

Regression models project

Executive summary

First of all, we load the required packages

```
library(ggplot2)
library(reshape2)
```

For convenience, we convert the 'am' variable, indicating the type of transmission, to a factor variable. The new variable will have two levels:

- 'auto': automatic transmission
- 'manual': manual transmission

```
mtcars$am = as.factor(1 - mtcars$am)
levels(mtcars$am) = c("manual", "auto")
mtcars$am
```

```
## [1] manual manual manual auto auto auto auto auto auto auto
## [11] auto auto auto auto auto auto auto auto manual manual manual
## [21] auto auto auto auto auto manual manual manual manual manual
## [31] manual manual
## Levels: manual auto
```

We fit a model for miles per gallon, using the type of transmission as the only predictor

```
model = glm(mpg ~ am, data = mtcars)
summary(model)
```

```
##
## Call:
## glm(formula = mpg ~ am, data = mtcars)
##
## Deviance 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)   24.392     1.360   17.941 < 2e-16 ***
## amauto        -7.245     1.764   -4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 24.02989)
##
##      Null deviance: 1126.0  on 31  degrees of freedom
## Residual deviance:  720.9  on 30  degrees of freedom
## AIC: 196.48
##
## Number of Fisher Scoring iterations: 2
```

The model suggests that the type of transmission is a good predictor for the miles per gallon variable. On average, the following results could be reported:

- 'manual' transmission: 24.392 mpg
- 'automatic' transmission: 17.147 mpg

For the convenience of the reader, a boxplot is attached in the Appendix.

Nonetheless, examining the correlation between the variables (in Appendix), the type of transmission seems to be negatively correlated to the weight of the car. That is, in the proposed dataset only 'big', heavy car are equipped with automatic transmission. Typically, the heavier the car, the lower the mpg.

If we fit another model for mpg, with both weight and transmission as predictors, we found out that the type of transmission is no longer significant, with a p-value of 0.988, whereas weight is a good predictor for miles per gallon.

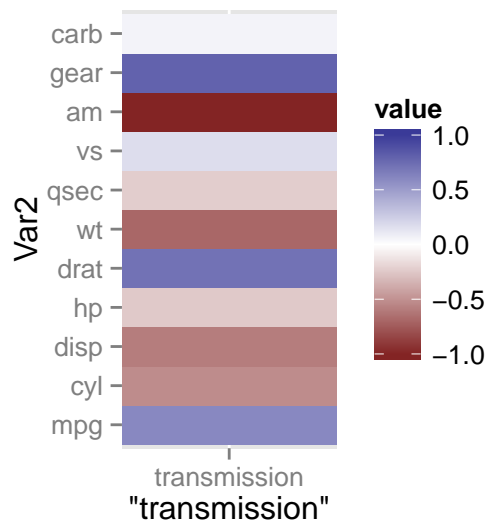
```
model = glm(mpg ~ am + wt, data = mtcars)
summary(model)

##
## Call:
## glm(formula = mpg ~ am + wt, data = mtcars)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.5295  -2.3619  -0.1317   1.4025   6.8782
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.29794    2.08566  17.883  < 2e-16 ***
## amauto        0.02362    1.54565   0.015   0.988
## wt           -5.35281    0.78824  -6.791 1.87e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 9.597231)
##
##      Null deviance: 1126.05  on 31  degrees of freedom
## Residual deviance:  278.32  on 29  degrees of freedom
## AIC: 168.03
##
## Number of Fisher Scoring iterations: 2
```

