STA-371G
Homework 8

This assignment contains 3 questions; there are 121 total points.

Instructions

Submit your assignment as a PDF/Word document and an R script to Canvas by **11/10/2017 by 7:00PM**. It is your responsibility to follow the homework guidelines on Canvas and ensure that the files you upload are complete and uncorrupted.

Grading

A small number of the problems on each assignment will be graded for correctness, and the remainder graded for completeness. A complete response answers the question posed and also shows your work. This means showing the steps of a mathematical calculation, or including the R code you used to arrive at your answer. For questions that are not just calculations (e.g., more than computing an expected value from a table) you should answer in complete, concise sentences. Detailed solutions will be available – you should always check your work against these solutions.

---

1. (25 points) Using our baseball data (**baseball.csv**), regress R/G on a binary variable for league membership (League = 0 if National and League = 1 if American) and OBP.

$$R/G = \beta_0 + \beta_1 League + \beta_2 OBP + \epsilon$$

   (a) (6 points) Based on the model assumptions, what is the expected value of R/G given OBP for teams in the AL? How about the NL? (Your answer should be in terms of the β's, not the estimated values.)

   (b) (9 points) Interpret β_0, β_1 and β_2.

   (c) (5 points) After running the regression and obtaining the results, can you conclude with 95% confidence that the marginal effect of OBP on R/G (after taking into account the League effect) is non-zero?

   (d) (5 points) Is there evidence of a difference in runs per game between the American and National leagues, comparing teams with the same on base percentage? State and perform the appropriate hypothesis test.

2. (55 points) Read the case "Orion Bus Industries: Contract Bidding Strategy". Orion Bus Industries wants to develop a method for determining how to bid on specific bus contracts to maximize expected profits. In order to do this, it needs to develop a model of winning bids that takes into account such factors as the number of buses in the contract,

the estimated cost of the buses and the type of bus (e.g. length, type of fuel used, etc.). The data set is available in the course website. This data set only includes the bus contracts from Exhibit 1 in the case where Orion did not win the contract. This eliminates 28 of the 69 observations and leaves a sample of size n = 41 observations.

(a) (14 points) Run a regression of $WinningBid$ against $NumberOfBusesInContract$, $OrionsEstimatedCost$, $Length$, $Diesel$ and $HighFloor$, ie, the following regression model:

$$WinningBid_i = \beta_0 + \beta_1 NumberOfBusesInContract_i + \beta_2 OrionsEstimatedCost_i + \beta_3 Length_i + \beta_4 Diesel_i + \beta_5 HighFloor_i + \epsilon_i$$

What is the estimated regression model? How would you interpret the estimated coefficient associated with the dummy variable Diesel?

(b) (3 points) What is the estimate of σ^2 in the model in part (a)?

The city of Louisville, Kentucky is putting out a contract for bid for five 30-foot, low-floor, diesel-fuelled buses. Orion estimates their cost to manufacture these buses to be $234,229 per bus.

(c) (5 points) Using the model in part (a), what is the estimated distribution representing the uncertainty about the amount of the winning bid per bus for this contract (plug-in the parameter estimates)? In particular, what are the mean and standard deviation of the distribution?

(d) (8 points) Given the distribution in part (c), what is the probability that Orion wins the contract if it bids $240,000 per bus? If it wins the contract, what is its profit per bus per bus?

(e) (5 points) What is the probability that Orion loses the contract if it bids $240,000 per bus? If it loses the contract, what is its profit per bus? (You do not need to take into account the cost of putting the bid together when determining the profit for a lost contract.)

(f) (10 points) Why is there uncertainty about the profit per bus that Orion will obtain if it bids $240,000 per bus? What is the probability distribution representing this uncertainty? In particular, what is the mean of the distribution (i.e. what is the expected profit per bus if it bids $240,000 per bus)?

(g) (10 points) What is the expected profit if Orion bids $250,000 per bus? Describe how Orion could use this regression model to find the bid that maximizes their expected profit.

3. (41 points) Listen to this podcast:

`http://www.npr.org/blogs/money/2013/04/23/178635250/episode-453-what-causes-what`

(There is also a transcript you can read.)

(a) (8 points) Thinking about the discussion about crime, why can't I just get data from many different cities and run the regression of "Crime" on "Police" to understand how more cops in the streets affect crime? ("Crime" refers to some measure of crime rate and "Police" measures the number of cops in a city)

(b) (15 points) How were the researchers from Penn able to isolate this effect? Briefly describe their approach and discuss their result in "Table 2" below.

(c) (10 points) Why did they have to control for METRO ridership? What was term in their model trying to capture?

(d) (8 points) A fellow student states that they don't buy the claim of the paper (more police on the streets leads to fewer crimes) because the R^2 for the two fitted models is so low. Do you agree? Why or why not?

EFFECT OF POLICE ON CRIME

TABLE 2

TOTAL DAILY CRIME DECREASES ON HIGH-ALERT DAYS

	(1)	(2)
High Alert	−7.316*	−6.046*
	(2.877)	(2.537)
Log(midday ridership)		17.341**
		(5.309)
R^2	.14	.17

Figure 1: The dependent variable is the daily total number of crimes in D.C. This table present the estimated coefficients and their standard errors in parenthesis. The first column refers to a model where the only variable used in the High Alert dummy whereas the model in column (2) controls form the METRO ridership. * refers to a significant coeficient at the 5% level, ** at the 1% level.