

Homework 3: Sections 2.5 - 2.6

STA209-04: Applied Statistics

February 8, 2019

From the Book:

Due Date: Friday, February 15

Questions: 2.179, 2.185, 2.194, 2.212. 2.216

- 2.179** Multiple studies in both animals and humans show the importance of a mother's love (or the unconditional love of any close person to a child) in a child's brain development. A recent study shows that children with nurturing mothers had a substantially larger area of the brain called the hippocampus than children with less nurturing mothers. This is important because other studies have shown that the size of the hippocampus matters: People with large hippocampus area are more resilient and are more likely to be able to weather the stresses and strains of daily life. These observations come from experiments in animals and observational studies in humans.
- Is the amount of maternal nurturing one receives as a child positively or negatively associated with hippocampus size?
 - Is hippocampus size positively or negatively associated with resiliency and the ability to weather the stresses of life?
 - How might a randomized experiment be designed to test the effect described in part (a) in humans? Would such an experiment be ethical?
 - Can we conclude that maternal nurturing in humans causes the hippocampus to grow larger? Can we conclude that maternal nurturing in animals (such as mice, who were used in many of the experiments) causes the hippocampus to grow larger? Explain.
- 2.185** In Exercise 1.23, we learned of a study in which participants were randomly assigned to receive or not receive cognitive behavioral therapy (CBT), and then reported whether or not they experienced any sleep improvement. One news magazine reporting this study said "Sleep improvements were strongly correlated with CBT." Why is this an incorrect use of the statistics word *correlation*?
- 2.194** In Exercise 2.120 on page 92, we discuss a study in which the Nielson Company measured connection speeds on home computers in nine different countries in order to determine whether connection speed affects the amount of time consumers spend online. Table 2.29 shows the percent of Internet users with a "fast" connection (defined as 2Mb or faster) and the average amount of time spent online, defined as the total hours connected to the Web from a home computer during the month of February 2011. The data are also available in the dataset [GlobalInternet](#).
- What would a positive association mean between these two variables? Explain why a positive relationship might make sense in this context.
 - What would a negative association mean between these two variables? explain why a negative relationship might make sense in this context.
 - Make a scatterplot of the data, using connection speed as the explanatory variable and time online as the response variable. Is there a positive or negative relationship? Are there any outliers? If so, indicate the country associated with each outlier and describe the characteristics that make it an outlier for the scatterplot.
 - If we eliminate any outliers from the scatterplot, does it appear that the remaining countries have a positive or negative relationship between these two variables?

- (e) Use technology to compute the correlation. Is the correlation affected by the outliers?
- (f) Can we conclude that a faster connection speed causes people to spend more time online?

2.212 Exercise 2.143 on page 102 introduces a study that examines several variables on collegiate football players, including the variable *Years*, which is number of years playing football, and the variable *Cognition*, which gives percentile on a cognitive reaction test. Exercise 2.182 shows a scatterplot for these two variables and gives the correlation as -0.366. The regression line for predicting *Cognition* from *Years* is:

$$\widehat{Cognition} = 102 - 3.34 \cdot Years$$

- (a) Predict the cognitive percentile for someone who has played football for 8 years and for someone who has played football for 14 years.
- (b) Interpret the slope in terms of football and cognitive percentile.
- (c) All of the participants had played between 7 and 18 years of football. Is it reasonable to interpret the intercept in context? Why or why not? (Hint: When is the predicted cognition 102?)

2.216 Using the data in the [StudentSurvey](#) dataset, we use technology to find that a regression line to predict weight (in pounds) from height (in inches) is

$$\widehat{Weight} = -170 + 4.82(Height)$$

- (a) What weight does the line predict for a person who is 5 feet tall (60 inches)? What weight is predicted for someone 6 feet tall (72 inches)?
- (b) What is the slope of the line? Interpret it in context.
- (c) What is the intercept of the line? If it is reasonable to do so, interpret it in context. If it is not reasonable, explain why not.
- (d) What weight does the regression line predict for a baby who is 20 inches long? Why is it not appropriate to use the regression line in this case?

Miscellaneous:

Due Date: Friday, February 15

- S1** A 2005 study found the correlation between insulin and BMI to be 0.67. Furthermore, they found that insulin had a mean of 93.78 and a SD of 9.36 while BMI had a mean of 25.01 and a SD of 4.68.
- Suppose an individual's insulin is 0.4 standard deviations below average. How many standard deviations above/below average is that individual's BMI likely to be?
 - Suppose an individual's insulin is 103.1. What is that individual's BMI likely to be?
 - What is the regression slope coefficient for predicting insulin based on BMI?
- S2** A 1998 study found the correlation between fatty acid and uric acid to be 0.75. Furthermore, they found that fatty acid had a mean of 271.99 and a SD of 28.67 while uric acid had a mean of 5.19 and a SD of 0.96.
- Suppose an individual's fatty acid is 0.7 standard deviations above average. How many standard deviations above/below average is that individual's uric acid likely to be?
 - Suppose an individual's fatty acid is 274.9. What is that individual's uric acid likely to be?
 - What is the regression coefficient for predicting uric acid based on fatty acid?
- S3** Read [this blog post](#) which criticizes the **Bradford Hill Criteria**, which are criteria often used to argue causality in the absence of randomized experiments. Briefly summarize the article. Do you agree with the author's criticisms of some of the criteria? Explain.