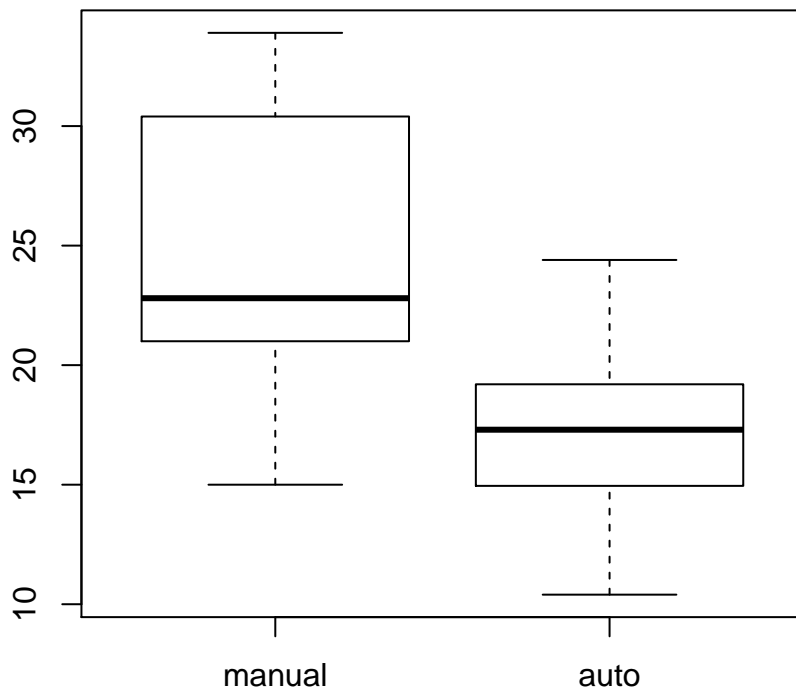
In conclusion, a naive analysis would highlight a strong correlation between the type of transmission and fuel consumption. Nonetheless, the dataset seems to be biased so that bigger cars have automatic transmission, whereas smaller ones have manual transmission. If weight is considered, the type of transmission is no longer significant in predicting the output variable.

Appendix

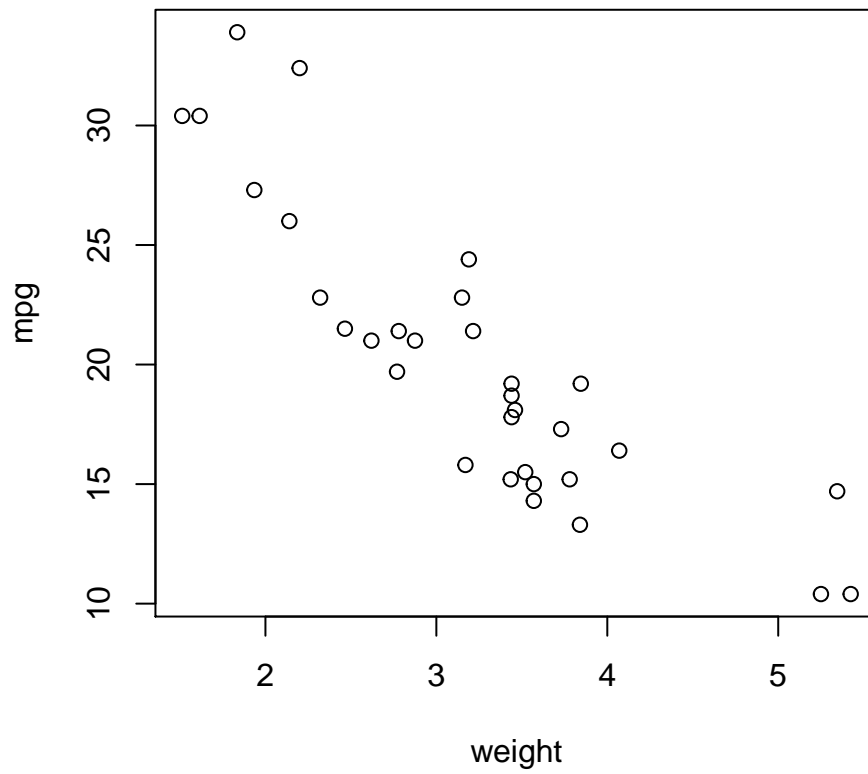
Correlation between type of transmission and other variables in the dataset.



Boxplot of miles per gallon as a function of the type of transmission:



Plot of miles per gallon as a function of the weight of the car:



Boxplot of car weight as a function of the type of transmission. Bigger cars tend to have automatic transmission:

