# The reationship between mpg and transmission type

## Introduction (executive summary)

This assignment calls for an analysis of the “mtcars” dataset to find out the relationship between a set of variables and miles per gallon (mpg). Specifically, whether different types of transmission (“automatic or manual) affect mpg of a car; if yes, how much effect it has. After multiple linear regression analysis, it seems the type of transmission indeed affects the mpg, but the most critical factor for mpg is the weight of the vehicle. ### Data preprocessing By loading and taking a look at the dataset, one can see that the data are very clean already–there are only 32 observations and 11 variables. No pre-processing is needed.

## Data Analsysis

The simplest way to look at the relationship between mpg and transmission type could be: run a linear regression (mpg~am) and see whether the coefficient of the ‘am’ is significantly from zero:

```
fit1 <- lm(mpg ~ am, data = mtcars)
anova(fit1)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq F value  Pr(>F)
## am         1    405     405    16.9 0.00029 ***
## Residuals 30    721      24
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA of the linear model shows that this simple linear regression coefficent is significantly different from 0. Manual transmission (am=1) is better for mpg (7.2) than automatic transmission (am=0).

However, given this dataset is so small, we are not sure how strong this effect could be. The other important factor to consider is that there are several other variables which could affect mpg. It is entirely possible that some other factors are more critical for mpg. This could be illustrated by plotting the residuals of this model using different color for different vehicle weights (See figure 1 in appendix).

Therefore, in the next model, all available predictors were included. The ANOVA showed that this is indeed a better model. The comparison of residuals is plotted in figure 2, Appendix.

```
fit2 <- glm(mpg~., data = mtcars)
anova(fit1,fit2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##   Res.Df RSS Df Sum of Sq    F  Pr(>F)
## 1     30 721
## 2     21 147  9       573 9.07 1.8e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The second model clearly did a better job at predicting the mpg of cars – the residuals are much smaller. Also comparing with the first model, the coefficent of “am” variable in the second model is much smaller – 2.5 versus 7.2 in the first model. This shows that once considering other variables, the type of transimission plays a less critical role in determing a car’s mpg.

Lastly, not all variables might be necessary for deciding the mpg of cars. Therefore, we use back elimination to remove insignificant variables.

```
step(fit2,direction="backward",test="F"); #the output is too long. suppressed it.
```

