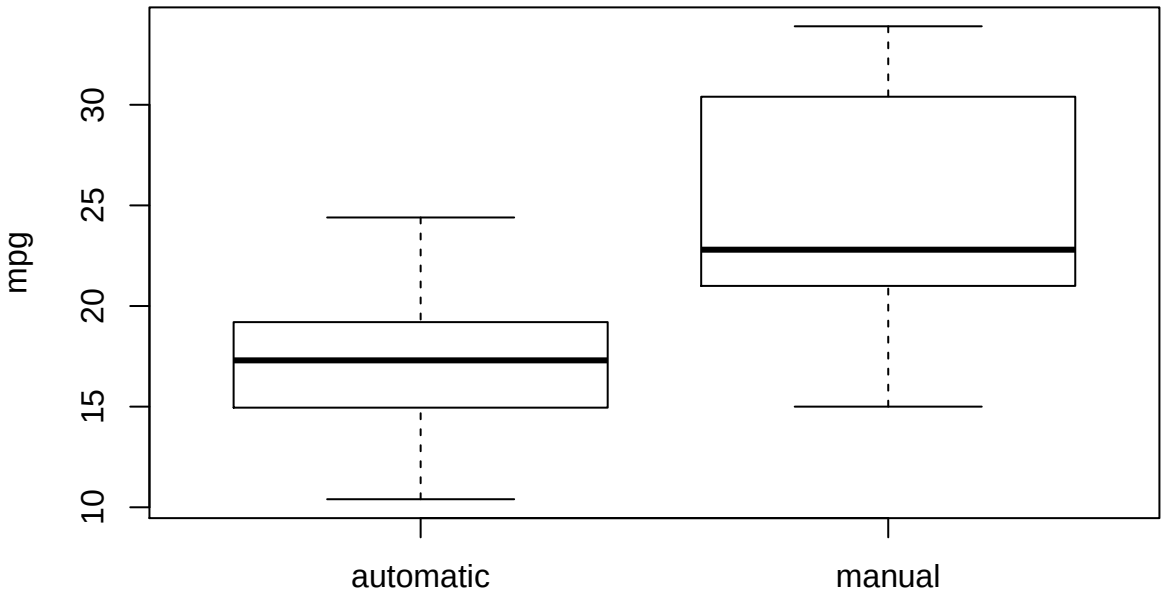
Regression Models Course Work

In this work we answer the questions "Do we get better mileage with automatic or manual transmission?" and "What is the size of the difference between automatic and manual?" drawing our conclusions from the *mtcars* dataset. In summary, the answers to these questions are "Manual" and "It depends."



We load the *mtcars* dataset, convert the *am* column to a factor and plot. As can be seen, although the *whiskers* overlap, the boxes don't, indicating some sort of effect.

```
fit1 <- lm(mpg ~ am, data=mtcars)
summary(fit1)$coef
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.134e-15
## ammanual       7.245      1.764   4.106 2.850e-04
```

From this simple linear fit, we note that the t-values for the coefficients are very high indicating that the probability of getting them by chance is small. Furthermore, the fit indicates that there is an expected gain for manual transmission over automatic of about 7.2 mpg.

Let's see if we can do better and add more predictors to the model. Let's try horsepower, *hp*, and weight, *wt*:

```
fit2 <- lm(mpg ~ am + hp, data=mtcars)
summary(fit2)$coef
```

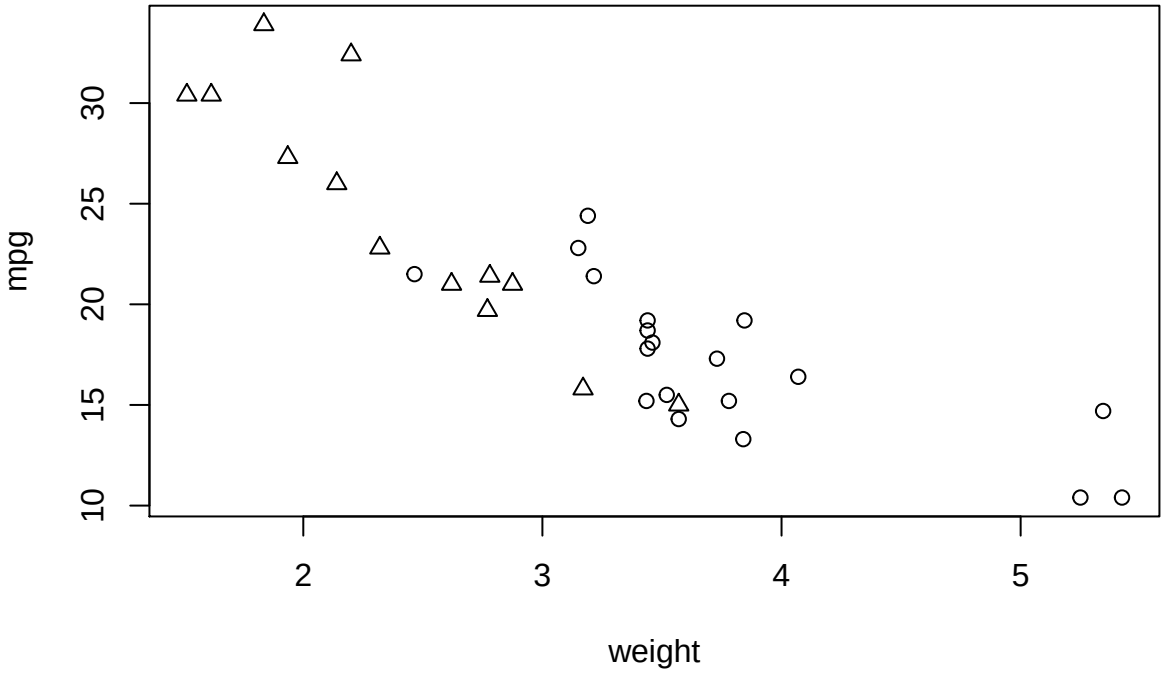
```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 26.58491   1.425094  18.655 1.074e-17
## ammanual     5.27709   1.079541   4.888 3.460e-05
## hp          -0.05889   0.007857  -7.495 2.920e-08
```

Adding in horsepower lowers the effect of manual transmission and also indicates that every additional horsepower costs about 0.06 mpg. Again the t-values indicate significance.

```r
fit3 <- lm(mpg ~ am + hp + wt, data=mtcars)
summary(fit3)$coef
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 34.00288   2.642659  12.867 2.824e-13
## ammanual     2.08371   1.376420   1.514 1.413e-01
## hp          -0.03748   0.009605  -3.902 5.464e-04
## wt          -2.87858   0.904971  -3.181 3.574e-03
```

Considering weight further lowers the effect of manual transmission beyond significance, indicating a possible correlation between weight and transmission-type. Indeed, plotting mpg versus weight and indicating the transmission type shows that lighter cars tend to have manual transmission.



Let's verify with with an "analysis of variance" on three *different* nested models:

```r
anova(lm(mpg ~ wt, data=mtcars), lm(mpg ~ wt + hp, data=mtcars), lm(mpg ~ wt + hp + am, data=mtcars))
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt
## Model 2: mpg ~ wt + hp
## Model 3: mpg ~ wt + hp + am
##   Res.Df RSS Df Sum of Sq     F Pr(>F)
## 1     30 278
## 2     29 195  1      83.3 12.93 0.0012 **
## 3     28 180  1      14.8  2.29 0.1413
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The conclusion we draw from all of this is that weight is the main predictor of fuel consumption, followed by horsepower. In so far as transmission type has an effect, it is as a proxy for weight.