Are Miles per Gallon different for manual or automatic transmission cars?

Sebastian Gomez

Executive Summary

This report attempts to answer two questions about the relationship between a set of variables and miles per gallon (mpg) as the outcome for a collection of cars represented in the mtcars data set: 1. "Is an automatic or manual transmission better for MPG", and 2. "Quantify the MPG difference between automatic and manual transmissions". In order to do this, exploratory data analysis, and statistical inference are conducted to assess the correlation between transmission types and mpg. Also, several regression models were tried to estimate the actual differences in mpg values for both automatic and manual transmission types.

Preparing and Loading the Data

```
data("mtcars"); mtcarsold = mtcars; mtcars[mtcars$am == 0, ]$am = 'automatic'
mtcars[mtcars$am == 1, ]$am = 'manual'; mtcars$am = as.factor(mtcars$am)
```

Exploratory Data Analysis

In the Appendix section you will find a detailed box-plot showing the relationship between mpg and transmission variables. It indicates an apparent difference in mpg for automatic and manual transmission types. We will find out in our Inference section. In that section also, the absolute value vector shows the variables with high correlation (>= 0.8) to mpg are cyl, disp and wt.

Inference

The initial approach to answering question number 1 (Is an automatic or manual transmission better for MPG?) is by executing a t.test to identify whether manual transmissions have higher miles per gallon (alternative hypothesis - true difference in means is not equal to 0):

```
ttest = t.test(mpg ~ am, mtcars); ttest$conf.int
```

```
## [1] -11.280194  -3.209684
## attr(,"conf.level")
## [1] 0.95
```

The 95% confidence interval indicates (see the ttest full details in the Appendix section) the difference for mpg values in transmission types is statistically significant. In summary, manual transmission performs significantly for MPG.

Regression

The following models were evaluated to answer question number 2 (Quantify the MPG difference between automatic and manual transmissions):

```
fitlineonlyam = lm(mpg ~ am, data = mtcars)
fitlineall = lm(mpg ~ ., data = mtcars)
fitcorrvar = lm(mpg ~ cyl + disp + wt + am, data = mtcars)
fitstep <- step(fitlineall, direction = "both", trace = FALSE)
```

1. Linear model including am factor variable only (see full summary in the Appendix section). This model indeed indicates a significantly valid relationship between transmission type and miles per gallon but a very poor model fit that only explains 0.36 of the outcome's variability (as per the multiple R-squared)
2. Linear model with all available variables (see full summary in the Appendix section). This one indicates a very good fit, with an R-squared value of 0.87. However, none of the terms included in the model appear to be significant, which is not a good indicator either.
3. Linear model with the variables showing more than 0.8 absolute correlation value with mpg (cyl, disp and wt), and am. This model shows an 0.83 R-squared value, which has a very good fit, but the associated term for am manual is not statistically significant. However, the intercept is highly significant, and that is associated with am automatic level.
4. Finally, a step-wise automatic regression approach using the step() function in R (see full summary in the Appendix section). This model ended up selecting wt, qsec and am (manual) as statistically significant regressors, and an R-squared value of 0.85.

Now, since models the last two models have relatively high R-squared values, model 4 (step automatic regression) not only is just a little bit higher but simpler as it has one less regressor variable. For this reason, the selected model is number 4, the step-wise approach including wt, qsec and am variables (see coefficients).

```
fitstep$coefficient
```

```
## (Intercept)          wt        qsec    ammanual
##    9.617781   -3.916504    1.225886    2.935837
```

The above terms indicate that for a given constant weight (wt) in 1000lbs, and constant quarter mile time (qsec), manual transmissions have on average 2.94 more miles per gallon than automatic ones, with the following 95% confidence interval:

```
confint(fitstep)['ammanual', ]
```

```
##      2.5 %     97.5 %
## 0.04573031 5.82594408
```

Residuals and Diagnostics

After fitting the model, further analysis of the residuals (see details in the Appendix section) is made indicating the following very important assumptions are valid for inferential purposes:

- Residuals do not show an apparent trend (points above and below 0). Although there is a small decrease towards the middle of the fitted values
- Normal Q-Q plot of standardized residuals indicates a pretty good fit with a normalized theoretical distribution

