Hypothesis Testing For Automobile Prices

Classical vs. Bootstrap Methods

Greg DeVore

July 29th, 2017

Overview

The purpose of this assignment is to compare classical and bootstrap methods for hypothesis testing in statistics. Bootstrap methods involve repeatedly sampling a population with replacement in order to compute statistics such a mean, median, or even statistics like the difference in means between two populations. They are useful when the assumptions required for using a classical method (with regards to distribution, variance, etc.) cannot be guaranteed. This comparison will be done by looking at automobile data, where the response variable is the price of the automobile, and the explanatory variables are different attributes of each automobile. In the previous write-up, the following conclusions were reached using classical methods:

1. A statistically significant relationship exists for the price of an automobile when stratified by aspiration (standard vs. turbo), but not fuel type (gas vs. diesel).
2. A statistically significant relationship exists between the price of a sedan versus a hatchback, but not between wagons and hatchbacks, or sedans and wagons.

These same hypotheses will be tested, but this time using a bootstrapped approach. After performing the same tests, it will be found that:

1. Using bootstrap methods, a statistically significant relationship exists for the price of an automobile when stratified by both aspiration (standard vs. turbo) and fuel type (gas vs. diesel).
2. Using bootstrap methods, a statistically significant relationship exists between the price of sedans and hatchbacks, and wagons and hatchbacks, but not between sedans and wagons.

In both bivariate and multivariate tests, classical and bootstrap methods produce different results for the automobile data. When the alternative hypothesis for a test was rejected by a close margin using a classical method, it was accepted using a bootstrap method.

Data Preparation

As in the previous write-up, we'll need to convert the automobile price to a numeric variable. Also, we'll want several of the explanatory variables to be factors, as this will be helpful later on. Finally, we'll remove any incomplete observations (rows containing one or more NA's).

```
# Coerce some factors to numeric and retain only complete observations
auto$price <- as.numeric(auto$price)
auto$fuel.type <- as.factor(auto$fuel.type)
auto$aspiration <- as.factor(auto$aspiration)
auto$body.style <- as.factor(auto$body.style)
auto <- auto[complete.cases(auto),]
```

Additionally, in the previous write-up, we determined that the log of the automobile price more closely resembled a normal distribution when compared to the price alone. We'll add it as a new column to the data.

```
# Add logPrice as additional column
auto$logPrice <- log(auto$price)
```

Significance Tests For Two Samples

In this section, we'll use bootstrap methods to test whether or not there is a significant difference in automobile price when stratified by fuel type and aspiration, and compare the results to those obtained using a classical method (Welch's t-test). In all cases, the null hypothesis will be that there is no relationship between the price of an automobile when stratified by the given explanatory variable (the difference in the means of the two samples is zero). The alternative hypothesis will be that a relationship exists (the difference in the means of the two samples is not zero). Since the direction of the difference is not specified (greater than or less than), a two-tailed test will be used. Also, for all tests a confidence level of 95% will be used. This means $\alpha = 0.05$, and the null hypothesis will be rejected in favor of the alternative hypothesis if the p-value is less than α. In all plots, a solid red line indicates the mean of the distribution, and dotted red lines indicate the 95% confidence interval for the mean.

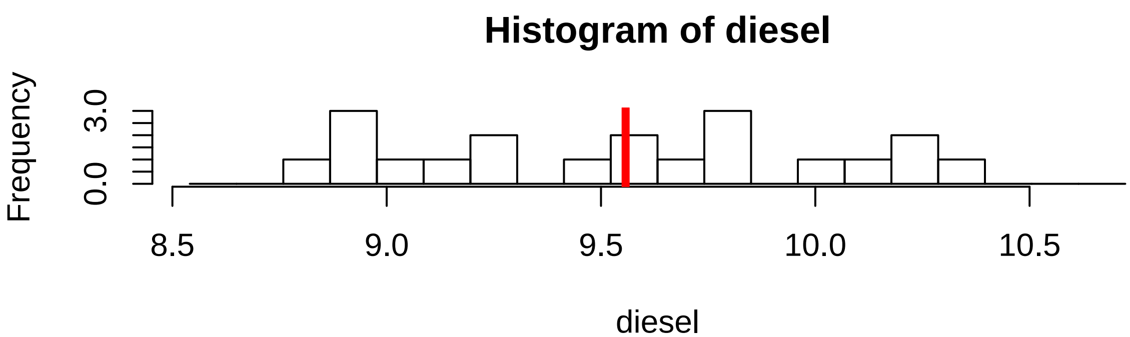
Fuel Type

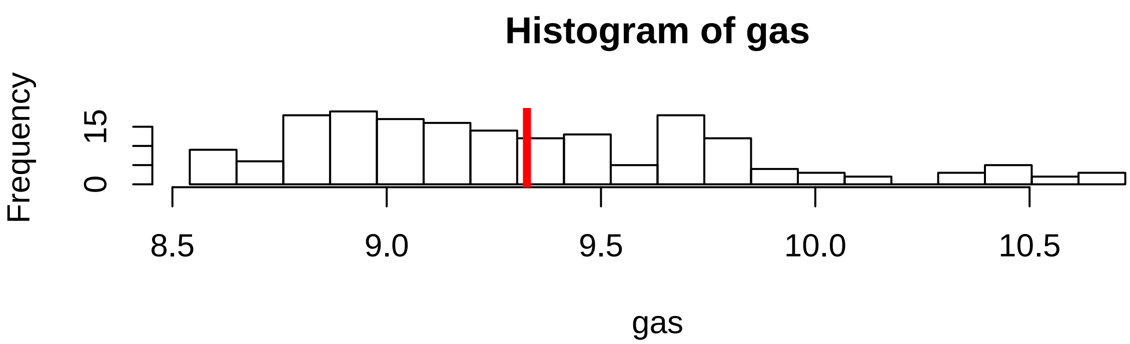
First, let's look at price stratified by fuel type. To do that, we'll create variables by splitting the price of automobiles by the two fuel types, gas and diesel.

```
diesel_price <- filter(auto, fuel.type == 'diesel') %>% pull('logPrice')
gas_price <- filter(auto, fuel.type == 'gas') %>% pull('logPrice')
```

Let's look at the histogram of each population:

```
plot.dists(diesel_price, gas_price, cols = c('diesel', 'gas'))
```

