SDS 315 Homework Assignment 3

Karthikeya Kasa

2025-02-13

UTEID: kk38378

Github Link:

https://github.com/karthikeyakasa/SDS2025.git

Problem 1

Theory A: Gas Stations could charge more if they lacked competition that are visible

Evidence:

Table 1: Confidence Interval for Price Difference (With vs. Without Competitors)

name	lower	upper	level	method	estimate
result	-0.055	0.008	0.95	percentile	-0.05

The mean difference in prices between stations with and without competitors is - 0.0329, which suggests that gas stations without competitors usually charge lower prices(by about 3.3 cents), than those with competitors, thus proving my theory wrong.

The 95% confidence interval ranges from -5.54 cents to 0.75 cents. Since the interval includes zero, this means that difference in prices could be zero or that visible competition could have no effect on the prices, or it could suggest a small difference in either direction.

Conclusion: The theory that gas stations charge more if they lack direct competition isn't strongly supported by this data, as while the estimate suggests that stations without competitors charge less, the 95% confidence interval still includes 0, thus this data may not be statistically significant. Based on the data, I can't conclude that the absence of visible competitors leads to higher prices, as the difference could simply be random variation.

Theory B) The richer the area, the higher the gas prices

Claim: My theory is that gas stations in wealthier areas charge higher prices than those in poorer areas. The reasoning is that people in higher-income areas may be less sensitive to price differences, allowing gas stations to charge more. Additionally, the operating costs in wealthier areas might be higher, which could also contribute to higher gas prices.

Table 2: Confidence Interval for Price Difference (Richer vs. Poorer Areas)

name	lower	upper	level	method	estimate
result	-0.009	0.07	0.95	percentile	0.052

Evidence: The results from the confidence interval show that the difference in prices between richer and poorer areas is pretty small. The interval goes from -0.008 to 0.069, which means the actual difference could be slightly negative or positive, but it's pretty close to zero. This suggests that the price difference isn't huge, so the theory that richer areas always charge higher prices doesn't seem to hold up. The prices might be a little higher in wealthier areas, but it's not a big enough difference to say it's always true.

Conclusion: Based on the evidence, it seems like the theory that richer areas always have higher gas prices isn't fully supported. The confidence interval is pretty close to zero, which shows that the price difference between richer and poorer areas isn't that big. While wealthier areas might charge a little more, the difference isn't strong enough to make a solid conclusion. So, this theory isn't really proven right by the data.

Theory C) Gas Stations at stoplights charge more

Claim: My theory suggests that gas stations that are located near stoplights tend to charge higher prices for gas based on the assumption that these gas stations have more visbility or traffic due to the stoplight. Drivers often have to stop at traffic signals, thus making it easier for them to notice the gas station. As a result of the increased visibility, the gas stations may feel that they can charge higher prices, as the assumed customer flow will continue being higher.

Table 3: Confidence Interval for Price Difference (At Stoplights vs. Without Stoplights)

name	lower	upper	level	method	estimate
result	-0.037	0.03	0.95	percentile	0.021

Evidence: In this case, I compared the average prices at gas stations with stoplights to the gas stations without spotlights. The claim I made expects that gas at spotlights should have higher gas prices, so I should look for a positive difference in price. I've found that the estimate for the difference in gas stations between gas stations with stoplights and those without stoplights are very close to zero, which means that there isn't any significant difference between prices at gas stations with stoplights and thos without.

Furthermore, the confidence interval ranged from -3.87 cents to 2.99 cents, which means that based on the data, I can be 95% confident that the true difference in prices can be anywhere in between 3.87 cents less at stations with spotlights to 2.99 cents more. Since this interval contains 0, it suggests that the presence of a spotlight doesn't lead to higher or lower prices.

Conclusion:

While the theory suggests that spotlights may increase gas prices, the data shows that there is no strong evidence of this effect. The price difference is so small that I can't say that the presence of a spotlight affects gas prices in anyway, thus I reject the theory that gas stations at stoplights charge higher prices.

Theory D) Gas Stations with and without highway access

Claim:

My claim is that gas stations with direct access to the highway are likely to charge higher prices for gasoline. My reasoning is that these gas stations are more accessible to drivers on the highway, which means that they might experience more demand from long distance travelers who need gas. As a result, they may be able to raise there prices, as the drivers will keep up demand due to the necessity of gas.

