Transmission analysis

Chris BC

19 October 2015

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

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

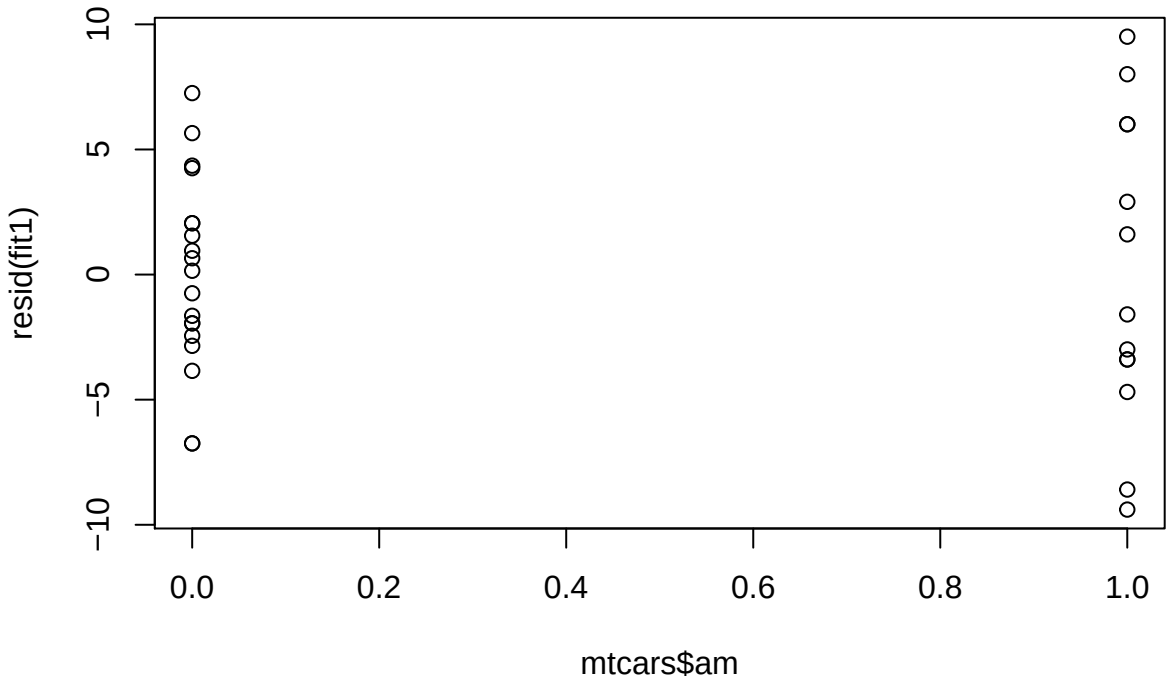
```
fit1 <- lm(mpg ~ am, mtcars)
summary(fit1)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## am             7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

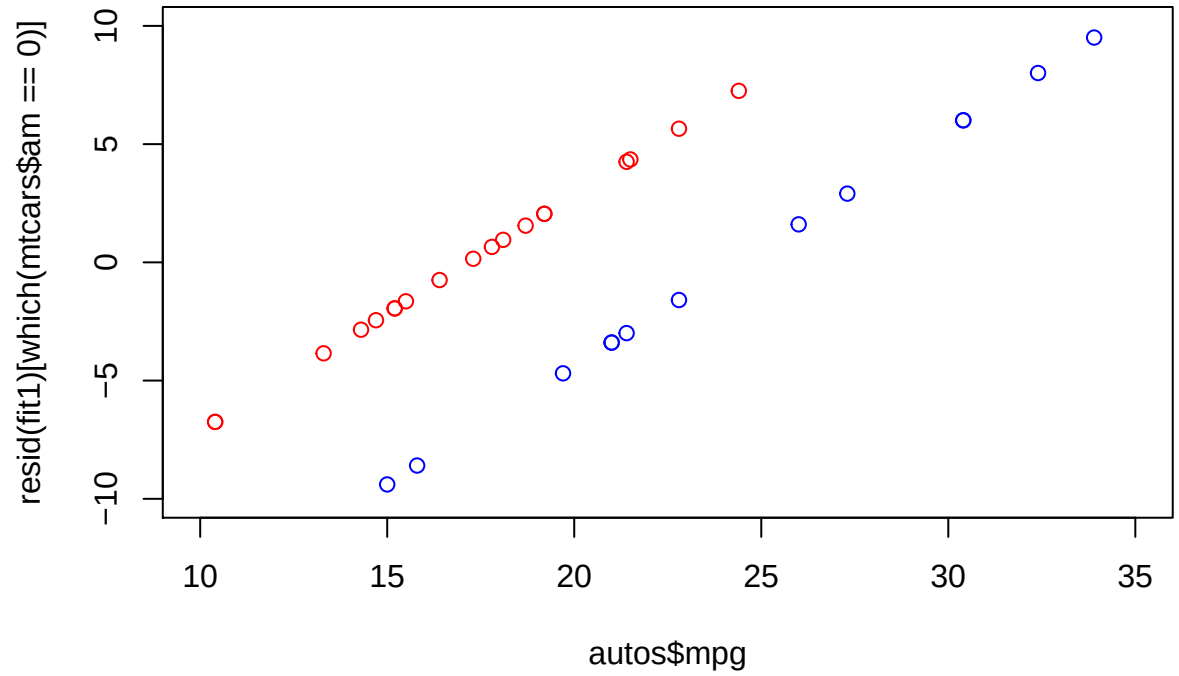
This suggests a strong positive relationship between the two variables. am coefficient 7.245 is well removed from 2*std. error (1.764), and t-value probability approaches zero, indicating a strong probability.

Residual standard error appears significant so let's look at residuals.

