ST 502 R Project 2

Elizabeth Burke Bruce Campbell, John Davidson

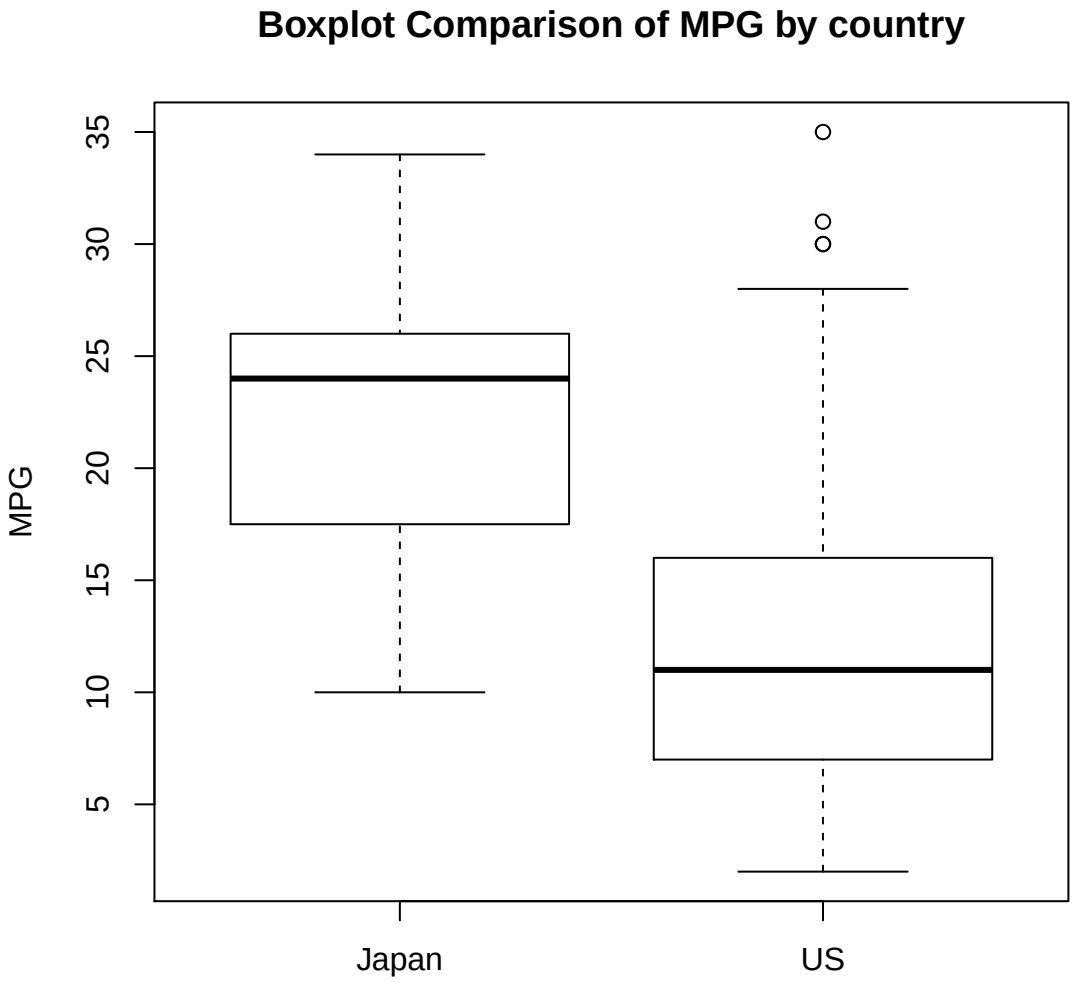
April 14, 2017

This report considers independent samples of miles per gallon measurements of US and Japanese manufactured cars. For this analysis we assume the measurements come from normally distributed populations. To test the hypothesis that means of the two populations are different we will perform two-sample t-tests. We will be performing a test where equal variance is assumed (pooled) and one where unequal variances are assumed.

Plot the data

```
# install.packages('pander') install.packages('plyr')
# install.packages('dplyr') install.packages('readr')
# install.packages('ggplot2')
library(pander)
library(plyr)
library(dplyr)
library(readr)
library(ggplot2)

mpg <- read.table("mpg.txt")
# This makes the variables available in the name space
attach(mpg)
boxplot(MPG ~ Country, main = "Boxplot Comparison of MPG by country", ylab = "MPG")
```



Part 1 - calculation of the confidence intervals

a) Conduct both two-sample t-tests

In this section we calculate the 2 sample t-test on the data at the $\alpha = 0.05$ signicance level.