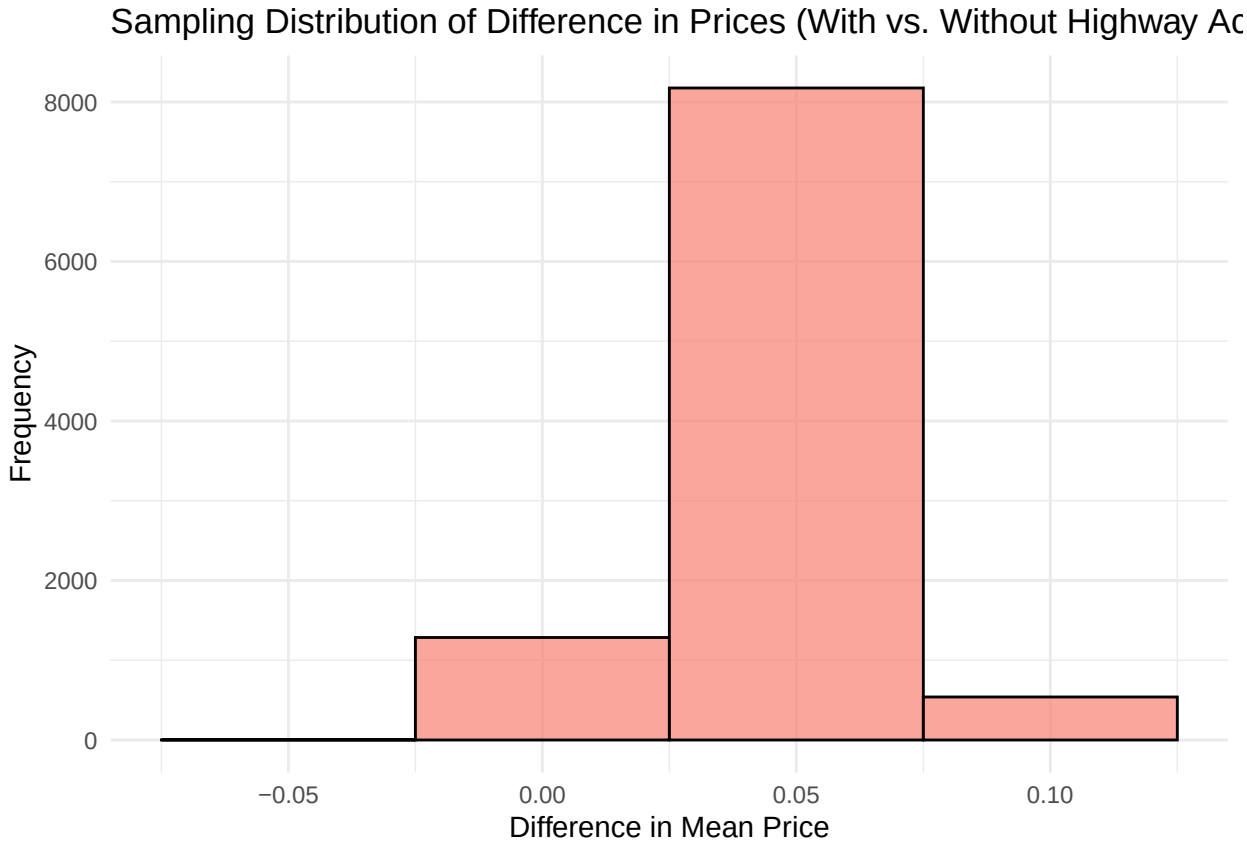

Table 4: Confidence Interval for Price Difference (With vs. Without Highway Access)

name	lower	upper	level	method	estimate
result	0.009	0.081	0.95	percentile	0.041

Evidence:

The estimate for the difference in mean prices between gas stations with and without highway access is about 3.25 cents, which suggests that on average, gas stations on the highway access charge 3.25 cents more than those without highway access.

The 95% confidence interval is between 0.93 cents more and 8.14 cents more at stations with highway access compared to those without highway access. Since the entire confidence interval lies above 0, this shows that the difference in prices is likely in favor of gas stations with highway access, which suggests that they do charge more on average than those without highway access.

Conclusion:

Based on the data and confidence interval, there is strong evidence to suggest that gas stations located along the highway charge more. The positive difference in price, along with the confidence interval that doesn't include 0, shows that gas stations near highways tend to charge higher prices.

Theory E) Gas Stations with and without highway access

Claim: My theory suggests that Shell gas stations charge higher prices for regular gasoline compared to other brands such as ExxonMobil, ChevronTexaco, or other companies. My idea is that Shell, may have

more pricing power compared to other companies due to their brand loyalty, possible assumed quality, or established locations, which could allow them to set higher prices than their competitors.

