Automatic or Manual - Which is Better for MPG?

Carlos Schuler

Executive Summary

At **Motor Trends**, we have been asked many times the question:

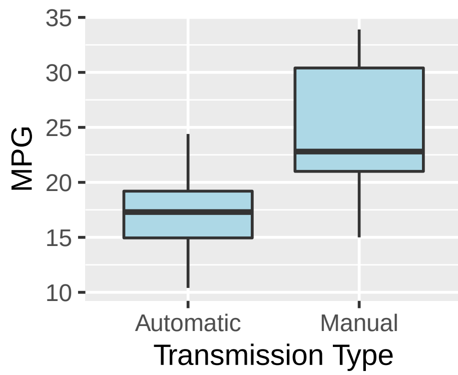
> *"Is an automatic or manual transmission better for MPG"*?

We will address this question by exploring the relationship of MPG to multiple automobile variables, such as number of cylinders, vehicle weight, block shape, etc. This will allow us to quantify the MPG difference between automatic and manual transmissions.

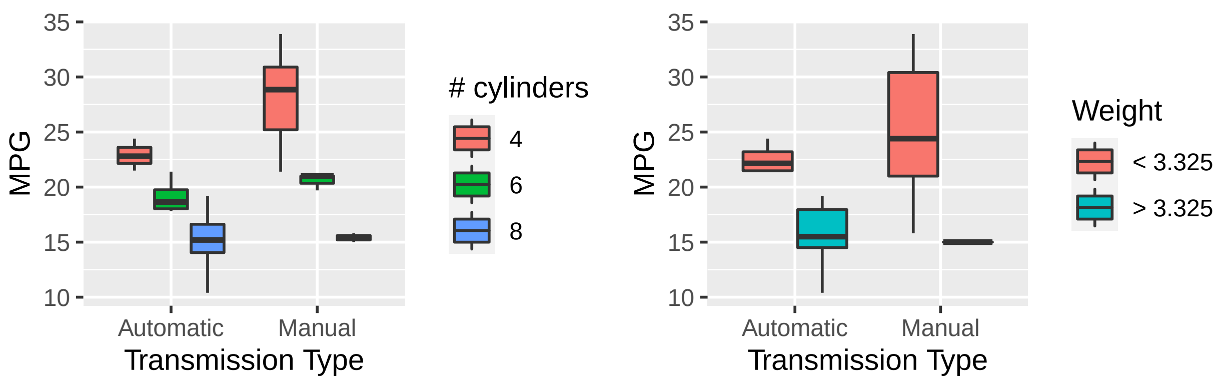
Although it would appear initially that a manual transmission yields better MPG, it turns out the answer is "*it depends*" on what other features the car has.

Data Exploration

We will use the **mtcars** dataset, described in the Appendix. From this dataset, we can plot how MPG depends on transmission type ("code chunks" are shown in the Appendix):



From this figure, it would appear that a manual transmission results in better MPG. Things look diferent, however, if we group the data by number of cylinders, or by vehicle weight:



For 4-cyl vehicles, a manual transmission appears to yield higher MPG. However, that appears not to be the case for 8-cyl vehicles. Also, for lighter vehicles (weighing less than the median for the sample), the relationship between MPG and transmission type appears not to be significant (*note*: there is only one vehicle weighing higher than the median weight with an manual transmission, the **Maserati Bora**. Its MPG does not appear to be better than other vehicles in its weight class.)

