# Manual or automatic: how does affect fuel performance?

*Javier Santibáñez*

*Saturday, January 24, 2015*

## Summary

In this report we present the results of a comparison between manual and automatic vehicles, measured in miles per gallon of fuel (mpg). The results shows that automatic cars has a better performance than manual cars. The expected difference was estimated in XXX mpg.

## Introduction

We used a data set that consists of the measurements of 11 variables from observations of cars from 32 cars. The variables are the following:

- `mpg`, miles(US)/gallon
- `cyl`, number of cylinders
- `dip`, displacement (cu.in.)
- `hp`, gross horsepower
- `drat`, rear axie ratio
- `wt`, weigth (lb/1000)
- `qsec` 1/4 mile time
- `vs` V/S
- `am` transmission (automatic, manual)
- `gear`, number of forward gears
- `carb`, number of carburetors

Additionaly, we created a new variable, called `am2`, which is the factor version of `am` with levels *automatic* and *manual*.

In this report we have to address the questions:

1. Is an automatic or manual transmission better for performance?
2. What is the difference in performance between automatic and manual transmissions?

To deal with this questions we used a statistical approach, in specific, we used regression models.

The first qustion can be easily answered. Figure *Performance vs. transmission* in Appendix suggests that performance, measured in miles per gallon, depends of the car transmissions.

Despite the previous graphic, we must verify that manual trasmission is better than automatic trasnmission. Also, it is important to consider other variables, because some of them can explain part of the variability.

Finally, we used the software statistical enviroment `R` and the packages `ggplot2` and `corrplot` for graphics.

## Methodology

The following is a brief description of the analysis performed:

1. Fit a regression model with `mpg` (performance) as output and `am` (transmission) as input.
2. Select other covariables from the data set, considering their correlation with `mpg`.
3. Fit a regression model with `mpg` as output and the set of variables from the preceding step as input.
4. Evaluate and compare models fitted in steps 1 and 4.

## Results

### Step 1. Basic model.

The basic model consists in only one input (`mpg`). We used the next code to fit the model.

```
fitMod1<-lm(mpg~am2,data=mtcars)
summary(fitMod1)
```

```
##
## Call:
## lm(formula = mpg ~ am2, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.392 -3.092 -0.297  3.244  9.508
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    17.15       1.12   15.25  1.1e-15 ***
## am2Manual       7.24       1.76    4.11  0.00029 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.9 on 30 degrees of freedom
## Multiple R-squared:  0.36,   Adjusted R-squared:  0.338
## F-statistic: 16.9 on 1 and 30 DF,  p-value: 0.000285
```

The previous results indicates that there is a highly significant difference in perfomance between automatic and manual transmission, estimated as 7.2449 miles per gallon, on average. With this information we can answer our two questions, but we improve them with the following steps.

We verify the accuracy of the normal assumption with a *qq-plot*. Figure 2 shows that the residuals are aproximately normal distributed.

### Step 2. Select new variables.

To select new variables to include in the model, we used the correlation matrix via the `corrplot` function. From Figure 3 we can identify a set of variables highly correlated:

- cyl
- disp
- hp
- drat

Despite this result, we cannot include the whole set of variables because there are a high correlation between them. Then, we have tho choose one variable from that set. We select the variable `hp` because we can use it as a confounder in a extended model, to reduce variability and to improve the estimation of difference in `mpg` due to transmission.

### Step 3. Extended model.

Now we extend our basic model adding the variable `hp`. First we fit a model with an interection term with the following code.

```
fitMod2<-lm(mpg~am2+hp+am2*hp,data=mtcars)
summary(fitMod2)
```

```
##
## Call:
## lm(formula = mpg ~ am2 + hp + am2 * hp, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.382 -2.270  0.134  1.706  5.875
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  26.624848   2.182943   12.20    1e-12 ***
## am2Manual     5.217653   2.665093    1.96     0.06 .
## hp           -0.059137   0.012945   -4.57    9e-05 ***
## am2Manual:hp  0.000403   0.016460    0.02     0.98
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.96 on 28 degrees of freedom
## Multiple R-squared:  0.782,  Adjusted R-squared:  0.759
## F-statistic: 33.5 on 3 and 28 DF,  p-value: 2.11e-09
```

The previous results suggest that there is evidence to support the hipotesis that the interaction term is not necessary.