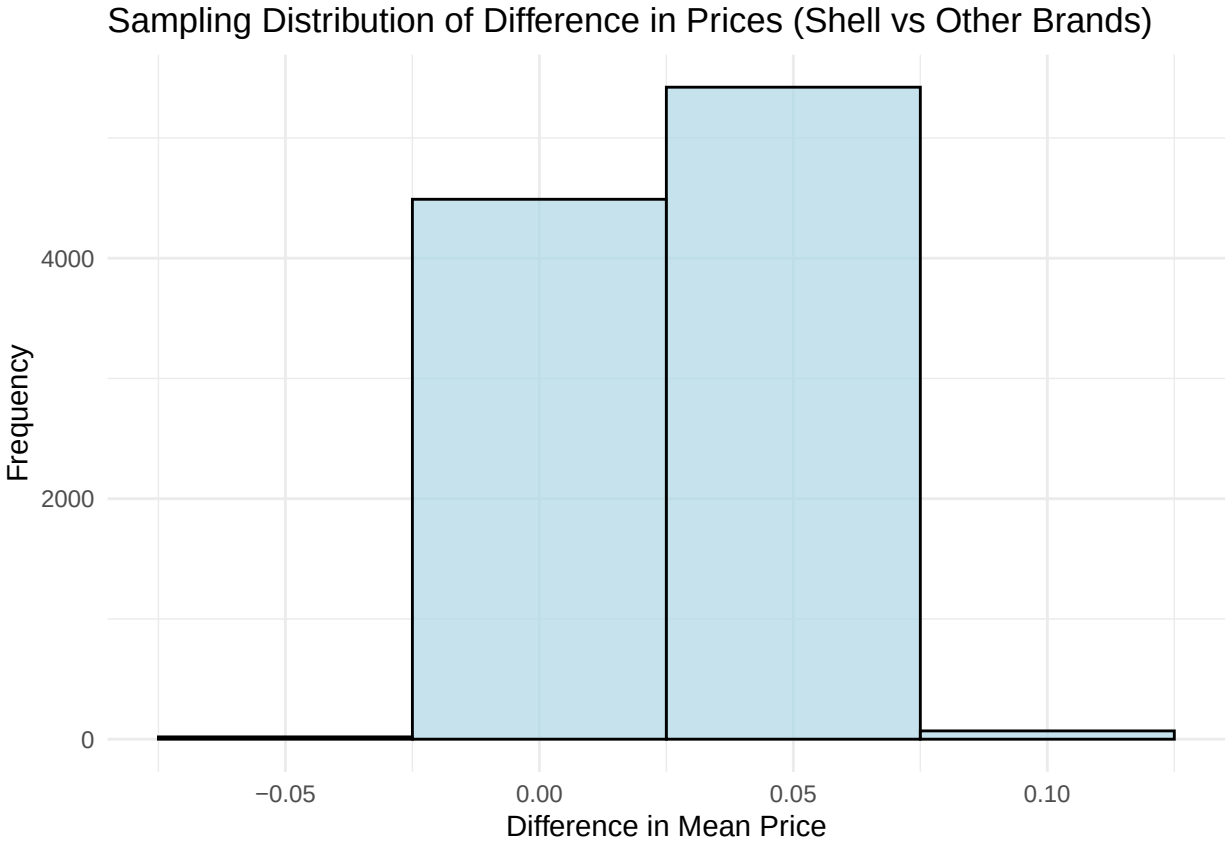

Table 5: Confidence Interval for Price Difference (Shell vs. Non-Shell Gas Stations)

name	lower	upper	level	method	estimate
result	-0.01	0.065	0.95	percentile	0.027

In this case, the estimate for the difference in average prices between Shell and non-Shell companies was 1.77 cents. This suggests that on average , Shell gas stations charge 1.77 cents more than non-Shell gas stations.

The 95% confidence interval lies between -0.91 and 6.52 cents. This means that while the estimate points to Shell charging more, the confidence interval also includes negative values, indicating that there is a possibility that non-shell companies charge higher, or the difference could be 0.

Conclusion:

Based on the data and the confidence interval, there is no strong evidence to suggest that Shell gas stations charge more than non-Shell gas stations. The confidence interval includes negative values which means that, there is a possibility that the difference is 0. This shows that my claim that Shell gas stations charge higher is unsupported by data, and there is no clear evidence for my claim.

Problem 2

Part A)

Table 6: Bootstrap Confidence Interval for the Average Mileage of 2011 S-Class 63 AMGs

name	lower	upper	level	method	estimate
result	26359.1	31772.22	0.95	percentile	28961.76

Part B)

Table 7: Bootstrap Confidence Interval for the Proportion of 2014 S-Class 550s Painted Black

name	lower	upper	level	method	estimate
result	0.417	0.453	0.95	percentile	0.427

Problem 3

Part A)

Claim: I want to figure out which show, "Living with Ed" or "My Name is Earl," makes viewers happier. The goal is to see if there's any strong evidence that one of these shows consistently leads to higher happiness ratings from viewers.

