

# Motor Trend Assignment

*Alan C Bonnici*

*Tuesday, October 21, 2014*

## Assignment

You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

“Is an automatic or manual transmission better for MPG”

“Quantify the MPG difference between automatic and manual transmissions”

## Data

The data source for this assignment comes from the **mtcars** dataset which is bundled with R. The dataset can be loaded into R by typing `data(mtcars)`. The effect of this command is to load into the environment **mtcars**. The resulting dataset consists of **32** rows, each row made up of **11** columns. The information about this data can be called up by typing `?mtcars` at the R prompt. A link to the online version of the page is available at (<https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/mtcars.html>).

The property *am* (indicating whether the transmission is manual or automatic) will be converted into a factor variable.

```
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c("Automatic", "Manual")
```

The boxplot (Fig 1) explains the different transmission types with respect to the mpg. From this chart, it seems that manual transmission have a higher mpg than automatic transmissions. One could also note that neither have any outliers and the whiskers of the box plot are balanced.

The histograms (Fig 2) compare the mpg of the two types of transmissions.

```
cars.Auto <- mtcars[mtcars$am == "Automatic",]
cars.Man <- mtcars[mtcars$am == "Manual",]
```

The average mpg for vehicles with automatic transmission is **17.15**. The average mpg for vehicles with manual transmission is **24.39**.

```
t.test(cars.Auto$mpg, cars.Man$mpg)
```

```
##
## Welch Two Sample t-test
##
## data: cars.Auto$mpg and cars.Man$mpg
## 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 of x mean of y
## 17.14737 24.39231
```

From the information presented in the Welch Two Sample t-test, it is clear that at a 95% confidence interval the difference between the mpg means of the automatic and petrol vehicles is always negative implying that manual will always be 3.21 mpg more efficient than automatic.

From the information presented above, the *null hypothesis: true difference in means is equal to 0* will be rejected at the 99% confidence level.

Even though the information shown above indicates that manual is more efficient than automatic, one needs to ensure that other factors are not confounding the results. For example, if automatic cars analysed here are heavier than manual ones then weight could be “upsetting” the results. Likewise if one category of vehicle have higher, hp or cylinder values this could be a factor and not the transmission.

A correlation analysis can help show the relationship of the various properties with respect to mpg shows that an increase in cylinders, displacement, hp, weight and number of carburetors all negatively impact mpg while rear axle ratio, 1/4 mile time and number of forward gears all contribute to higher mpgs.

```
##          wt          cyl          disp          hp          carb          qsec
## -0.8676594 -0.8521620 -0.8475514 -0.7761684 -0.5509251  0.4186840
##          gear          am          vs          drat
##  0.4802848  0.5998324  0.6640389  0.6811719
```

In a simple regression analysis from mpg given transmission

```
##
## 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 ***
## amManual       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
```

From the above data we can infer that, on average, manual transmission vehicle give 7.25 mpgs more than then automatic transmission. A value of  $R^2$  of 35.98% means that the model does not explain all of the variance and further investigation is required.

In order to investigate further a multivariate regression analysis is performed. In this case we adopt a stepwise linear regression. Stepwise regression is a semi-automated process of building a model by successively removing variables based solely on the t-statistics of their estimated coefficients.

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
```

```

##      Min      1Q  Median      3Q      Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337    18.71788   0.657  0.5181
## cyl         -0.11144     1.04502  -0.107  0.9161
## disp         0.01334     0.01786   0.747  0.4635
## hp          -0.02148     0.02177  -0.987  0.3350
## drat         0.78711     1.63537   0.481  0.6353
## wt          -3.71530     1.89441  -1.961  0.0633 .
## qsec         0.82104     0.73084   1.123  0.2739
## vs           0.31776     2.10451   0.151  0.8814
## amManual     2.52023     2.05665   1.225  0.2340
## gear         0.65541     1.49326   0.439  0.6652
## carb        -0.19942     0.82875  -0.241  0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869, Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF, p-value: 3.793e-07