```
autos<-mtcars[which(mtcars$am==0),]
manuals<-mtcars[which(mtcars$am==1),]
library(ggplot2)
plot(mtcars$am, resid(fit1))
```




```
plot(autos$mpg, resid(fit1)[which(mtcars$am==0)], col="red", xlim=c(10,35),ylim=c(-10,10))
points(manuals$mpg, resid(fit1)[which(mtcars$am==1)], col="blue")
```



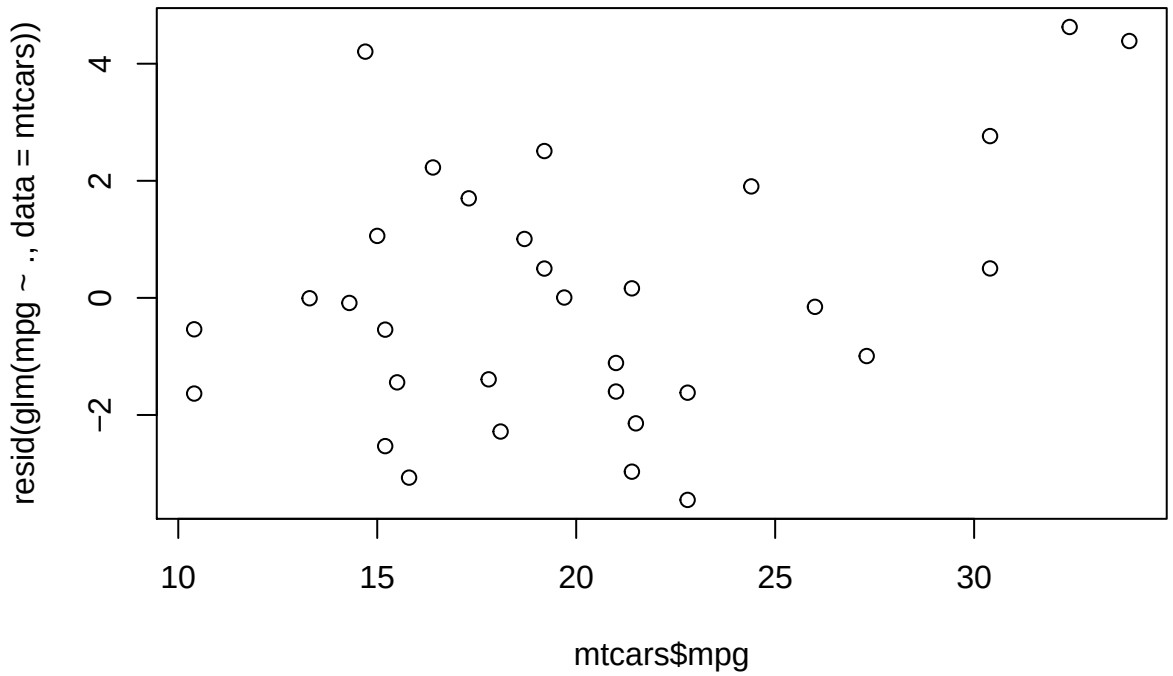
It is clear that there's a strong correlation between the two variables, but with residuals increasing as mpg increases there's more to the story so a more detailed model is required.

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