From above results we found out that only three coefficients are significant: that of weight (“wt”), 1/4 mile time (“qsec”) and type of transmission (“am”). The type of transmission does affect the mpg a bit – with manual transmission cars have a slightly higher mpg (+2.9). So in the last model we only included these three variables.

```
fit3 <- glm(mpg ~ am + qsec + wt, data = mtcars)
anova(fit1,fit2,fit3)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
## Model 3: mpg ~ am + qsec + wt
##   Res.Df RSS Df Sum of Sq    F  Pr(>F)
## 1     30 721
## 2     21 147  9       573 9.07 1.8e-05 ***
## 3     28 169 -7       -22 0.44    0.86
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

# Conclusion

With the regression analysis, it seems that the type of transmission does play some role in predicting the mpg of cars. Manual transmission has a slightly higher mpg (+2.9) than automatic transmission. However, the most significant factor in determing a car’s mpg is its weight.

# Appendix

Figure1: the residuals of the simplest model (mpg~am) for different weight:

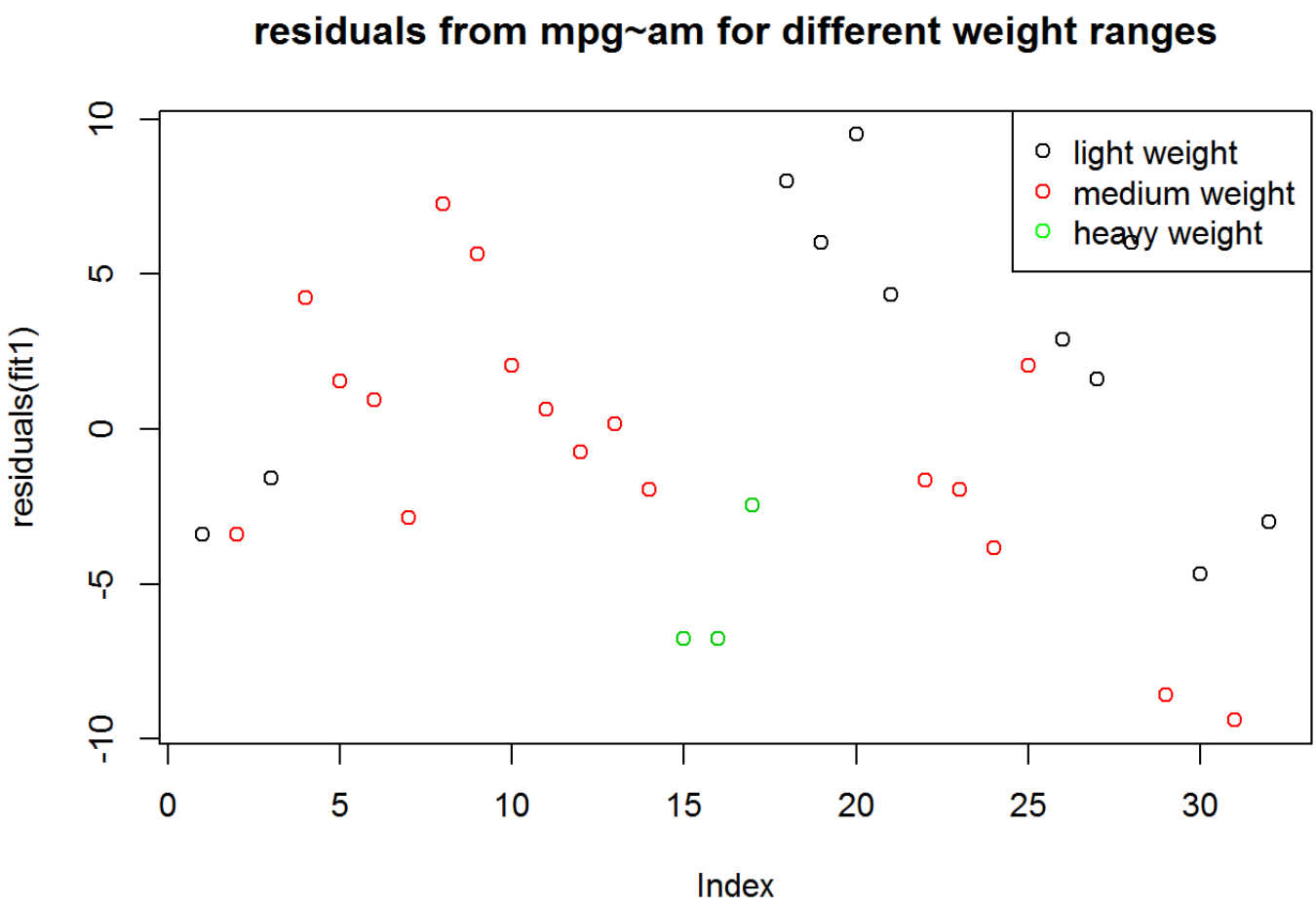
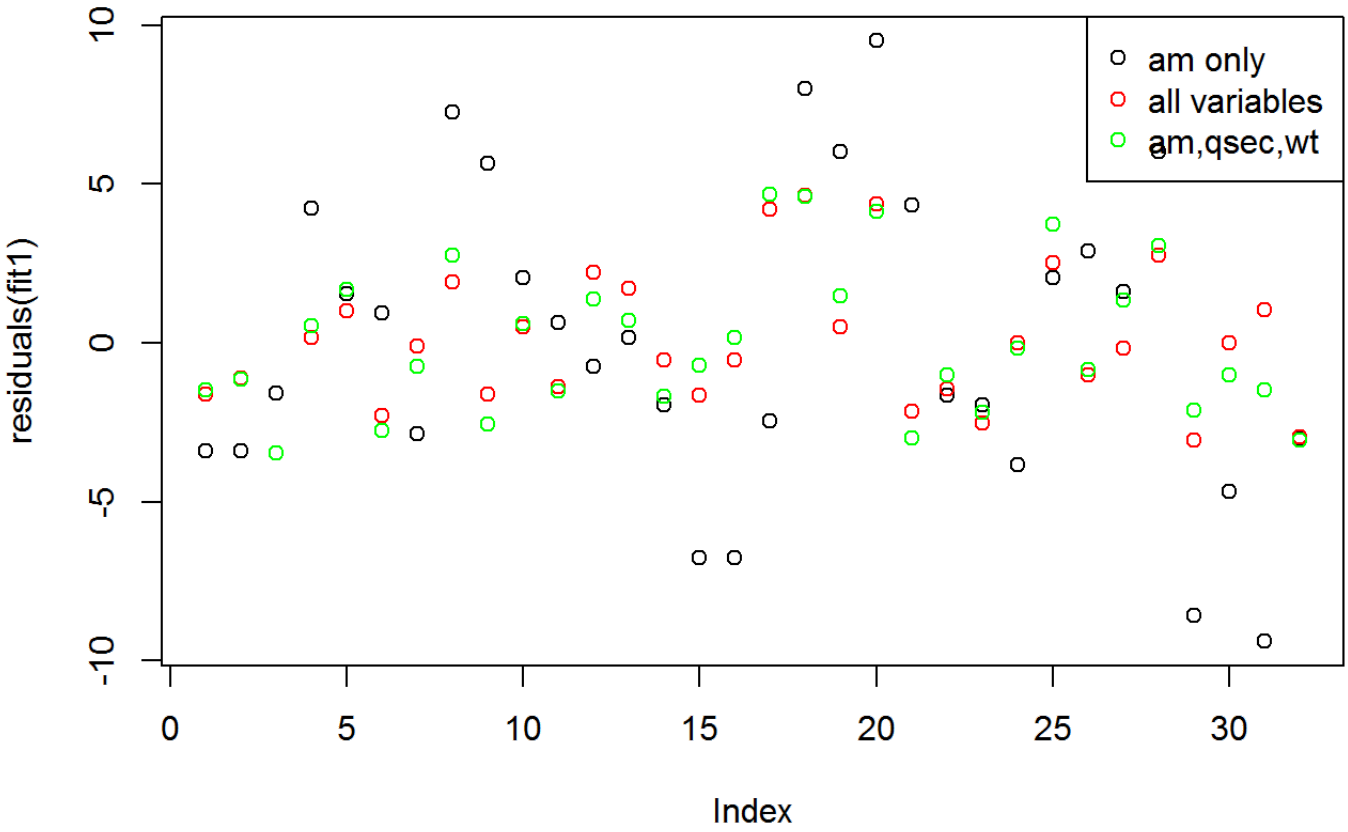



Figure2: the residuals of 3 different models


residuals from 3 models
residuals(fit1)
Index
am only
all variables
am,qsec,wt
10
5
0
-5
-10
0
5
10
15
20
25
30