Sangar_RegressionModels

Vineet Sangar

August 21, 2015

Executive Summary: MotorCar Trend magazine are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome) and particularly the following two questions:

1. "Is an automatic or manual transmission better for MPG" Result: Analysis clearly shows that manual transmission provides better mpg.
2. "Quantify the MPG difference between automatic and manual transmissions" Result: Manual transmission vehicle on an average gives 7.425 more miles/gallon as compared to automatic transmission

___Exploring the data:___ The mtcars data were explored for determining the relationship between transmission and MPG. For this the transmission variable was converted into a factor variable and a box plot was generated. The boxplot (Fig 1) clearly showed that manual transmission had a higher mpg (y-axis) as compared to automatic transmission.

```
data(mtcars)
library(ggplot2)
library(GGally)
transmission<-factor(c("automatic", "manual"))
```

Regression models Analysis:

A regression model was fit in which "mpg" was the predicted and transmission was the regressor. The model presented the mean mpg for automatic transmission as 17.147 miles/gallon(intercept), however the manual presentation had higher (24.752) mean mpg (intercept + coeff = 17.147 + 7.425). The significance value ($p < 0.001$) showed that type of transmission certainly influences the mpg however, the lower R-squared (0.3598) indicated that transmission might not be the only influencer of mpg.

```
modelMpgAm<-lm(mpg ~ am, data = mtcars)
summary(modelMpgAm)
```

```
##
## 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 ***
## am             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
```

Understanding models through residual variation:

Exploring the residual variation (Fig 2) of the model clearly showed a pattern in which there was some linear relationship between the residuals of the model with the transmission. This suggests that change in mpg cannot entirely explained by transmission alone and confirms the low r-squred value of the model.

Exploring relationship between additional variables and mpg:

Correlation plot (Fig 3) suggested that all variables have some relationship with the mpg. So it will be worth generating linear regression models with adding more variables and examine if the model gets better at explaing the changes in mpg as influenced by other variables. So, two more models were generated.

modelMpgAll: uses all the variables as the regressors

modellMpgAll2 : uses all variables except am. This model was generated to explore if am is providing more accuracy to the model.