```
##
```

```
## Call:
## glm(formula = mpg ~ ., data = mtcars)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.4506  -1.6044  -0.1196   1.2193   4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## cyl         -0.11144    1.04502  -0.107   0.9161
## disp         0.01334    0.01786   0.747   0.4635
## hp          -0.02148    0.02177  -0.987   0.3350
## drat         0.78711    1.63537   0.481   0.6353
## wt          -3.71530    1.89441  -1.961   0.0633 .
## qsec         0.82104    0.73084   1.123   0.2739
## vs           0.31776    2.10451   0.151   0.8814
## am           2.52023    2.05665   1.225   0.2340
## gear         0.65541    1.49326   0.439   0.6652
## carb        -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 7.023544)
##
##     Null deviance: 1126.05  on 31  degrees of freedom
## Residual deviance:  147.49  on 21  degrees of freedom
## AIC: 163.71
##
## Number of Fisher Scoring iterations: 2
```

```
plot(mtcars$mpg, resid(glm(mpg~.,data=mtcars)))
```



This shows a series of residuals lacking a pattern, indicating a candidate model, but with all of our coefficients being within 2 standard errors of zero fails to provide any insight.

```
fit2<-glm(mpg~am+cyl+wt,data=mtcars)
fit3<-glm(mpg~am+hp+wt,data=mtcars)
summary(fit2)
```

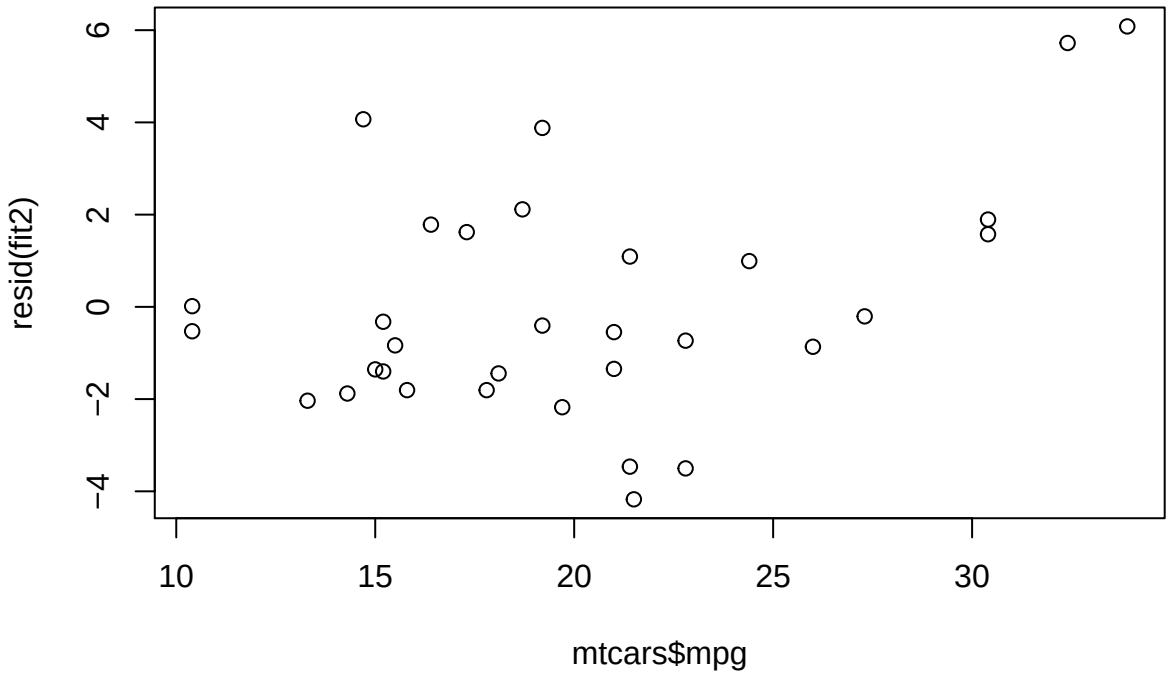
```
##
## Call:
## glm(formula = mpg ~ am + cyl + wt, data = mtcars)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.1735  -1.5340  -0.5386   1.5864   6.0812
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  39.4179     2.6415  14.923 7.42e-15 ***
## am            0.1765     1.3045   0.135  0.89334
## cyl          -1.5102     0.4223  -3.576  0.00129 **
## wt           -3.1251     0.9109  -3.431  0.00189 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 6.823109)
##
##     Null deviance: 1126.05  on 31  degrees of freedom
## Residual deviance:  191.05  on 28  degrees of freedom
## AIC: 157.99
##
## Number of Fisher Scoring iterations: 2
```

