STA-371G
Homework 7

This assignment contains 3 questions; there are 91 total points.

Instructions

Submit your assignment as a PDF/Word document and an R script to Canvas by **11/3/2017 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. (20 points)

   Suppose we are modeling house price as depending on house size, the number of bedrooms in the house and the number of bathrooms in the house. Price is measured in thousands of dollars and size is measured in thousands of square feet.

   Suppose our model is:

   $$P = 20 + 50\,\text{size} + 10\,\text{nbed} + 15\,\text{nbath} + \epsilon, \quad \epsilon \sim N(0, 10^2).$$

   (a) (4 points) Suppose you know that a house has size =1.6, nbed = 3, and nbath =2.

   What is the distribution of its price given the values for size, nbed, and nbath.

   (hint: it is normal with mean = ?? and variance = ??)

(b) (4 points) Given the values for the explanatory variables from part (a), give the 95% predictive interval for the price of the house.

(c) (4 points) Suppose you know that a house has size =2.6, nbed = 4, and nbath =3. Give the 95% predictive interval for the price of the house.

(d) (4 points) In our model the slope for the variable nbath is 15. What are the units of this number?

(e) (4 points) What are the units of the intercept 20? What are the units of the the error standard deviation 10?

2. (45 points)

For this problem the data is the file **Profits.csv** - see the R script accompanying the homework code to read the data into R.

There are 18 observations.
Each observation corresponds to a project developed by a firm, with the following attributes:
y = Profit: profit on the project in thousands of dollars.
x1 = RD: expenditure on research and development for the project in thousands of dollars.
x2 = Risk: a measure of risk assigned to the project at the outset.

We want to see how profit on a project relates to research and development expenditure and the risk measure.

(a) (8 points) Plot profit versus each of the two x variables. That is, do two plots y vs. x1 and y vs x2. You can't really understand the full three-dimensional relationship from these two plots, but it is still a good idea to look at them. Does it seem like the y is related to the x's?

(b) (5 points) Suppose a project has risk = 7 and research and development = 76. Give the 95% plug-in predictive interval for the profit on the project.

(c) (10 points) Now all you knew was risk=7. Run the **simple** linear regression of profit on risk and compute the 68% and 95% plug-in predictive intervals for profit.

(d) (10 points) How does the size of your interval in (c) compare with the size of your intervals in (b)? What does this tell us about our variables? In particular, what does it tell us about the value of using R&D in addition to risk when predicting future profits?

(e) (12 points) Compute R's more accurate predictive intervals for the three scenarios in parts b and c above. How do they compare to the plug-in intervals you computed earlier? Why are they different?

3. (26 points)

The data for this question is in the file **zagat.csv**. (See the R script accompanying the assignment for code to load it into R.)

The dataset was compiled from the Zagat restaurant guide. There are 114 observations, where each observation corresponds to a restaurant.

There are 4 variables:
price: the price of a typical meal
food: the zagat rating for the quality of food.
service: the zagat rating for the quality of service.
decor: the zagat rating for the quality of the decor.

We want to see how the price of a meal relates the quality characteristics of the restaurant experience as measured by the variables food, service, and decor.

(a) (12 points) Plot price vs. each of the three x's. Does it seem like our y (price) is related to the x's (food, service, and decor) ?

(b) (5 points) Suppose a restaurant has food = 18, service=14, and decor=16. Run the multiple regression of price on food, decor, and service and give the 95% predictive interval for the price of a meal.

(c) (5 points) What is the interpretation of the coefficient estimate for the explanatory variable food in the multiple regression from part (b) ?

(d) (4 points) Suppose I asked you to use the multiple regression results to predict the price of a meal at a restaurant with food = 20, service = 3, and decor =17. Would you feel comfortable making that prediction? Why or why not?