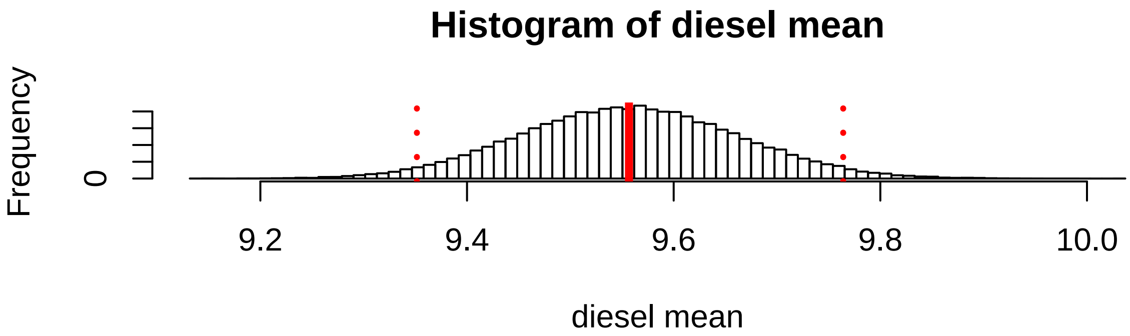


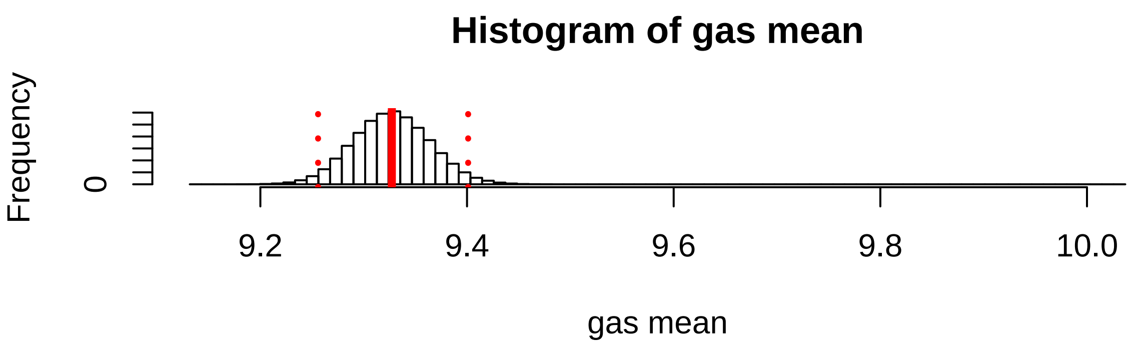


The populations have a noticeable difference in means, so let's bootstrap the means and plot the results for a more direct comparison.

```
# Calculate bootstrap means
mean.boot.diesel = one.boot(diesel_price, mean, R = 100000)
```

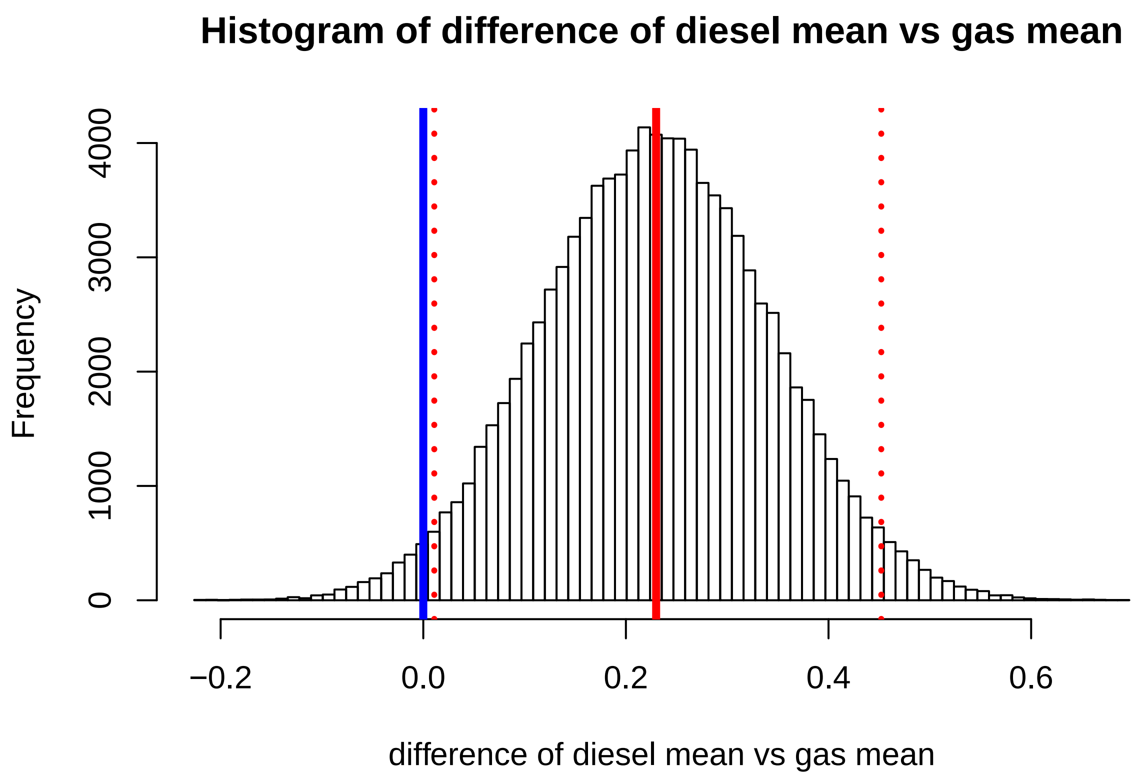
```r
mean.boot.gas = one.boot(gas_price, mean, R = 100000)
# Plot bootstrap means
plot.t(mean.boot.diesel$t, mean.boot.gas$t, cols = c('diesel mean','gas mean'))
```





The 95% confidence intervals overlap, which suggests the populations might not be different enough to reject the null hypothesis. To be sure, let's bootstrap the difference in means and plot the results.

```r
# Calculate bootstrap difference of means
mean.boot.diesel_gas = two.boot(diesel_price, gas_price, mean, R = 100000)
plot.diff(mean.boot.diesel_gas$t, 'difference of diesel mean vs gas mean')
```



Note that the distribution appears to closely follow a normal distribution. The 95% confidence interval does not include zero, which means we can reject the null hypothesis and accept the alternative hypothesis. That is, we can say that a statistically significant relationship does exist for the price of an automobile when stratified by fuel type. Let's compare this result to the result obtained using Welch's t-test:

```
# Welch's t-test
t.test(diesel_price, gas_price, alternative = 'two.sided')
```

```
##
##  Welch Two Sample t-test
##
## data:  diesel_price and gas_price
## t = 1.9971, df = 23.627, p-value = 0.05746
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.007901153  0.468325711
## sample estimates:
## mean of x mean of y
##  9.557420  9.327208
```

Here, the p-value is just over 0.05, and the confidence interval just barely includes zero. This means we cannot reject the null hypothesis and cannot say that a statistically significant relationship exists between the price of an automobile when stratified by fuel type. This result directly disagrees with the result obtained using the bootstrap method.