```
summary(fit3)
```

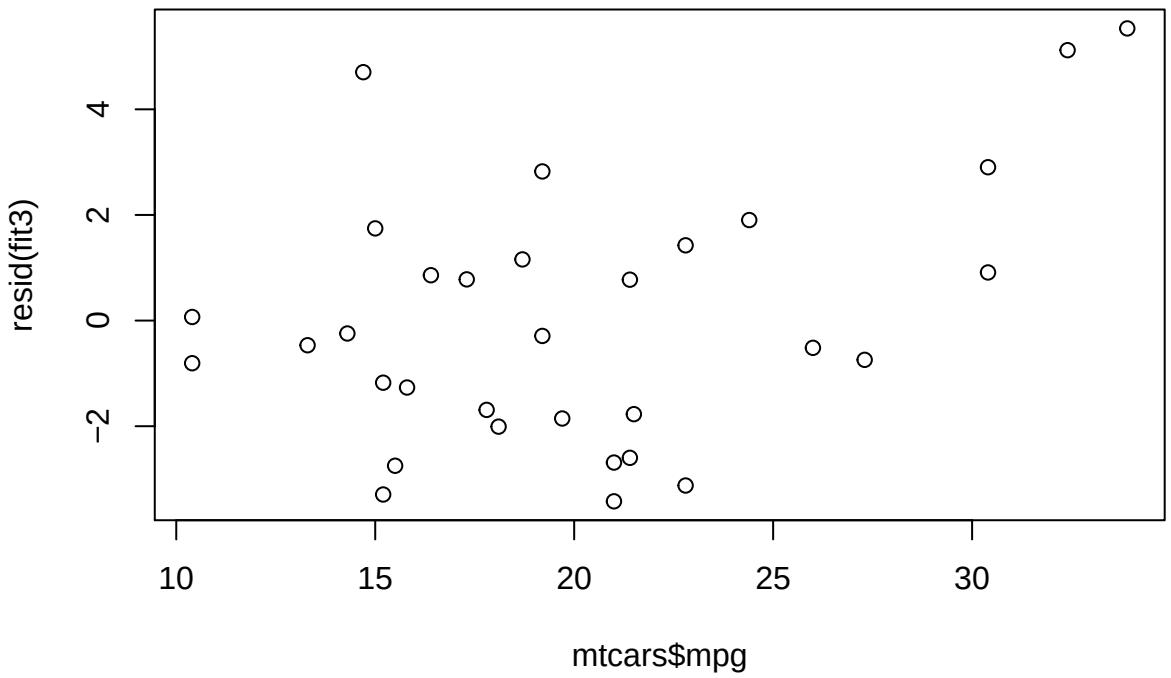
```
##
## Call:
## glm(formula = mpg ~ am + hp + wt, data = mtcars)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.4221  -1.7924  -0.3788   1.2249   5.5317
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.002875   2.642659  12.867 2.82e-13 ***
## am           2.083710   1.376420   1.514 0.141268
## hp          -0.037479   0.009605  -3.902 0.000546 ***
## wt          -2.878575   0.904971  -3.181 0.003574 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 6.438967)
##
```

