ECO394D Probability and Statistics Homework 4

Sibo Ding

Summer 2023

## Problem 1

### Part A

Question: Does one of “Living with Ed” and “My Name is Earl” have a higher mean Q1\_Happy than the other?  
Approach: 2-sample two-sided t-test  
Results:

##   
## Welch Two Sample t-test  
##   
## data: lwe and mni  
## t = 1.1676, df = 162.57, p-value = 0.2447  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.1030341 0.4011371  
## sample estimates:  
## mean of x mean of y   
## 3.926829 3.777778

Conclusion: p-value is greater than 0.05 (or 95% confidence interval includes 0). We cannot reject so no one show has a higher mean Q1\_Happy at 5% significance.

### Part B

Question: Does one of “The Biggest Loser” and “The Apprentice: Los Angeles” have a higher mean Q1\_Annoyed than the other?  
Approach: 2-sample two-sided t-test  
Results:

##   
## Welch Two Sample t-test  
##   
## data: tbl and tal  
## t = -2.1032, df = 300.66, p-value = 0.03628  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.52455614 -0.01743792  
## sample estimates:  
## mean of x mean of y   
## 2.036232 2.307229

Conclusion: p-value is smaller than 0.05 (or 95% confidence interval does not include 0). We reject so one show has a higher mean Q1\_Annoyed at 5% significance.

### Part C

Question: Use a filtered data set to infer the proportion of 4 or grater Q2\_Confusing of “Dancing with the Stars” with its 95% confidence interval.  
Approach 1: 1-sample proportions test  
Results:

##   
## 1-sample proportions test with continuity correction  
##   
## data: sum(dwt$Q2\_Confusing >= 4) out of length(dwt$Q2\_Confusing >= 4), null probability 0.5  
## X-squared = 127.65, df = 1, p-value < 2.2e-16  
## alternative hypothesis: true p is not equal to 0.5  
## 95 percent confidence interval:  
## 0.04453431 0.12893254  
## sample estimates:  
## p   
## 0.07734807

Approach 2: Normal approximation for binomial distribution (based on C.L.T.)  
Results:  
Actual proportion

## [1] 0.07734807

Sample size

## [1] 181

## Lower limit: 0.03842989

## Upper limit: 0.1162662

Approach 3: Use bootstrap and Monte Carlo simulation to generate many samples and estimate C.I..

## Problem 2

Question: Whether the revenue ratios are the same in the treatment and control groups?  
Approach: 2-sample two-sided t-test  
Results:

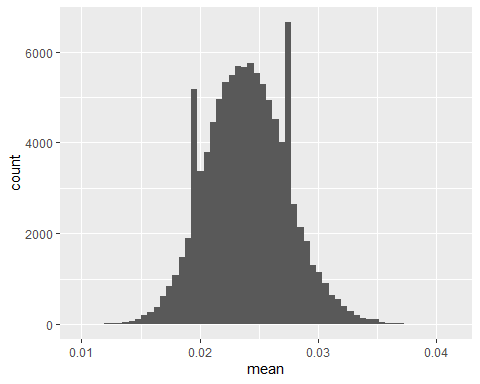
##   
## Welch Two Sample t-test  
##   
## data: control and treatment  
## t = 2.6367, df = 110.2, p-value = 0.00958  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.01298714 0.09157576  
## sample estimates:  
## mean of x mean of y   
## 0.9488775 0.8965961

Conclusion: p-value is smaller than 0.05 (or 95% confidence interval does not include 0). We reject so the revenue ratios in the treatment and control groups are different at 5% significance.

## Problem 3

: The proportion of flagged trades from Iron Bank is 2.4%.  
Test statistic:   
When plotting simulation results, histogram rather than p.d.f. makes more sense:

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

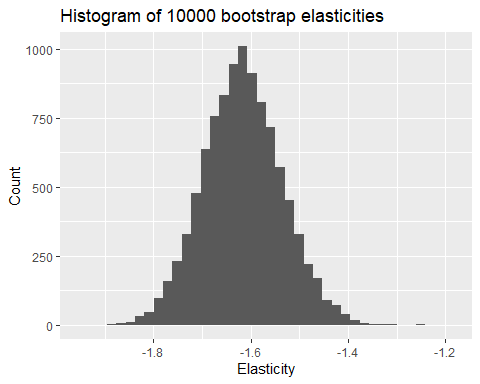


p-value for observing

## [1] 0.00129

Conclusion: is not plausible in light of the data because is pretty small, which is very rare to happen.

## Problem 4



## Confidence Interval from Bootstrap Distribution (10000 replicates)

## 2.5% 97.5%  
## percentile -1.77 -1.45

## Problem 5

### Part A

1. Similarly, , thus
2. Since having the same finite mean and variance, as , based on C.L.T.,
3. ,   
   Based on C.L.T., ,

### Part B

Similar to Part A,