Motortrend Report

Marcelo Bohrer

Summary of Findings

The purpose of this report is to investigate the relationshiph between miles per galon and the type of transmition in different cars. The data comes from the "mtcars" dataset. We found that transmition type and horse power explain a good part of the variance in miles per galon, roughly 75% of it. Holding hp constant switching from automatic to manual transmition accounts for a 5.3 mpg increase.

Exploratory Data Analysis

The variables taken into consideration are: mpg (miles per gallon); hp (horse power); am (transmition type) am=0 for automatic and am=1 for manual; qsec (1/4 mile time); vs (engine type); and carb (number of carburators). Additionally the dataset has information on: number of cylinders (cyl); displacement (disp); rear axle ratio (drat); and number of forward gears (gear).

The two variables we are mostly interested in are "mpg" (miles per galon) and "am"(automatic or manual transmition). Let's look at summary statistics for them. See also the first figure in appendix.

```
summary(mtcars$mpg)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.40   15.42   19.20   20.09   22.80   33.90
```

```
table(mtcars$am)
```

```
##
##  0  1
## 19 13
```

Now that we have some intuition about the dataset we can proceed to running some regressions.

Regression Analysis

The explanatory variable we are interested in is am. But we can still make use of aditional variables. In order to avoid biases we have to make use of variables that have little or no correlation to am. The following correlation table guides the choice of variables (correlation with am.

```
##            mpg       cyl      disp         hp      drat         wt
## [1,] 0.5998324 -0.522607 -0.591227 -0.2432043 0.7127111 -0.6924953
##            qsec        vs am      gear       carb
## [1,] -0.2298609 0.1683451  1 0.7940588 0.05753435
```

As we can see in the table above, hp (horse power), qsec(1/4 mile time), VS(Engine type) and carb (number of carburators) are variables that might help transmition type explain miles per gallon.

Now we have to determine which of these 4 variables to use. We can look at the correlation between these variables to make a better choice.

```
##             mpg         hp       qsec         vs       carb
## mpg   1.0000000 -0.7761684  0.4186840  0.6640389 -0.5509251
## hp   -0.7761684  1.0000000 -0.7082234 -0.7230967  0.7498125
## qsec  0.4186840 -0.7082234  1.0000000  0.7445354 -0.6562492
## vs    0.6640389 -0.7230967  0.7445354  1.0000000 -0.5696071
## carb -0.5509251  0.7498125 -0.6562492 -0.5696071  1.0000000
```

As the table shows, hp is higly correlated with mpg, as well as with all othe variables, hence hp should not be use in conjunction with the other 3 variables. The number of carburators (carb) and the engine type vs have the smallest correlation, hence they could be used in conjunction. We proceed to investigate these models.

```r
fit1 <- lm(mpg~am+hp, mtcars)
fit2 <- lm(mpg~am + carb +vs, mtcars)
summary(fit1)$coefficients
```

```
##               Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) 26.5849137 1.425094292 18.654845 1.073954e-17
## am           5.2770853 1.079540576  4.888270 3.460318e-05
## hp          -0.0588878 0.007856745 -7.495191 2.920375e-08
```

```r
summary(fit2)$coefficients
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 19.517399  1.6090815 12.129528 1.155904e-12
## am           6.797956  1.1014890  6.171606 1.154742e-06
## carb        -1.430783  0.4081085 -3.505890 1.552505e-03
## vs           4.195736  1.3245867  3.167581 3.695735e-03
```

Note that the models provide very similar overall fits. Hence, preference should be given to the simpler model, namely fit1. We can compare the variance inflation factors to support our choice.

```r
vif(fit1)
```

```
##       am       hp
## 1.062867 1.062867
```

```r
vif(fit2)
```

```
##       am     carb       vs
## 1.067446 1.535339 1.574889
```

Now that we have decided on a model we can run some basic inference and interpreation, after which we can run diagnostics analysis.

```r
confint(fit1)
```

```
##                   2.5 %      97.5 %
## (Intercept) 23.67026866 29.49955884
## am           3.06917692  7.48499369
## hp          -0.07495665 -0.04281896
```

We can see that both predictors are significantly different from zero. The intercept of 26.58 means that cars with automatic transmition (am=0) run an average of 26.58 miles per galon, controling for horse power. The am conefficient of 5.27 implies that switching from automatic to manual transmition goes along with an expected 5.27 aditional miles per gallon (totalling 31.862), controlling for horse power. The hp coefficient implies that by each additional horse power, a car is expected to reduce it's miles per galon by .059, controlling for transmition type.