2-Sample t-test equal variance

```
alpha <- 0.05

pooled.var <- t.test(x = mpg[Country == "Japan", ]$MPG, y = mpg[Country == "US",
    ]$MPG, alternative = "two.sided", var.equal = TRUE, conf.level = alpha)
pooled.var
```

```
##
##  Two Sample t-test
##
## data:  mpg[Country == "Japan", ]$MPG and mpg[Country == "US", ]$MPG
## t = 12.238, df = 326, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 5 percent confidence interval:
##  10.02165 10.12496
## sample estimates:
## mean of x mean of y
##  22.35443  12.28112
```

We see that for the 2 sample t-test with pooled variance the $p - value < 2.2e - 16$ this means there is not enough evidence to support the null hypothesis that the means are the same. We reject the null hypothesis and claim that the evidence supports that the population mean mpg of Japanese and US manufactured cars are different.

2-Sample t-test unequal variances

```
unequal.var <- t.test(x = mpg[Country == "Japan", ]$MPG, y = mpg[Country ==
    "US", ]$MPG, alternative = "two.sided", var.equal = FALSE, conf.level = alpha)
unequal.var
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg[Country == "Japan", ]$MPG and mpg[Country == "US", ]$MPG
## t = 12.99, df = 145.45, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 5 percent confidence interval:
##  10.02459 10.12202
## sample estimates:
## mean of x mean of y
##  22.35443  12.28112
```

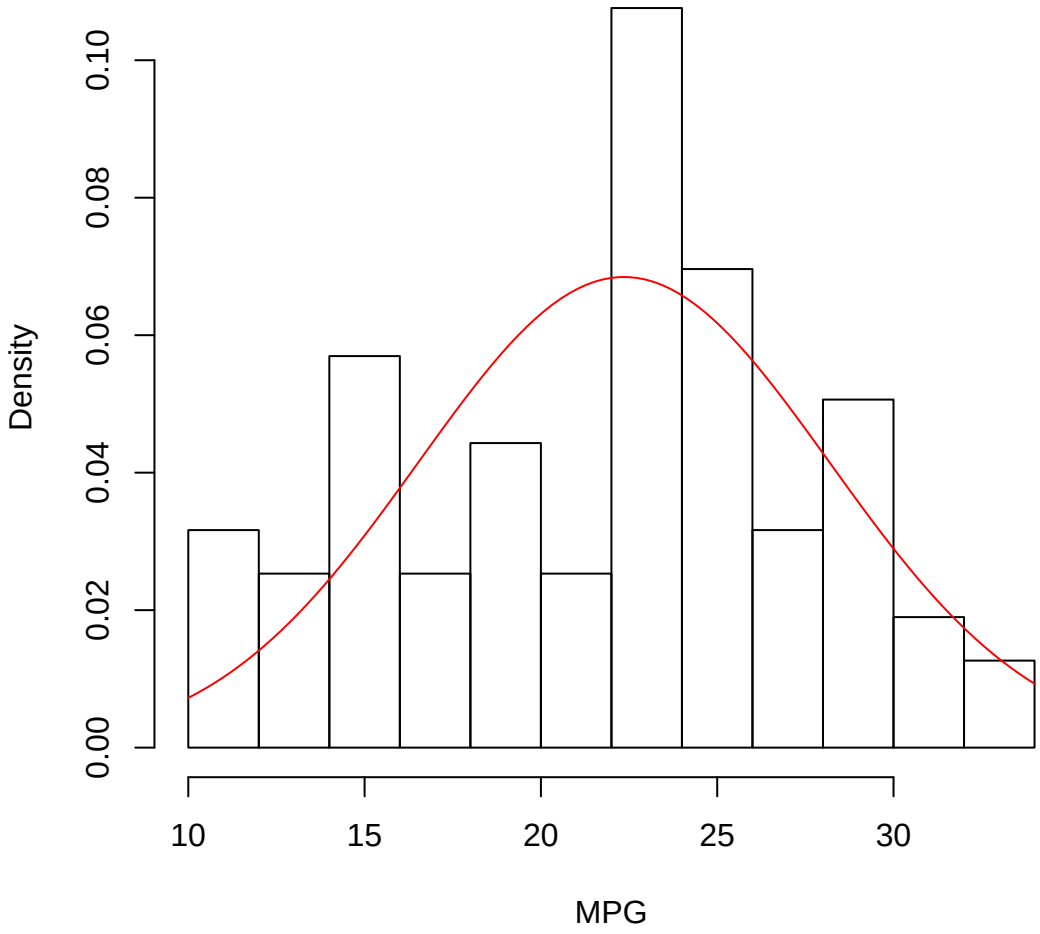
We see that for the 2 sample t-test with unequal variances assumet that the $p - value < 2.2e - 16$ this means there is not enough evidence to support the null hypothesis that the means are the same. We reject the null hypothesis and claim that the evidence supports that the population mean mpg of Japanese and US manufactured cars are different.

