STA 309

Homework 3

Question: Did the survey accurately reflect a population level trend? Build a model to assess support for any male/female differences

Approach: To answer the question, I used the bootstrapping with resampling technique in attempts to make the populations for each Monte-Carlo simulation more accurate. I repeated the simulation 10,000 times and regressed each of those models.

Results:

	Left Hand			
-	Lower	Upper	Level	Estimate
Intercept	0.33	0.515	0.95	0.423
Sex Male	-0.085	0.182	0.95	0.048

In this table, the intercept and the dummy variable (being male) are shown to depict whether or not you will fold your left arm over the other. This also shows the upper and lower levels of the confidence intervals, at a 95% level and the corresponding estimate.

Conclusion: At a 95% confidence level, the overall estimation that one will cross their left arms over is somewhere between 33% and 51.5%. If someone is a male, they have between a -8.5 to 18.2% estimate to cross their left hand over their right, at a 95% confidence interval, with the estimate being a 4.8% chance; thus, the survey (resulting in about 5%) did accurately reflect the level trend, which is a 4.8% estimate, with 95% confidence.

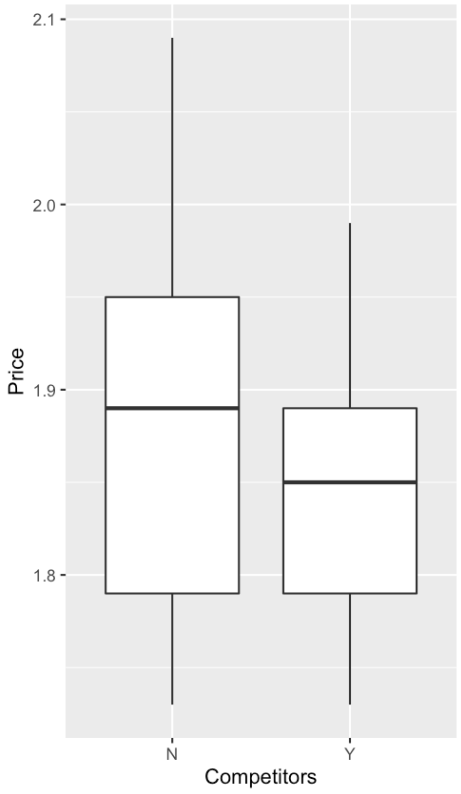
2-A

Question: Gas stations charge more if they lack direct competition in sight

Approach: To answer this question, I used a bootstrap to create a linear regression model, to make the Austin sample more representative of overall gas stations. I repeated the Monte Carlo simulations 10,000 times. I also used box plots to represent the data and a histogram to plot the regression models.

Results:

	Competition			
-	Lower	Upper	Level	Estimate
Intercept	1.852	1.9	0.95	1.876
Competitors	-0.005	0.008	0.95	-0.023



In the table, the intercept and the confidence intervals are provided as well as the estimate for the dummy variable (competitors), based on Monte Carlo simulation. In addition, the boxplot shows the differences in price based on competitors, solely based on the data.

Conclusion: Based on the results, the base price will be between $1.852 and $1.9, at a 95% interval. If a gas station were to have nearby competitors, they will charge between ($0.005) and $0.008 more, estimated to be ($0.023) at 95% confidence. Thus, gas stations do not charge more with competition nearby, on average; this theory is **not supported** by the data.

2-B

Question: The richer the area, the higher the gas price?

Approach: To answer this question, I regressed a model with the response variable being income and the predicted variable being price. (i.e. price = intercept + income(x)). I used Monte Carlo simulation, 10,000 times to bootstrap the gas station to be representative of overall gas prices.

Results:

	Income			
-	Lower	Upper	Level	Estimate
Intercept	1.759616703	1.829440561	0.95	1.793442499
Income	0.000000661	0.000001769	0.95	0.000001248

In this table, the price is predicted as a function of income, with the base price being between $1.759 and $1.829, estimated to be $1.793. For each dollar of income, the regressed model predicted that the price would increase between $0.000000661 and $0.000001769, estimated to be $0.000001248 per dollar.

Conclusion: Based on the results, with a 95% confidence level, for every dollar more in income for the respective zip codes, the gas price increased by $0.000001248, thus the richer the area, the higher the gas price. This theory **is supported** by the data.