Table 8: 95% Confidence Interval for the Difference in Viewer Happiness (Q1_Happy) Between 'Living with Ed' and 'My Name is Earl'

name	lower	upper	level	method	estimate
result	-0.1	0.398	0.95	percentile	-0.185

Evidence: I constructed a 95% confidence interval for the difference in mean viewer happiness (Q1_Happy) between the two shows. The confidence interval I got was from -0.106 to 0.398, with an estimated difference of 0.002. Since the interval includes 0, it suggests that the difference in viewer happiness between the two shows could be close to zero. This means there's no clear indication that one show makes viewers significantly happier than the other.

Conclusion: Based on the data and the confidence interval, I can conclude that there's no strong evidence that one show makes people consistently happier than the other. Since the interval includes 0, it suggests the difference in happiness could just be due to random chance. So, whether you watch "Living with Ed" or "My Name is Earl," the happiness levels seem to be pretty similar.

Part B)

Claim: I want to find out which reality show, "The Biggest Loser" or "The Apprentice: Los Angeles," made viewers feel more annoyed. The goal is to see if one show caused significantly higher annoyance ratings compared to the other.

Table 9: 95% Confidence Interval for the Difference in Viewer Annoyance (Q1_Annoyed) Between 'The Biggest Loser' and 'The Apprentice: Los Angeles

name	lower	upper	level	method	estimate
result	-0.527	-0.023	0.95	percentile	-0.315

Evidence: The 95% confidence interval for the difference in viewer annoyance (Q1_Annoyed) between the two shows ranged from -0.524 to -0.013, with an estimated difference of -0.264. Since the entire confidence interval is negative, it indicates that "The Biggest Loser" was perceived as more annoying than "The Apprentice: Los Angeles."

Conclusion: The data provides evidence that "The Biggest Loser" caused more annoyance among viewers compared to "The Apprentice: Los Angeles." The negative confidence interval suggests that this difference is statistically significant, supporting the conclusion that "The Biggest Loser" produced higher annoyance ratings.

Part C)

Claim: I want to assess the proportion of viewers who found "Dancing with the Stars" confusing. Specifically, I'm interested in the proportion of viewers who gave a response of 4 or 5 to the Q2_Confusing question, indicating that they agreed or strongly agreed with the statement that the show was confusing.

Table 10: Viewer Responses to 'Dancing with the Stars': Analyzing the Q1_Happy and Q1_Annoyed Scores for the Show

name	lower	upper	level	method	estimate
result	0.039	0.116	0.95	percentile	0.066

Evidence: The 95% confidence interval for the proportion of viewers who found "Dancing with the Stars" confusing ranges from 0.039 to 0.116, with an estimated proportion of 0.083. This means that between 3.9% and 11.6% of viewers are likely to have found the show confusing, based on the survey data.

Conclusion: The data suggests that a small but notable proportion of viewers found "Dancing with the Stars" confusing, with an estimated 8.3% of viewers likely agreeing or strongly agreeing that the show was confusing. The confidence interval shows that this proportion is statistically significant, supporting the idea that some viewers found the show difficult to follow.

Problem 4

Claim: The question being tested is whether Ebay's paid search advertising on Google creates more revenue for the company. EBay ran an experiment where paid search advertising was paused in a random subset of 70 out of 210 designated market areas (DMAs). I expect that the revenue ratio for the DMAs in the treatment group (where advertising was halted) to be lower than that of the control group(where advertising was continued). If the advertising has a positive impact on the revenue, the revenue ratio in the treatment group should be lower than the ratio in the control group.

Table 11: 95% Confidence Interval for the Difference in Revenue Ratios Between EBay DMAs with and without Google AdWords Advertising

name	lower	upper	level	method	estimate
result	-0.091	-0.013	0.95	percentile	-0.052

Evidence: The data was analyzed using bootstrapping to calculate the 95% confidence interval for the difference in revenue rations between the treatment and control groups. The negative estimate of -0.052 means that on average, the revenue ratio in the treatment group (where advertising was paused) is lower than in the control group (where advertising continued). The entire confidence interval lies below zero, indicating that the difference in revenue ratios is statistically significant. With 95% confidence, I can say that the true difference in revenue ratios is somewhere between -0.090 and -0.013, favoring the control group.

Conclusion: The data shows that EBay's paid search ads on Google seem to help with making more money. The confidence interval is all negative, which means the DMAs where the ads were paused made less money than those where the ads kept running. This points to the fact that EBay's paid ads are probably helping boost revenue. Since the revenue ratio in the treatment group is lower than in the control group, it's clear that the ads are helping drive more money for EBay.

So, basically, the numbers back up the idea that EBay's paid search ads aren't a waste. They're actually a good investment that helps the company make more money. This shows how important it is for EBay to keep using these ads to grow their revenue over time.