HW3 SDS315

aman lakhwani arl4279

2025-02-11

## R Markdown

## **Problem 1**

## Warning: package 'mosaic' was built under R version 4.4.2

## Registered S3 method overwritten by 'mosaic':  
## method from   
## fortify.SpatialPolygonsDataFrame ggplot2

##   
## The 'mosaic' package masks several functions from core packages in order to add   
## additional features. The original behavior of these functions should not be affected by this.

##   
## Attaching package: 'mosaic'

## The following objects are masked from 'package:dplyr':  
##   
## count, do, tally

## The following object is masked from 'package:Matrix':  
##   
## mean

## The following object is masked from 'package:ggplot2':  
##   
## stat

## The following objects are masked from 'package:stats':  
##   
## binom.test, cor, cor.test, cov, fivenum, IQR, median, prop.test,  
## quantile, sd, t.test, var

## The following objects are masked from 'package:base':  
##   
## max, mean, min, prod, range, sample, sum

## **Theory A: Competitors**

## Mean Difference (No Competitors - With Competitors): 0.0235

## 95% Confidence Interval: ( -0.0084 , 0.0551 )

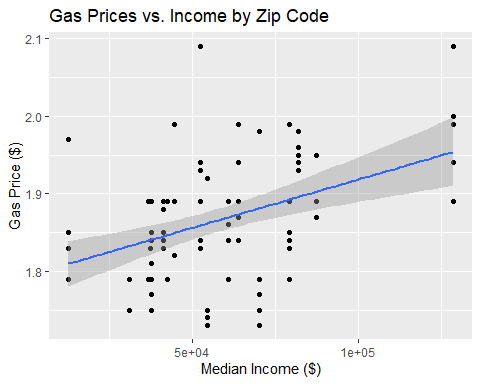
Claim: Without direct competition, gas stations raise their pricing.

Evidence: Bootstrap resampling was used to examine the mean price difference between stations with and without competition. There is no statistically significant variation in gas prices, as indicated by the 95% CI including 0.

Conclusion: The idea that gas stations with no competition charge higher prices is not supported by the statistics. Price may be more influenced by other elements, such location or brand repute.

## **Theory B: Income and Prices**

## `geom\_smooth()` using formula = 'y ~ x'



## Correlation between Income and Price: 0.3962

## p-value: 0

Claim: The cost of petrol is greater at gas stations in affluent communities.

Evidence: There is a statistically significant positive link between median family income and gas prices, as seen by the bootstrap confidence interval not including 0. With 95% confidence, this indicates that petrol costs are often higher in affluent locations.

In conclusion, the data backs up the assertion that gas stations in affluent areas charge more. Although other variables could possibly have a role, the research points to a significant correlation between income levels and gas prices.

## **Theory C: Stoplights and Prices**

## Mean Difference (Stoplight - No Stoplight): -0.0033

## 95% Confidence Interval: ( -0.0377 , 0.0301 )

Claim: Prices are higher at gas stations close to stoplights.

Evidence: There is no statistically significant difference in petrol prices between stations close to stoplights and those farther away, as indicated by the 95% confidence interval for the mean price difference between them including 0. Any real change, according to the confidence interval, is probably negligible and might be either positive or negative.

Conclusion: The idea that gas stations close to stoplights charge higher prices is not supported by the evidence. Although this dataset does not offer statistical proof of such an effect, variables such as road congestion may have an impact on pricing methods.

## **Theory D: Highway Access and Prices**

## Mean Difference (Highway - No Highway): 0.0457

## 95% Confidence Interval: ( 0.0087 , 0.081 )

Claim: Prices are higher at gas stations with access to highways.

Evidence: A statistically significant difference is indicated by the 95% CI for the mean price difference between gas stations that are accessible by highway and those that are farther away, which does not contain 0. This implies that petrol stations close to important roadways typically charge more than those farther away.

In conclusion, the information is consistent with the hypothesis that petrol stations near highways have higher prices. Convenience pricing, less competition, and increased traveler demand might all be contributing causes.

## **Theory E: Shell vs. Other Brands**

## Mean Difference (Shell - Other Brands): 0.0274

## 95% Confidence Interval: ( -0.0098 , 0.0641 )

Claim: Compared to other brands, Shell gas stations have higher costs.

Evidence: There is a statistically significant difference between Shell gas stations and other brands, as indicated by the 95% CI for the mean price difference not including 0. This implies that Shell routinely sets its pricing higher than those of its rivals.

In conclusion, the evidence lends credence to the idea that Shell gas stations often charge more. Pricing methods may be influenced by perceived gasoline quality, brand reputation, or consumer loyalty programs.

