

# Peer Assessment on Regression Models - Data Science Specialization

**Michael Karp**

**1/22/2015**

## Executive Summary

In this project I will investigate the mtcars dataset that comes with R. As noted in the data dictionary for this problem “The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).”

I will address the following points: “Is an automatic or manual transmission better for MPG?” - manual is better

“Quantify the MPG difference between automatic and manual transmissions” - on average a car with manual transmission will get 7.245 mpg

```
# always set seed to make research reproducible
set.seed(1000)
# inspect dataset
str(mtcars)
```

‘data.frame’: 32 obs. of 11 variables: \$ mpg : num 21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ... \$ cyl : num 6 6 4 6 8 6 8 4 4 6 ... \$ disp: num 160 160 108 258 360 ... \$ hp : num 110 110 93 110 175 105 245 62 95 123 ... \$ drat: num 3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ... \$ wt : num 2.62 2.88 2.32 3.21 3.44 ... \$ qsec: num 16.5 17 18.6 19.4 17 ... \$ vs : num 0 0 1 1 0 1 0 1 1 1 ... \$ am : num 1 1 1 0 0 0 0 0 0 0 ... \$ gear: num 4 4 4 3 3 3 3 4 4 4 ... \$ carb: num 4 4 1 1 2 1 4 2 2 4 ...

```
summary(mtcars)
```

mpg

cyl

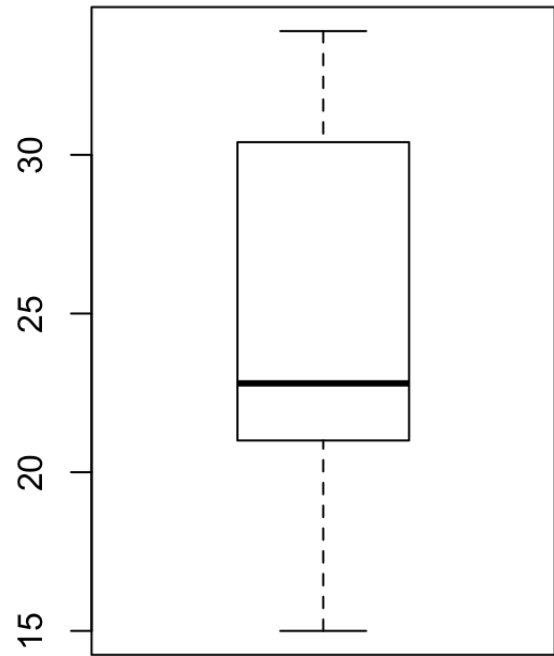
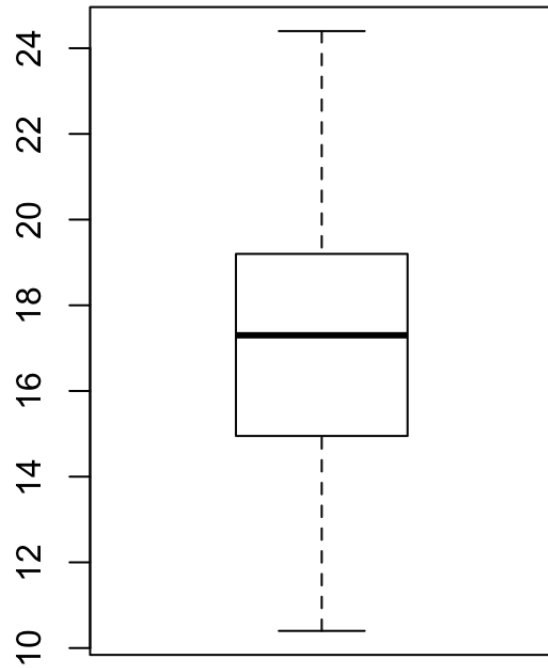
disp

