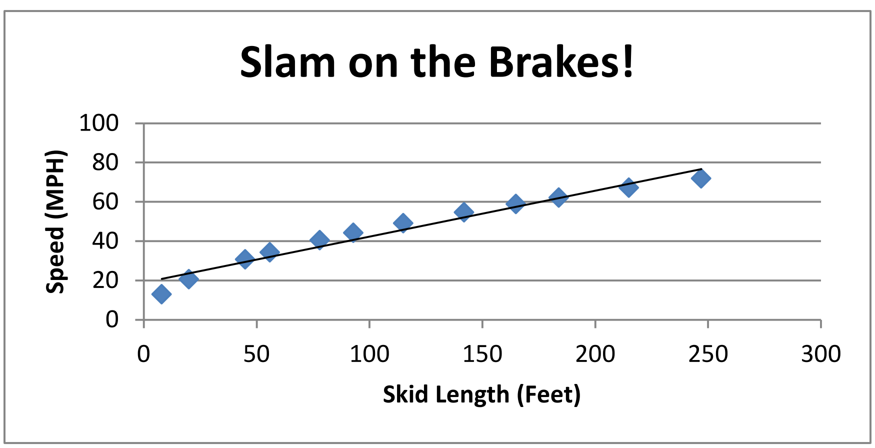
0.1435 = 0.0259x

5.54 = x

This suggests that someone with a BAC of 0.122 will have consumed 5 to 6 beers.

20)

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



b) ŷ = 18.825 + 0.2341x where 'x' is the length of the skid mark and ŷ is the estimated speed.

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 appears to be curved with no outliers. Longer skid marks are associated with faster traveling vehicles.

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

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

g) 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. In addition, we should be cautious of any conclusions here because we would be extrapolating.