We fit a new model wihtout the interaction term between `am` and `hp`, these are the results:

```
fitMod3<-lm(mpg~am2+hp,data=mtcars)
summary(fitMod3)
```

```
##
## Call:
## lm(formula = mpg ~ am2 + hp, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.384 -2.264  0.137  1.697  5.866
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 26.58491    1.42509   18.65  < 2e-16 ***
## am2Manual    5.27709    1.07954    4.89  3.5e-05 ***
## hp          -0.05889    0.00786   -7.50  2.9e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.91 on 29 degrees of freedom
## Multiple R-squared:  0.782,  Adjusted R-squared:  0.767
## F-statistic:   52 on 2 and 29 DF,  p-value: 2.55e-10
```

These results suggest support again that there is a difference in performace between automatic and manual transmission, estimated in 5.2771 miles per gallon, but in the case, we take into account a confounder that explains some of the residual variability.

We evaluated the assuption of normality with a *qq-plot*. Figure 4 shows that there is no problems with normality. Also, to check corralation between residuals an covariables we made a scatter plot. In Figure 5 we showed that the slope of a regression line of residuals over the covariable `hp` is aproximately zero, which suggests that there is no correlation problems.

#### Step 4. Comparing models.

We estimated two models. An analysis of variance (ANOVA) shows that the extended model is better than the basic model, with a high level of significance (p-value less than 0.05).

```
kable(anova(fitMod1,fitMod3))
```

| Res.Df | RSS | Df | Sum of Sq | F | Pr(>F) |
|---:|---:|---:|---:|---:|---:|
| 30 | 720.9 | NA | NA | NA | NA |
| 29 | 245.4 | 1 | 475.5 | 56.18 | 0 |

Also, the ratio of the sum of squared residuals of the extended model to the basic model is 0.34, which means that the residual variation of the extended model is 34% smallet than the residual variation of the basic model.

## Conclusions

According to the preceding results we can conclude that:

1. A manual transmission is better than an automatic one, in terms of performance, measured in miles per gallon.
2. There is a difference of 5.2771 miles per gallon between manual and automatic transmissions.

