

# Regression Models - Course Project

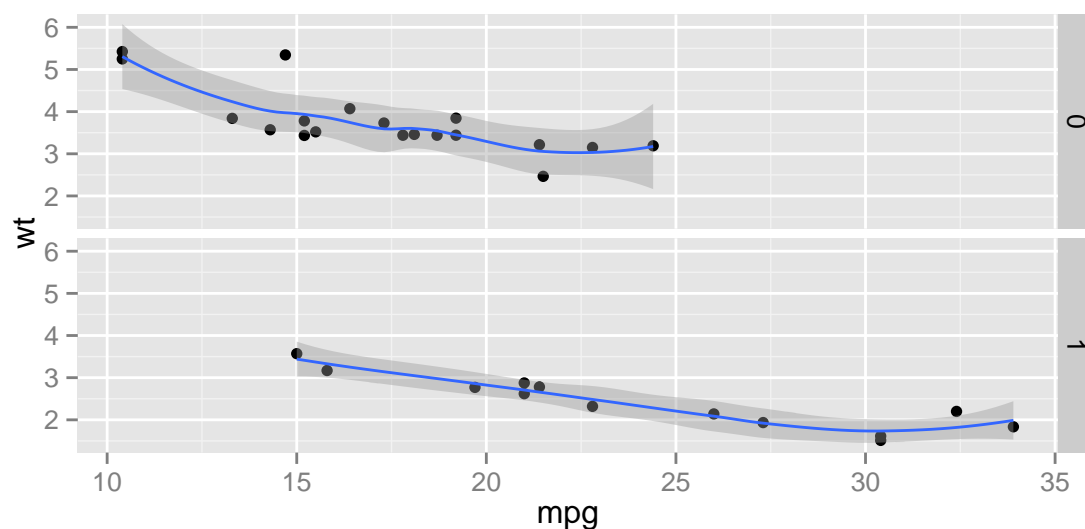
Gurmeet Singh

## Executive Summary

This report attempts to answer the question “Is a car with automatic transmission perform better on MPG compared to a manual one?” Although a simpler analysis might point to a “Yes,” further analysis shows that *there is no significant effect of presence or absence of an automatic transmission on the MPG of a car* withing the `mtcars` dataset. The average MPG of cars with automatic and manual transmissions is indeed significantly different but the difference is explained by the fact that cars with manual transmission in the dataset get lower MPG presumably due to them being heavier and not because they have manual transmissions.

## Coefficients

Let’s plot the MPG of the cars as function of their weight with manual (`am=0`) and auto (`am=1`) transmissions:



Looks like the cars with automatic transmission have higher MPG, but they also weigh less on an average. This is confirmed analytically in the appendix. Now let’s fit a linear model between ‘`am`’ as the predictor and ‘`mpg`’ as the outcome.

```
summary(lm(mpg ~ am,data=mtcars))
```

```
##
## 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
```

The ‘intercept’ and ‘am’ intercept are respectively, the mean MPG (17.147) of cars with manual and the increase (7.245) in MPG by moving to an automatic transmission. We can see that there is a *significant* (**>95% CI**) relationship between the two (am and mpg) indicated by the ‘\*\*\*’. The estimate column confirms the MPG means we computed in the appendix for cars with manual and automatic transmissions.

## Diagnostics

The residual plots for the fitted model are in Figure A1 in the appendix. There doesn’t seem to be any obvious pattern in the residual plots, but a correlation between residuals and weight of the cars suggests that the model may improve by adding ‘wt’ as a predictor (See appendix).

The  $R^2$  value indicates that about 33.3% of the variance in MPG is explained by the ‘am’ variable.

## Model Selection using ANOVA

Investigate if additional predictors improve model accuracy.

```
fit1 <- lm(formula = mpg ~ wt, data = mtcars)
fit2 <- lm(formula = mpg ~ wt + am, data = mtcars)
fit3 <- lm(formula = mpg ~ wt + am + cyl, data = mtcars)
anova(fit1,fit2,fit3)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt
## Model 2: mpg ~ wt + am
## Model 3: mpg ~ wt + am + cyl
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      30 278.32
## 2      29 278.32  1      0.002 0.0003 0.985671
## 3      28 191.05  1     87.273 12.7907 0.001292 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Both from anova and summary(fit3) (see appendix) outputs, we can quantifiably say that ‘am’ is not a significant predictor of MPG (in presence of ‘wt’ predictor), whereas the ‘wt’ **is** a significant predictor. In the summary output, the p-value for ‘am’ is 0.3142 > 0.05 and in the anova output, the Model2 (Pr>F) = 0.985 > 0.05 indicating that adding ‘am’ to the predictors didn’t significantly improve the model prediction.

## Appendix

### Content Creation

The content of this report is authored in RStudio using R Markdown format and converted to PDF format using the **knitr** package. The R Markdown file itself can be found on [GitHub](#)

The report can be generated by copying the “Project.Rmd” file into the working directory of an R session and running these commands:

