RegMods Course Project - MPG analysis based on transmission type

cafkhami

Executive Summary

Using regression models and statistical inference this paper answers the following questions:

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

The analysis will show that `Manual` transmission vehicles have better MPG. and that we are 97.5% certain that the increase in MPG from automatic to manual is greater than `3.21` MPG for a given vehicle.

Problem Definition

You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are 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"

We will use the `mtcars` dataset in order to answer these questions to the best of our ability.

Exploratory Analysis

Let's explore the `mtcars` dataset to attempt to gain some intuition about the questions posed.
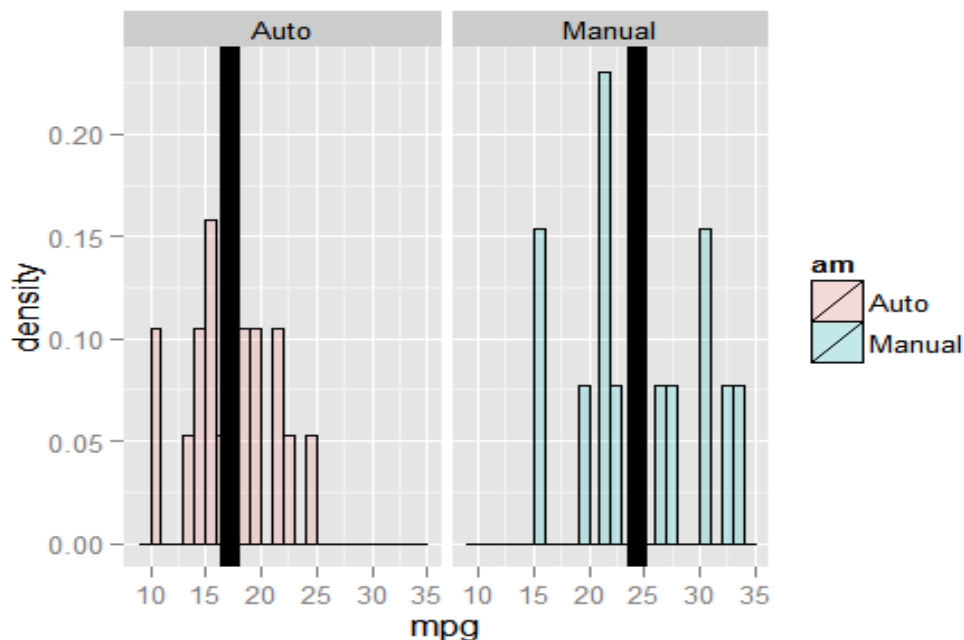
These histograms with the averages shown by the black vertical line help us to understand the intuition about the questions. My initial observations are that the average mpg is higher for the manual vehicles than for the automatics, but the automatics have a more consistent mpg across vehicles while the manuals have a greater spread.

Students T.Test

Let's perform a t.test on the mpg of Auto and Manual vehicles, testing the hypothesis that the difference between the two is greater than 0.

```
mpgAuto <- mtcars$mpg[mtcars$am == 0]
mpgMan <- mtcars$mpg[mtcars$am == 1]
t.test(mpgMan,mpgAuto, paired = F)

##
##   Welch Two Sample t-test
##
## data:  mpgMan and mpgAuto
## t = 3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##    3.209684 11.280194
## sample estimates:
## mean of x mean of y
##   24.39231  17.14737
```

Again, this test confirms our intuition. The 95% confidence interval does not contain 0 and the p-value for the test is very small (<.005). These two things allow us to also conclude that Manual vehicles have higher mpg than automatic vehicles.