## Appendix

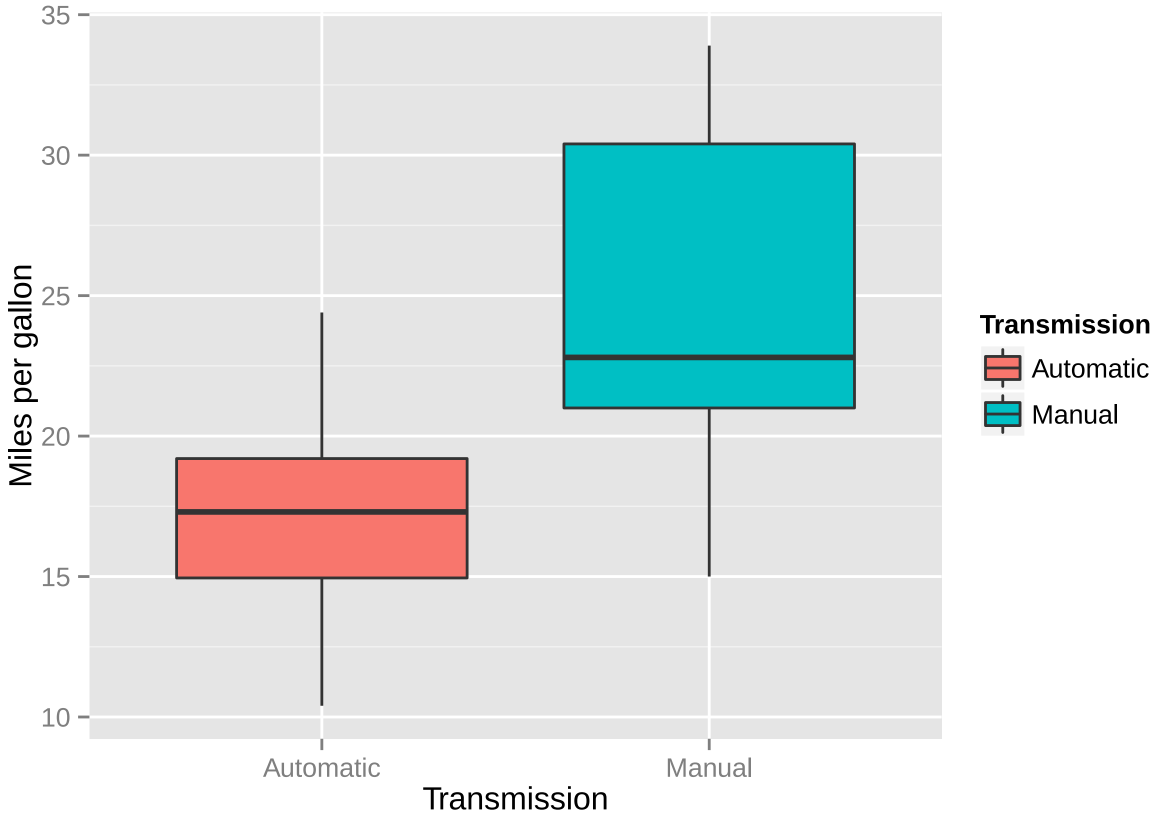


Figure 1: Performance vs. transmision

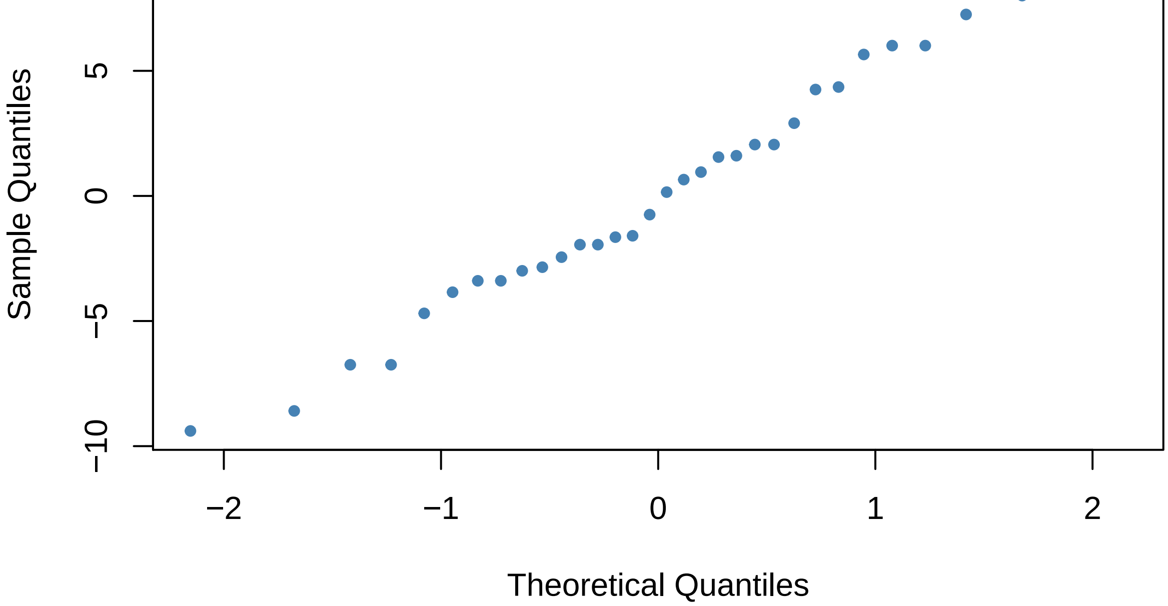