```
library(knitr)
knit2html("Project.Rmd")
```

### Mean MPG of manual and automatic cars

Let’s calculate the mean MPG and weight of manual and auto cars:

```
#Mean MPG of cars with manual transmission
manualcars <- mtcars[mtcars$am == 0,]
mean(manualcars$mpg)
```

```
## [1] 17.14737
```

```
mean(manualcars$wt)
```

```
## [1] 3.768895
```

```
#Mean MPG of cars with automatic transmission
autocars <- mtcars[mtcars$am == 1,]
mean(autocars$mpg)
```

```
## [1] 24.39231
```

```
mean(autocars$wt)
```

```
## [1] 2.411
```

### T-test for comparing mean MPG for cars with auto and manual transmissions

We can quantify the MPG data with a t-test:

```
t.test(manualcars$mpg, autocars$mpg, var.equal = T)
```

```
##
## Two Sample t-test
##
## data: manualcars$mpg and autocars$mpg
## t = -4.1061, df = 30, p-value = 0.000285
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

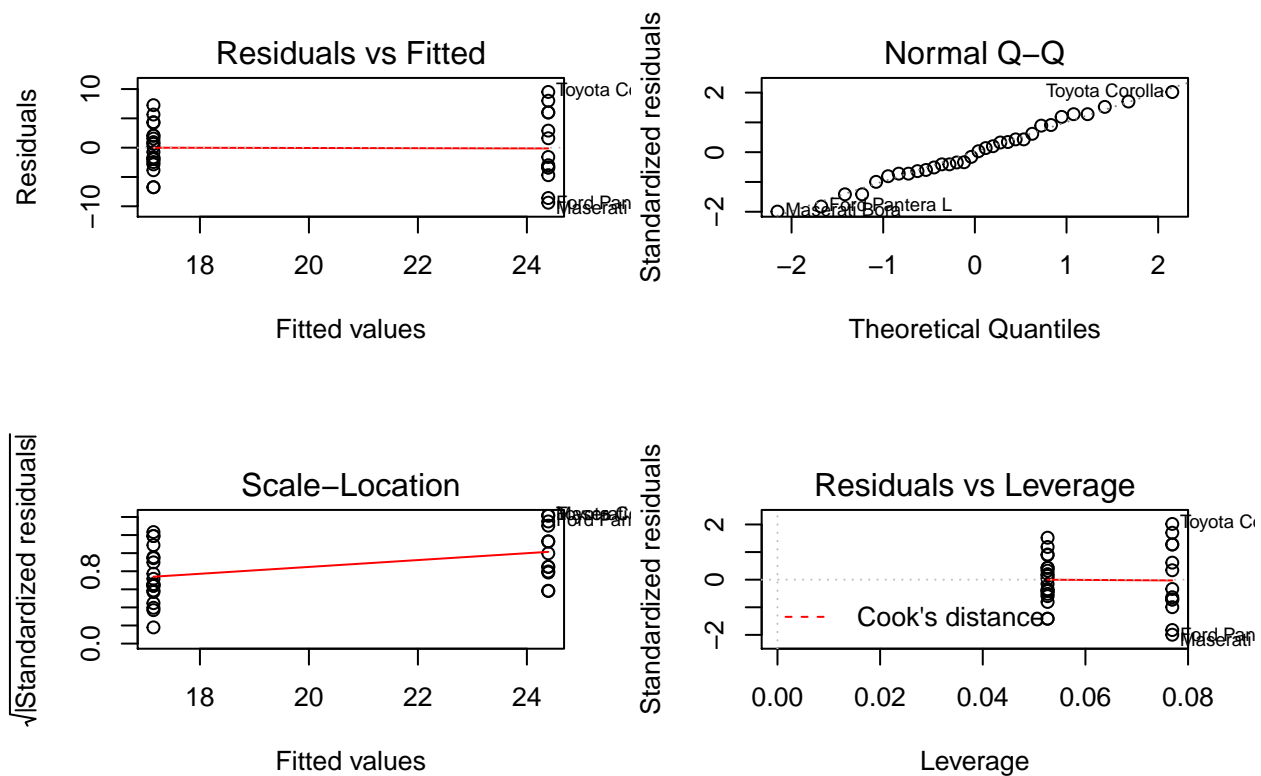
```
## -10.84837 -3.64151
## sample estimates:
## mean of x mean of y
## 17.14737 24.39231
```

The difference is significant with  $> 95\%$  confidence since the p-value  $0.000285 < 0.05$ . Also, the p-value for 'am' matches the p-value we obtained earlier with the  $(\text{mpg} \sim \text{am})$  model.