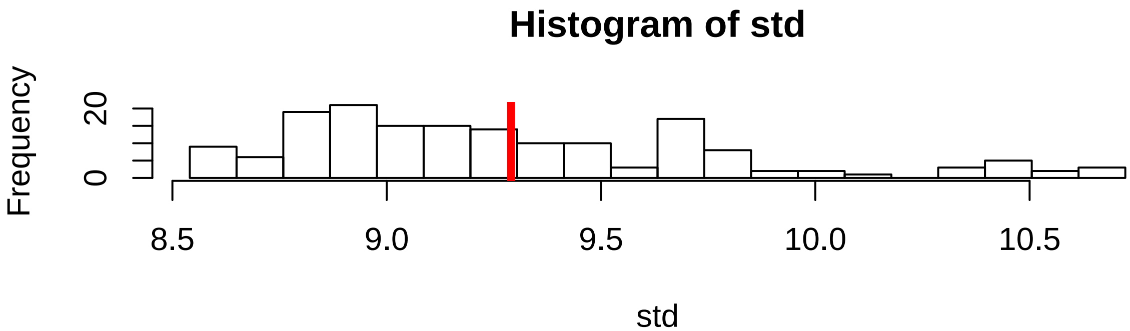
Aspiration

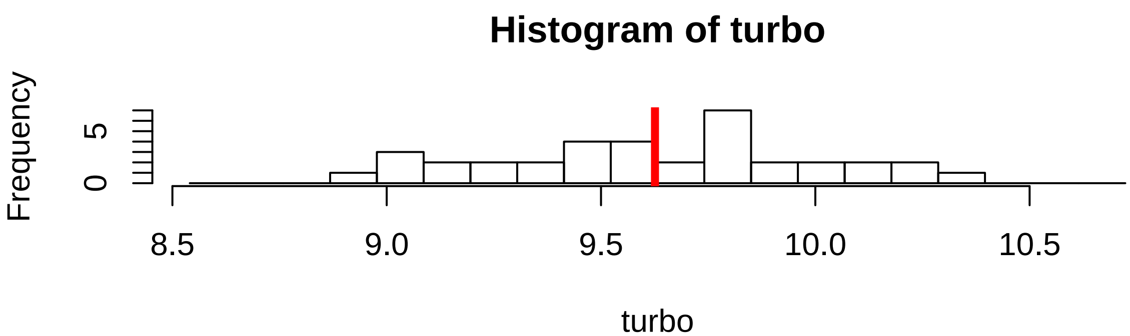
Next, let's look at price stratified by aspiration. To do that, we'll create variables by splitting the price of automobiles by the two types of aspiration, standard and turbo.

```
std_price <- filter(auto, aspiration == 'std') %>% pull('logPrice')
turbo_price <- filter(auto, aspiration == 'turbo') %>% pull('logPrice')
```

Let's look at the histogram of each population:

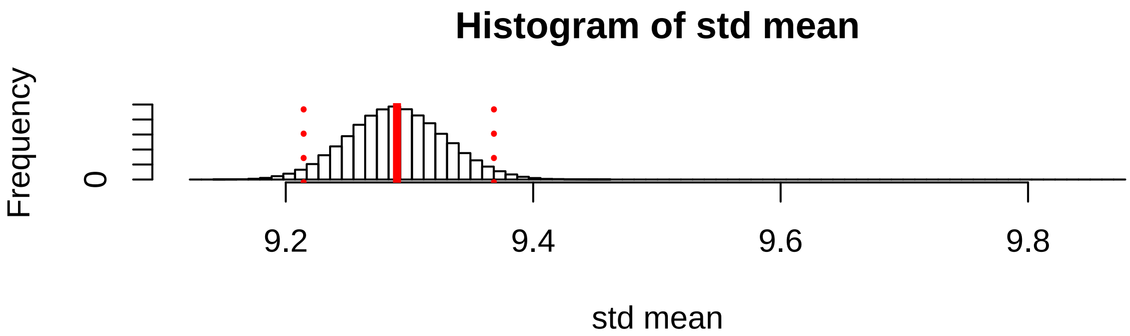
```
plot.dists(std_price, turbo_price, cols = c('std', 'turbo'))
```

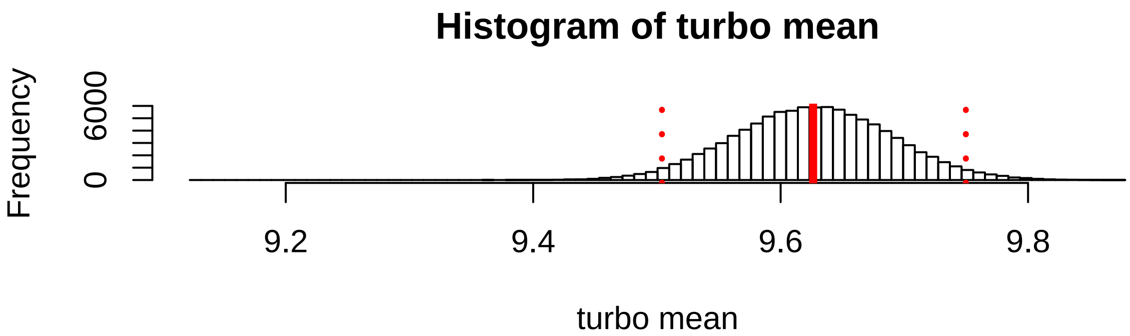




Once again, the populations have a noticeable difference in means, so let's bootstrap the means and plot the results for a more direct comparison.