Figure 2: QQ plot for residuals of basic model

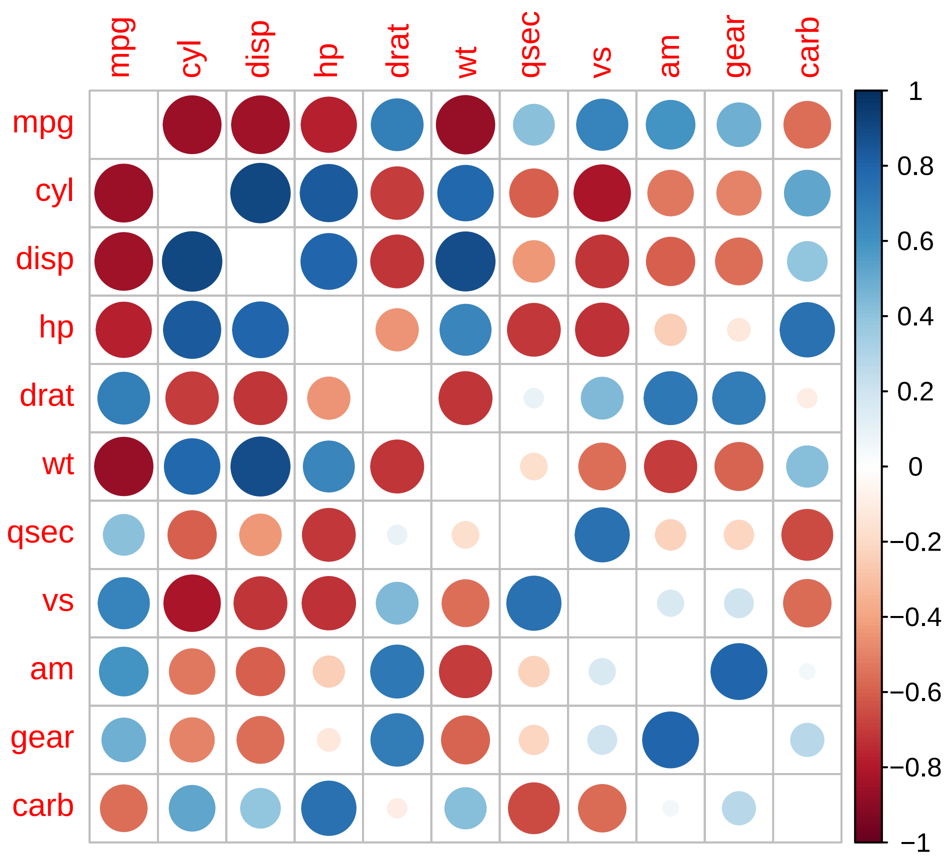


Figure 3: Correlation matrix

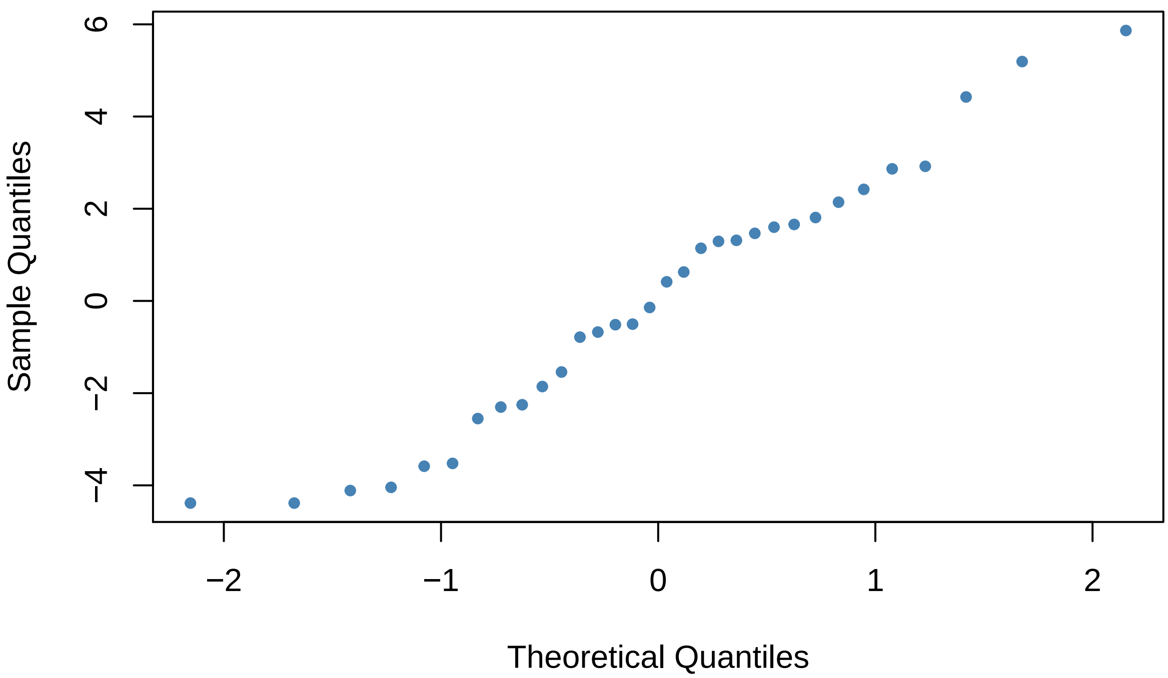