hp

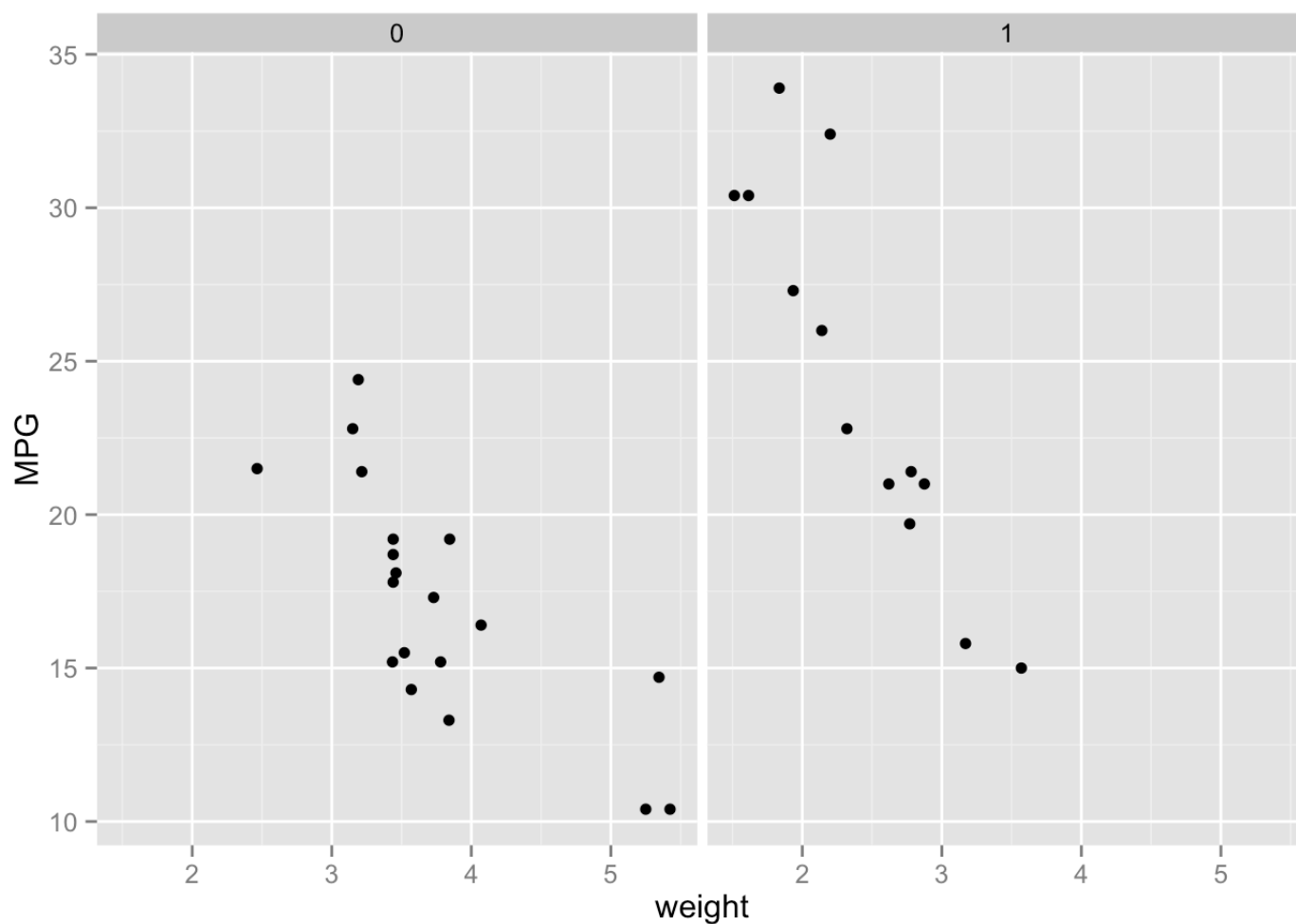
Min. :10.40 Min. :4.000 Min. : 71.1 Min. : 52.0  
 1st Qu.:15.43 1st Qu.:4.000 1st Qu.:120.8 1st Qu.: 96.5  
 Median :19.20 Median :6.000 Median :196.3 Median :123.0  
 Mean :20.09 Mean :6.188 Mean :230.7 Mean :146.7  
 3rd Qu.:22.80 3rd Qu.:8.000 3rd Qu.:326.0 3rd Qu.:180.0

Max. :33.90 Max. :8.000 Max. :472.0 Max. :335.0  
drat wt qsec vs  
Min. :2.760 Min. :1.513 Min. :14.50 Min. :0.0000  
1st Qu.:3.080 1st Qu.:2.581 1st Qu.:16.89 1st Qu.:0.0000  
Median :3.695 Median :3.325 Median :17.71 Median :0.0000  
Mean :3.597 Mean :3.217 Mean :17.85 Mean :0.4375  
3rd Qu.:3.920 3rd Qu.:3.610 3rd Qu.:18.90 3rd Qu.:1.0000  
Max. :4.930 Max. :5.424 Max. :22.90 Max. :1.0000  
am gear carb  
Min. :0.0000 Min. :3.000 Min. :1.000  
1st Qu.:0.0000 1st Qu.:3.000 1st Qu.:2.000  
Median :0.0000 Median :4.000 Median :2.000  
Mean :0.4062 Mean :3.688 Mean :2.812  
3rd Qu.:1.0000 3rd Qu.:4.000 3rd Qu.:4.000  
Max. :1.0000 Max. :5.000 Max. :8.000

```
par(mfrow = c(1,2))  
# boxplot for automatic transmission  
boxplot(mtcars$mpg[mtcars$am == 0])  
# boxplot for manual transmission  
boxplot(mtcars$mpg[mtcars$am == 1])
```



```
par(mfrow = c(1,1))  
library(ggplot2)  
mpg_trans_plot <- ggplot(data = mtcars, aes(x = wt, y = mpg)) + geom_point() + xlab("weight  
") + ylab("MPG") + facet_grid(.~am)  
mpg_trans_plot
```



```
# clearly lower weight manual transmission cars have the highest mpg ratio
```

```
mpg_fit <- lm(data = mtcars, formula = mpg ~ am)
summary(mpg_fit)
```

Call: `lm(formula = mpg ~ am, data = mtcars)`

Residuals: Min 1Q Median 3Q Max -9.3923 -3.0923 -0.2974 3.2439 9.5077

Coefficients: Estimate Std. Error t value Pr(>|t|)

(Intercept) 17.147 1.125 15.247 1.13e-15 **am 7.245 1.764 4.106 0.000285** — Signif. codes: 0 “**0.001**” 0.01  
” 0.05 “.” 0.1 “ ” 1

Residual standard error: 4.902 on 30 degrees of freedom Multiple R-squared: 0.3598, Adjusted R-squared:  
0.3385 F-statistic: 16.86 on 1 and 30 DF, p-value: 0.000285

```
# in quantifying the difference between manual and automatic transmission I will assess the  
mean, median, and standard  
# deviations for each of two optins  
  
tapply(mtcars$mpg, mtcars$am, median)
```

0 1 17.3 22.8

```
tapply(mtcars$mpg, mtcars$am, mean)
```

0 1

17.14737 24.39231

```
tapply(mtcars$mpg, mtcars$am, sd)
```

0 1

3.833966 6.166504