##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##      Min      1Q  Median      3Q      Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt            -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec           1.2259     0.2887   4.247 0.000216 ***
## amManual       2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF, p-value: 1.21e-11

```

From the information presented above we can conclude that manual transmission contributes 2.94 mpg over automatic. This is far less than the value reported using the linear regression of mpg given transmission on its own or the t-test and can be explained by the weight and 0.25 mile time that play an important role in overall mpg value. See Fig 3 are a series of charts relating to the step wise regression.

Fig 1

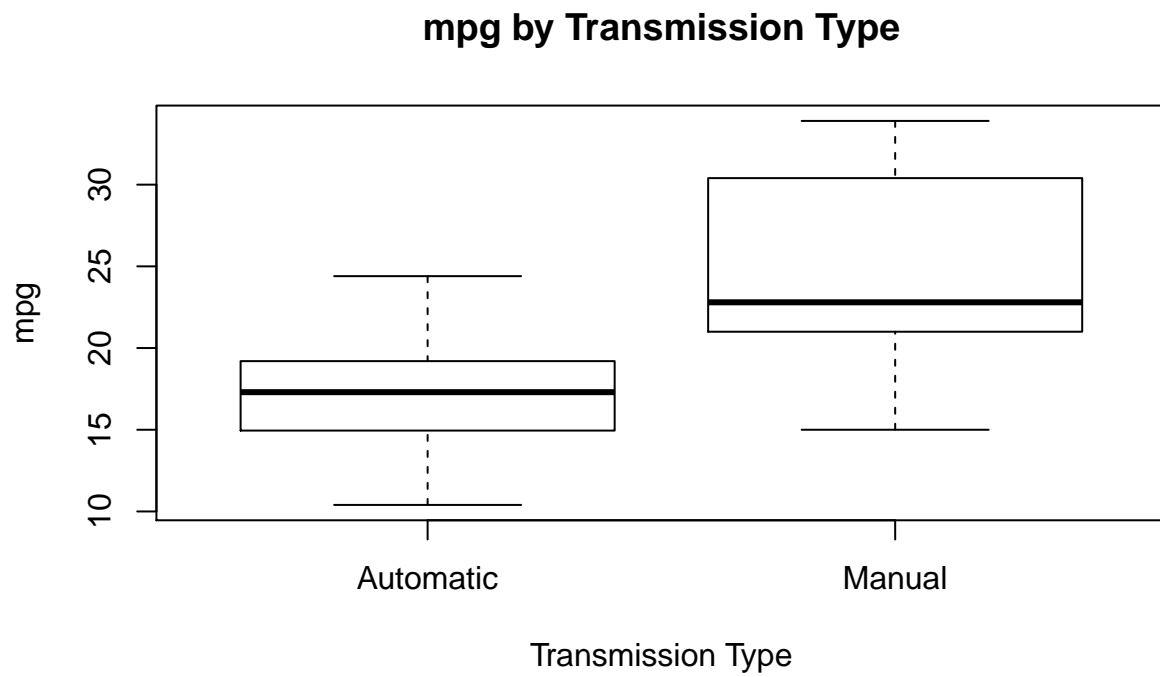


Fig 2

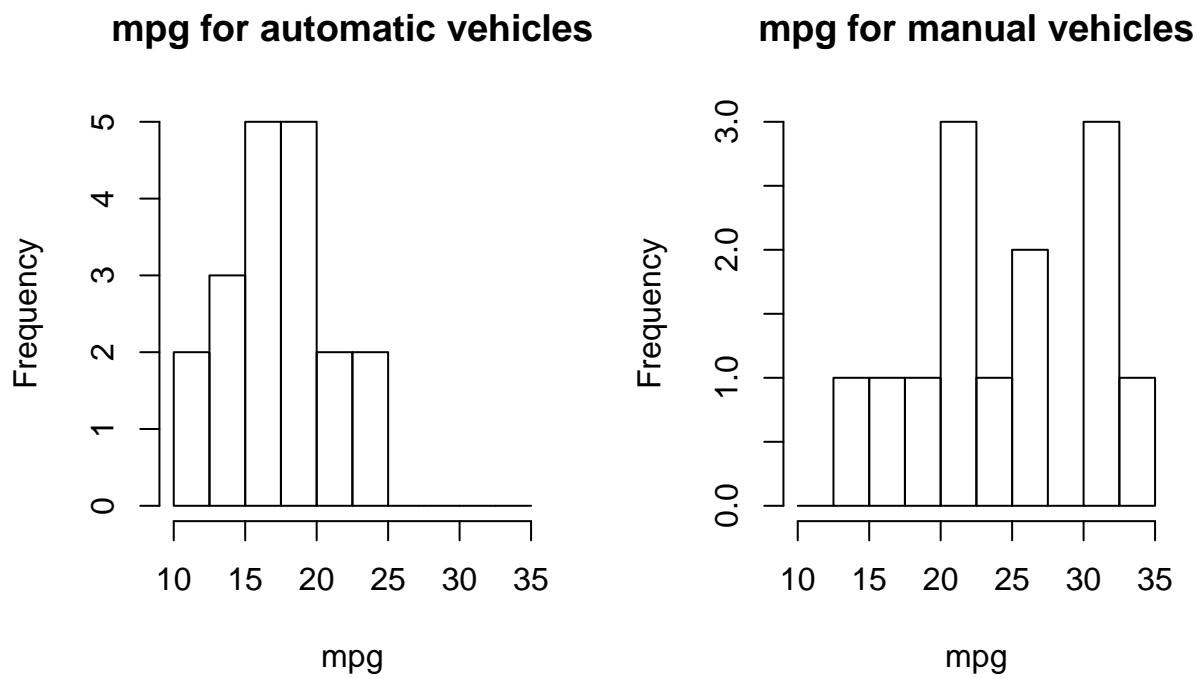


Fig 3

