Transmission analysis

Chris BC

19 October 2015

Objective: Analyse any relationship between transmission type and fuel efficiency as measured in miles per gallon.

Executive Summary

An analysis of the mtcars dataset indicates a superficial relationship between fuel efficiency, as measured in miles per gallon, and transmission type (automatic or manual), however closer investigation shows this relationship to be inseparable from other factors within an appropriate confidence interval (in other words, the relationship found is within two standard deviations of zero). The most appropriate model found a coefficient of **2.08** associated with a manual transmission, however a standard error of **2.64** which indicates no clear association.

Investigation

Initially let's look at a simple linear regression between the two variables:

```
fit1 <- lm(mpg ~ am, mtcars)
summary(fit1)$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
```

This suggests a strong positive relationship between the two variables. am coefficient 7.245 is well removed from 2*std. error (1.764), and t-value probability approaches zero, indicating a strong probability.

A residual plot is depicted in *figure 1.*

It is clear that there's a strong correlation between the two variables, but with residuals increasing as mpg increases there's more to the story so a more detailed model is required.

```
summary(glm(mpg~.,data=mtcars))$coef
```

```
##               Estimate  Std. Error    t value   Pr(>|t|)
## (Intercept) 12.30337416 18.71788443  0.6573058 0.51812440
## cyl         -0.11144048  1.04502336 -0.1066392 0.91608738
## disp         0.01333524  0.01785750  0.7467585 0.46348865
## hp          -0.02148212  0.02176858 -0.9868407 0.33495531
## drat         0.78711097  1.63537307  0.4813036 0.63527790
## wt          -3.71530393  1.89441430 -1.9611887 0.06325215
## qsec         0.82104075  0.73084480  1.1234133 0.27394127
## vs           0.31776281  2.10450861  0.1509915 0.88142347
## am           2.52022689  2.05665055  1.2254035 0.23398971
## gear         0.65541302  1.49325996  0.4389142 0.66520643
## carb        -0.19941925  0.82875250 -0.2406258 0.81217871
```

This shows all coefficients being within 2 standard errors of zero, failing to provide any insight.

```
fit2<-glm(mpg~am+cyl+wt,data=mtcars)
summary(fit2)$coef
```

```
##               Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) 39.4179334  2.6414573 14.9227979 7.424998e-15
## am           0.1764932  1.3044515  0.1353007 8.933421e-01
## cyl         -1.5102457  0.4222792 -3.5764148 1.291605e-03
## wt          -3.1251422  0.9108827 -3.4308942 1.885894e-03
```

```
fit3<-glm(mpg~am+hp+wt,data=mtcars)
summary(fit3)$coef
```

```
##                Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) 34.00287512 2.642659337 12.866916 2.824030e-13
## am           2.08371013 1.376420152  1.513862 1.412682e-01
## hp          -0.03747873 0.009605422 -3.901830 5.464023e-04
## wt          -2.87857541 0.904970538 -3.180850 3.574031e-03
```

Figure 2 and *Figure 3* show residual plots of these models,depicting a good spread of residuals but failing to show a correlation (within 2s.d. of 0) between mpg and transmission.

As shown below and in *figure 4* both poisson and binomial models olso present models that appear to have high certainty but which have a strong pattern in their residuals. Similar variants of these models controlling for additional predictor variables fail to identify a significant pattern.

```
fitP<-glm(mpg~am,data=mtcars,family="poisson")
summary(fitP)$coef
```

```
##              Estimate Std. Error   z value     Pr(>|z|)
## (Intercept) 2.8418447 0.05540187 51.295101 0.000000e+00
## am          0.3524231 0.07888566  4.467518 7.913225e-06
```

```
fitP2<-glm(mpg~.,data=mtcars,family="poisson")
summary(fitP2)$coef
```

```
##                  Estimate  Std. Error     z value   Pr(>|z|)
## (Intercept)  2.6091973640 1.523199204  1.71297185 0.08671772
## cyl          0.0155953923 0.090303090  0.17270054 0.86288682
## disp         0.0003463133 0.001607143  0.21548384 0.82939011
## hp          -0.0012379436 0.001881961 -0.65779447 0.51067021
## drat         0.0123698809 0.135007763  0.09162348 0.92699720
## wt          -0.1955312643 0.163113619 -1.19874272 0.23062800
## qsec         0.0387876859 0.058922842  0.65827928 0.51035869
## vs           0.0086356612 0.168201129  0.05134128 0.95905358
## am           0.0667101329 0.168616226  0.39563294 0.69237583
## gear         0.0696854414 0.124203111  0.56106035 0.57475640
## carb        -0.0135886004 0.070541306 -0.19263324 0.84724621
```

```r
fitB<-glm(am~mpg,data=mtcars,family="binomial")
summary(fitB)$coef
```

```
##               Estimate Std. Error   z value    Pr(>|z|)
## (Intercept) -6.6035267  2.3514416 -2.808289 0.004980557
## mpg          0.3070282  0.1148416  2.673493 0.007506579
```

No suitable model can be found predicting miles per gallon based on transmission type that is free from other significant influences and with significant confidence.