Appendix

ttest details

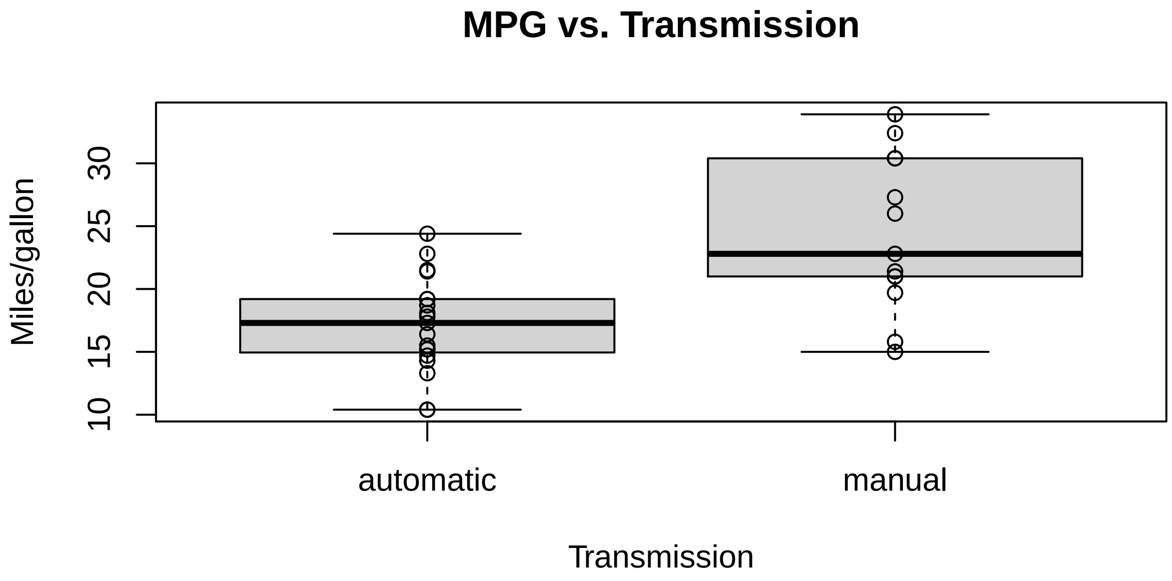
```
ttest$estimate; ttest$conf.int
```

```
## mean in group automatic    mean in group manual
##                17.14737                24.39231
```

```
## [1] -11.280194  -3.209684
## attr(,"conf.level")
## [1] 0.95
```

Box-plot of the mpg outcome vs. transmission type

```
plot(mtcars$am, mtcars$mpg, main = 'MPG vs. Transmission', ylab = 'Miles/gallon',
     xlab = 'Transmission'); points(mtcars$am, mtcars$mpg)
```



Absolute correlation value vector for all mtcars variables vs the mpg outcome

```
abs(cor(mtcarsold))[-1, 1]
```

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

Model 1 summary: mpg ~ am

```
summary(fitlineonlyam)$coefficient; summary(fitlineonlyam)$r.squared
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## ammanual     7.244939   1.764422  4.106127 2.850207e-04
```

```
## [1] 0.3597989
```

Model 2 summary: mpg ~ .

```
summary(fitlineall)$coefficient; summary(fitlineall)$r.squared
```

```
##                Estimate  Std. Error    t value   Pr(>|t|)
## (Intercept) 12.30337416 18.71788443  0.6573058 0.51812440
## cyl         -0.11144048  1.04502336 -0.1066392 0.91608738
## disp         0.01333524  0.01785750  0.7467585 0.46348865
## hp          -0.02148212  0.02176858 -0.9868407 0.33495531
## drat         0.78711097  1.63537307  0.4813036 0.63527790
## wt          -3.71530393  1.89441430 -1.9611887 0.06325215
## qsec         0.82104075  0.73084480  1.1234133 0.27394127
## vs           0.31776281  2.10450861  0.1509915 0.88142347
## ammanual     2.52022689  2.05665055  1.2254035 0.23398971
## gear         0.65541302  1.49325996  0.4389142 0.66520643
## carb        -0.19941925  0.82875250 -0.2406258 0.81217871
```

```
## [1] 0.8690158
```

Model 3 summary: Highly correlated values + am (transmission types)

```
summary(fitcorrvar)$coefficient; summary(fitcorrvar)$r.squared
```

```
##                 Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) 40.898313414 3.60154037 11.3557837 8.677574e-12
## cyl         -1.784173258 0.61819218 -2.8861142 7.581533e-03
## disp         0.007403833 0.01208067  0.6128661 5.450930e-01
## wt          -3.583425472 1.18650433 -3.0201537 5.468412e-03
## ammanual     0.129065571 1.32151163  0.0976651 9.229196e-01
```

```
## [1] 0.8326661
```