Regression Models

Let's evaluate a couple regression models to attempt to further quantify the difference in mpg between automatic and manual vehicles. First lets simply consider the regression model of transmission type, to mpg.
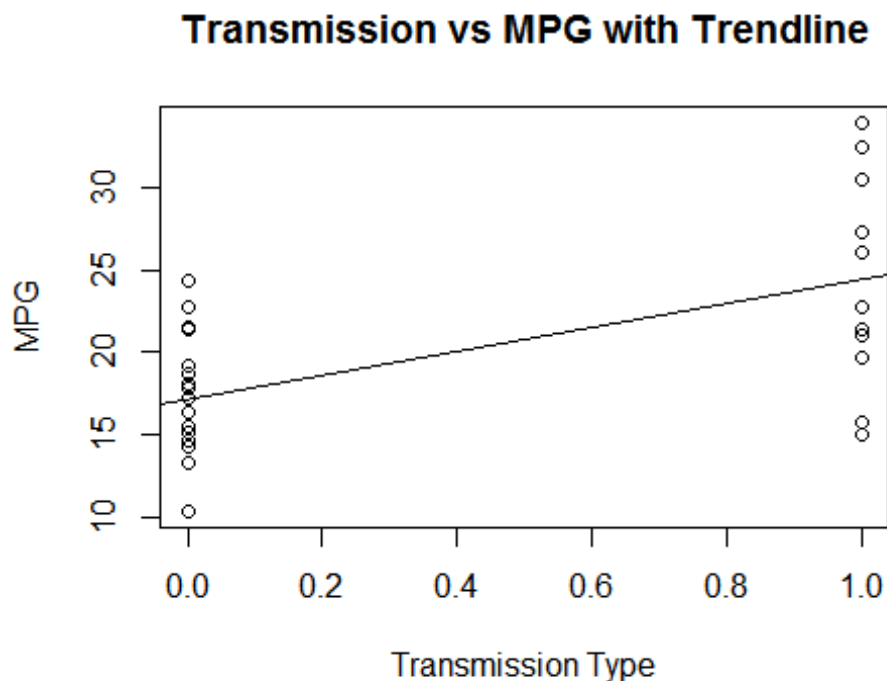
```
fit1 <- lm(mpg ~ factor(am),data = mtcars)
summary(fit1)$coeff

##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## factor(am)1  7.244939   1.764422  4.106127 2.850207e-04

summary(fit1)$r.squared

## [1] 0.3597989

plot(mtcars$am, mtcars$mpg, xlab="Transmission Type",ylab="MPG",
main="Transmission vs MPG with Trendline")
abline(fit1)
```


This model shows a positive correlation between transmission type and mpg, but it is simplistic and simply shows the difference between the means of the two groups. Thus it suggests that switching from automatic to manual will provide a 7.2 mpg increase but the model only accountsfor ~36% of the variation in MPG (R^2 = .3598).

Let's consider a slightly more advanced model that also considers horsepower as an input variable.

```
fit2 <- lm(mpg ~ hp + factor(am),data = mtcars)
summary(fit2)$coeff

##               Estimate  Std. Error   t value      Pr(>|t|)
## (Intercept) 26.5849137 1.425094292 18.654845 1.073954e-17
## hp          -0.0588878 0.007856745 -7.495191 2.920375e-08
## factor(am)1  5.2770853 1.079540576  4.888270 3.460318e-05

summary(fit2)$r.squared

## [1] 0.7820346

## Warning in abline(fit2): only using the first two of 3 regression
## coefficients
```
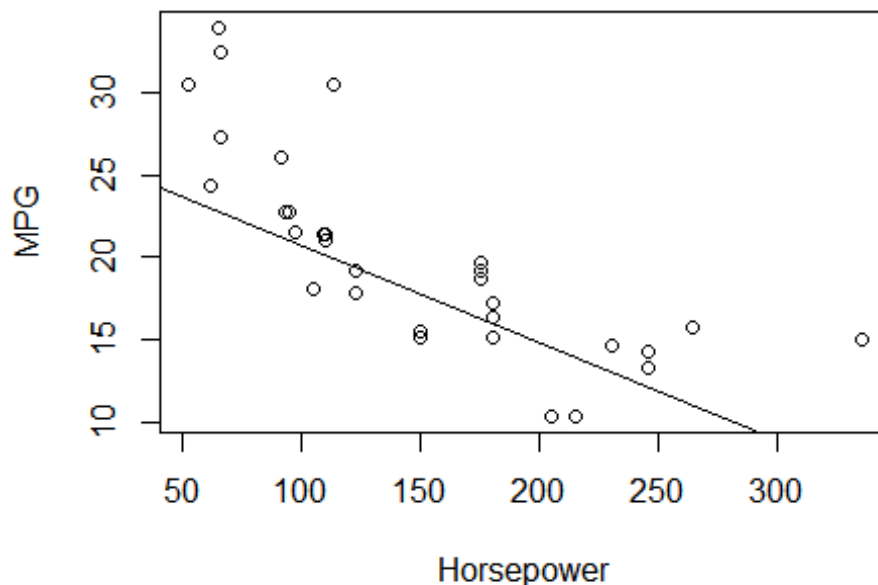
Transmission vs MPG with Trendline

