

# Mileage Analysis on mtcars Data

*Arkosnato Neogy*

*Saturday 23 May 2015*

## Executive Summary

An analysis of the `mtcars` dataset shows that mileage is affected by several variables, including but not limited to transmission mode. Overall, manual transmission appears to give a higher mileage by about 7 miles per gallon. However, this variation can also be explained by other variables; for example, after adjusting for the effect of weight, no significant variation is left between the mileages of the 2 transmission modes. Therefore, it is not possible on the basis of this data to draw a *causal inference* on the effect of transmission mode on mileage.

## Analysis Details

In this report we attempt to investigate the effect of transmission mode (automatic or manual) on the mileage (in miles per gallon) of cars. This investigation will be based on the `mtcars` dataset from Motor Trend.

## Exploratory Plots

We initiate the analysis by looking at some basic plots that show the distribution of mileage for a given transmission mode. Fig.1 shows basic scatter plots of the data columns in `mtcars`. Fig.2 shows notched box plots of the mileage for the two transmission modes. Fig.3 shows the milage figures against cylinder count for each transmission mode.

From these plots a few inferences may be drawn. Firstly, Fig.1 evidences that mileage is affected by several variables (most significantlt wt, disp, hp, cyl) and not just transmission (am). The boxplot shows that on an average, manual transmission cars have higher mileage than automatic transmission. This analysis has the caveat that other variables are *not held constant* and therefore this can be called an *observational* but not *inferential* statement. As an example, we see that mileage can also thought to be explained by cylinder count in Fig.3 for a given transmission mode. Fig.2 further shows that the data can be thought to be approximately normally distributed but the variances are definitely not equal.

## Hypothesis Testing

We can run a t-test to try to quantify the difference in means between automatic and manual transmission cars. In the following test, automatic transmission is treated treatment while manual is the reference. The test shows that manual transmission cars have a clearly higher mileage at 5% significance level of the t-test by about 7 miles per gallon. Note however that in this t-test other variables are not held constant, so this cannot be inferred as a causal explanation of the mileage.

```
##
## Welch Two Sample t-test
##
## data: mpg by am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
## -11.280194 -3.209684
## sample estimates:
## mean in group 0 mean in group 1
##      17.14737      24.39231
```

## Factor Regression Models

To investigate this issue further we build regression models on mileage on transmission mode (as a factor), but we include, in turns, some of the key variables as evidenced in the scatter plots in Fig.1. The following panels show the details of the models, while their plots (modeled and marginal effects) are shown in the appendix

```
##           Estimate Std. Error    t value    Pr(>|t|)
## (Intercept) 37.32155131  3.0546385 12.21799285 5.843477e-13
## wt          -5.35281145  0.7882438 -6.79080719 1.867415e-07
## factor(am)1 -0.02361522  1.5456453 -0.01527855 9.879146e-01
```

```
##           Estimate Std. Error    t value    Pr(>|t|)
## (Intercept) -1.949883   7.073285 -0.2756687 0.78475740
## drat         5.811143   2.129833  2.7284496 0.01069548
## factor(am)1  2.807061   2.282159  1.2300023 0.22858143
```

```
##           Estimate Std. Error    t value    Pr(>|t|)
## (Intercept) 27.84808111  1.834071377 15.183750 2.452658e-15
## disp        -0.03685086  0.005781896 -6.373490 5.747528e-07
## factor(am)1  1.83345825  1.436099585  1.276693 2.118396e-01
```

From these figures, it is clear that `wt`, `drat` and `disp` can individually explain the mileage quite well (especially `wt`). The difference in means that is visible without adjusting for the added variables (marked by horizontal lines) is either absent or not significant after the adjustment has been made (visible from the intercepts of the regression lines for two the transmission modes).

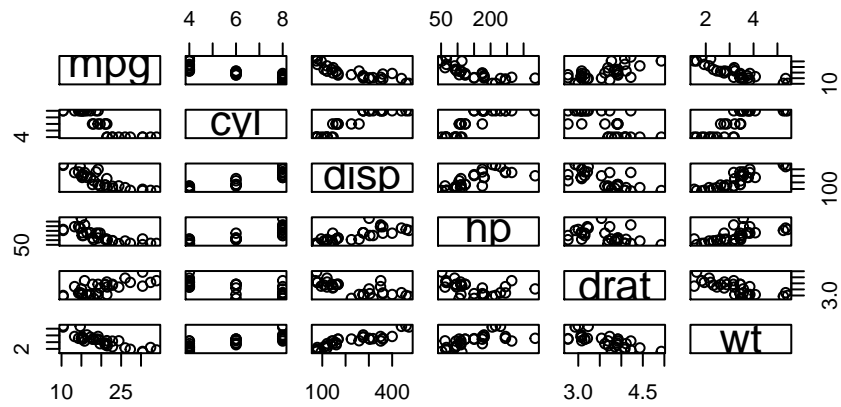
Further we show the residual plot for the best model (of `wt + factor(am)`). From the residuals it appears that the model is a reasonably good fit through there is a slightly parabolic trend in the residuals, potentially indicating the need of a square term in `wt`.