Model 4 summary: Step

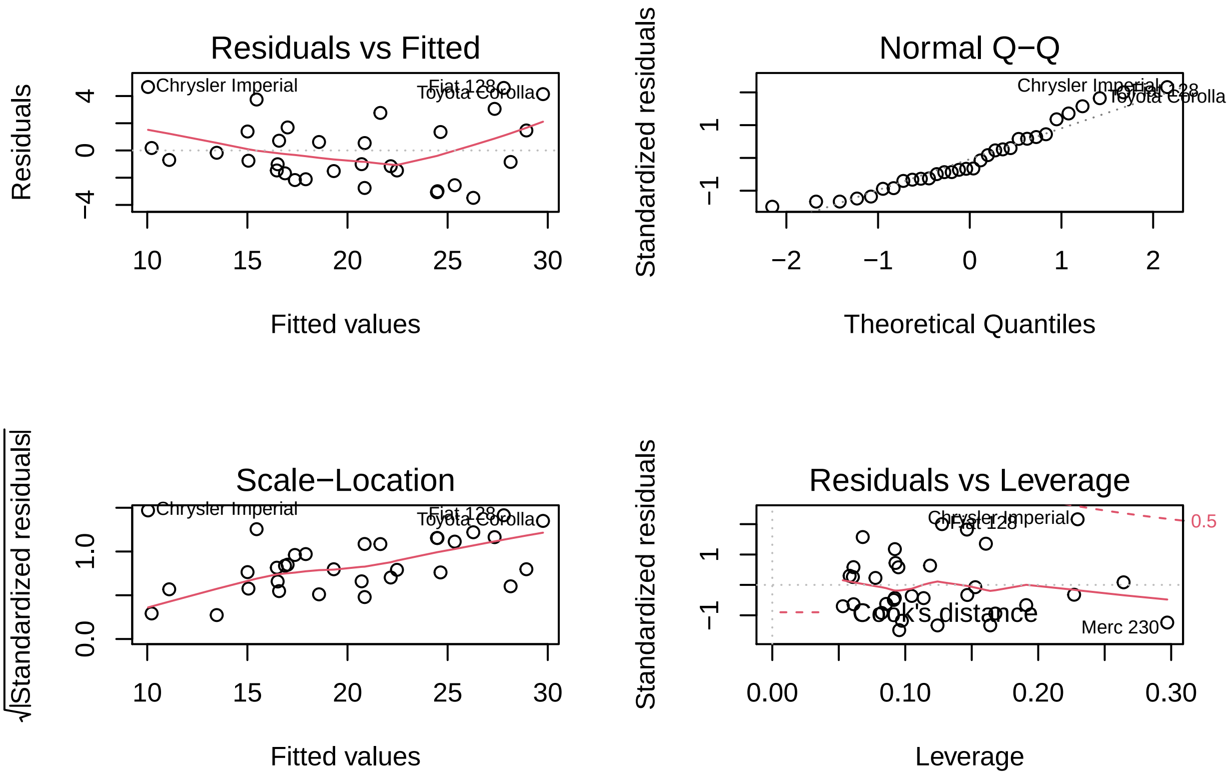
```
summary(fitstep)$coefficient; summary(fitstep)$r.squared
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  9.617781  6.9595930  1.381946 1.779152e-01
## wt          -3.916504  0.7112016 -5.506882 6.952711e-06
## qsec         1.225886  0.2886696  4.246676 2.161737e-04
## ammanual     2.935837  1.4109045  2.080819 4.671551e-02

## [1] 0.8496636
```

Residuals and diagnostics plots

```
par(mfrow = c(2,2))
plot(fitstep)
```


Residuals vs Fitted
Residuals
−4
0
4
Chrysler Imperial
Fiat 128
Toyota Corolla
10
15
20
25
30
Fitted values
Normal Q−Q
Standardized residuals
−1
1
Chrysler Imperial
Fiat 128
Toyota Corolla
−2
−1
0
1
2
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
0.0
1.0
Chrysler Imperial
Fiat 128
Toyota Corolla
10
15
20
25
30
Fitted values
Residuals vs Leverage
Standardized residuals
−1
1
Chrysler Imperial
Fiat 128
Merc 230
0.5
Cook's distance
0.00
0.10
0.20
0.30
Leverage