Exponential Distribution Analysis Project

Kaylee Walsh

April 25, 2015

Overview

Motor Trends is investigating the effect transmission type has on the fuel efficiency in a vehicle. To do this, we will explore the `mtcars` data set and build a regression model, then show regression adjusted with transmission type.

Exploration

```
data(mtcars)
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

This dataset contains 32 observations with 11 variables. For a parsimonious model I am choosing to regress with the variable with the highest correlation to mpg, `wt`.

```
mpg_cors <- cor(mtcars)[1, ]; mpg_cors
```

```
##        mpg        cyl       disp         hp       drat         wt
##  1.0000000 -0.8521620 -0.8475514 -0.7761684  0.6811719 -0.8676594
##       qsec         vs         am       gear       carb
##  0.4186840  0.6640389  0.5998324  0.4802848 -0.5509251
```

```
regressor <- mpg_cors[abs(mpg_cors) == max(abs(mpg_cors[mpg_cors != 1]))]; regressor
```

```
##         wt
## -0.8676594
```

Likelihood Ratio Test

```r
fit1 <- lm(mpg ~ wt, mtcars)
fit2 <- lm(mpg ~ wt + factor(am), mtcars)
```

Here we've built two models, one unadjusted and one adjusted with the `am` variable. Using a likelihood ratio test through the `anova()` function we can compare the significance of adding in transmission type to the model.

```r
anova(fit1, fit2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt
## Model 2: mpg ~ wt + factor(am)
##   Res.Df    RSS Df Sum of Sq     F Pr(>F)
## 1     30 278.32
## 2     29 278.32  1 0.0022403 2e-04 0.9879
```

```r
summary(fit1)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 37.285126   1.877627 19.857575 8.241799e-19
## wt          -5.344472   0.559101 -9.559044 1.293959e-10
```

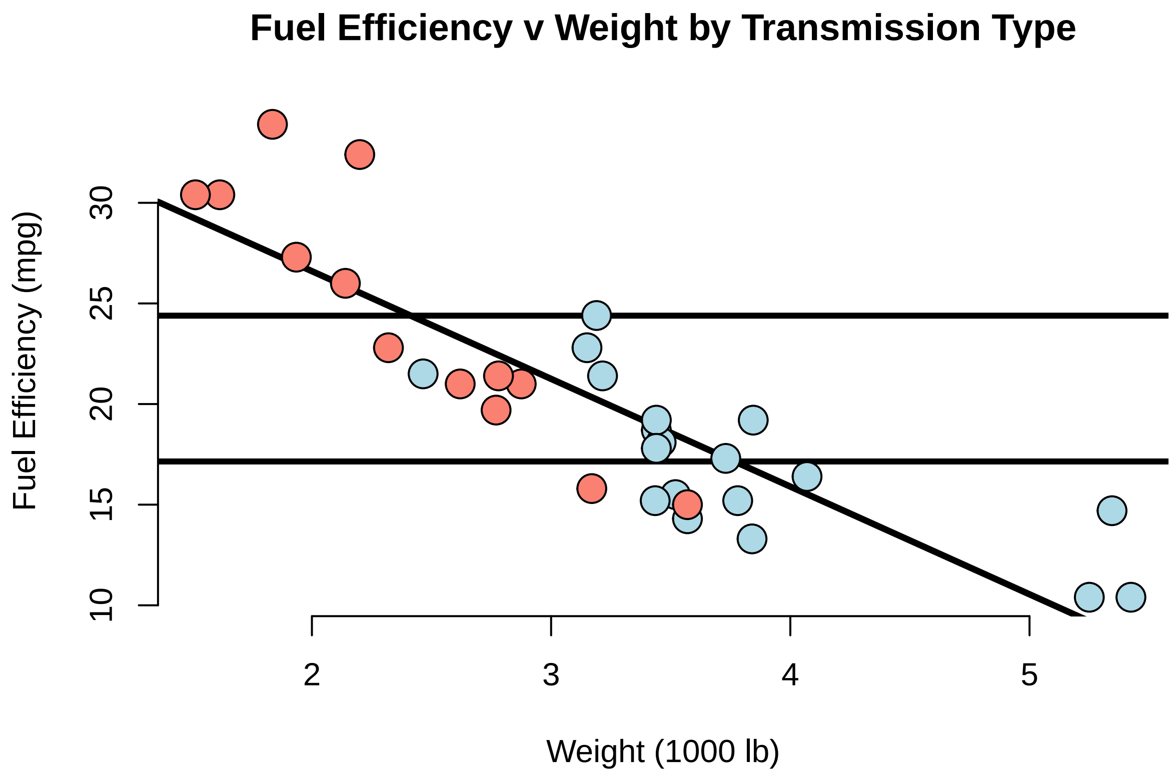
```r
summary(fit2)$coef
```