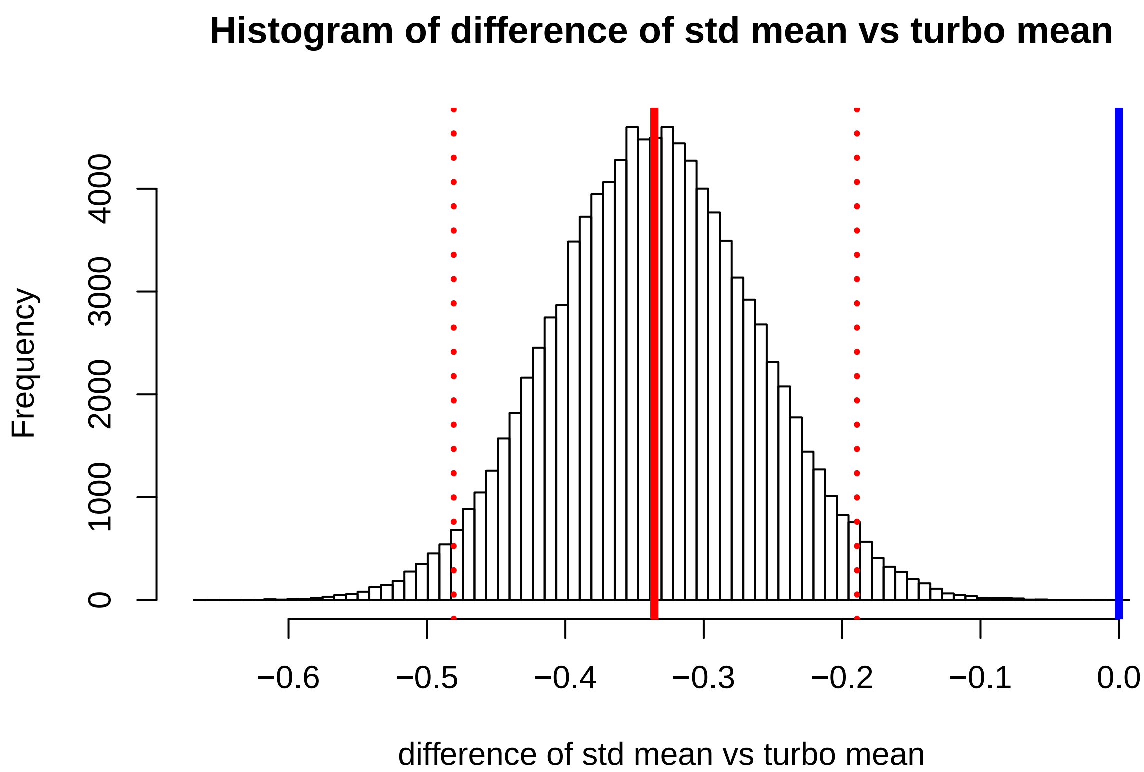
```
# Calculate bootstrap means
mean.boot.std = one.boot(std_price, mean, R = 100000)
mean.boot.turbo = one.boot(turbo_price, mean, R = 100000)
# Plot bootstrap means
plot.t(mean.boot.std$t, mean.boot.turbo$t, cols = c('std mean','turbo mean'))
```





Here, the 95% confidence intervals do not overlap, which suggests the populations might be different enough to reject the null hypothesis. To be sure, let's bootstrap the difference in means and plot the results.

```
# Calculate bootstrap difference of means
mean.boot.std_turbo = two.boot(std_price, turbo_price, mean, R = 100000)
plot.diff(mean.boot.std_turbo$t,  'difference of std mean vs turbo mean')
```



Once again, note that the distribution appears to closely follow a normal distribution. The 95% confidence interval definitely excludes zero, which means we can again reject the null hypothesis and accept the alternative hypothesis. That is, we can say that a statistically significant relationship does exist between the price of an automobile when stratified by aspiration.

Let's compare this result to the result obtained using Welch's t-test:

```
# Welch's t-test
t.test(std_price, turbo_price, alternative = 'two.sided')
```

```
##
##  Welch Two Sample t-test
##
## data:  std_price and turbo_price
## t = -4.4777, df = 64.681, p-value = 3.137e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.4858704 -0.1861209
## sample estimates:
## mean of x mean of y
##  9.289936  9.625932
```

Here, the p-value is significantly less than 0.05, and the confidence interval definitely excludes zero. This means we can reject the null hypothesis and say that a statistically significant relationship exists between the price of an automobile and the aspiration. This result agrees with the result obtained using the bootstrap method.