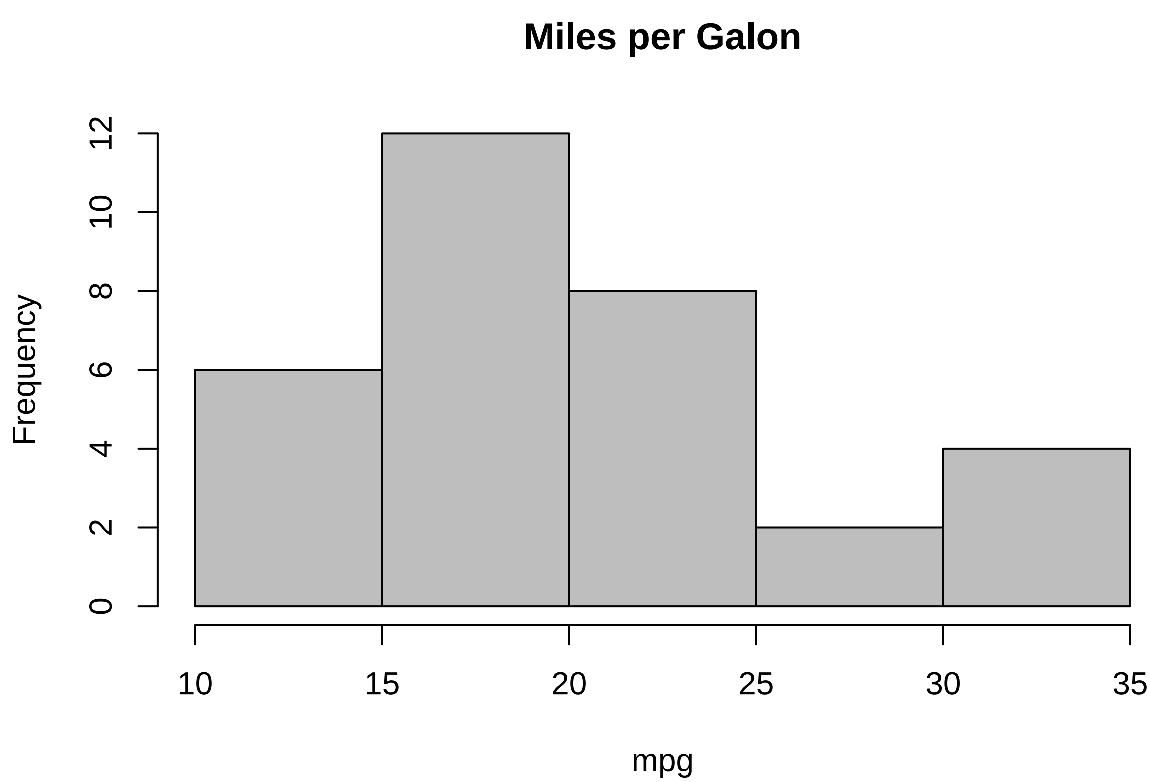
Basic diagnostics analsysis can now be run. See appendix.

As the first panel shows, there seems to be no trend in the residuals.Despite the small sample (32), the residuals seem to be somewhat normal, panel 2.

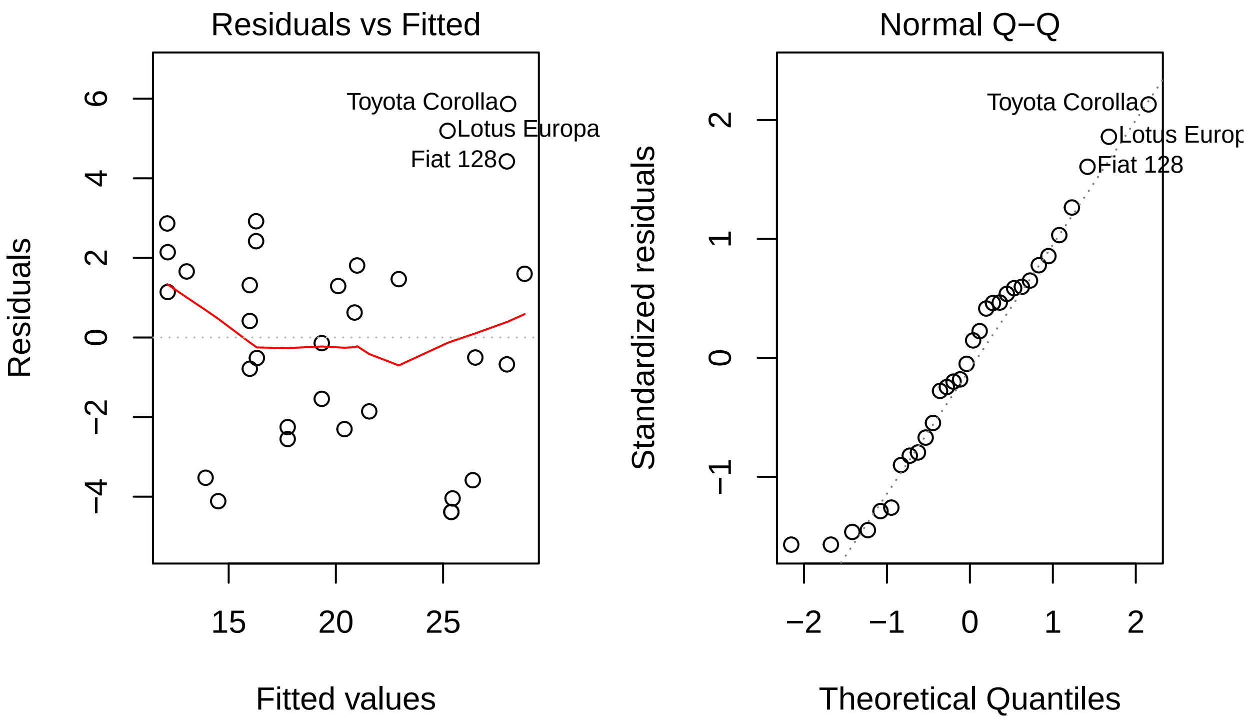
The main limitation concearning this report is the small sample size, only 32 observations. Increasing the number of observation could potentially reduce the correlation among the independent variables, which would allow for a richer model.

Apendix

```
hist(mtcars$mpg, main = "Miles per Galon", xlab = "mpg", col = "grey")
```




```
par(mfrow=c(1,2))
plot(fit1, which = 1)
plot(fit1, which = 2)
```


Residuals vs Fitted
Residuals
Fitted values
6
4
2
0
−2
−4
15
20
25
Toyota Corolla
Lotus Europa
Fiat 128
Normal Q−Q
Standardized residuals
Theoretical Quantiles
2
1
0
−1
−2
−1
0
1
2
Toyota Corolla
Lotus Europ
Fiat 128