This model shows that for a given horsepower, we can expect to see an increase of about 5.28 mpg when evaluating a manual vehicle over an automatic. This model also accounts for ~78% of the variation in MPG, which is more that twice that of the previous model.

While there may be more advanced models that better represent the difference and account for an even larger percentage of MPG variation, we will accept this standard of model for answering the posed questions. To verify that this model doesnt have additional dependencies that we cannot see, lets look at the residual plot.

Residual of fit2 plot vs Horsepower
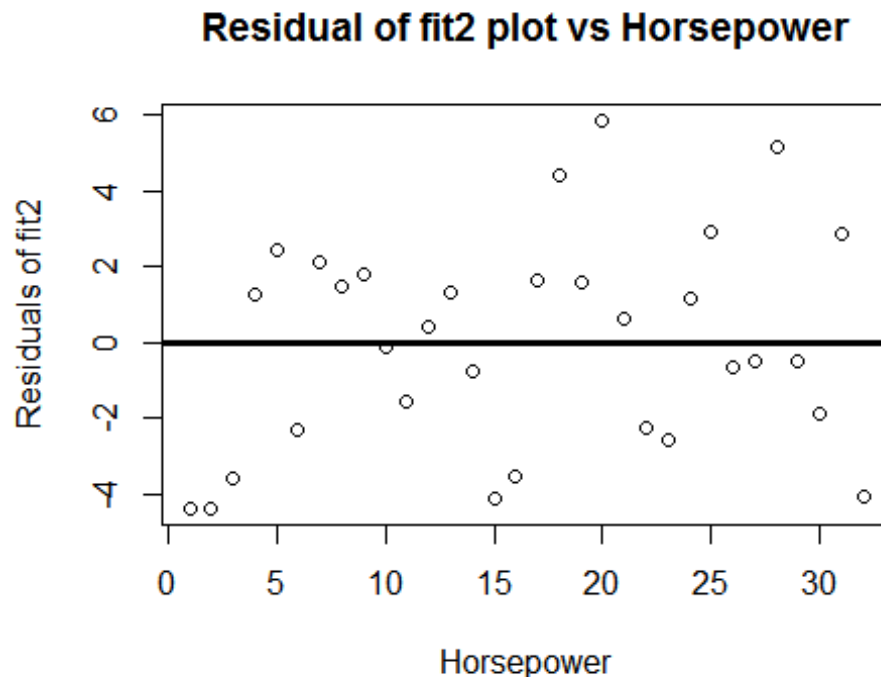
Finally, let's perform a T Test to determine the uncertainty in our estimate of 5.28 MPG increase

```
t.test(mpgMan,mpgAuto, paired = F, mu = 5.28)$p.value
```

```
## [1] 0.3202165
```

```
t.test(mpgMan,mpgAuto, paired = F, mu = 5.28)$conf
```

```
## [1]  3.209684 11.280194
## attr(,"conf.level")
## [1] 0.95
```

This large of a pvalue will tell you that there is some uncertainty in our measurement. Therefore let's use a more certain response. The lower end of the 95% confidence interval of the difference in the means of the populations, 3.21 MPG.

Results

- "Is an automatic or manual transmission better for MPG"

The answer is Manual transmission.

- "Quantify the MPG difference between automatic and manual transmissions"

Using our model that also considers horsepower to have a significant effect on MPG the expected increase for a vehicle of the same horsepower when switching to a Manual transmission from automatic is 3.21 MPG.