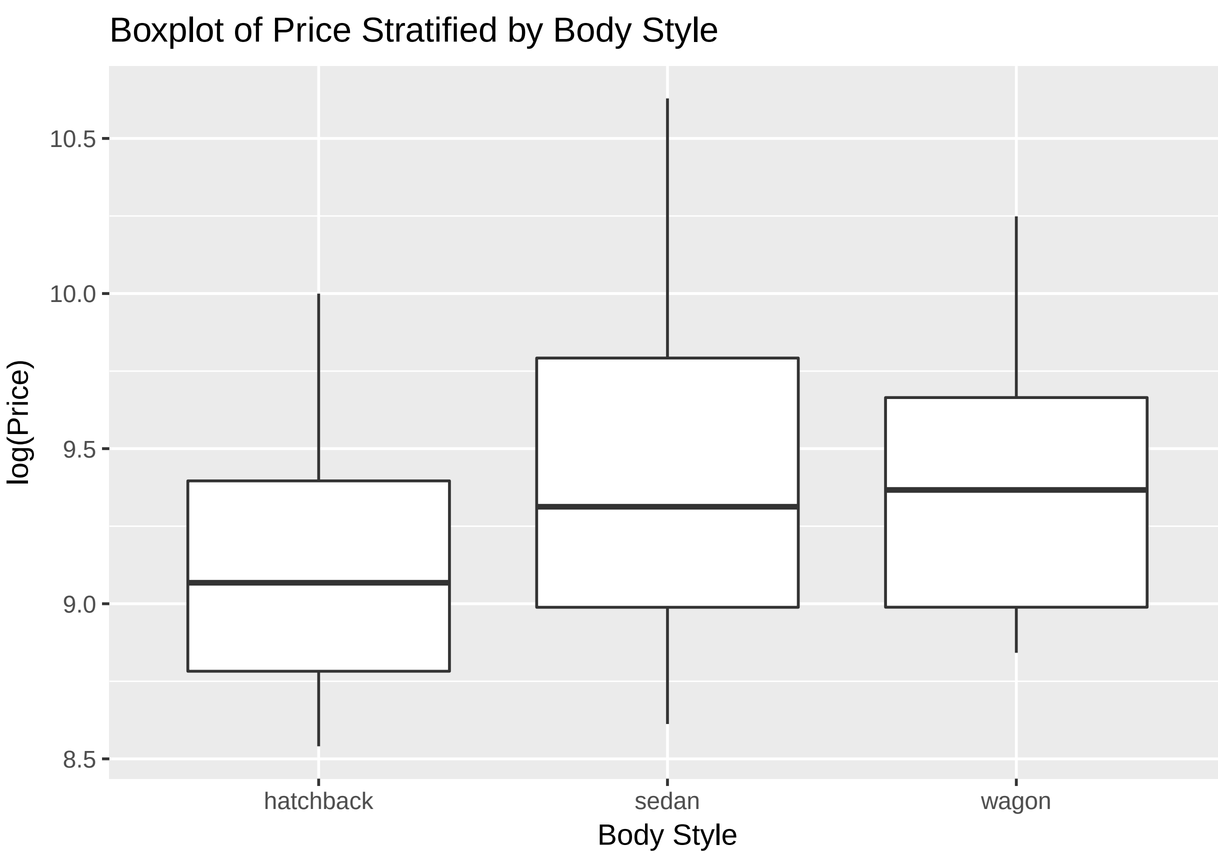
Multi-Sample Significance Tests

In this section, we'll use bootstrap methods to see whether or not there is a significant difference in automobile price when stratified into more than two groups (using body style). Once again, we'll compare the results to those obtained using classical methods (Tukey's HSD). As with the two-sample tests, the null hypothesis will be that all of the groups come from the same population (i.e., there is no difference in means), and the alternative hypothesis will be that at least one of the groups has a statistically different mean. Also, as before a 95% confidence level will be used.

In the previous write-up, we determined that we could only perform comparisons between the body styles of sedan, hatchback, and wagon, as only they had sufficient data for comparison.

To review, let's look at the boxplots of price for each of three body types.

```
# Create subset by dropping convertible and hardtop (insufficient data)
auto_bodystyle <- subset(auto, auto$body.style %in% c('hatchback','sedan','wagon'))
# Visual
ggplot(auto_bodystyle, aes(body.style,logPrice)) + geom_boxplot() + xlab('Body Style') +
  ylab('log(Price)') + ggtitle('Boxplot of Price Stratified by Body Style')
```



Note that hatchbacks appear to be the most different when compared to sedans and wagons. Let's also look at the results of the classical method (Tukey's HSD) before using bootstrap methods.

```
# ANOVA
df_aov <- aov(logPrice ~ body.style, data = auto_bodystyle)
# Tukey HSD
tukey_anova <- TukeyHSD(df_aov)
tukey_anova
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = logPrice ~ body.style, data = auto_bodystyle)
##
## $body.style
##                        diff         lwr       upr     p adj
## sedan-hatchback  0.31088820  0.13945380 0.4823226 0.0000870
## wagon-hatchback  0.22291209 -0.02895758 0.4747818 0.0944466
## wagon-sedan     -0.08797611 -0.33030122 0.1543490 0.6675193
```

Note that only sedans and hatchbacks have a statistically significant difference in price (p-value less that 0.05). Moving on to the bootstrap methods, we first need to compute the bootstrap means for each body style.

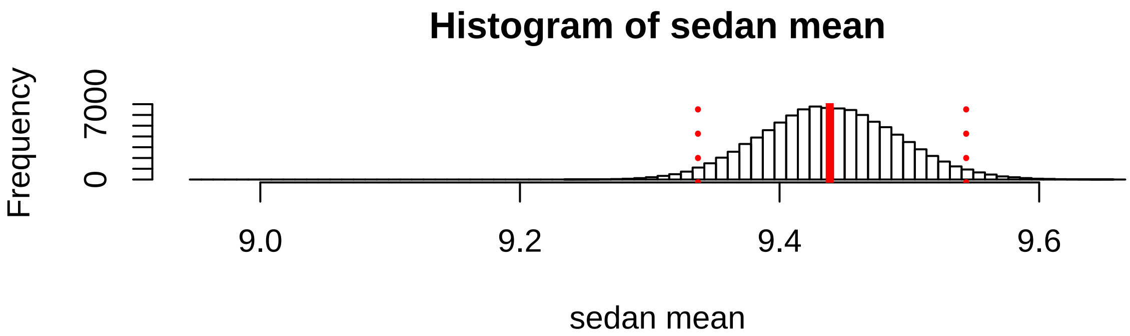
```r
# Get price for each body style
hatch_price <- filter(auto, body.style == 'hatchback') %>% pull('logPrice')
sedan_price <- filter(auto, body.style == 'sedan') %>% pull('logPrice')
wagon_price <- filter(auto, body.style == 'wagon') %>% pull('logPrice')
mean.boot.hatch = one.boot(hatch_price, mean, R = 100000)
mean.boot.sedan = one.boot(sedan_price, mean, R = 100000)
mean.boot.wagon = one.boot(wagon_price, mean, R = 100000)
```

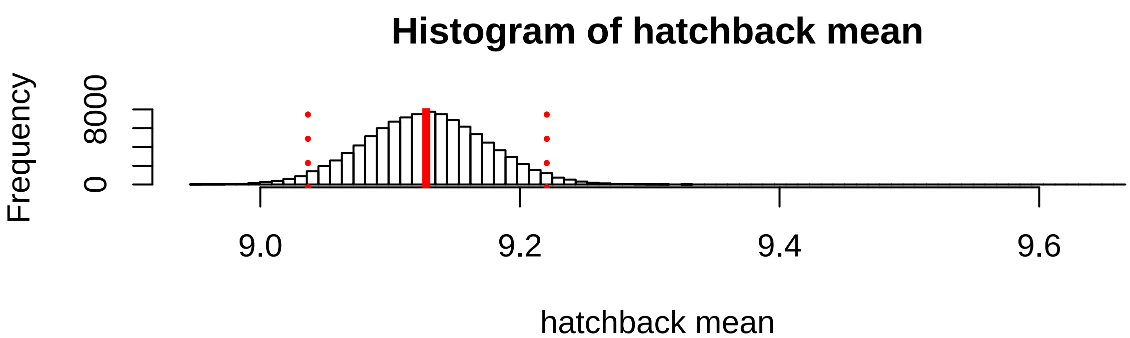
Now, we can do pairwise comparisons between the three groups.