Regression Model

We will use stepwise (forward + backward) regression to find the subset of variables in the data set based on the *Akaike information criterion* (AIC) resulting in the best performing model (see *http://www.sthda.com/english/articles/37-model-selection-essentials-in-r/154-stepwise-regression-essentials-in-r/*)

```
mtcarsF <- mtcarsF %>% select(-wtQ) #delete the wtQ field we added earlier
baseModel <- lm(data=mtcarsF, mpg ~ .)
optimalModel <- step(baseModel, direction= "both")
```

The step function sequentially eliminated *carb*, *gear*, *drat*, *disp*, *qsec*, and *vs* as model parameters. The start AIC was 76.4, and it was reduced to AIC=61.65; the final model is *mpg ~ cyl + hp + wt + am*. We anticipated *wt* and *am* as possible important factors in the model. We also anticipated (see Appendix) the elimination of several highly correlated variables. The final model parameters are shown in the Appendix. The coefficients for the model are: *(Intercept)=33.71, cyl6=-3.03, cyl8=-2.16, hp=-0.03, wt=-2.5, amManual=1.81*. This model explains about 87% of the variance. The *(Intercept)* represents the predicted MPG for the base configuration: *4-cyl*, *hp=0*, *wt=0*, *automatic*. For the continuous variables, *hp* and *wt*, their coefficients represent the change in MPG per unit change in that variable, holding all other factors in the model constant. For the factor variables, *cyl* and *am*, their coefficients represent the change in MPG when the factor changes to that configuration from the base configuration.

Thus, the model predicts a change of 1.81 in MPG when switching from *automatic* to *manual* transmission, that is an increase. However, this change is not to be statistical significant, at p=0.2 (see full model results in the Appendix.) Thus, according to this model, the difference in MPG is not significantly different from zero, when switching from *automatic* to *manual* transmission, **neither is BETTER according to this model!**

We briefly explore the effect of interactions:

```
newModel <- lm(data=mtcarsF, mpg ~ cyl + hp + wt + am + wt*am)
```

The coefficients for this new model are: *(Intercept)=30.65, cyl6=-2.38, cyl8=-2.9, hp=-0.02, wt=-2.21, amManual=9.9, wt:amManual=-3.14*. This model explains about 88% of the variance. **This new model predicts an increase of 9.9 in MPG switching from *automatic* to *manual* transmission**, significant at p=0.03. This matches our observation that automatic-cars are heavier with have lower MPG, while manual-cars are lighter and have higher MPG. The *wt:amManual= -3.14* negative change in slope term shows that *wt* has a higher impact on MPG for manual cars than for cars with an automatic transmission.

Residuals and Diagnostics

Residuals and other diagnostic plots are included in the Appendix for our final model, including interactions. No patterns are apparent in the residuals and the errors appear to be normally distributed. Data for **Maserati Bora** appears to have unusual leverage. However, as this is the only sample of a *heavy* car with a *manual* transmission, we have decided to keep this sample in the dataset.

Conclusion and Limitations

It appears that **Manual transmissions yield higher MPG in light cars, but not in heavy cars**, with the following caveats:

- The sample size studied is small
- Today's vehicles have significant new technologies that may change the outcome
- The population selected is not random, and thus may be biased
- The analysis assumes t-distribtions for the models

APPENDIX

Data was extracted from the 1974 Motor Trend magazine: MPG and 10 design and performance aspects for 32 1973-74 automobiles:

Field Name	Description
mpg	Miles/(US) gallon
cyl	Number of cylinders
disp	Displacement (cu.in.)
hp	Gross horsepower
drat	Rear axle ratio
wt	Weight (1000 lbs)
qsec	1/4 mile time in seconds
vs	Engine (0 = V-shaped, 1 = straight)
am	Transmission (0 = automatic, 1 = manual)
gear	Number of forward gears
carb	Number of carburetors