Performing anova on the all the models clearly showed that including all the variables as regressors made the model better (p value < 0.001)

```
modelMpgAll<-lm(mpg ~ ., data = mtcars)
modelMpgAll2<-lm(mpg ~ cyl + disp + hp + drat + wt + qsec + vs + gear + carb, data = mtcars)
anova(modelMpgAm, modelMpgAll, modelMpgAll2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
## Model 3: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + gear + carb
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     21 147.49  9    573.40 9.0711 1.779e-05 ***
## 3     22 158.04 -1    -10.55 1.5016     0.234
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appendix: Figure1: Boxplot showing differences in mpg (yaxis) as a factor of transmission(x-asis) You can also embed plots, for example:

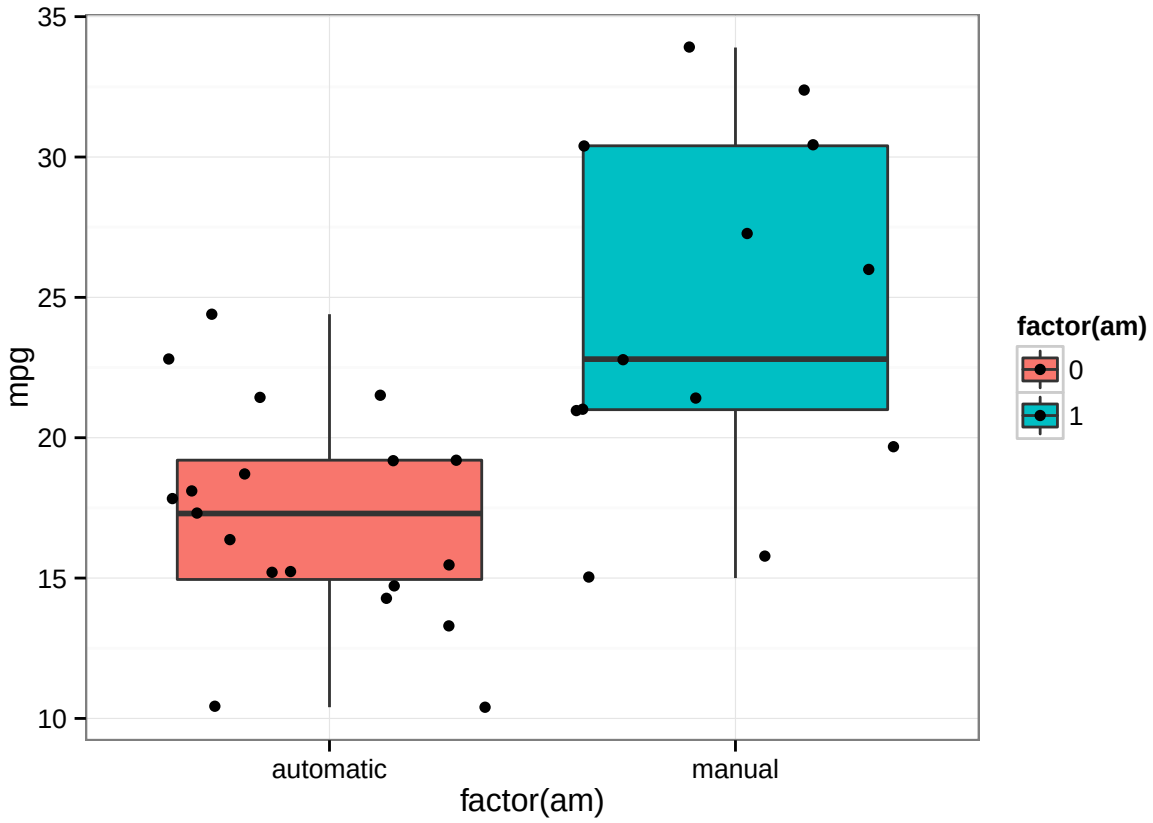

Figure 2: Plot presenting the relationship between residuals (only transmission) and the transmission.

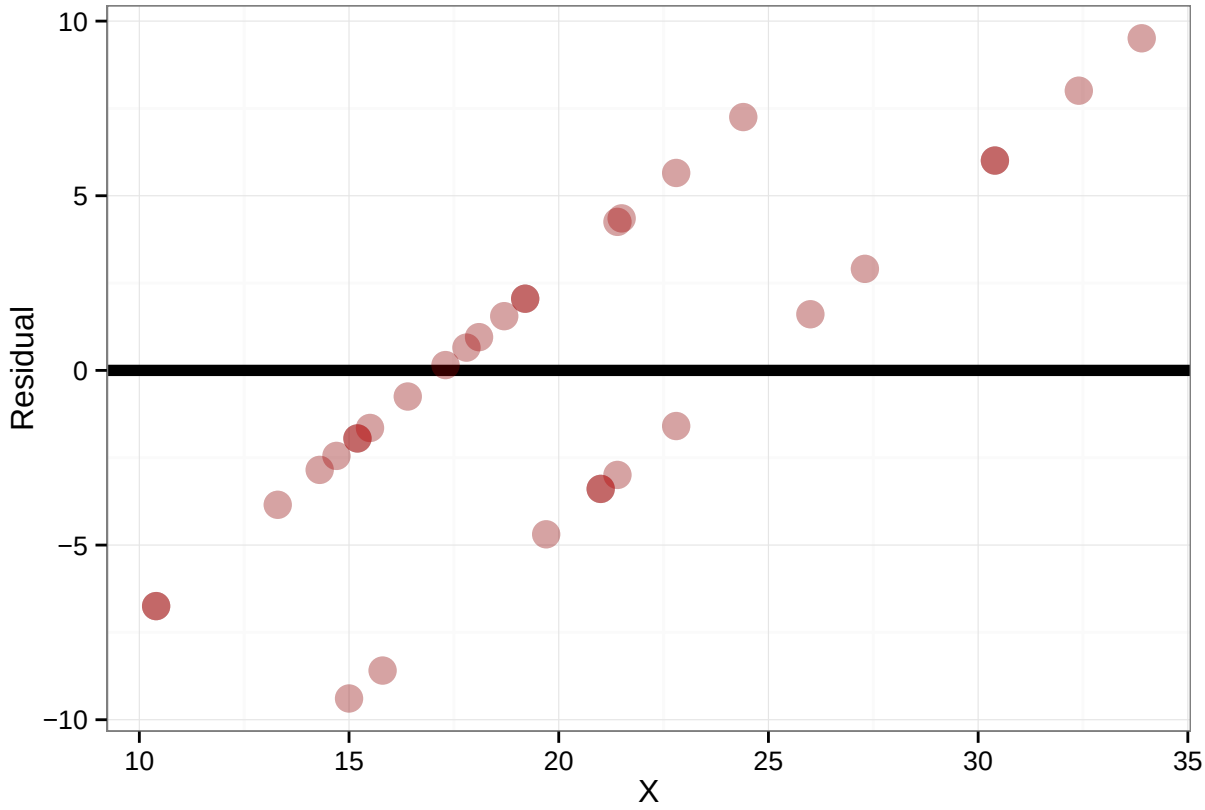

Figure3: Correlation plots of all variables.

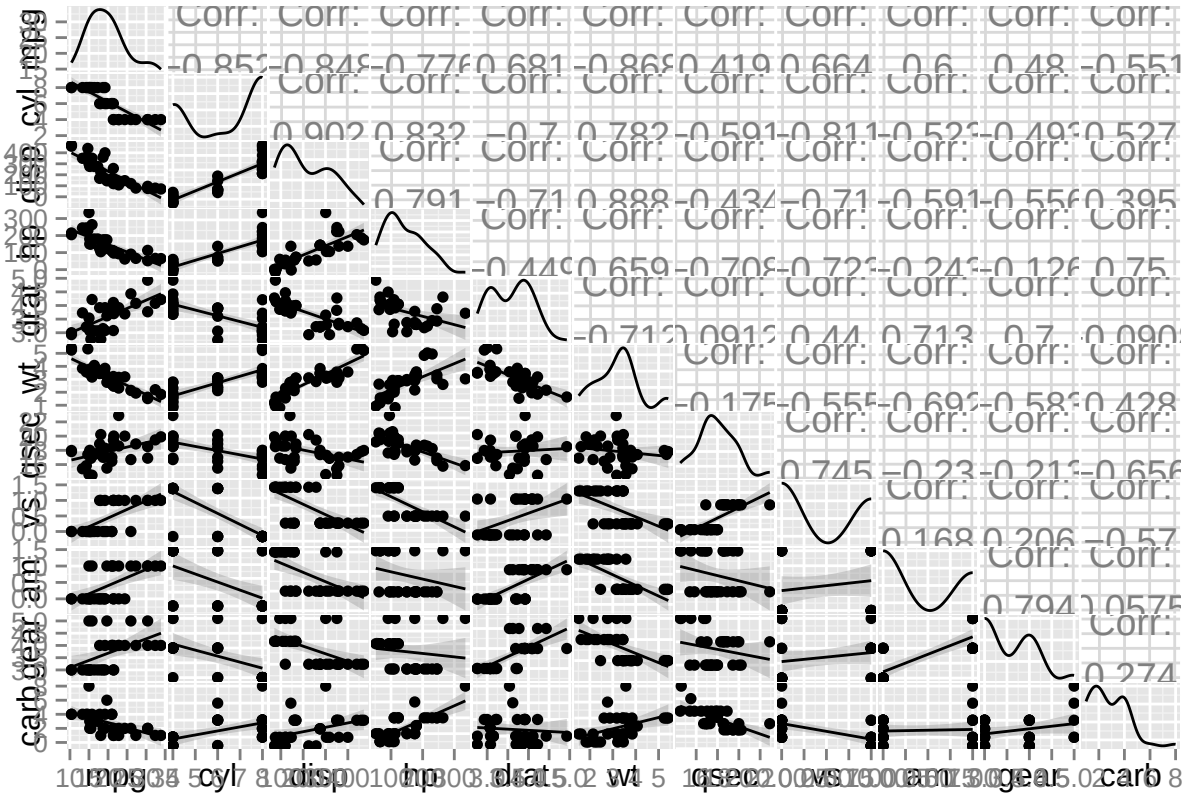

Figure 4: Plot presenting the relationship between residuals (all variable model) and transmission