Sedan versus Hatchback

First, let's look at sedans versus hatchbacks and plot the bootstrapped means:

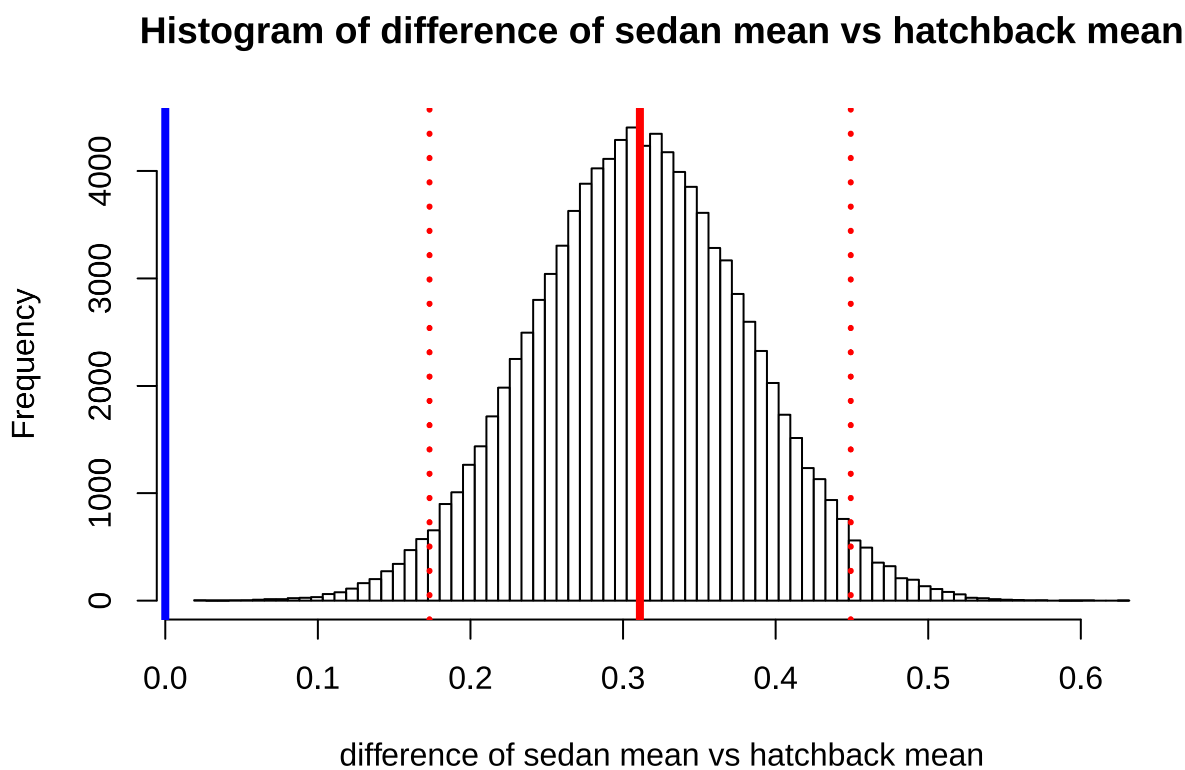
```r
# Bootstrap means
plot.t(mean.boot.sedan$t, mean.boot.hatch$t, cols = c('sedan mean','hatchback mean'))
```





Note that the 95% confidence intervals for the means do not overlap. Let's compute the bootstrapped value for the difference in means and plot the result.

```
# Bootstrap difference of means
mean.boot.sedan_hatch = two.boot(sedan_price, hatch_price, mean, R = 100000)
plot.diff(mean.boot.sedan_hatch$t,  'difference of sedan mean vs hatchback mean')
```



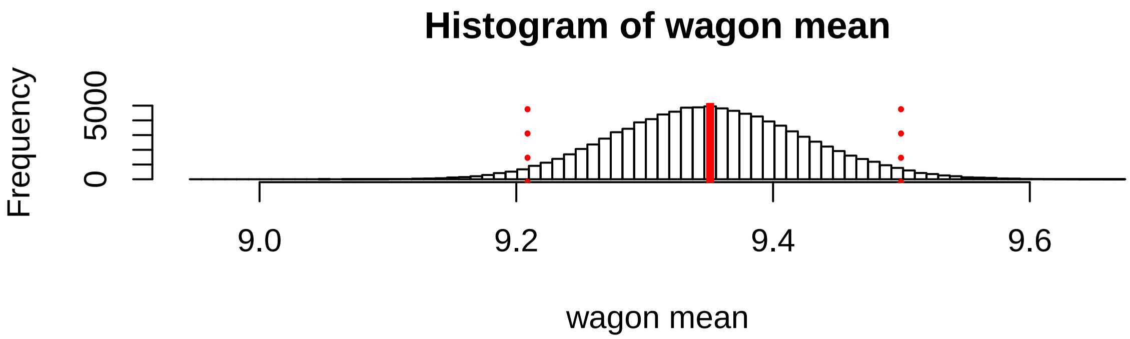
The 95% confidence interval does not include zero, so we can reject the null hypothesis and say that a statistically significant relationship exists between the price of a sedan versus a hatchback.

