

SDS315_HW4

Neil Sebastian

2024-02-20

HW4

UT EID: ns36965

[GitHub Link](#)

Problem One

Claim A

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

To test this theory, I compared the average prices of gas of gas stations with competitors and gas stations without competitors. To do this I used bootstrapping from the data and created the a 95% confidence interval for average prices of gas stations with competitors and gas stations without competitors. Using diffmean to compare the averages.

##	name	lower	upper	level	method	estimate
## 1	N	1.851567	1.900211	0.95	percentile	1.875882
## 2	Y	1.832406	1.872401	0.95	percentile	1.852400

##	name	lower	upper	level	method	estimate
## 1	diffmean	-0.05546772	0.007865699	0.95	percentile	-0.02348235

Using the diffmean function we compared the average prices of of gas stations with competitors and gas stations without competitors. From this comparison I can determine that the price of gas between these 2 categories of gas stations range from about 5 cents for gas stations without competitors and about 4 cents for gas stations with competitors in sight. Then comparing the estimate we can see a differences of about 2 cents more for gas stations without competitors in sight.

Conclusion: Using a 95% confidence interval we can conclude that gas stations with competitors in sight are more likley to charge less compared to gas stations with no surrounding competitors who tend to charge more.

Claim B

The richer the area, the higher the gas prices.

Evidence: In this theory we are analyzing the correlation between the average wealth (Income) in an area and if that has a positive correlation with gas prices. To determin if there is a correlation between the established variables we can bootstrap a linear regression model that will help us visualize the relationship between gas prices and wealth in an area.

##	name	lower	upper	level	method	estimate
## 1	Intercept	-307492.5	-52115.98	0.95	percentile	-177642.8
## 2	Price	56531.0	196980.39	0.95	percentile	125717.6

##	name	lower	upper	level	method	estimate
## 1	cor	0.1970376	0.5695179	0.95	percentile	0.3961546

From the model above we can see that on average for every \$1 change in gas prices the change in wealth of an area can be from about \$56,000 to \$198,000. This range for wealth is on a 95% confidence intercal again. From the second result we can see that the correlation on a 95% interval is between .2 and .5. With an estimate of about .4 we can say there is a positive correlation between gas prices and wealth of an area.

Conclusion:Using a 95% confidence interval we can conclude, the price of gas has a positive correlation with the wealth (income) of an area. Meaning a wealthier area may tend to have higher gas prices.

Claim C

Gas stations at stoplights charge more.

Evidence: This theory is asking gas stations at stoplights charge more or less compared to gas stations not at stoplights. To do this we do it in the same fashion as the first theory. We bootstrap the relationships and then compare the means prices at a 95% confidence level.

##	name	lower	upper	level	method	estimate
## 1	N	1.838542	1.895611	0.95	percentile	1.866316
## 2	Y	1.844477	1.881344	0.95	percentile	1.863916

##	name	lower	upper	level	method	estimate
## 1	diffmean	-0.03847728	0.03034283	0.95	percentile	-0.003299916

In the first table returned we see the range of mean prices at a 95% confidence interval. For gas stations with stoplights we are at about a .04 range and withouth stoplights a .06 range. This means that the price averages of the two categories of gas stations vary by a very small margin.

Conclusion: Using a 95% confidence interval we can see that the average price of gas stations is rather un-related to wheter or not they are located at stoplights.

Claim D

Gas stations with direct highway access charge more

Evidence: This theory is determining the relation of gas station prices to wheter or not they have highway access. We can do this in the same fashion as theory A & C. I will bootstrap the relationships and compare the mean prices of both kinds of gas stations within a 95% confidence level.

##	name	lower	upper	level	method	estimate
## 1	N	1.836666	1.872052	0.95	percentile	1.854304
## 2	Y	1.867220	1.932108	0.95	percentile	1.900000

##	name	lower	upper	level	method	estimate
## 1	diffmean	0.01012218	0.08027445	0.95	percentile	0.0456962

From the first table we can see the range of mean prices for gas stations with and withouth highway access. For highway access gas stations we have a range of about .06 cents and for withouth highway access it is about .03 cents, both these ranges are determined on a 95% confidence interval. This means that the price of gas dependent on highway acces does not vary greatly. We can also see that the estimated diff mean is only about .04 cents, further supporting the idea that highway access and price do not have a correlation.

Conclusion: Using a 95% confidence interval we can determine that the price of gas does not increase with highway access.

Claim E

Shell charges more than all other non-Shell brands.

Evidence: This claim states that shell charges more compared to all other gas brands. To find out if this claim is true we can compare the average price of all gas brands to the average price of gas at Shell stations. To do this we will bootstrap the relationships and compare the means between the 2 categories using a 95% confidence interval.

