Effects on consumption of Automatic or Manual Transmissions

David Tonarini

03/02/2020

Executive summary

The effect of transmission on consumption, in Miles(US) / gallon, is analyzed. Data analysis indicates that cars with automatic transmission tend to have lower consumptions, reaching more Miles per gallon compared to cars with manual transmission.

Exploratory analysis

The analysis is done on the dataset *mtcars*, extracted from the 1974 Motor Trend US magazine, which comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models). The dataset contains information about 13 models with automatic transmission, and 19 models with manual transmission, for a total of 32 models. To understand the effect of transmission on consumption, we plotted the datapoints for the mpg of the two groups (Figure 1) and a boxplot with the quartile of the two groups (Figure 2), which suggests an effect of the car's transmission type on consumption. For cars with automatic transmission, the mean consumption is 24.39 mpg $\pm$ 6.17 For cars with automatic transmission, the mean consumption is 17.15 mpg $\pm$ 3.83

Modelling

We are going to create a linear model to predict how different parameters affect consumption, giving particular attention to transmission type. In order to assess which parameters to include, we checked how each of the variables in the dataset correlates with consumptions.

This results in the following correlation vector:

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

As we can see, a few variables in the dataset have a very high correlation with consumption (|rho| > 0.8). These variables and their respective correlations with consumption are Number of cylinders (-0.85), Displacement (-0.85), Weight (-0.87). On the other hand, these variables are also highly correlated one with an other, so they may not be all useful predictors within a linear model. The correlation of transmission type with consumption is 0.6, suggesting that its effect as a predictor is less than other variables.

Eight different linear models are tested. The first one uses only transmission type as a predictor.

The next three add to the model one of the three variables which are the most correlated to consumption.

The following two progressively add the three most highly correlated variables to the model.

Finally, the last model uses all the other variables in the dataset to predict mpg.

We run an aNova to compare the models, which results in the following:

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt
## Model 3: mpg ~ am + disp
## Model 4: mpg ~ am + cyl
## Model 5: mpg ~ am + wt + disp
## Model 6: mpg ~ am + wt + cyl
## Model 7: mpg ~ am + wt + disp + cyl
## Model 8: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 63.0133 9.325e-08 ***
## 3     29 300.28  0    -21.96
## 4     29 271.36  0     28.92
## 5     28 246.56  1     24.81  3.5318   0.07415 .
## 6     28 191.05  0     55.51
## 7     27 188.43  1      2.62  0.3732   0.54782
## 8     21 147.49  6     40.93  0.9713   0.46840
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The results indicate that the main benefit was obtained by including in the regression the effect of weight, which was the most thighly correlated variable. Adding further variables to the model had a lesser impact, but still helped to progressively reduce the Residual Sum of Squares. The simple model including all variables was found to be best, so we will choose this one, and observe the effect of transmission type on consumption from within this model.

Results

The coefficients for transmission type of the multivariate regression with the chosen model is 2.52 ± 2.06. Thus, when other parameters are held constant, a car which implements automatic transmission is expected to drive 2.52 extra miles per gallon compared to a similar car with manual transmission. At a 0.95 level of statitical significance, the confidence intervals are -1.6 and 6.64 As the interval contains zero, statistical significance at this level is thus not reached. Residual analysis and diagnostics do not indicate problems within our model (Figure 4 and 5)