Appendix: Additional figures

Figure 1 - Initial residuals plot

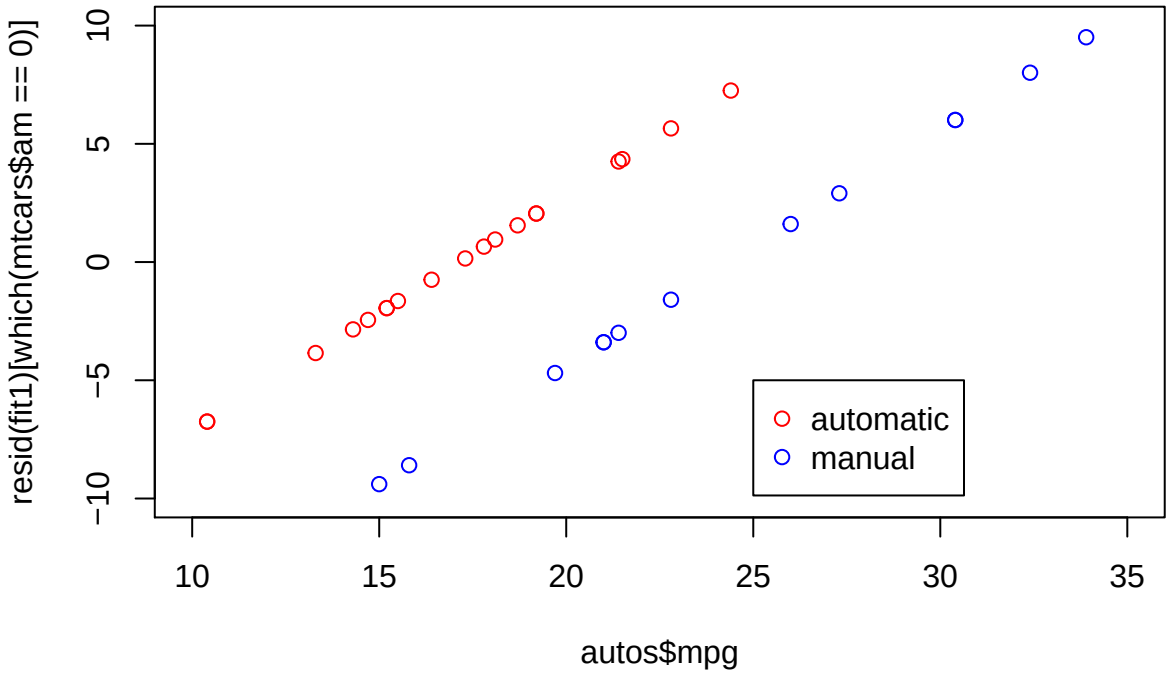

Figure 2 - Residuals with additional predictors transmission, cylinders, weight

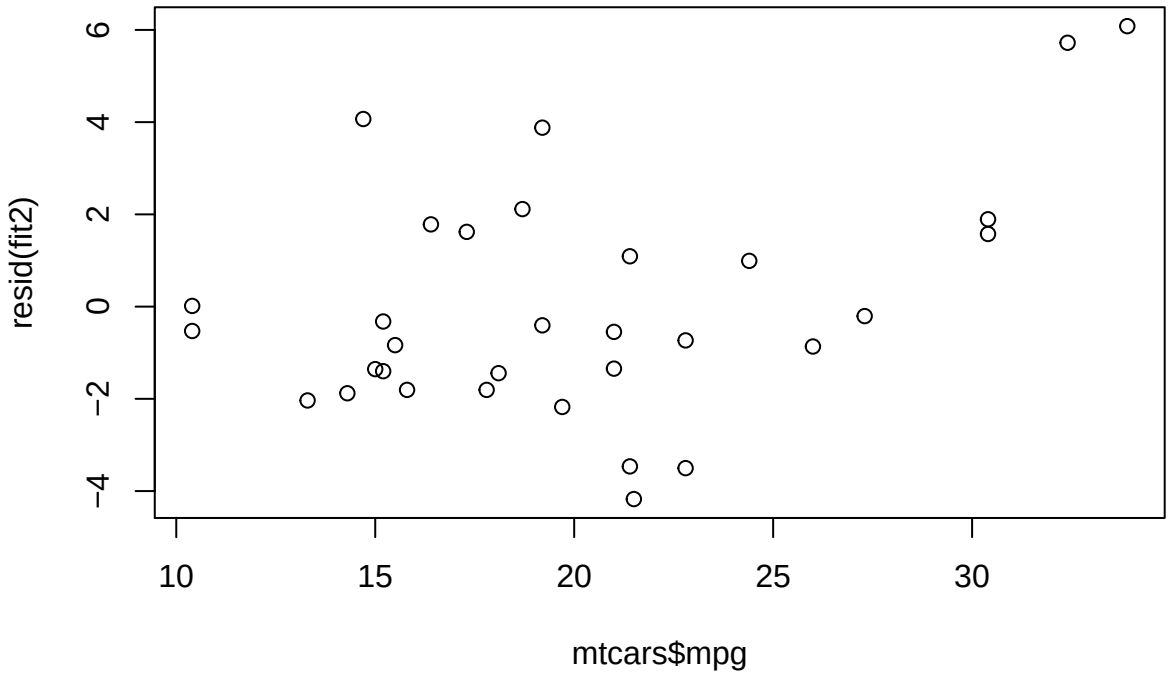

Figure 3 - Residuals with additional predictors transmission, horsepower, weight

