Homework 8

Jared S. Murray
Statistics
McCombs School of Business

Instructions

Submit your assignment at the **beginning** of class on **Monday Feb 4**. Please make sure that the pages of your homework assignment are in the correct order and securely stapled together. You may (but are not required to) work in groups of up to four students. Each group should hand in one copy of the homework.

Grading

A small number of problems 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 R output that justifies your conclusions. 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.

Problem 1: Gas Prices (70 points)

For this exercise, you'll need to download the GasPrices.csv data set from the class website. This data set came from a student project in spring of 2016. (It was a pretty awesome project.) I'll let the students who did the project describe things in their own words:

Have you ever been driving through town looking to make a quick stop to fill up your car with gas and noticed that different gas stations are advertising different gas prices? Have you ever stopped to wonder why this might be the case? Could there be some underlying factors responsible for this noticeable difference in price, specifically for the same, regular unleaded mix of gas on the same day at the same time?

These data were collected by visiting 101 gas stations in the Austin area. Once the researchers had visited all 101 gas stations, they used the US Census Bureau's American Fact Finder to input the median income for each zip code.

The variables in the data set are as follows:

- ID: Order in which gas stations were visited
- Name: Name of gas station
- Price: Price of regular unleaded gasoline, gathered on Sunday, April 3rd, 2016
- Pumps: How many pumps does the gas station have?

- Interior: Does the gas station have an interior convenience store?
- Restaurant: Is there a restaurant inside the gas station?
- CarWash: Does the gas station have a car wash attached?
- Highway: Is the gas station accessible from either a highway or a highway access road?
- Intersection: Is the gas station located at an intersection?
- Stoplight: Is there a stoplight in front of the gas station?
- IntersectionStoplight: three-way variable for if the gas station was at an intersection and/or a stoplight (None, Intersection (only), or Both).
- Gasolines: How many types of gasoline are offered? (Regular, midgrade, etc.)
- Competitors: Are there any gas stations in sight?
- Zipcode: Zip code in which gas station is located
- Address: Physical location of gas station
- Income: Median Household Income of the ZIP code where the gas station is located based on 2014 data from the U.S. Census Bureau
- Brand: is the gas station branded by one of the major oil companies (ExxonMobil, ChevronTexas, Shell) or not (Other)?

Use this dataset to answer the questions below:

1. Run a simple linear regression of prices against the highway variable. Does it seem that gas stations located off the highway charge more, less, or about the same as other gas stations?
2. Run a simple linear regression of prices against the income variable. Interpret the estimated coefficient. Is there a relationship between the local income and gas prices?
3. Now let's build a multiple regression model: Add the Brand, CarWash, and Restaurant variables, and include both the Price and Highway variables. Compare the estimated coefficients in this multiple regression model to those in parts 1 and 2. Are they similar? If not, describe how they changed, and briefly explain why they might have shifted.
4. We suspect that the relationship between local income and gas prices might be different when the gas station is located off the highway, because those gas stations get more business from non-locals. Using the model from part 3, add an interaction between Income and Highway. Is there evidence in the data to support this hypothesis? If so, describe how the relationship between income and gas prices differs between stations on and off the highway.

Problem 2: Estimating Demand (40 points)

This question considers data on sales volume, price, and advertisting display activity for packages of Borden sliced cheese, available as "cheese.csv" (see the R code accompanying this assignment). For each of 88 stores (store) in different US cities, we have repeated observations of the weekly sales volume (vol, in terms of packages sold), unit price (price), and whether the product was advertised with an in-store display during that week (disp = 1 for display). Altogether there are 5,555 observations in the data set. For this problem we are interested in the relationship between price and sales volume, i.e. the demand curve.

- Plot sales volume against price. Does the relationship appear to be linear?
- Regress volume against price, and plot the residuals versus price and the fitted values. Do you see any problems here?
- Regress the log of volume against the log of price, and plot the residuals. Does this plot look better than the plot in part 2?
- Interpret the estimated coefficient in part 3.