Plotting the fitted model: residuals etc..

```
fit <- lm(mpg ~ am, data=mtcars)
par(mfrow=c(2,2))
plot(fit)
title("Figure A1", outer=T, line=-1)
```

**Figure A1**

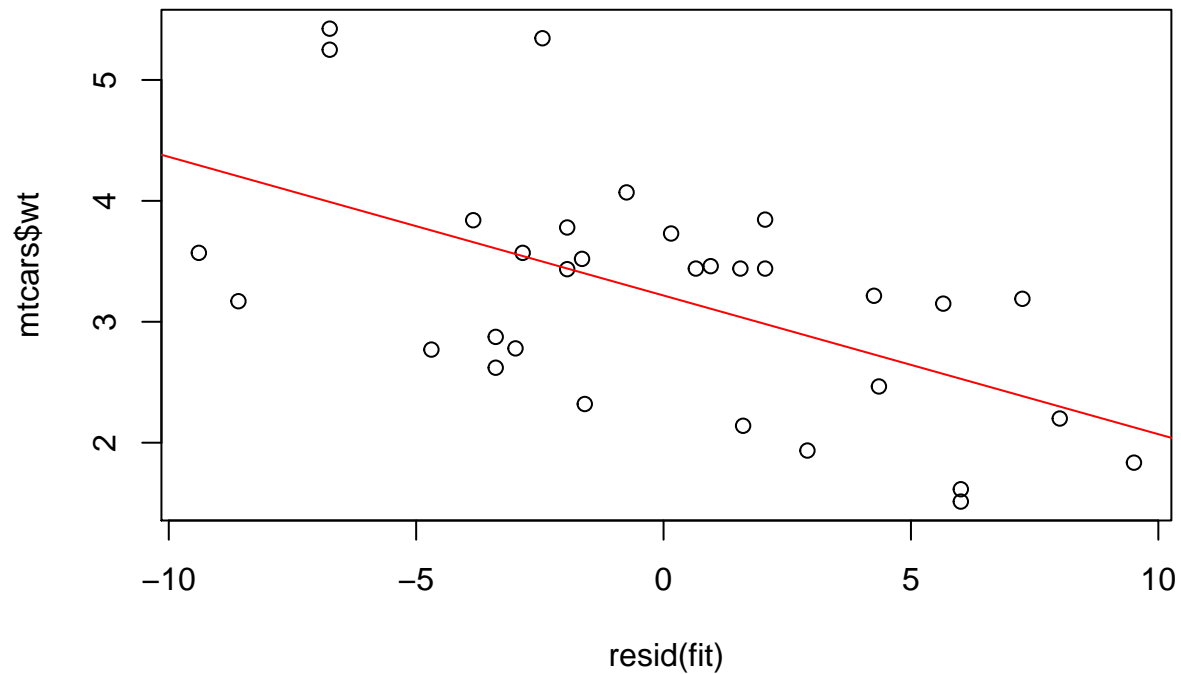


Plot model residuals against 'wt': significant correlation shows that residuals may decrease by addition of 'wt' predictor to the model.

```
cor(resid(fit), mtcars$wt)
```

```
## [1] -0.565259
```

```
plot(resid(fit), mtcars$wt)
abline(lm(mtcars$wt ~ resid(fit)), col=2)
```



```
summary(fit3)
```

```
##
## Call:
## lm(formula = mpg ~ wt + am + cyl, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.1735 -1.5340 -0.5386  1.5864  6.0812
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  39.4179     2.6415   14.923 7.42e-15 ***
## wt          -3.1251     0.9109   -3.431 0.00189 **
## am           0.1765     1.3045    0.135 0.89334
## cyl         -1.5102     0.4223   -3.576 0.00129 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.612 on 28 degrees of freedom
## Multiple R-squared:  0.8303, Adjusted R-squared:  0.8122
## F-statistic: 45.68 on 3 and 28 DF,  p-value: 6.51e-11
```