

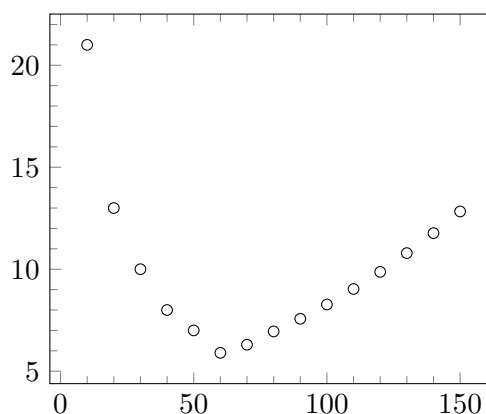
Homework №4

Author: David Oniani
Instructor: Dr. Eric Westlund

February 14, 2019

- 4.1 (a) The numbers of lectures attended will, in most cases (there might be some outliers such as people who can study on their own), affect the overall grade for the course. Therefore, it is reasonable to view the number of lectures attended as an explanatory variable and the grade as the response variable.
- (b) The number of hours per week spent exercising will, in most cases, affect the calories burned per week. Therefore, it is reasonable to view the number of hours per week spent exercising as an explanatory variable and the calories burned per week as the response variable.
- (c) Hours per week spent online using Facebook will, in most cases, affect the grade point average. Therefore, it is reasonable to view the hours per week spent online using Facebook as the explanatory variable and the grade point average as the response variable.
- (d) Hours per week spent online using Facebook will, in most cases, have no effect on the IQ. Therefore, it would be unreasonable to view one as the explanatory variable and the other as the response variable. Hence, in this case, it is better to simply explore the relationship between these two variables.

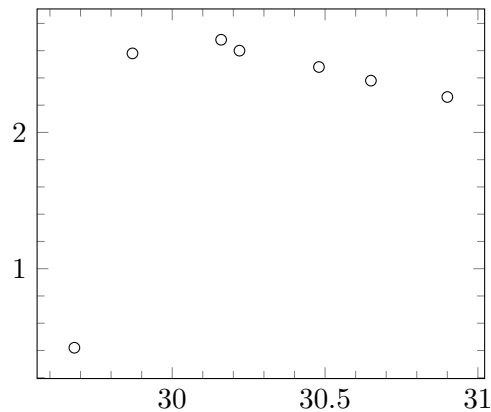
- 4.8 (a) Below is the scatterplot for the data.



In most cases, speed will affect the fuel consumption and therefore, speed is the explanatory variable.

- (b) It seems to be curved. More specifically, it looks like a parabolic plot. It makes sense since if the speed is too low, the fuel consumption goes up and same is for the very high speed. For the optimal speed, however, which seems to be approximately 60, the fuel consumption seems to be very low (which is good).
- (c) Unfortunately, there is no way describe variables as strictly positively or negatively associated. The reason is that the plot is parabolic and the association changes from negative to positive. In other words, we have a mixed association.
- (d) I would say it is reasonably strong since there are no points that have a big deviation from the pattern. Also, the plot represents quite tight rugby ball, which prompts us that the relationship is strong.

4.10 (a) Below is the scatterplot for the data.



The temperature will have some effects on the coral growth. Therefore, the temperature is an explanatory variable.

- (b) The formula for r is $r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \times \left(\frac{y_i - \bar{y}}{s_y} \right)$.

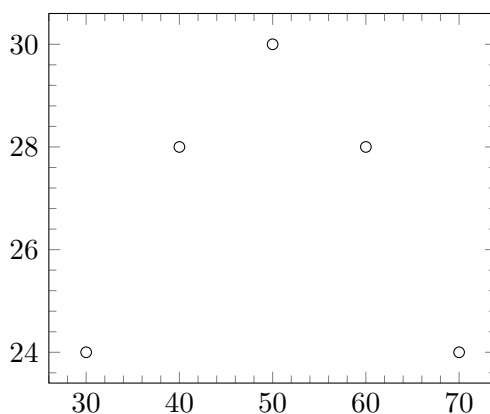
Therefore, we first need to find the averages for x and y . We get that $\bar{x} = 30.28$ and $\bar{y} \approx 2.51$. We now need to find the standard deviations for x and y . We get $s_x \approx 0.42$

and $s_y \approx 0.15$. Then we have:

$$\begin{aligned}
 r &= \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \times \left(\frac{y_i - \bar{y}}{s_y} \right) \\
 &= \frac{1}{7-1} \sum_{i=1}^n \left(\frac{x_i - 30.28}{0.42} \right) \times \left(\frac{y_i - 2.51}{0.15} \right) \\
 &= \frac{1}{6} \times \left(\left(\frac{29.68 - 30.28}{0.42} \times \frac{2.63 - 2.51}{0.15} \right) \times \left(\frac{29.87 - 30.28}{0.42} \times \frac{2.58 - 2.51}{0.15} \right) \right. \\
 &\quad + \left(\frac{30.16 - 30.28}{0.42} \times \frac{2.68 - 2.51}{0.15} \right) \times \left(\frac{30.22 - 30.28}{0.42} \times \frac{2.60 - 2.51}{0.15} \right) \\
 &\quad + \left(\frac{30.48 - 30.28}{0.42} \times \frac{2.48 - 2.51}{0.15} \right) \times \left(\frac{30.65 - 30.28}{0.42} \times \frac{2.38 - 2.51}{0.15} \right) \\
 &\quad \left. + \left(\frac{30.90 - 30.28}{0.42} \times \frac{2.26 - 2.51}{0.15} \right) \right) \\
 &\approx \frac{1}{6} (-1.14 + (-0.45) + (-0.32) + (-0.08) + (-0.09) + (-0.76) + (-2.46)) \\
 &\approx \frac{-5.3}{6} \approx -0.88.
 \end{aligned}$$

Finally, we got that $r = -0.13$. The correlation coefficient shows that the plot gives us a negative linear pattern which is easy to see in the plot of exercise (a).

4.13 Below is the scatterplot for the data.



Using the correlation calculator at <https://www.socscistatistics.com/tests/pearson/Default2.aspx>, we get that $r = 0$ (this is the verification, the problem does not ask to do it by hand again). Although there is a strong relationship between speed and mileage, the correlation is still 0 since the shape of the scatterplot is not linear - it is curved. Correlation can be used as a measure only in linear associations and since this is not linear and rather curved, it does not make a lot of sense (other than maybe it is since the curve is half positive-half negative in terms of slope of the line that approximates the branches but generally speaking, still has nothing to do with it).

- 4.37 (a) Recall that the lower correlation, the lower association. Therefore, Rachel should go with small-cap stocks because $0.21 < 0.50$.
- (b) She should look for a negative correlation (since the association between variables is negative).
- 4.38 The Psychologist talks about the correlation being 0 while the newspaper reports that the correlation is negative.
- 4.39 (a) Sex is not a number.
- (b) Correlation ranges from -1 to 1. Therefore, it cannot be 1.09 which is larger than 1.
- (c) Correlation is a unitless measure. Centimeter per kilogram is a unit and therefore, the statement contains a blunder.