```
##                Estimate Std. Error     t value     Pr(>|t|)
## (Intercept) 37.32155131  3.0546385 12.21799285 5.843477e-13
## wt          -5.35281145  0.7882438 -6.79080719 1.867415e-07
## factor(am)1 -0.02361522  1.5456453 -0.01527855 9.879146e-01
```

We observe that this variable does not bear significance to a good fit for the model. Indeed we see in the coefficients after adjustment there is little change to the model in introducing `am`. Let's explore further through plotting the adjustment.

Adjustment

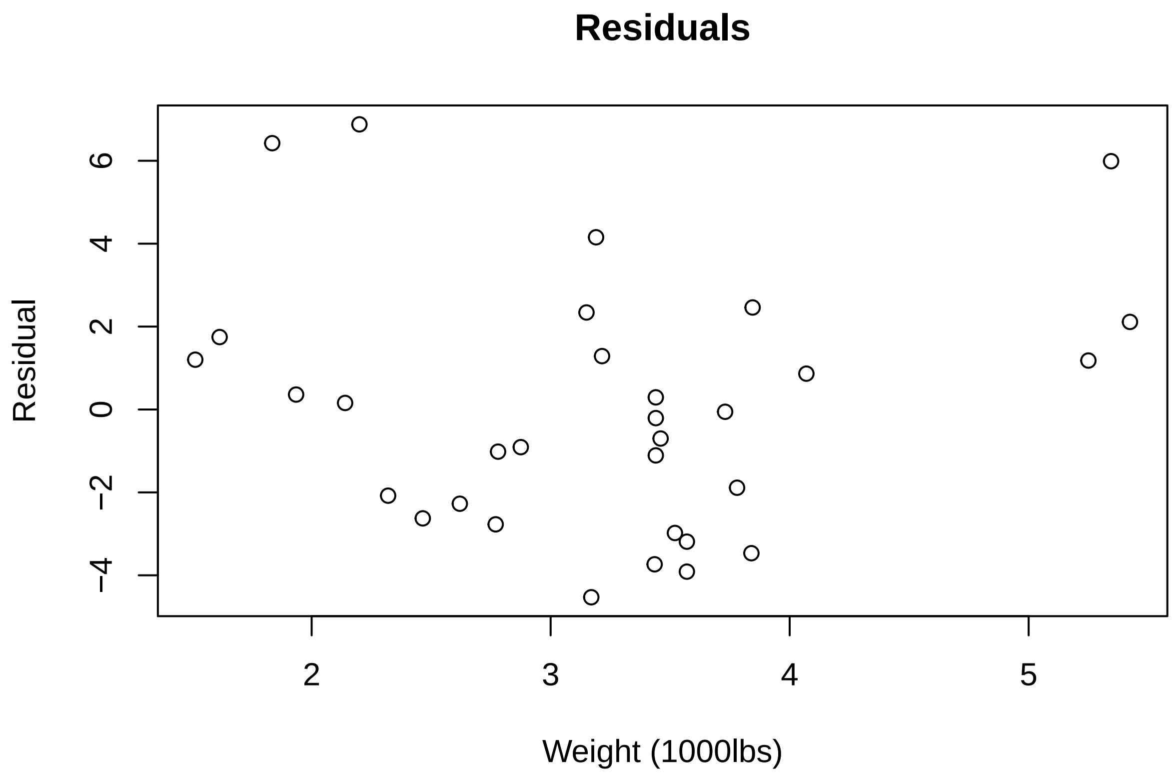
```r
plot(mtcars$wt, mtcars$mpg, type = 'n', frame = FALSE, main = "Fuel Efficiency v Weight by Transmission
abline(lm(mpg ~ wt, mtcars), lwd = 2)
abline(h = mean(mtcars[mtcars$am == 0, ]$mpg), lwd = 3)
abline(h = mean(mtcars[mtcars$am == 1, ]$mpg), lwd = 3)
abline(coef(fit2)[1], coef(fit2)[2], lwd = 3)
abline(coef(fit2)[1] + coef(fit2)[3], coef(fit2)[2], lwd = 3)
points(mtcars[mtcars$am == 0,]$wt, mtcars[mtcars$am == 0,]$mpg, pch = 21, col = 'black', bg = 'lightblu
points(mtcars[mtcars$am == 1,]$wt, mtcars[mtcars$am == 1,]$mpg, pch = 21, col = 'black', bg = 'salmon',
```



As shown here the two groups really are not comparable since they do not overlap domains significantly. I conclude an answer cannot be drawn from this model even though the difference in means between the two groups is 7.24 mpg.

Residuals

```
plot(fit2$residuals ~ mtcars$wt, main = "Residuals", xlab = "Weight (1000lbs)", ylab = "Residual")
```



From the residual plot we see that there is no clear pattern in error, thus the fit is not exhibiting any clear bias or heteroscedasity.

Executive Summary

Overall there is clear relation between weight and mpg, however between the transmission groups there is not sufficient overlap between the group's domains in order to make reasonable comparisons. As a straight difference, the difference between the means is 7.24 mpg with the manual transmission having a higher average in mpg. The model does not create any sort of biasness or heterscedasity as was shown by the residual plot.