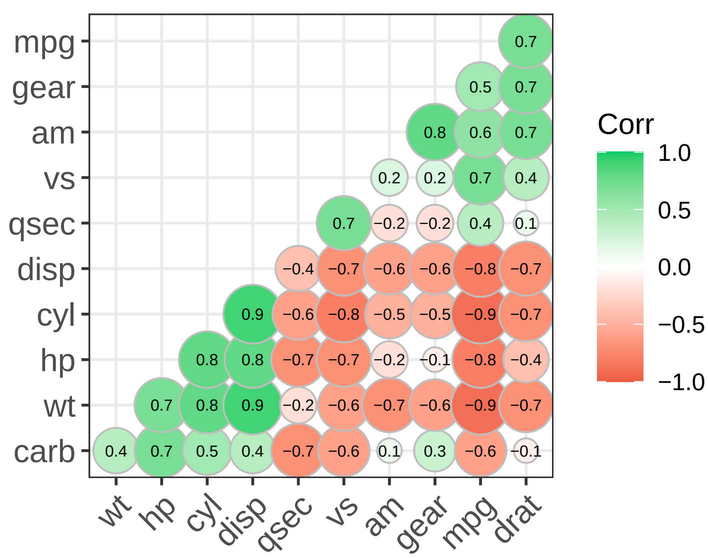
Data Exploration

The following code chunk generates the figures shown in the main report:

```
library(tidyverse); data("mtcars"); mtcars <- as_tibble(mtcars); mtcarsF<-mtcars
mtcarsF$cyl <-  factor(mtcarsF$cyl); mtcarsF$vs <-   factor(mtcarsF$vs,labels = c("V block","S block"))
mtcarsF$am <-   factor(mtcarsF$am,labels = c("Automatic","Manual")); mtcarsF$gear <- factor(mtcarsF$gear)
mtcarsF$carb <- factor(mtcarsF$carb)
wtMedian <- median(mtcarsF$wt); lightWeight=paste("<",wtMedian); heavyWeight=paste(">",wtMedian)
mtcarsF <- mtcarsF %>% mutate(wtQ=factor(ntile(wt,2),labels=c(lightWeight,heavyWeight)))
rownames(mtcarsF)<-rownames(mtcars)
ggplot(mtcarsF, aes(x=am,y=mpg)) + geom_boxplot(fill="lightblue") + labs(x="Transmission Type", y="MPG")
p2 <- ggplot(mtcarsF, aes(x=am,y=mpg)) + geom_boxplot(aes(fill=factor(cyl))) +
        labs(x="Transmission Type", y="MPG", fill = "# cylinders")
p3 <- ggplot(mtcarsF, aes(x=am,y=mpg)) + geom_boxplot(aes(fill=factor(wtQ))) +
        labs(x="Transmission Type", y="MPG", fill="Weight")
library(gridExtra); grid.arrange(p2,p3,nrow=1)
```

