Regression Models Course Project

Tomasz Dworowy

```
knitr::opts_chunk$set(warning=FALSE, message=FALSE)
```

```
if (!require("ggplot2"))
  install.packages("ggplot2")
library(ggplot2)
```

```
data("mtcars")
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$vs <- as.factor(mtcars$vs)
mtcars$am <- factor(mtcars$am,labels = c("automatic","manual"))
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
```

Summary

Goal of this analysis is to explore relationship between miles per gallon and transition type (automatic and manual). The `mtcars` data set was used,

Exploratory analisis

Data summary

Data summary:

```
summary(mtcars)
```

```
##       mpg        cyl         disp             hp             drat
##  Min.   :10.40   4:11   Min.   : 71.1   Min.   : 52.0   Min.   :2.760
##  1st Qu.:15.43   6: 7   1st Qu.:120.8   1st Qu.: 96.5   1st Qu.:3.080
##  Median :19.20   8:14   Median :196.3   Median :123.0   Median :3.695
##  Mean   :20.09          Mean   :230.7   Mean   :146.7   Mean   :3.597
##  3rd Qu.:22.80          3rd Qu.:326.0   3rd Qu.:180.0   3rd Qu.:3.920
##  Max.   :33.90          Max.   :472.0   Max.   :335.0   Max.   :4.930
##        wt             qsec       vs             am     gear   carb
##  Min.   :1.513   Min.   :14.50   0:18   automatic:19   3:15   1: 7
##  1st Qu.:2.581   1st Qu.:16.89   1:14   manual   :13   4:12   2:10
##  Median :3.325   Median :17.71                         5: 5   3: 3
##  Mean   :3.217   Mean   :17.85                                4:10
##  3rd Qu.:3.610   3rd Qu.:18.90                                6: 1
##  Max.   :5.424   Max.   :22.90                                8: 1
```

```
auto_mpg_mean <- mean(subset(mtcars, am==0)$mpg)
manual_mpg_mean <- mean(subset(mtcars, am==1)$mpg)
```

Mean miles per gallon for automatic transmission is NaN and for manual transmission is NaN. So at first glance we can see that mqp for manual transmission seems to be better, the **Appendix Figure I** and **Appendix Figure II** seem to confirm the same observations.

Statistical Inference

T-Test transmission type and MPG

```
results <- t.test(mpg ~ am, data = mtcars)
p_value <- results$p.value
```

P-value 0.0013736 is smaller than 0.5 so we should rejects the null hypothesis, so the difference between in mpg between transmissions is statistically significant.

```
estimate <- results$estimate
```

The difference estimate between the 2 transmissions is 7.2449393 mpg better for manual.

Regression Analysis

Fit the full model of the data

```
full_fit <- glm(mpg ~ ., data = mtcars)
summary(full_fit)$coeff
```

```
##                Estimate  Std. Error     t value   Pr(>|t|)
## (Intercept) 23.87913244 20.06582026  1.19004018 0.25252548
## cyl6        -2.64869528  3.04089041 -0.87102622 0.39746642
## cyl8        -0.33616298  7.15953951 -0.04695316 0.96317000
## disp         0.03554632  0.03189920  1.11433290 0.28267339
## hp          -0.07050683  0.03942556 -1.78835344 0.09393155
## drat         1.18283018  2.48348458  0.47627845 0.64073922
## wt          -4.52977584  2.53874584 -1.78425732 0.09461859
## qsec         0.36784482  0.93539569  0.39325050 0.69966720
## vs1          1.93085054  2.87125777  0.67247551 0.51150791
## ammanual     1.21211570  3.21354514  0.37718957 0.71131573
## gear4        1.11435494  3.79951726  0.29328856 0.77332027
## gear5        2.52839599  3.73635801  0.67670068 0.50889747
## carb2       -0.97935432  2.31797446 -0.42250436 0.67865093
## carb3        2.99963875  4.29354611  0.69863900 0.49546781
## carb4        1.09142288  4.44961992  0.24528452 0.80956031
## carb6        4.47756921  6.38406242  0.70136677 0.49381268
## carb8        7.25041126  8.36056638  0.86721532 0.39948495
```

See **Appendix Figure III** for residuals plot.

Fit model only with transition type and mpg.

```
min_fit <- glm(mpg ~ am, data = mtcars)
summary(min_fit)$coeff
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## ammanual     7.244939   1.764422  4.106127 2.850207e-04
```

See **Appendix Figure IV** for residuals plot.

Conclusion

Difference in MPG based on transmission type is significant. At least in this data set cars with manual transition have slightly better miles per gallon usage. However, weight, horsepower and number of cylinders have significant influence to.