##	name	lower	upper	level	method	estimate
## 1	N	1.839275	1.874367	0.95	percentile	1.856389
## 2	Y	1.851363	1.916970	0.95	percentile	1.883793

##	name	lower	upper	level	method	estimate
## 1	diffmean	-0.008782587	0.06519746	0.95	percentile	0.02740421

After compiling this data we can see in the first tabelle the range of average gas prices at Shell and all the other gas brand stations. This ranges from about .6 for Shell gas stations and .04 for others. This means that the price for stations do not vary greatly. We can then compare the estimate within our 95% confidence interval which have a difference of about 3 cents higher for Shell. This means on average Shell stations charge about 3 cents more than non-Shell stations.

Conclusion: Using a 95% confidence interval, Shell stations are slightly more expensive (charge more) than non-Shell stations.

Problem Two

Part A

Filter the data set down to include only those cars where year == 2011 and trim == “63 AMG”. Based on these 116 cars, compute a 95% bootstrap confidence interval for the average mileage of 2011 S-Class 63 AMGs that were hitting the used-car market when this data was collected.

##	name	lower	upper	level	method	estimate
## 1	mean	26302.3	31811.52	0.95	percentile	28997.34

The table above, calculated using a 95% confidence interval shows a range from about 26,200 to 31,800 average miles on 2011 S-Class AMG's hitting the used car market.

Part B

Filter the data set down to include only those cars where year == 2014 and trim == “550”. Based on this sample of 2889 cars, compute a 95% bootstrap confidence interval for the proportion of all 2014 S-Class 550s that were painted black. Hint: you might find this easiest if you use mutate to first define a new variable, isBlack, that is either TRUE or FALSE depending on whether the car is black.

##	name	lower	upper	level	method	estimate
## 1	prop_TRUE	0.4174455	0.4520595	0.95	percentile	0.4347525

The table above, calculated using a 95% confidence interval shows the range of the proportion of 2014 550 S-Class AMG's that were painted black. This range is from about 41.7% to 45.3%.

Problem Three

Part A

Are viewers happier watching “Living with Ed” or “My Name is Earl”?

Approach: To do this I will create a bootstrapped relationship ship comparing the difference in average happiness values of viewers based on what show they were watching. I will do this using a 95% confidence interval.

##	name	lower	upper	level	method	estimate
## 1	diffmean	-0.4001312	0.1045087	0.95	percentile	-0.1490515

The table above shows the diff mean on average using the 95% interval as about .15 points higher. Meaning that My Name is Earl is viewed as a happier show.

Conclusion: Calculated on a 95% confidence interval “My Name is Earl” is a happier show than “Living with Ed” by a difference of about 15 points.

Part B

Are viewers more annoyed watching “The Biggest Loser” or “The Apprentice: Los Angeles”?

Approach:In a similar fashion to part A I will create a bootstrapped relationship ship comparing the difference in average annoyance values of viewers based on what show they were watching. I will do this using a 95% confidence interval.

##	name	lower	upper	level	method	estimate
## 1	diffmean	-0.5360384	-0.01734375	0.95	percentile	-0.270997

The table above shows the difference in mean annoyance on average using the 95% interval is about .27 points lower for The Biggest loser than it is for The Apprentice.

Conclusion: Using a 95% confidence interval “The Biggest Loser” is seen on average about .27 points less annoying than “The Apprentice: Los Angeles”.

Part C

Are viewers confused while watching “Dancing with the Stars”?

Approach: In a similar fashion to part A & B I will create a bootstrapped relationship ship, except instead of comparing means I will be looking for the proportion of confused viewers. I will do this based on confusion scores and a 95% confidence interval.

##	name	lower	upper	level	method	estimate
## 1	prop_N	0.8839779	0.961326	0.95	percentile	0.9226519

The table shows the range of the range of proportion of non-confused viewers watching “Dancing with The Stars” On the confidence interval of 95% the proportion ranges from about 87% to 96% of viewers not being confused during the show. Even though a large majority of the viewers are not confused there is still about 8% viewers confused (taken from estimate).

Conclusion: On the confidence interval of 95% the show “Dancing with the Stars” is not confusing for about 92% percent of viewers. Therefore the show is not considered confusing.

Problem Four

Is Ebay's paid search advertising more or less profitable than regular search suggestions and are they worth it?

Approach: To test if Ebay's search advertising is more or less profitable than regular search suggestions AND to see if they are worth it, we have to compare the amount of money earned before and after Ebays search experimnt. When we compare these rations we can tell if areas that used paid searching were more profitable than areas that did not. To do this we will bootstrap the relationship and compare the average values of revenue ratios.

##	name	lower	upper	level	method	estimate
## 1	diffmean	-0.09023835	-0.01372769	0.95	percentile	-0.05228145

From the calculations and table above we can see that the mean score of the non paid search suggestiona range between .0905 and .012 points on average. This means that the expirment group had a slighly greater ratio when compared to the control on the 95% confidence interval.

Conclusion: On a 95% confidence interval, Ebay's paid advertising strategies were both slightly less effective AND less profitable than a strategy using regular search suggestions.