Appendixes

Figure 1

Distribution of the values

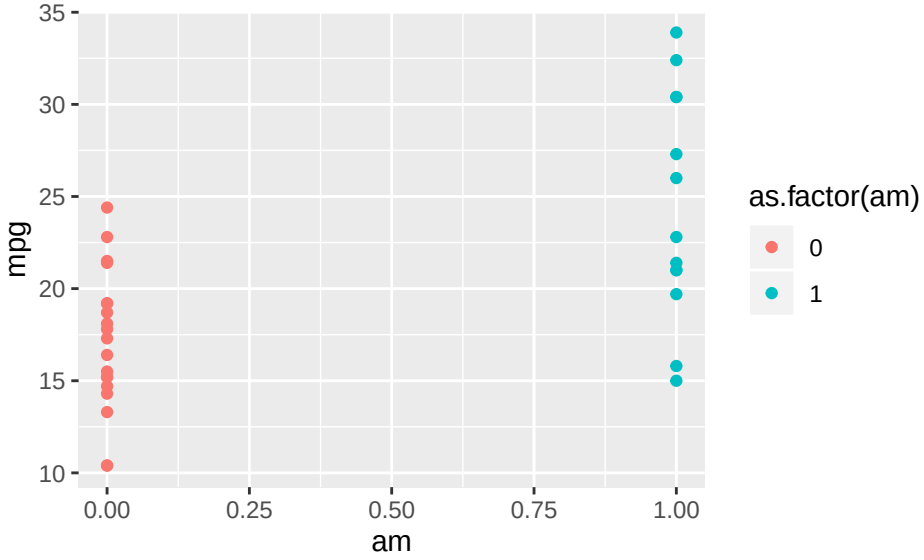

Figure 2

Boxplot to visualize the differences in the distribution of the values

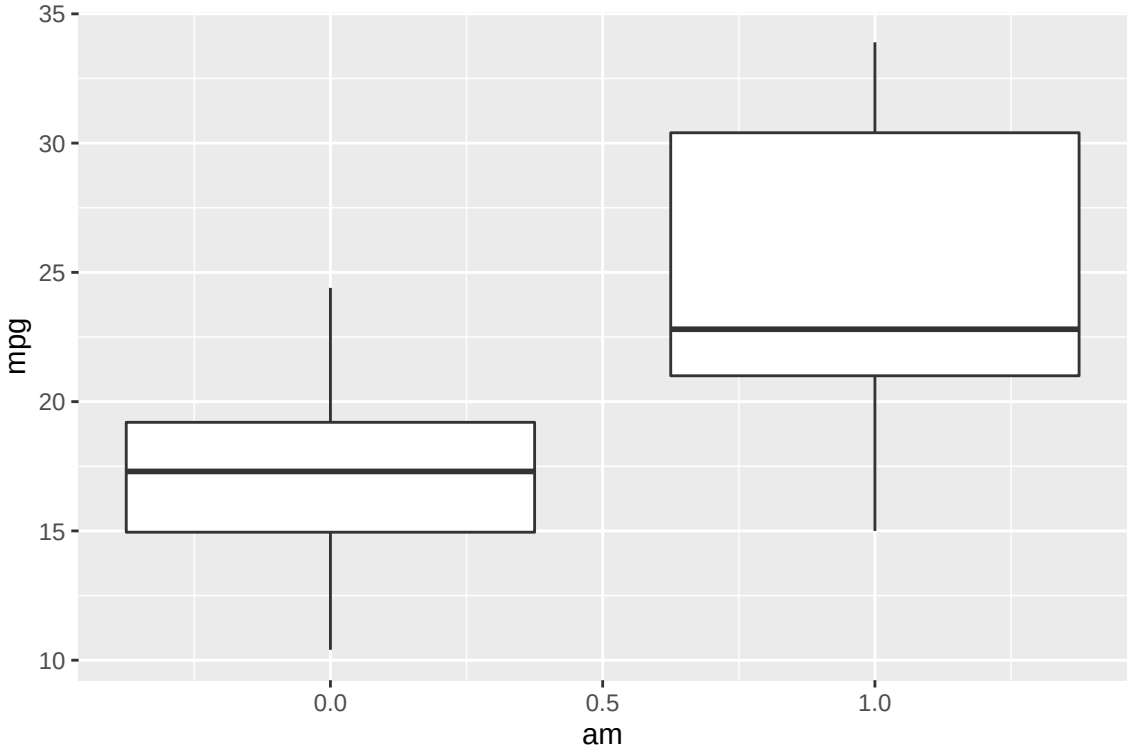

Figure 3

Model showing mpg plotted against the strogest predictor (weight), using color code for transmission type, and showing our chosen linear model with its confidence interval

