SDS315 Homework 3 Report

Prady Kandi, EID: prk599

2025-02-13

Contents

The link to the Github repo containing the R file can be found here.

Problem 1

Part A

Table 1: 95% Confidence Interval for Difference in Means of Gas Prices based on Direct Competition

Percentile	Difference in Means
2.5%	-0.0554615
97.5%	0.0083577

The first theory states that gas stations without direct competition charge more on their prices. The 95% confidence interval spans from around -6 cents to 1 cent. Since this interval includes 0, it can be concluded that there is no significant difference, and so gas stations do not charge higher prices due to lack of direct competition. The stated theory is unsupported by evidence.

Part B

Table 2: 95% Confidence Interval for Correlation between Price of Gas and Area Income

Percentile	Correlation
2.5%	0.1937627
97.5%	0.5681092

The second theory states that the prices of gas stations correlate with average income in the area. The 95% confidence interval spans from around .19 to .57. This interval only contains positive measurements, and with the average around .38, it can be concluded that there is a moderate correlation between gas prices and income. The data supports the theory that richer areas have higher gas prices.

Part C

Table 3: 95% Confidence Interval for Difference in Means of Gas Prices based on Stoplight Proximity

Percentile	Difference in Means
2.5%	-0.0385649
97.5%	0.0304900

The third theory states that gas stations at stoplights charge higher prices. The 95% confidence interval spans from around -4 cents to 3 cents. Since this interval includes 0, it can be concluded that there is no significant difference, and so gas stations do not charge higher prices if they are located at a stoplight. The stated theory is unsupported by evidence.

Part D

Table 4: 95% Confidence Interval for Difference in Means of Gas Prices based on Highway Proximity

Percentile	Difference in Means
2.5%	0.0089272
97.5%	0.0818940

The fourth theory states that gas stations located directly off of highways charge more on their prices. The 95% confidence interval spans from less than a cent to 8 cents. The interval does not include 0 and only contains positive measurements, so it can be concluded that there is a significant difference, and so gas stations do in fact charge higher prices if located directly next to a highway. The theory is supported by the data.

Part E

Table 5: 95% Confidence Interval for Difference in Means of gas Prices based on Shell Brand

Percentile	Difference in Means
2.5%	-0.0091541
97.5%	0.0656198

The fifth theory states that Shell gas stations charge higher prices than other brands. The 95% confidence interval spans from around -1 cents to 6 cents. Since this interval includes 0, it can be concluded that there is no significant difference, and so Shell gas stations do not charge higher prices. The stated theory is unsupported by evidence.

Problem 2

Part A

Table 6: 95% Confidence Interval for Average Mileage

Percentile	Mean Mileage
2.5%	26273.19
97.5%	31813.32

The lower bound of the 95% interval is around 26000 miles and the upper bound is 32000 miles, implying that the true mean is within this set with a 95% chance; that is, the average mileage of all given cars made in 2011 and with a 63 AMG trim.

Part B

Table 7: 95% Confidence Interval for Proportion of Black Cars

Percentile	Proportion
2.5%	0.4170993
97.5%	0.4527518

The lower bound of the 95% interval is around 41.6% and the upper bound is 45.3%, implying that the true mean is within this set with a 95% chance; that is, the chance that all cars made in 2014 and with a 550 trim are black.

Problem 3

Part A

Table 8: 95% Confidence Interval for Difference in Means of Happiness

Percentile	Difference in Means
2.5%	-0.3962999
97.5%	0.1033589

The question we are trying to answer is whether the pilot episode of Living with Ed or My Name is Earl received higher happiness ratings. To answer the question, a Monte Carlo simulation was utilized with a 95% confidence interval, where the simulation was run 10000 times. The results were then compiled into a table to observe the results. The difference ranges from -0.4 to 0.1, indicating there is no significant difference. It can be said that the true difference between the two happiness ratings is insignificant and lies within the given range with 95% certainty, implying that neither show received substantially better or worse happiness ratings than the other.

Part B

Table 9: 95% Confidence Interval for Difference in Means of Annoyance

Percentile	Difference in Means
2.5%	-0.5267926
97.5%	-0.0218691

The question we are trying to answer is whether the pilot episode of The Biggest Loser or The Apprentice: Los Angeles received higher annoyance ratings. To answer the question, a Monte Carlo simulation was utilized with a 95% confidence interval, where the simulation was run 10000 times. The results were then compiled into a table to observe the results. The difference ranges from -.55 to -.01, indicating there is a significant difference. It can be said that the true difference between the two happiness ratings is indeed significant, and that The Apprentice: LA pilot received higher annoyance ratings. The data for this conclusion lies within the given range with 95% certainty, showing that it is very likely the viewers as a population were more annoyed with the The Apprentice: LA pilot than the The Biggest Loser pilot.

Part C

Table 10: 95% Confidence Interval for Confused Viewers of Dancing with the Stars

Percentile	Mean Viewers Confused
2.5%	0.038674
97.5%	0.121547

The question we are trying to answer is how many viewers were confused with the pilot of Dancing with the Stars, ie. agreed or strongly agreed that they were confused. To answer the question, a Monte Carlo simulation was utilized with a 95% confidence interval, where the simulation was run 10000 times. The results were then compiled into a table to observe the results. The mean ranges from 3.5% to 12%. It can be said that the true percentage of viewers confused lies within this range with 95% certainty. In conclusion, there is a 95% certainty that the average proportion of confused viewers ranges from 3.5% to 12%. It can be concluded that a slim minority of the population found the pilot confusing, with most viewers able to understand it.

Problem 4

Part A

Table 11: 95% Confidence Interval for Difference in Means of Revenue Ratio

Percentile	Difference in Means
2.5%	-0.0910114
97.5%	-0.0126256

The question we are trying to answer is whether there was a substantial difference in eBay's relative revenue depending on whether or not they ran ads on Google. To solve the question, a Monte Carlo simulation was utilized with a 95% confidence interval, where the simulation was run 10000 times. The results were then compiled into a table to observe the results. The difference ranges from -.09 to -.01, indicating there is a significant difference. The data indicates that eBay did make less money when not running ads, and that the relative amount lost is within the aforementioned interval with 95% certainty. Therefore, it is more efficient for eBay to run ads on Google as opposed to not doing so if they wish to generate the most revenue.