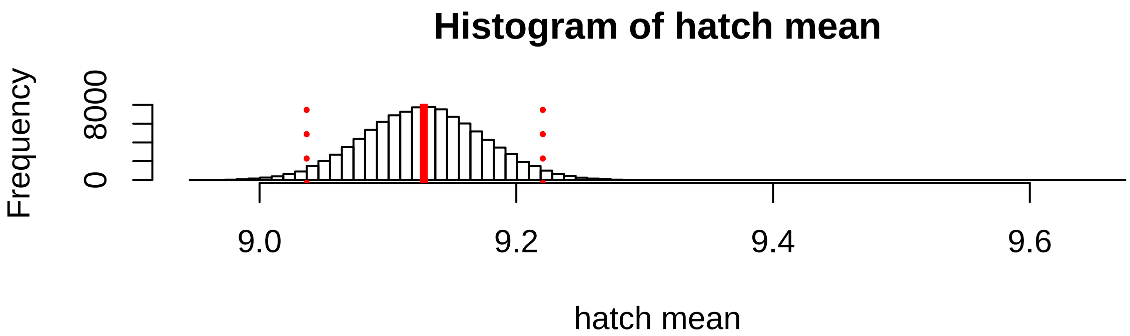
Using Tukey's HSD for this pairwise comparison resulted in a p-value of 0.0000870, which is well below the cutoff of 0.05. In that case, we could reject the null hypothesis, so here the classical and bootstrap methods agree.

Wagon versus Hatchback

Next, let's look at wagons versus hatchbacks and plot the bootstrapped means:

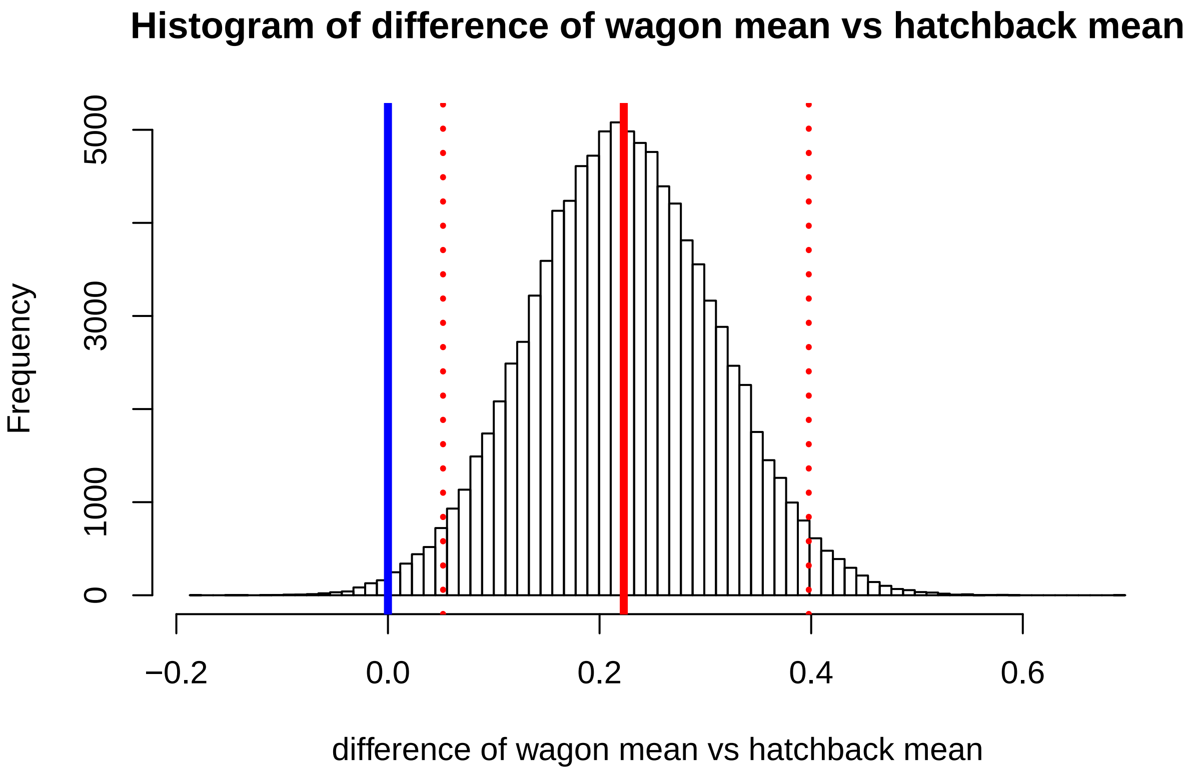
```r
# Bootstrap means
plot.t(mean.boot.wagon$t, mean.boot.hatch$t, cols = c('wagon mean','hatch mean'))
```





Now, the 95% confidence intervals for the means overlap slightly. Let's compute the bootstrapped value for the difference in means and plot the result.

```
# Bootstrap difference of means
mean.boot.wagon_hatch = two.boot(wagon_price, hatch_price, mean, R = 100000)
plot.diff(mean.boot.wagon_hatch$t,  'difference of wagon mean vs hatchback mean')
```



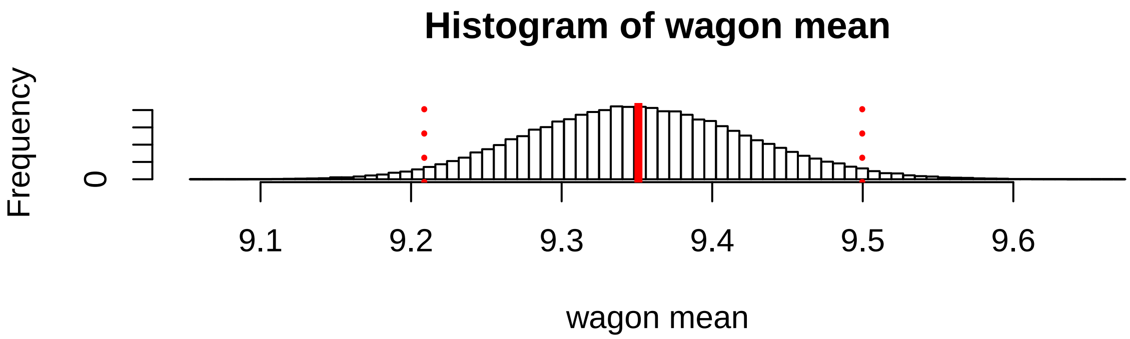
Once again, the 95% confidence interval does not include zero, so we can again reject the null hypothesis and say that a statistically significant relationship exists between the price of a wagon versus a hatchback.