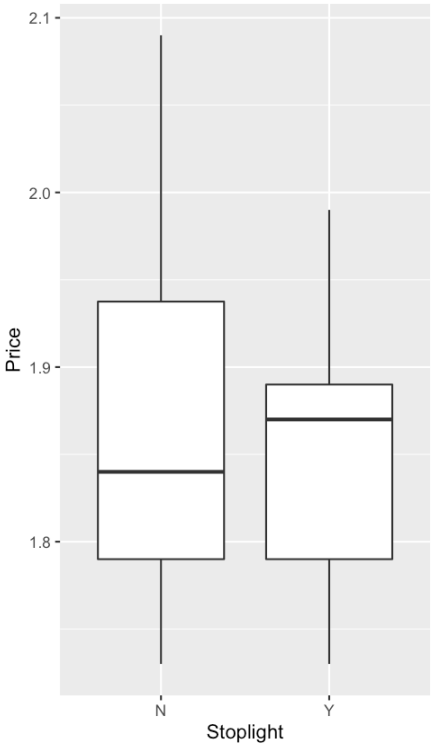
2-C

Question: Gas stations at stoplights charge more

Approach: To solve this question, I first graphed a bar plot to see what the data would predict. Then I bootstrapped the data to make it more representative of overall gas stations and regressed the model in a Monte Carlo simulation for 10,000 times.

Results:

Stoplights				
-	Lower	Upper	Level	Estimate
Intercept	1.838	1.896	0.95	1.866
Stoplight	-0.039	0.031	0.95	-0.003



The bar plot is faceted to show the difference between being near a stoplight and not near a stoplight, with price on the y axis. The table shows the lower, upper, and estimate values for the intercept and the stoplight, under the Monte Carlo simulation. The base price is estimated to be between $1.83 and $1.90 dollars, at a 95% confidence, with the price changing by between ($0.039) and 0.031 if at a stoplight.

Conclusion: Based on the results, I would say that the theory is **partially unsupported** by the data, due to the fact that, at a 95% confidence level, the price decreased, but only by $0.003.

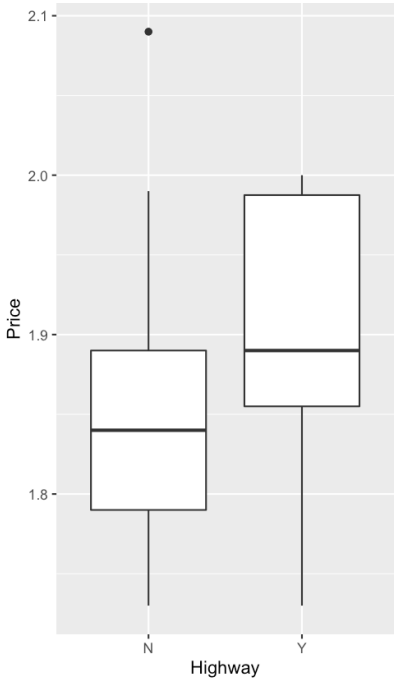
2-D

Question: Gas stations with direct highway access charge more

Approach: To solve this question, I created a dummy variable (highway) and regressed the model, while bootstrapping, over a Monte Carlo simulation, for 10,000 times. I additionally created a bar plot to see the data beforehand.

Results:

	Highway			
-	Lower	Upper	Level	Estimate
Intercept	1.838	1.872	0.95	1.854
Highway	0.009	0.081	0.95	0.046



The bar plot shows the distribution of price faceted by whether or not the station was near a highway, with price on the y-axis. The table shows the intercept and highway offset, with their respective upper and lower bounds, and estimate. The base price is between $1.838 and $1.872 with 95% confidence. The offset, if near a highway, is predicted to increase the price by between $0.009 and $0.081.

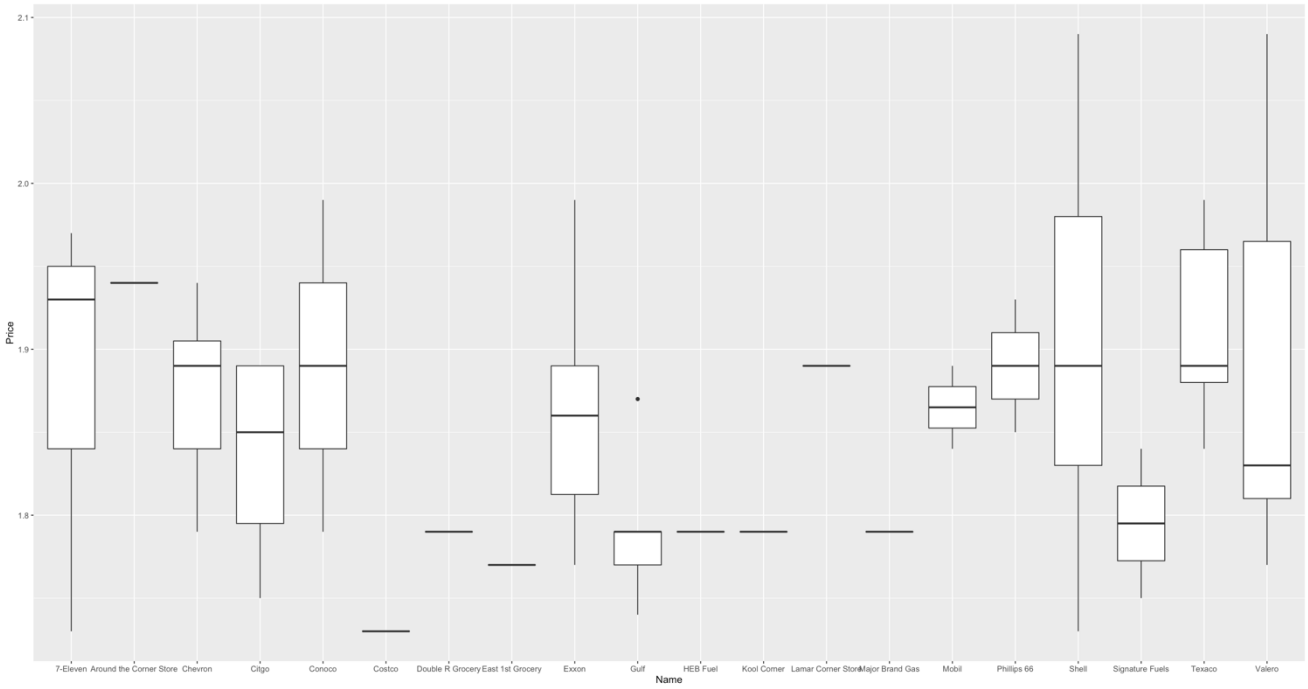
Conclusion: The theory **is supported** by the data, in that when bootstrapped, the data shows that the highway's lower and upper bounds will both increase the price of gasoline, at a 95% confidence interval.

