Regression

Summary

I work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, my editor interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). He is particularly interested in the following two questions:

- "Is an automatic or manual transmission better for MPG"
- "Quantify the MPG difference between automatic and manual transmissions"

A Look at the coefficients

Let's take a quick look at the coefficients shows us that the average mpg for automatic transmissions (in this case the intercept since automatics are given a default value of zero, and manuals one):

```
summary(lm(mpg ~ am, data=mtcars))$coef
```

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

From this we can see that the automatics have an average of ~17 (since they are the intercept), while the manuals have an average of ~24 (intercept plus am) Please see lecture on dummy variables time stamp 20:49 if this does not make sense.

So clearly, manual transmissions are clearly better in terms of mileage, right? Well, let's take a closer look.

Exploratory Data Analysis

First, let's take a look at a graph of vehicles, divided by transmission types plotting weight (a likely factor in a car's mileage) and miles per gallon with the number of cylinders divided by color. What do we see? It appears that almost all the manuals weigh less, and that at similar weights automatics perform better. Please see figure one in appendix.

Models (and their Selection)

A quick look at the variables from mtcars yields the following:

```
names(mtcars)
```

```
##  [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"
## [11] "carb"
```

Now if we compare these predictors in regards to mpg, we see the following

```r
sort(abs(coef(lm(mpg~., mtcars))), decreasing = TRUE)
```

```
## (Intercept)          wt          am        qsec        drat        gear
## 12.30337416  3.71530393  2.52022689  0.82104075  0.78711097  0.65541302
##          vs        carb         cyl          hp        disp
##  0.31776281  0.19941925  0.11144048  0.02148212  0.01333524
```

Multiple Models Fit

From this, we can see that weight is far and away the best predictor, but, let's take a look at some multivariable models and their residuals as well:

```
##   Model Type     Mean of Residuals
## 1     Weight -5.11472374870037e-17
## 2     wt+cyl  5.55111512312578e-17
## 3    wt+drat -8.67361737988404e-18
## 4    wt+gear -5.20417042793042e-18
```

What we see is that all of the results are have tiny mean residuals, indicating to me that weight alone should be sufficient as a predictor. Using the linear regression with weight as a predictor and mpg as a product we can examine the residual plot in the appendix (figure 2).

Now, lets take a look at the models themselves :

```
## [1] "Automatics:"
```

```
## (Intercept)          wt
##   31.416055   -3.785908
```

```
## [1] "Manuals:"
```

```
## (Intercept)          wt
##   46.294478   -9.084268
```

Certainly appears to be food for thought, we will discuss these in the next section.

Conclusions

Looking at our models, we see that although the the intercept for manuals is very high, so is the slope. While with automatics, the intercept starts signifigantly lower but the slope is much lower as well. What this means is that at low weights manuals will tend to have better gas mileage, but as the number of pounds increases automatics tend to do better Examining our residual plot (with fitted models and confidence intervals), we see that this seems to be the case. What does this mean? Well, it means that the original question could have been more specific. If you are looking for a car and the sole criteria is gas mileage, you will want a manual transmission. However, perhaps you are looking for something with a bit more leg room. In that case, the odds are that you want an automatic.

What this means, at the very least is that the question is incomplete and that we should probably do articles (lightweight value autos and midsize or bigger) or find a new topic. Time to let the editor know.

Appendix

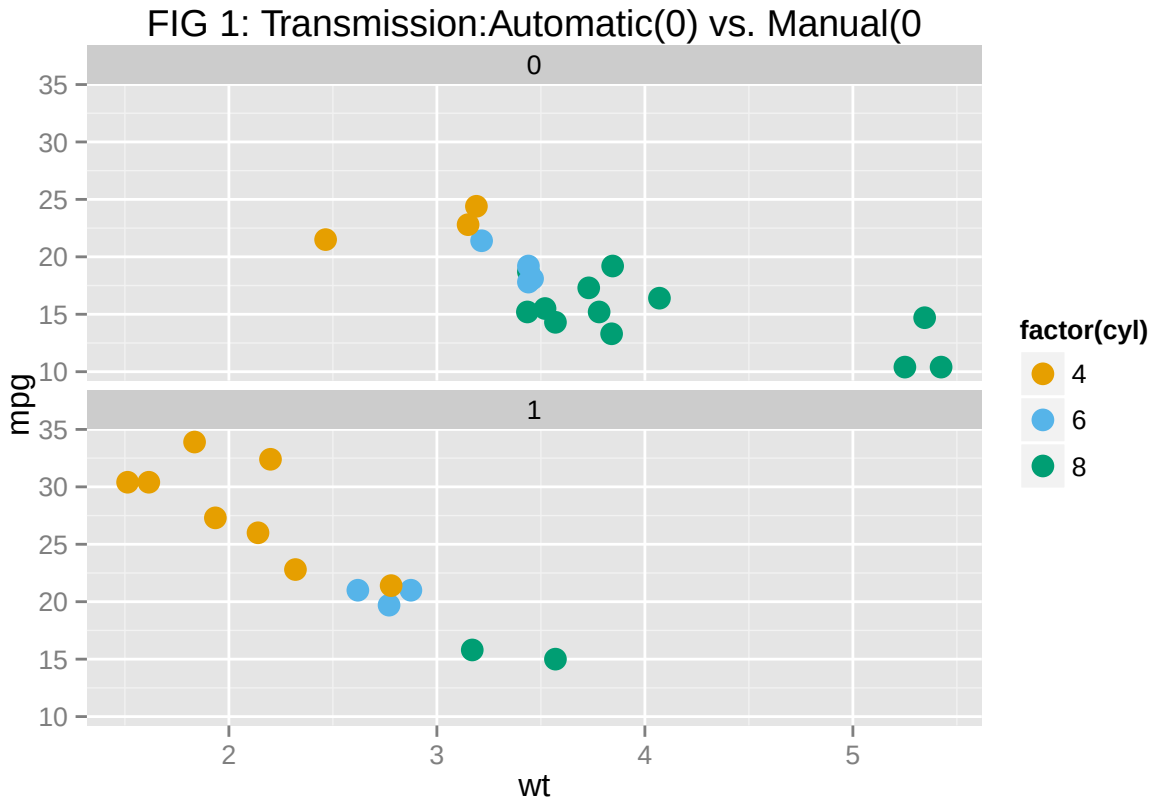
FIG 1: Transmission:Automatic(0) vs. Manual(0
0
1
mpg
wt
35
30
25
20
15
10
2
3
4
5
factor(cyl)
4
6
8

FIG 2: MPG vs. Weight model, plus residuals