Check for normality

Here we plot the

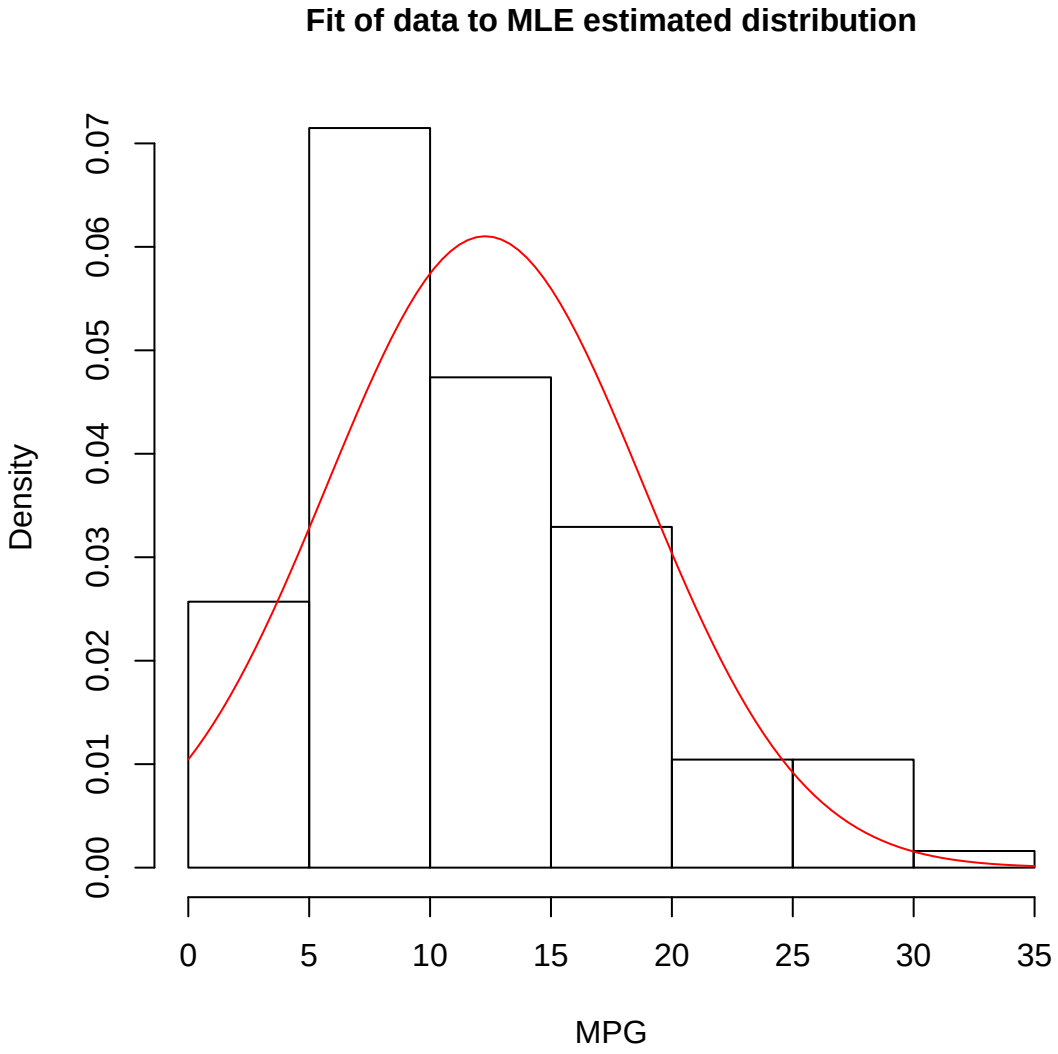
```
Japan <- mpg[Country == "Japan", ]$MPG
Japan.MLE.mean <- mean(Japan)
Japan.MLE.SD <- sd(Japan)

hist(Japan, 10, freq = FALSE, cex.main = 1, main = "Fit of data to MLE estimated distribution",
    xlab = "MPG")
curve(dnorm(x, Japan.MLE.mean, Japan.MLE.SD), add = TRUE, col = "red")
```




```
US <- mpg[Country == "US", ]$MPG
US.MLE.mean <- mean(US)
US.MLE.SD <- sd(US)

hist(US, 10, freq = FALSE, cex.main = 1, main = "Fit of data to MLE estimated distribution",
    xlab = "MPG")
curve(dnorm(x, US.MLE.mean, US.MLE.SD), add = TRUE, col = "red")
```



We see the US MPG data is right skewed.