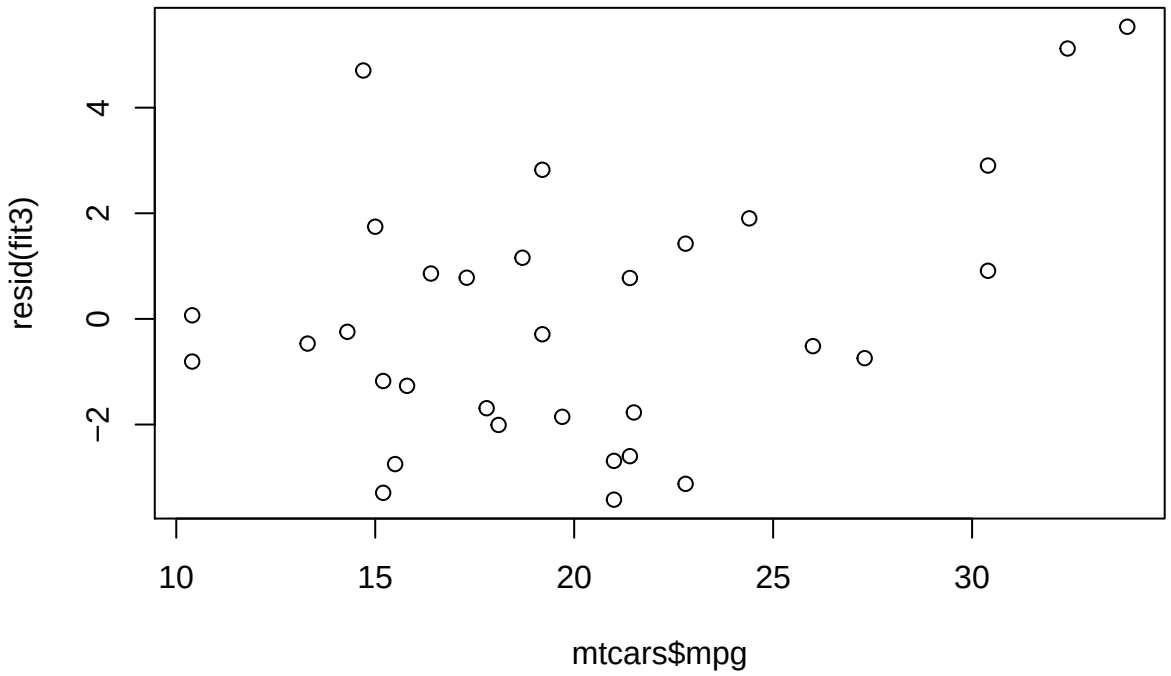

Figure 4 - Residual plots of poisson and binomial models

```r
plot(mtcars$mpg,resid(fitP), col="red")
points(mtcars$mpg,resid(fitB),col="blue")
legend(26,-1,c("Poisson model","Binomial model"),pch=c(1,1),col=c("red","blue"))
```