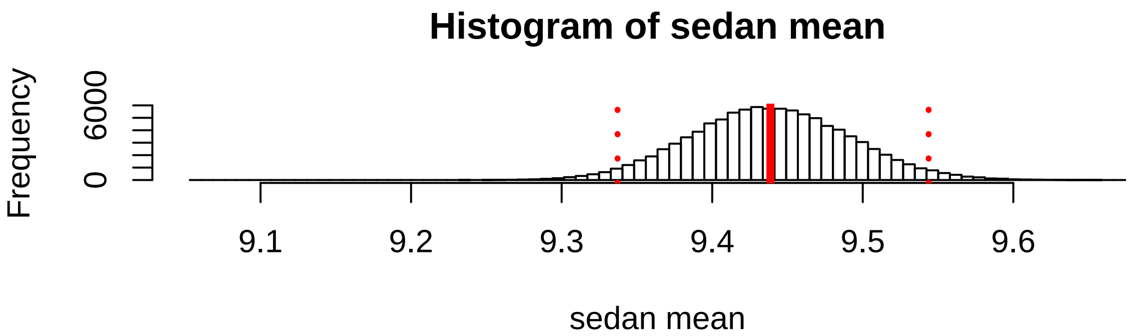
Using Tukey's HSD for this pairwise comparison resulted in a p-value of 0.094, which is above the cutoff of 0.05. In that case, we couldn't reject the null hypothesis, so here the classical and bootstrap methods disagree.

Wagon versus Sedan

Finally, let's look at wagons versus sedans and plot the bootstrapped means:

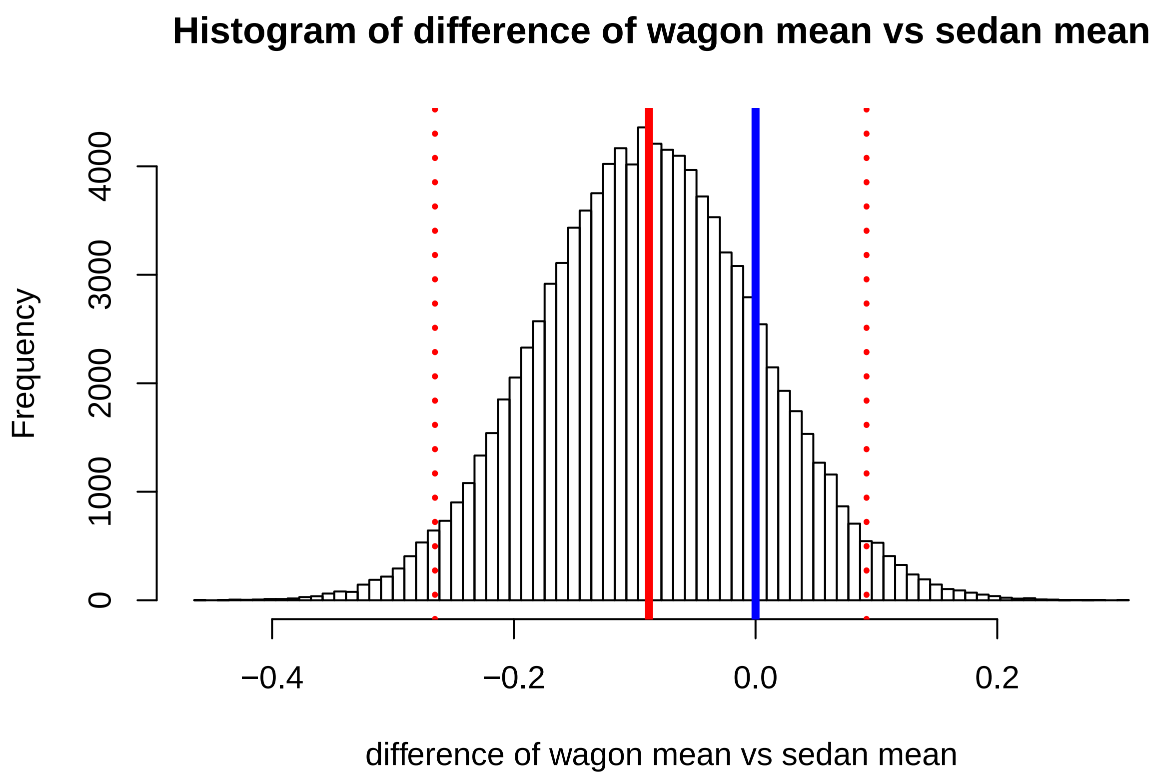
```
# Bootstrap means
plot.t(mean.boot.wagon$t, mean.boot.sedan$t, cols = c('wagon mean','sedan mean'))
```





This time, the 95% confidence intervals for the means overlap significantly. Let's compute the bootstrapped value for the difference in means and plot the result.

```
# Bootstrap difference of means
mean.boot.wagon_sedan = two.boot(wagon_price, sedan_price, mean, R = 100000)
plot.diff(mean.boot.wagon_sedan$t,  'difference of wagon mean vs sedan mean')
```



This time, the 95% confidence interval does include zero, so we cannot reject the null hypothesis and cannot say that a statistically significant relationship exists between the price of a wagon versus a sedan.

Using Tukey's HSD for this pairwise comparison resulted in a p-value of 0.668, which is well above the cutoff of 0.05. In that case, we couldn't reject the null hypothesis, so here the classical and bootstrap methods agree.

Summary and Conclusion

The purpose of this assignment was to compare classical and bootstrap methods for hypothesis testing in statistics. Specifically, we compared Welch's t-test to bootstrap methods for computing the difference in means between two populations (automobile price stratified by fuel type and aspiration). Also, we compared Tukey's HSD to bootstrap methods for computing the difference in means between three populations (automobile price stratified by body style). After comparing these various methods, it was found that:

1. Using bootstrap methods, a statistically significant relationship exists for the price of an automobile when stratified by both aspiration (standard vs. turbo) and fuel type (gas vs. diesel). Using Welch's t-test, a statistically significant relationship only exists when stratified by aspiration.
2. Using bootstrap methods, a statistically significant relationship exists between the price of sedans and hatchbacks, and wagons and hatchbacks, but not between sedans and wagons. Using Tukey's HSD, a statistically significant relationship only exists between the price of sedans and hatchbacks.

In both bivariate and multivariate tests, classical and bootstrap methods produce different results for the automobile data. When the alternative hypothesis for a test was rejected by a close margin using a classical method, it was accepted using a bootstrap method. A possible reason for this is that in the previous write-up, it was found that the distribution of the log of automobile price is right skewed. Welch's t-test and Tukey's HSD assume the populations being tested follow a normal distribution, and perhaps the violation of that assumption is the reason different outcomes were obtained. Non-parametric methods such as boostrapping do not assume any particular distribution.