Appendix Figures

```
ggplot(aes(x = am, y=mpg ),data=mtcars) +
  geom_boxplot(aes(fill=am))
```

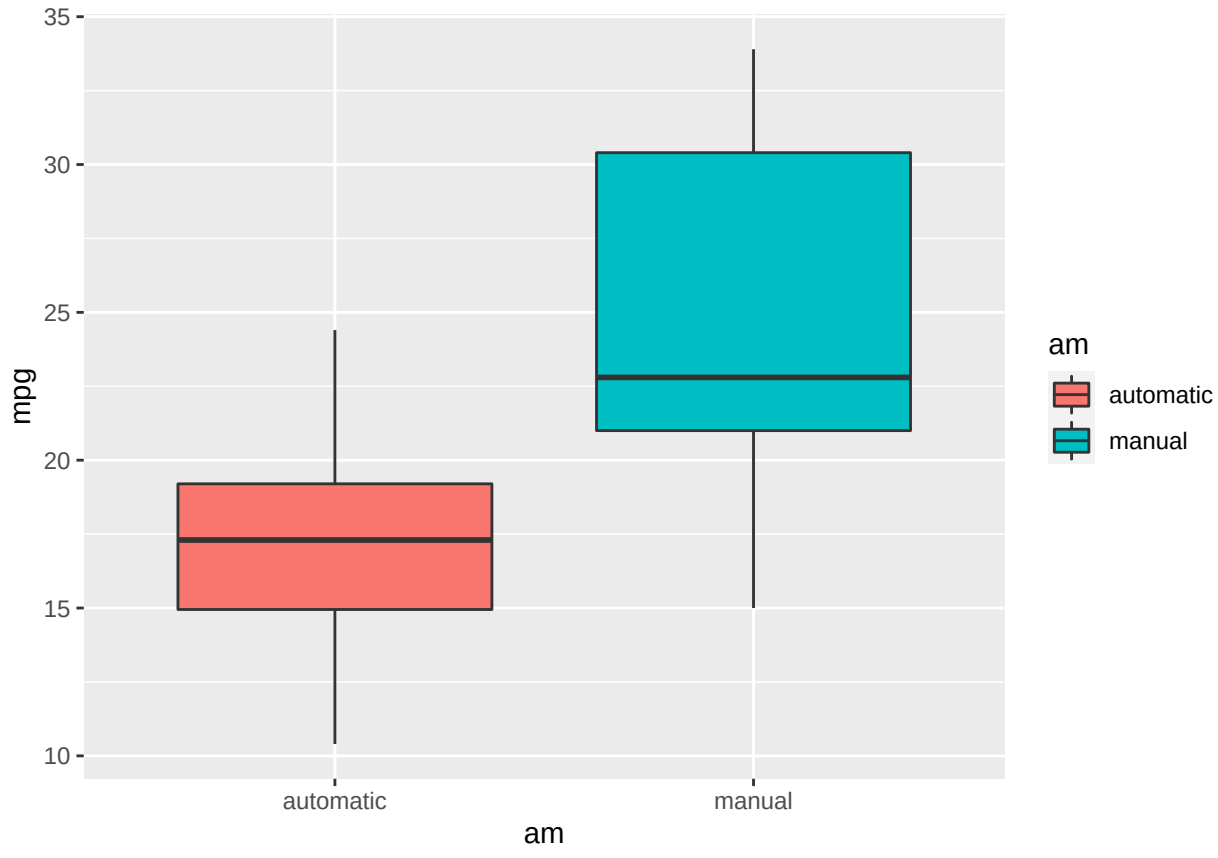

Figure I

```
ggplot(aes(x = mpg, y=am ),data=mtcars) +
  geom_point(aes(color=am))
```