2-E

Question: Shell charges more than all other non-Shell brands.

Approach: To solve this problem I created a bar plot for each named company, determining how Shell fell along the list. Then, I created a dummy variable (is Shell) to separate Shell from the other non-Shell companies and regressed the model while bootstrapping in a Monte Carlo simulation for 10,000 times.

Results:



The bar plot shows the prices on the y-axis and the names on the x-axis, with shell being the fourth to last.

	Shell			
-	Lower	Upper	Level	Estimate
Intercept	1.839	1.874	0.95	1.856
Shell	-0.01	0.066	0.95	0.027

The table shows the intercept and dummy variable confidence intervals and the estimated values for each. With the base price being between $1.839 and $1.847, and the offset (being Shell) being between ($0.01) and $0.066.

Conclusion: Based on the results, the theory that Shell charges more than other companies **is supported** due to the fact that, at a 95% confidence interval, shell will charge $0.027 more on average.

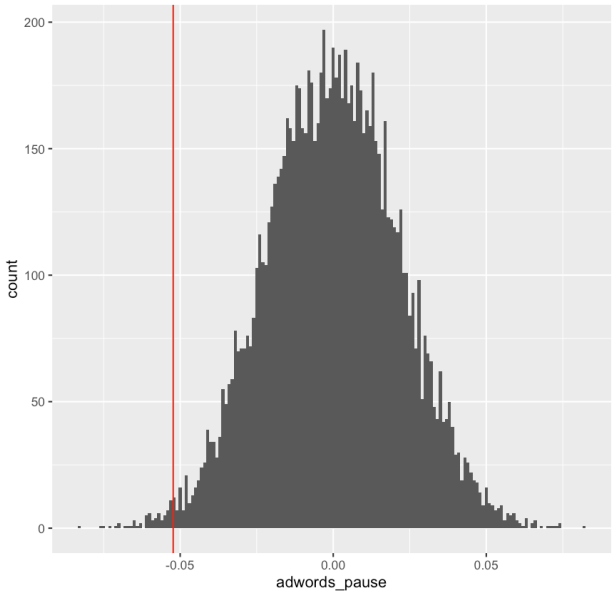
Question: Does the extra traffic brought to eBay's site through sponsored search results justify the cost? Assess whether the revenue ratio is the same in the treatment and control groups or if the data favors paid search advertising.

Approach: To determine whether the extra traffic played a part in increased revenue ratio, I conducted a randomized experiment. Using a Monte Carlo, randomized experiment to see if the decline in revenue ratio was due to chance or systematic to the pausing of paid searching.

Results:

Not Randomized

-	Amount
Intercept	0.9488775
Paused Ads	-0.0522815
==========	==========
P Value	0.0065



There are two results: one table shows the results for a not randomized regression model and the other is a distribution histogram showing the results of a randomized experiment to create a regression using Monte Carlo. In the non-randomized experiment, the intercept was determined to be 0.95 with pausing the ads to offset that amount by -0.052, this amount was plugged into the graph to determine the number of cases to be more or equally extreme. This ended with a p value of 0.0065 or 0.06%.

Conclusion: The P-Value tells us that the drop in the revenue ratio (-0.0522) is a 0.06% possibility; since it is less than 5%, we can throw away the null hypothesis that the decrease in revenue ratio is due to chance rather than the systematic effect of pausing ads. The extra traffic brought to eBay's sight justifies the cost due to the fact **that revenue ratio will decrease if the ads are paused**, not due to chance, and conversely will increase when continued.