## Conclusions

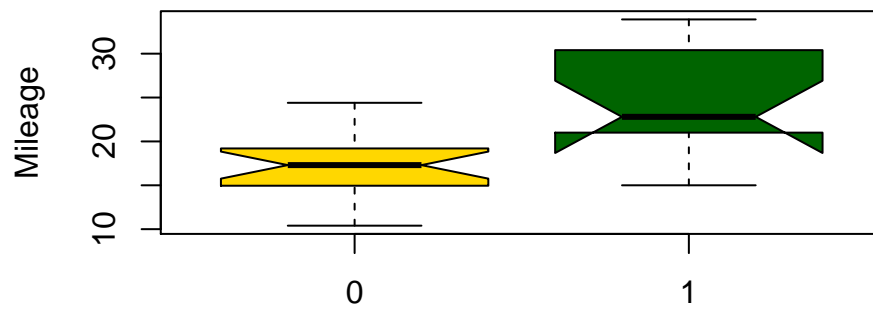
Therefore, while it is possible to say at an *observational* level that manual transmission cars have higher mileage than automatic, and the difference can be quantified through a t-test, we do not have sufficient data to conclude at an *inferential* level whether transmission mode *causes* a change in mileage.

## Appendix

**Fig.1: Scatter plot on mtcars data**

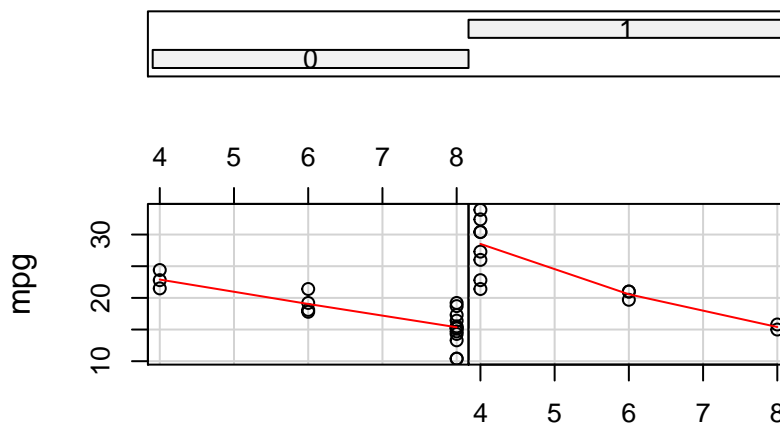


**Fig.2 Mileage vs Transmission**



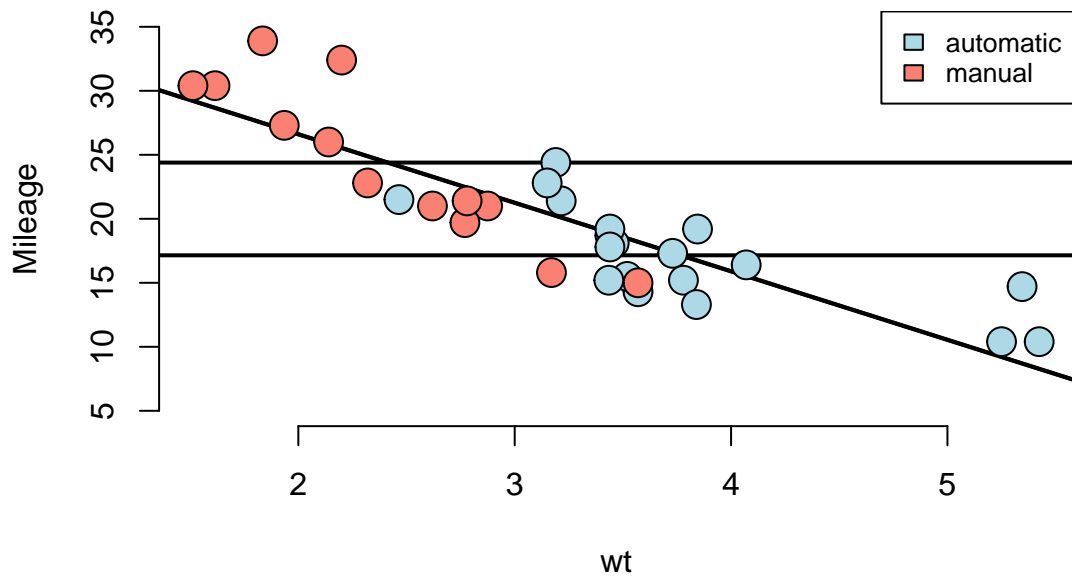
Transmission Mode: 0 (auto) and 1 (manual)

