Regression Models Course Project

Executive Summary

Based on data gathered from Motor Trends magazine about 32 cars, model year 1973-74, manual transmission cars have much better mileage per gallon than automatic transmission cars, by about 7.245 miles per gallon. Even after accounting for the fact that manual transmission cars tend to have less horsepower, most of the mileage gains remain.

Report

Description of Data

The data that we are using came from Motor Trend magazine in 1974. It details 32 cars from the model year 1973-74. Obviously, this data set is quite old and given the advances in automobile technology, we shouldn't draw strong conclusions about modern cars from 40-year-old data. Furthermore, we don't know how these cars were chosen, so we shouldn't assume that they are a representative sample of every available model of car. Finally, even if data shows a difference in mileage based on transmission, it only shows correlation, it doesn't show causation.

Luckily, this data does include a good divide between each type of transmission, with 59% (19/32) automatic transmission and 41% (13/32) manual.

```
manual <- 1
auto <- 0
sum(mtcars$am==manual)
```

```
## [1] 13
```

```
sum(mtcars$am==auto)
```

```
## [1] 19
```

Mileage by Transmission Type

When you regress miles per gallon by transmission type, the mileage difference is quite pronounced.

```
fit <- lm(mpg ~ am, data=mtcars)
summary(fit)$coefficients
```

```
##             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
```

The coefficient on the transmission type is `fit$coefficients[2]`, which implies that cars with manual transmission have much better mileage. Note that this number is simply the difference in the mean mileage for manual transmission cars and automatic ones.

```
mean(mtcars$mpg[mtcars$am==manual]) - mean(mtcars$mpg[mtcars$am==auto])
```

```
## [1] 7.244939
```

This is a large difference, it is a `fit$coefficients[2] / mean(mtcars$mpg[mtcars$am==auto]) * 100` percentage improvement compared to automatic transmission. And it is a significant difference, with a t-stat of `summary(fit)$coefficients[2,3]` and a p-value of `summary(fit)$coefficients[2,4]`, which is far below a typical 5% required alpha value.

A plot of the residuals (Figure 2 in the appendix) does not reveal any obvious unexplained features, like non-linearity or heteroskedasticity.

Considering horsepower

One possible explanation for improved mileage would be that people prefer automatic transmission for larger, more powerful cars, and designers of those cars are less concerned with mileage. If that if the case, then regressing mileage versus both horsepower and transmission should yield a low coefficient on transmission and a large coefficient on horsepower. However, when we look at that regression, the coefficient on horsepower is not only small, it's negative, and the coefficient on transmission is still fairly large.

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

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

So it appears that even when considering horsepower, the effect of transmission on mileage remains robust.

Conclusion

This data suggests that manual transmission vehicles have better mileage than automatic transmission, by as much as 7 miles per gallon, or about 40%. However, it would be better to have a data set where we could compare sets of cars which are the same except for the transmission.

Appendix

Fig 1. Histogram of mileage by transmission

Red is for manual transmission, green for automatic.

```
library(ggplot2)
ggplot(mtcars, aes(mpg, fill=as.factor(am))) + geom_histogram(position='stack', binwidth=2)
```

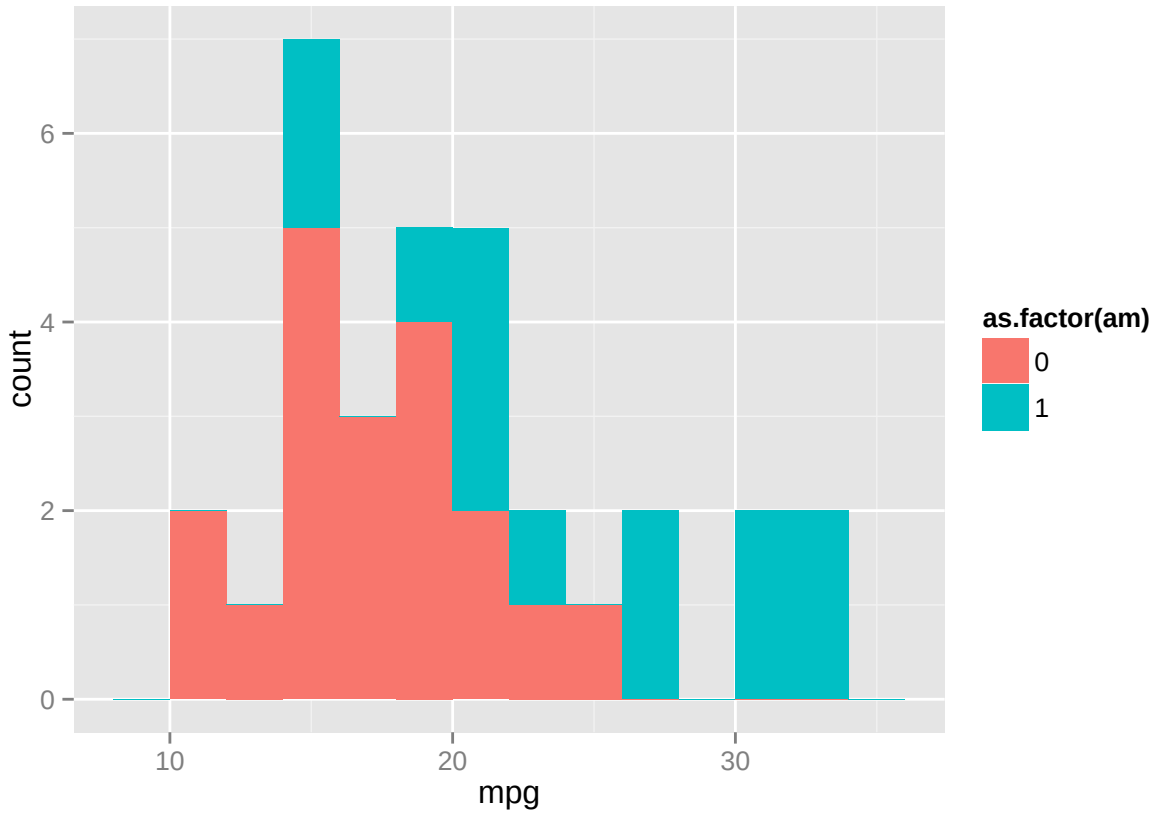

Fig 2. Residuals by tranmission

```
plot(resid(fit)[mtcars$am==F], col='red')
points(resid(fit)[mtcars$am==T], col='green')
```