Figure 4: QQ plot for residuals of extended model

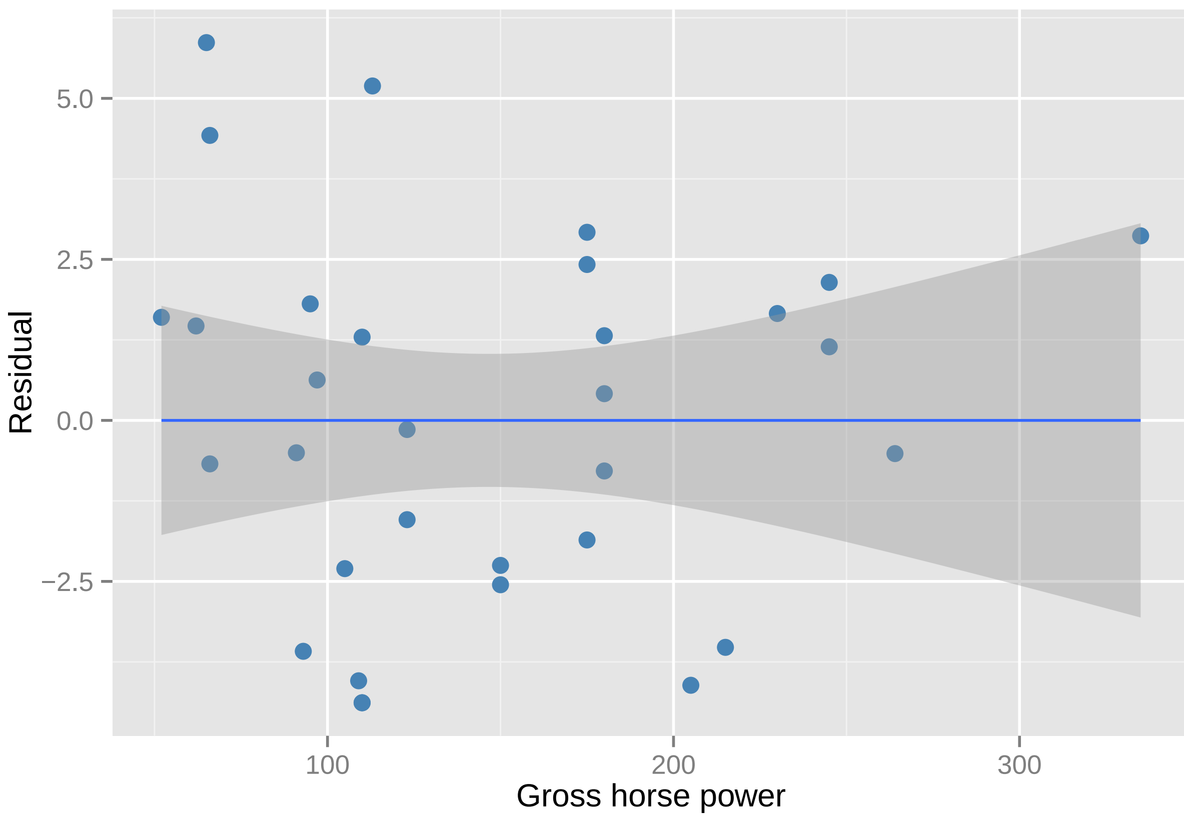


Figure 5: Scatter plot: residuals versus gross horse power