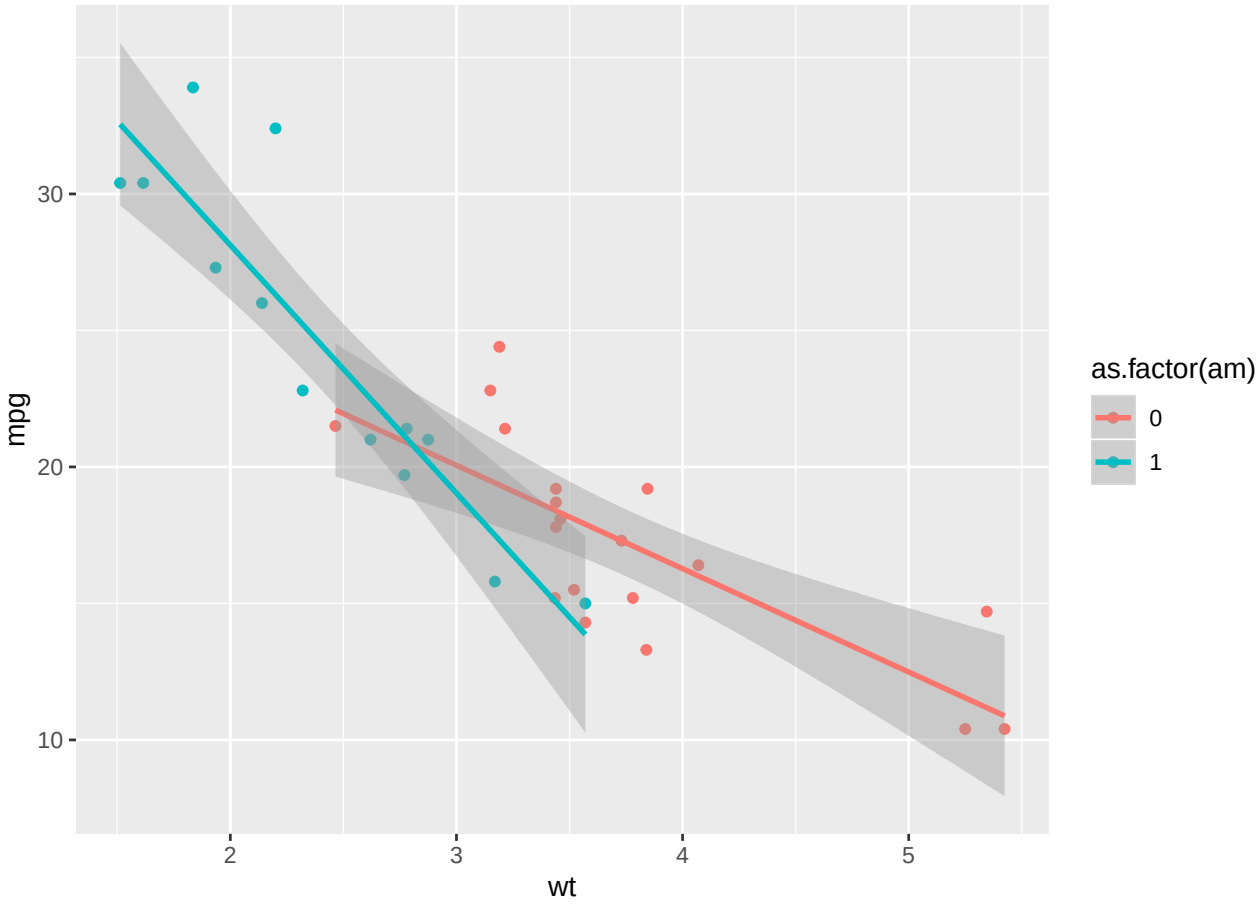

Figure 4

Plot of fitted values vs residuals

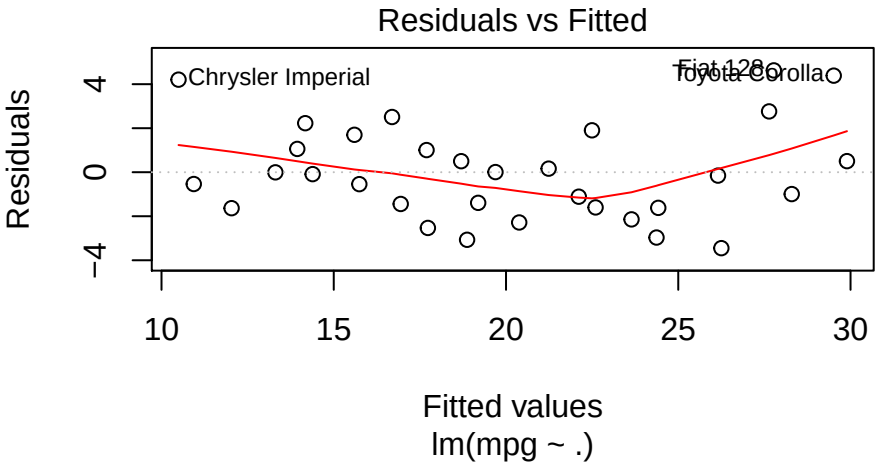

Figure 5

Scale-Location plot for analysis of standardized residuals