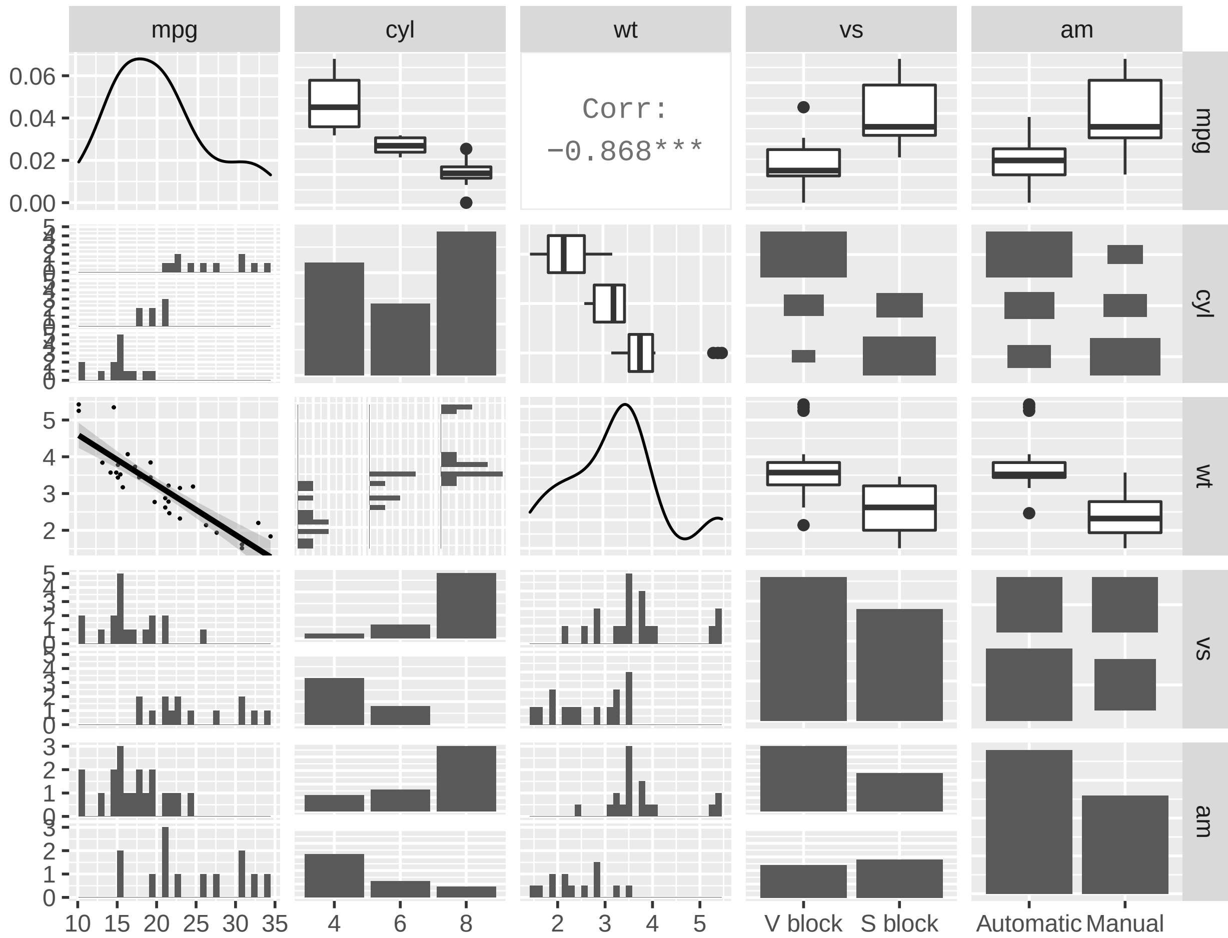
Many of the fields in mtcars are highly correlated to each other:

```
corr <- round(cor(mtcars), 1); library(ggcorrplot)
ggcorrplot(corr, hc.order = TRUE, type = "lower", lab = TRUE, lab_size = 2, method="circle",
           colors = c("tomato2", "white", "springgreen3"), ggtheme=theme_bw)
```



Eliminating fields that are highly cross-correlated, we can visualize some of the principal pair-wise relationships in the data:

```
library(GGally)
mtc<-mtcarsF %>% select(mpg, cyl, wt, vs, am)
ggpairs(mtc, lower = list(continuous = wrap("smooth", method = "lm", size=0.1)))
```



The optimal fit

```
summary(optimalModel)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcarsF)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## amManual     1.80921    1.39630   1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

Fit including wt*am interaction

```r
summary(newModel)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am + wt * am, data = mtcarsF)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6778 -1.5679 -0.5546  1.1531  4.7990
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 30.65247    2.90991  10.534 1.11e-10 ***
## cyl6        -2.38062    1.37364  -1.733   0.0954 .
## cyl8        -2.89911    2.19666  -1.320   0.1989
## hp          -0.01782    0.01484  -1.200   0.2412
## wt          -2.20687    0.85205  -2.590   0.0158 *
## amManual     9.89860    4.28571   2.310   0.0294 *
## wt:amManual -3.14499    1.58475  -1.985   0.0583 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.284 on 25 degrees of freedom
## Multiple R-squared:  0.8841, Adjusted R-squared:  0.8563
## F-statistic: 31.79 on 6 and 25 DF,  p-value: 1.555e-10
```

Residuals for Final Model, including interaction

```r
par(mfrow=c(2, 2))
plot(newModel)
```