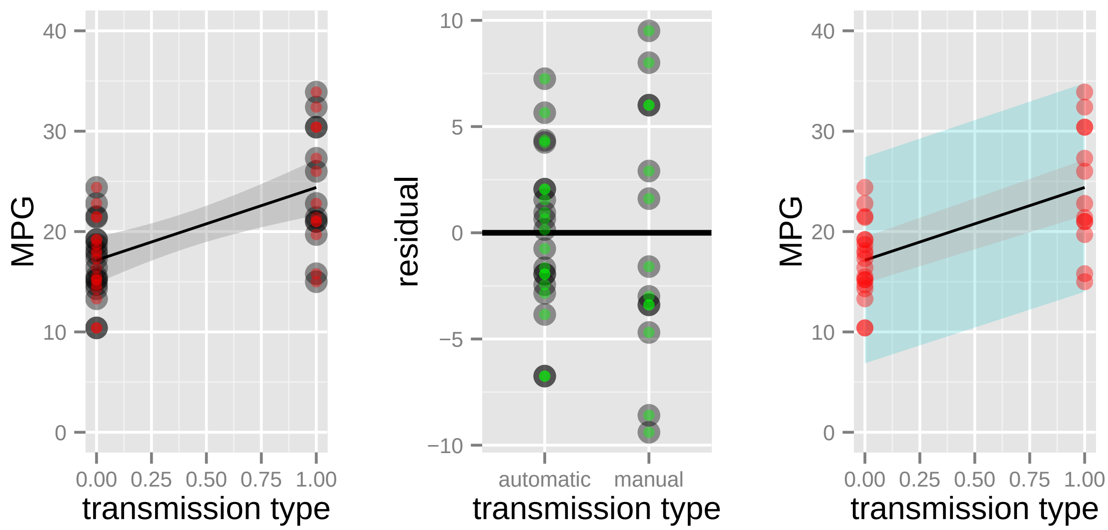
Regression Model Course Project

Liang

July 25, 2015

In this project, I studied the **mtcars** data-set. The two questions I tried to answer are "Is an automatic or manual transmission better for MPG" and "Quantify the MPG difference between automatic and manual transmissions". The R Makrdown file can be find here: https://github.com/LiangGao/MarkDownFiles. In spite of the seemingly linear relation between transmission type and MPG, the transmission type might be the least important feature affecting the MPG, and it failed to give a reasonable prediction for the MPG difference between two transmission types. While we cannot claim either transmission is better for MPG, the vehicle weight is the true feature determining the vehicle MPG, and the variable hiding in the background of the relation between MPG and transmission type.

I started with a simple linear fitting between MPG and transmission type (0 = automatic, 1 = manual). The result of the fitting, the residual, and the confidence (pink) + prediction (teal) interval figure are as following:



Although there is a seemingly linear relationship between the transmission type and MPG, the residual is too large to be a good model.

To quantify the MPG difference between automatic and manual transmissions, I could run the following code

```
fit0 = lm(mpg ~ factor(am), data = mtcars)
sumCoef = summary(fit0)$coefficients
sumCoef[2,1] + c(-1, 1) * qt(0.975, df = fit0$df) * sumCoef[2,2]
```

```
## [1]  3.64 10.85
```

The result indicate that the MPG difference between automatic and manual transmission vehicles is between 3.64 and 10.85 (95% confidence interval). Considering the standard deviation of the MPG itself is about 6.03. This answer really doesn't provide any useful information.

Next I performed the same liner fitting to other possible predictor variables, such as the number of cylinders (cyl), horsepower (hp), weight (lb/1000), and number of gears (gear), and wrote down the R-squared values and p-values:

```
##              transmission      cyl       hp       wt   gear
## R-squared        0.359799 7.26e-01 6.02e-01 7.53e-01 0.2307
## p-value          0.000285 6.11e-10 1.79e-07 1.29e-10 0.0054
```

It seems transmission is only better than gear when modeling the MPG. Everything considered, wt should be the best choice.

I compared the two situations as following:

1, to fit MPG (outcome) and transmission (am, predictor), then add wt:

```r
fit10 = lm(mpg ~ factor(am), data = mtcars)
fit11 = update(fit10, mpg ~ factor(am) + wt)
anova(fit10, fit11)$Pr[2]
```

```
## [1] 1.87e-07
```

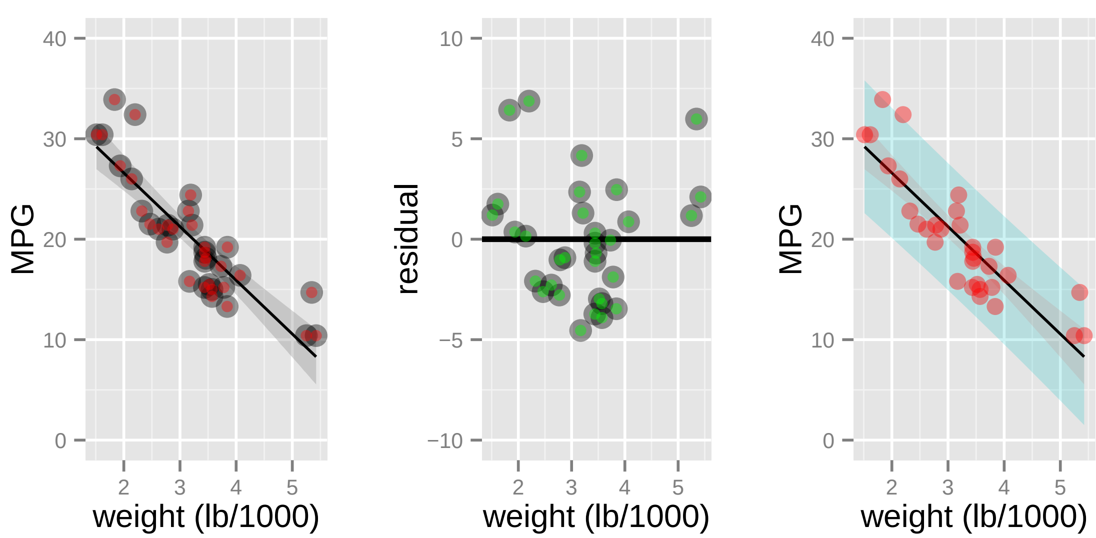
2, to fit MPG (outcome) and wt (predictor), then add am:

```r
fit20 = lm(mpg ~ wt, data = mtcars)
fit21 = update(fit20, mpg ~ wt + factor(am))
anova(fit20, fit21)$Pr[2]
```

```
## [1] 0.988
```

The large p value (0.988) of the second model indicates that it is unnecessary to add transmission type if the weight has been included in the model. On the other hand, the very small p value (~0) for the first situation means we should add the wt as another predictor in addition to transmission.

Below are the fitting, the residual, and the confidence (pink) + prediction (teal) interval plot for linear regression model of MPG ~ weight:



Apparently, vehicle weight is a proper predictor for MPG. The reason behind the correlation between the transmission type and MPG is probably that the vehicles with auto transmission were heavier during the time period (70s) when the data were taken.