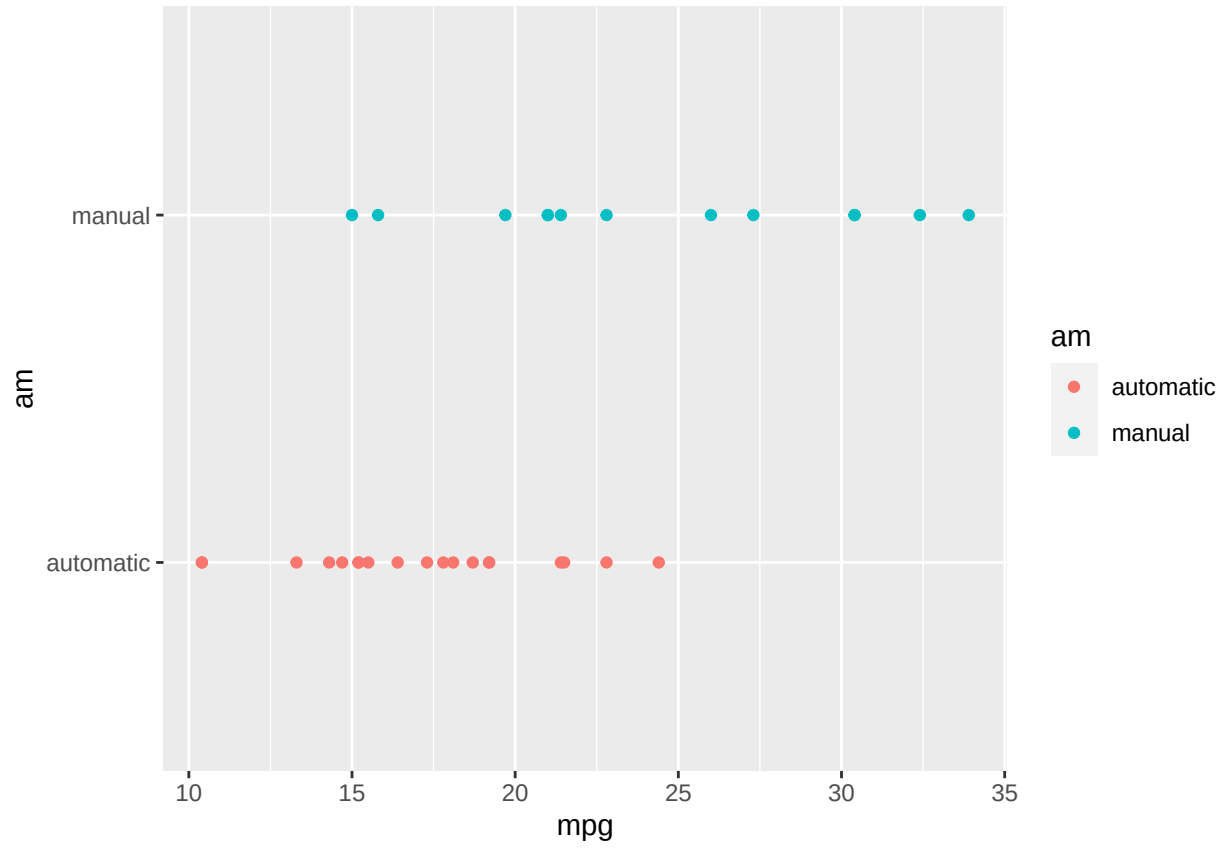

Figure II

```
par(mfrow=c(2,2))
plot(full_fit)
```

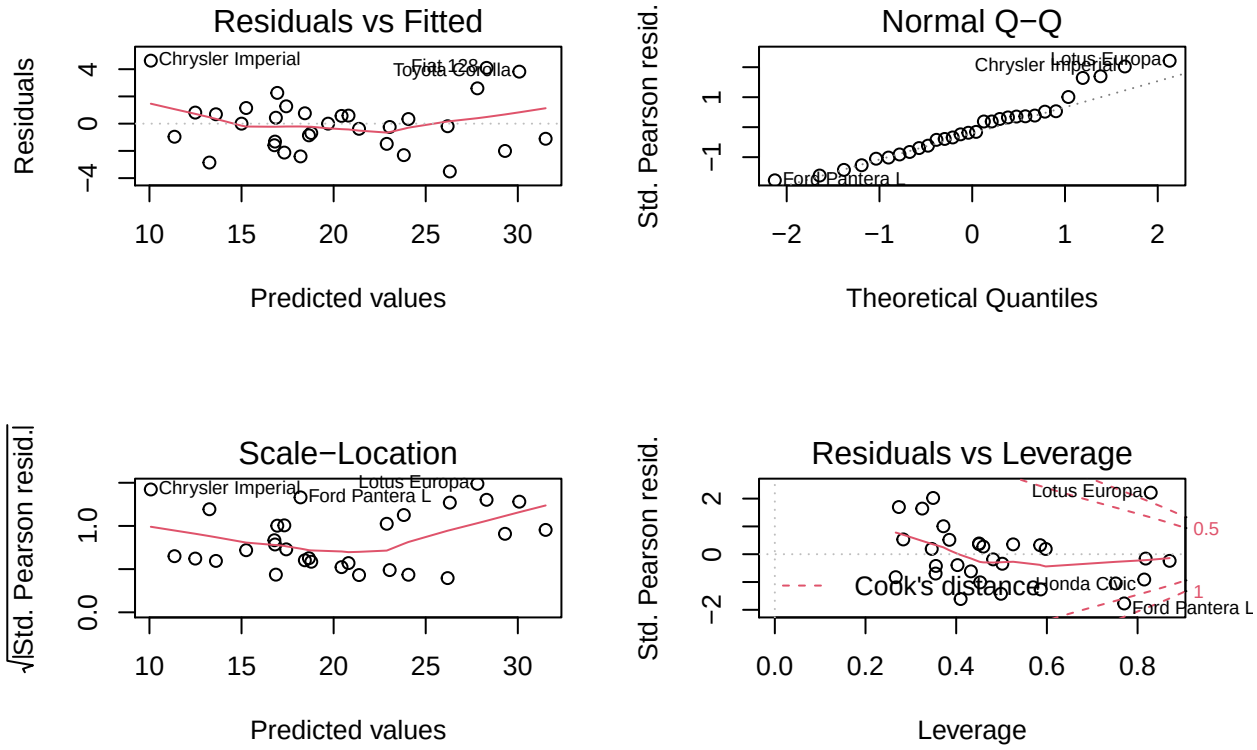

Figure III

```
par(mfrow=c(2,2))
plot(min_fit)
```

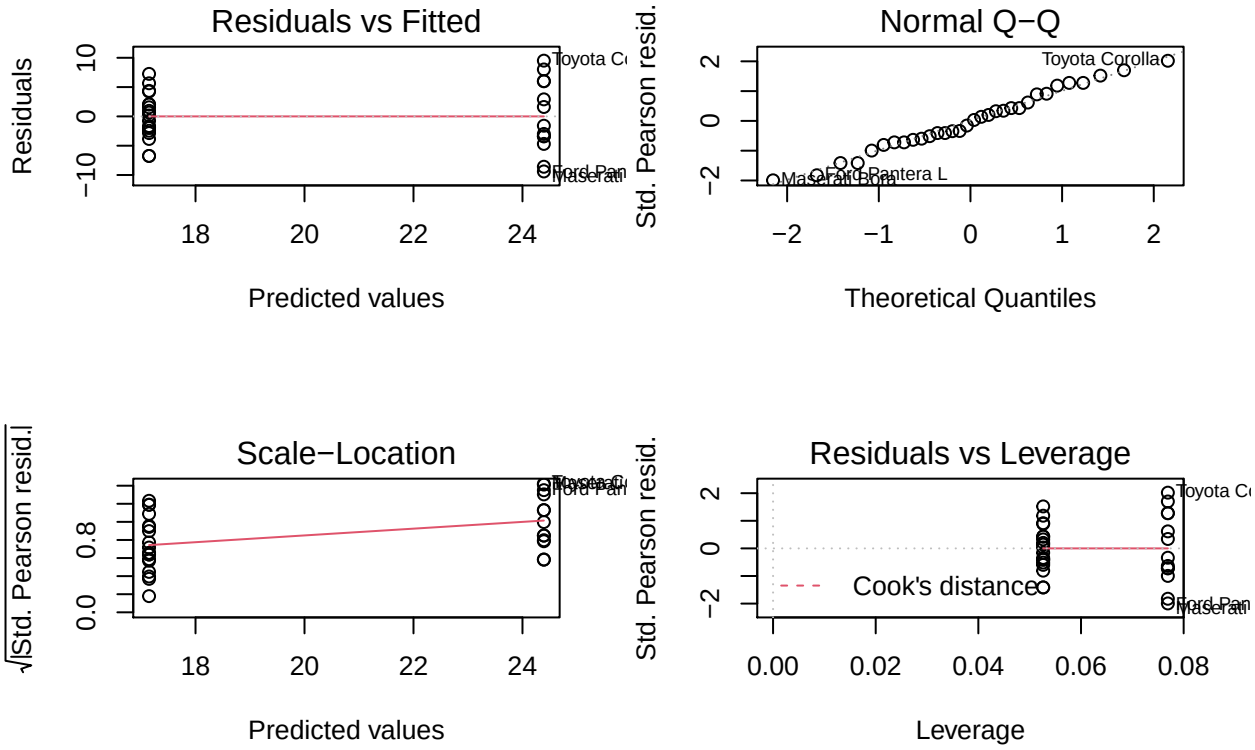
Residuals vs Fitted
Residuals
Predicted values
Toyota Co
Ford Pan
Maserati
Normal Q−Q
Std. Pearson resid.
Theoretical Quantiles
Toyota Corolla
Ford Pantera L
Maserati Bora
Scale−Location
√|Std. Pearson resid.|
Predicted values
Residuals vs Leverage
Std. Pearson resid.
Cook's distance
Leverage
Toyota Co
Ford Pan
Maserati

Figure IV