Given : am

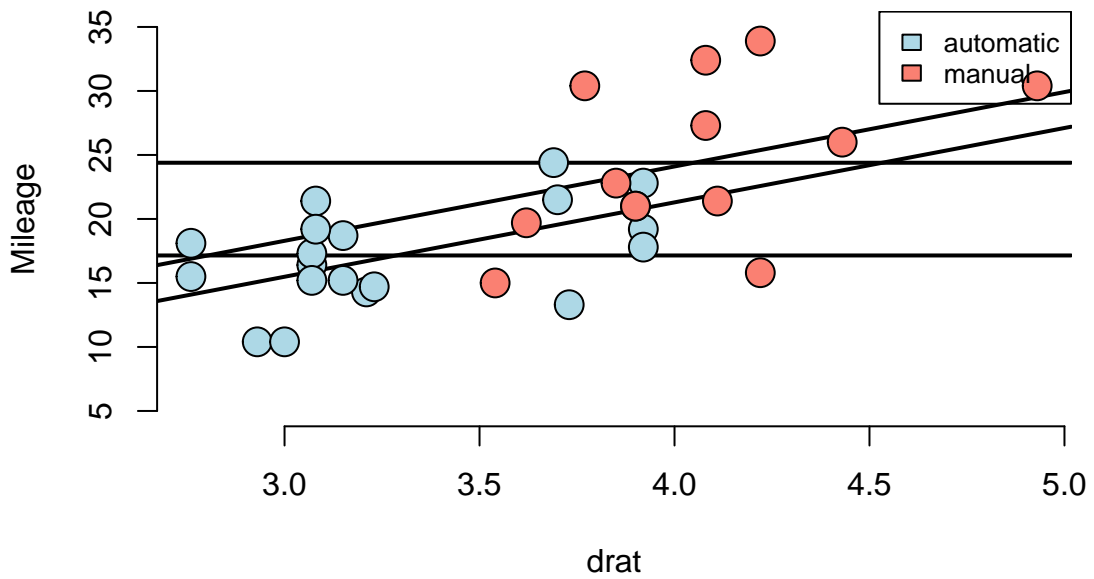


**Fig.3: Mileage vs Cylinder count given a transmission mode**

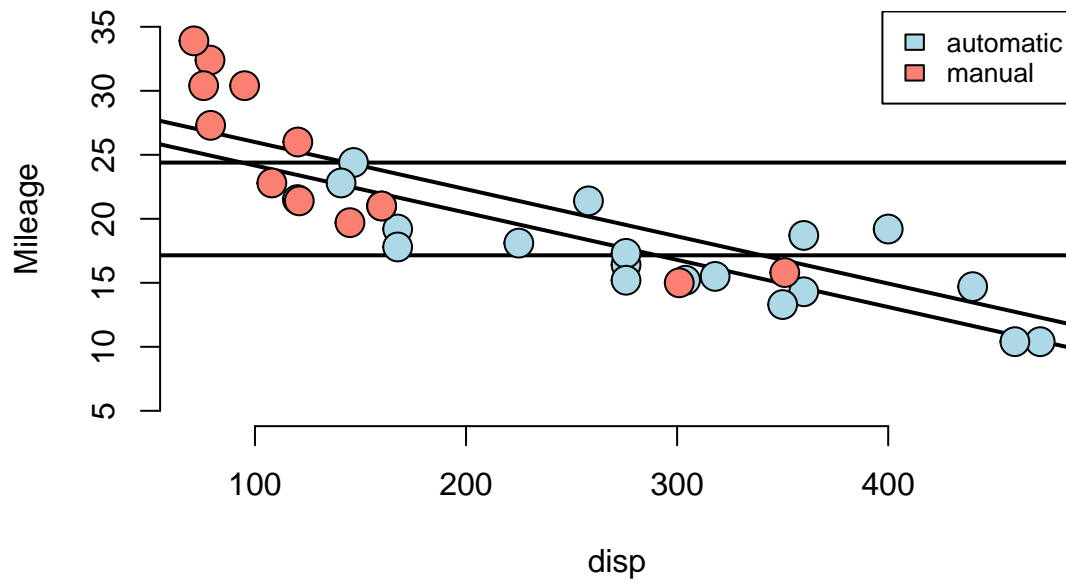
**Effect of Transmission after adjusting for wt**



**Effect of Transmission after adjusting for drat**



**Effect of Transmission after adjusting for disp**



**Residual plot for model with wt and am**

