HW 04
Stat140-04

Due date Nov 18 6PM EST on Gradescope

Notes on grading

In general, you will receive full points on the question if (1) there are no errors in your solution AND (2) the solution is written in highly articulate Statistical and English language. Point will be taken off if there are errors, your writing of the solution is incomplete, or there are issues with writing organization.

Problem 1

Here are the residuals for a regression of `Price` on `Capacity` for the hard drives of Exercise 1 in the Week 3 Day 3's tutorial.

Capacity	Residual
0.50	285.16
1.0	234.07
2.0	123.88
3.0	−20.27
4.0	−122.44
6.0	−133.43
8.0	−245.00
12.0	−329.80
32.0	207.81

a) Which residual contributes the most to the sum that is minimized by the Least Squares criterion?

b) Five of the residuals are negative. What does that mean about those drives? Be specific and use the correct units.

Problem 2

An online clothing retailer keeps track of its customers' purchases. For those customers who signed up for the company's credit card, the company also has information on the customer's `Total Yearly Purchases` and `Income`. The scatterplot of the two variables looks like this:
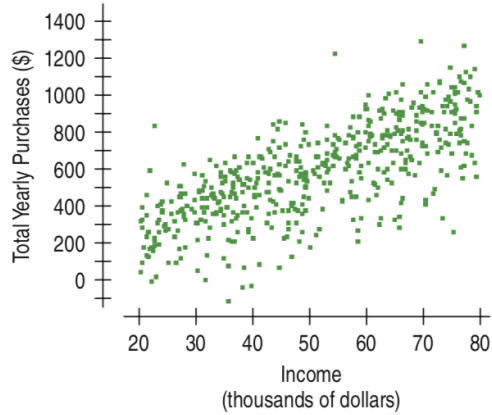
The correlation between *Total Yearly Purchases* and *Income* is 0.722. Summary statistics for the two variables are:

	Mean	SD
Income	$50,343.40	$16,952.50
Total Yearly Purchase	$572.52	$253.62

a) What is the linear regression equation for predicting Total Yearly Purchase from Income?

b) Do the assumptions and conditions for regression appear to be met?

2

c) What is the predicted Total Yearly Purchase for someone with a yearly Income of $20,000? For someone with an annual Income of $80,000?

d) What percent of the variability in Total Yearly Purchases is accounted for by this model?

e) Do you think the regression might be a useful one for the company? Comment.

Problem 3

In 2008, a study[1] was conducted measuring the impact that music volume has on beer consumption. The researchers went into bars, controlled the music volume, and measured how much beer was consumed. The article states that "the sound level of the environmental music was manipulated according to a randomization scheme." It was found that louder music corresponds to more beer consumption. Does this provide evidence that louder music causes people to drink more beer? Why or why not?

Problem 4

Several studies have been performed to examine the relationship between nut consumption and cholesterol levels. Here we consider two such studies. In Study 1, 47 participants were assigned into two groups: one group was given nuts to eat each day, and the other group was told to consume a diet without nuts. In Study 2, 48 participants were free to follow their own diet, and reported how many nuts they consumed. Cholesterol levels were measured for all participants, and both studies found that nut consumption was associated with lower levels of LDL ("bad") cholesterol. Based on the information above, which study do you think provides better evidence that nut consumption reduces LDL cholesterol? Explain your answer.

[1]Gueguen, N., Jacob, C., Le Guellec, H., Morineau, T., and Lourel, M., "Sound Level of Environmental Music and Drinking Behavior: A Field Experiment with Beer Drinkers," Alcoholism: Clinical and Experimental Research, 2008; 32: 1795–1798.

Problem 5

A headline in June 2015 proclaims "Infections can lower IQ."[2] The headline is based on a study in which scientists gave an IQ test to Danish men at age 19. They also analyzed the hospital records of the men and found that 35% of them had been in a hospital with an infection such as an STI or a urinary tract infection. The average IQ score was lower for the men who had an infection than for the men who hadn't.

  (a) What are the cases in this study?

  (b) What is the explanatory variable? Is it categorical or quantitative?

  (c) What is the response variable? Is it categorical or quantitative?

  (d) Does the headline imply causation?

  (e) Is the study an experiment?

  (f) Is it appropriate to conclude causation in this case?

---

[2]"Infections can lower IQ," The Week, June 12, 2015, p. 18.

Problem 6

Do Cat Videos Improve Mood? As part of an "internet cat videos/photos" study, Dr. Jessica Gall Myrick posted an on-line survey to Facebook and Twitter asking a series of questions regarding how individuals felt before and after the last time they watched a cat video on the Internet.31 One of the goals of the study was to determine how watching cat videos affects an individual's energy and emotional state. People were asked to share the link, and everyone who clicked the link and completed the survey was included in the sample. More than 6,000 individuals completed the survey, and the study found that after watching a cat video people generally reported more energy, fewer negative emotions, and more positive emotions.

(a) What is the population of interest?

(b) What is the sample?

(c) Would this be considered a simple random sample from a target population? Why or why not?

(d) Ignoring sampling bias, what other ways could bias have been introduced into this study?