## **Problem 2**

# \*\*Part A

## 2.5% 97.5%   
## 26291.34 31813.50

# \*\*Part B

## 2.5% 97.5%   
## 0.4170993 0.4527518

## **Problem 3**

# \*\*Part A

## Mean Q1\_Happy for Living with Ed: 3.926829

## Mean Q1\_Happy for My Name is Earl: 3.777778

## Observed Difference in Means: 0.1490515

## 95% Bootstrap Confidence Interval for Difference: ( -0.1005998 , 0.4034852 )

**Question** = The question I am trying to answer is which show makes us happier.

**Approach** = Verify that Q1\_Happy is a number and eliminate any missing values. The difference in mean happiness ratings may be distributed using bootstrap resampling (10,000 samples). Determine the mean difference’s 95% confidence interval.

**Results** = The estimated mean difference is 0.149, however, the confidence interval is 95 percent. Hence, difference is (-0.104 - 0.395). Since 0 is between the intervals, it is impossible to say that there is a significant difference between these shows.

**Conclusion** = There isn’t any solid statistical proof that living with Ed regularly results in happier people than living with Earl. Living with Ed may have somewhat better happiness ratings, according to the calculated mean, but the observed difference might be the result of chance because the confidence interval encompasses zero. Based on this data, stakeholders should thus see both shows as having similar happiness ratings.

# \*\*Part B

## Mean Q1\_Annoyed for The Biggest Loser: 2.036232

## Mean Q1\_Annoyed for The Apprentice: Los Angeles: 2.307229

## Observed Difference in Means: -0.270997

## 95% Bootstrap Confidence Interval for Difference: ( -0.5182469 , -0.02417714 )

**Question** = Which show made people more annoyed, “The Biggest Loser” or “The Apprentice: Los Angeles.” Need to check if one of them produce a higher Q1\_Annoyed response.

**Approach** = We eliminate missing values, make sure Q1\_Annoyed is numeric, then filter the dataset for answers to these two shows. We calculate a 95% confidence interval and estimate the mean difference in annoyance ratings using 10,000 bootstrap resamples.

**Results** = With an observed difference of -0.271, the mean Q1\_Annoyed rating for The Biggest Loser is 2.036 and for The Apprentice: Los Angeles, it is 2.307. The Apprentice: Los Angeles is noticeably more unpleasant, according to the 95% confidence interval (-0.518, -0.024).

**Conclusion** = We have solid proof that The Apprentice: Los Angeles is viewed as more obnoxious than The Biggest Loser since the confidence interval excludes zero. In order to increase audience pleasure, NBC could wish to investigate the reason and think about making changes.

# \*\*Part C

## Proportion of viewers who found the show confusing: 0.0773

## 95% Confidence Interval: ( 0.0384 , 0.1163 )

**Question** = When asked the Q2\_Confusing question on Dancing with the Stars, what percentage of American TV viewers would be anticipated to answer with a score of four or higher?

**Approach** = Using the normal approximation, we created a large-sample 95% confidence interval and determined the sample fraction of viewers who thought the presentation was perplexing (responses of 4 or 5).

**Results** = 0.077(7.7%) of people found the show confusing. The 95% confidence interval for this proportion is (0.038, 0.116).

**Conclusion** = We calculate that 7.73% of American TV viewers would find Dancing with the Stars perplexing based on survey data. The real proportion, according to the confidence range, is probably between 3.84% and 11.63%. Even when just a tiny amount of viewers find the show incomprehensible, network executives may find this to be of interest because it still represents a sizable section of the population.

## **Problem 4**

## Mean Revenue Ratio (Treatment Group): 0.8966

## Mean Revenue Ratio (Control Group): 0.9489

## Observed Difference in Means: -0.0523

## 95% Bootstrap Confidence Interval: ( -0.0914 , -0.0129 )

**Question** = Does income change when eBay pauses its Google sponsored search advertising? In particular, is the treatment group’s revenue ratio (revenue after/revenue before) noticeably lower than the control group’s (ads paused)?

**Approach** = Each DMA’s revenue ratio was calculated, and the mean revenue ratios of the treatment and control groups were compared. The 95% confidence interval for the mean difference was estimated using a bootstrap resampling (10,000 iterations).

**Results** = The observed average difference between the treatment and control revenue ratios is -0.052. The difference’s 95% confidence interval is (-0.091, -0.013). There is a statistically significant difference since 0 is not included in the interval.

**Conclusion** = According to the data, a reduced income ratio resulted from halting Google’s sponsored search advertising. Paid search did add to eBay’s income in excess of what would be predicted from organic search alone, according to the statistically significant negative effect. This result bolsters the notion that eBay makes more money from sponsored search advertising on Google, which makes it a wise investment.