```
##     Null deviance: 1126.05  on 31  degrees of freedom
## Residual deviance:  180.29  on 28  degrees of freedom
## AIC: 156.13
##
## Number of Fisher Scoring iterations: 2
```

```
plot(mtcars$mpg, resid(fit2))
```




```
plot(mtcars$mpg,resid(fit3))
```



These models show a good spread of residuals but fails to show a correlation (within 2s.d. of 0) between mpg and transmission.

As shown below both poisson and binomial models olso present models that appear to have high certainty but which have a strong pattern in their residuals.

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

```
## Warning in dpois(y, mu, log = TRUE): non-integer x = 22.800000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 21.400000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 18.700000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 18.100000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 14.300000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 24.400000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 22.800000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 19.200000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 17.800000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 16.400000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 17.300000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 15.200000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 10.400000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 10.400000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 14.700000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 32.400000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 30.400000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 33.900000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 21.500000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 15.500000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 15.200000
```

```
## Warning in dpois(y, mu, log = TRUE): non-integer x = 13.300000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 19.200000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 27.300000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 30.400000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 15.800000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 19.700000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 21.400000
```

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

```
## Warning in dpois(y, mu, log = TRUE): non-integer x = 22.800000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 21.400000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 18.700000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 18.100000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 14.300000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 24.400000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 22.800000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 19.200000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 17.800000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 16.400000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 17.300000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 15.200000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 10.400000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 10.400000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 14.700000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 32.400000
```

```
## Warning in dpois(y, mu, log = TRUE): non-integer x = 30.400000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 33.900000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 21.500000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 15.500000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 15.200000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 13.300000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 19.200000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 27.300000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 30.400000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 15.800000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 19.700000

## Warning in dpois(y, mu, log = TRUE): non-integer x = 21.400000
```

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

```
##
## Call:
## glm(formula = mpg ~ am, family = "poisson", data = mtcars)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.04892  -0.70378  -0.07251   0.68029   1.81681
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.84184    0.05540  51.295  < 2e-16 ***
## am           0.35242    0.07889   4.468 7.91e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 54.524  on 31  degrees of freedom
## Residual deviance: 34.699  on 30  degrees of freedom
## AIC: Inf
##
## Number of Fisher Scoring iterations: 4
```

```r
summary(fitP2)
```

```
##
## Call:
## glm(formula = mpg ~ ., family = "poisson", data = mtcars)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.60995  -0.32582  -0.06014   0.27589   0.95903
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.6091974  1.5231992   1.713   0.0867 .
## cyl          0.0155954  0.0903031   0.173   0.8629
## disp         0.0003463  0.0016071   0.215   0.8294
## hp          -0.0012379  0.0018820  -0.658   0.5107
## drat         0.0123699  0.1350078   0.092   0.9270
## wt          -0.1955313  0.1631136  -1.199   0.2306
## qsec         0.0387877  0.0589228   0.658   0.5104
## vs           0.0086357  0.1682011   0.051   0.9591
## am           0.0667101  0.1686162   0.396   0.6924
## gear         0.0696854  0.1242031   0.561   0.5748
## carb        -0.0135886  0.0705413  -0.193   0.8472
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 54.5238  on 31  degrees of freedom
## Residual deviance:  5.5362  on 21  degrees of freedom
## AIC: Inf
##
## Number of Fisher Scoring iterations: 4
```

```r
summary(fitB)
```

```
##
## Call:
## glm(formula = am ~ mpg, family = "binomial", data = mtcars)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5701  -0.7531  -0.4245   0.5866   2.0617
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -6.6035     2.3514  -2.808  0.00498 **
## mpg           0.3070     0.1148   2.673  0.00751 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
##     Null deviance: 43.230  on 31  degrees of freedom
## Residual deviance: 29.675  on 30  degrees of freedom
## AIC: 33.675
##
## Number of Fisher Scoring iterations: 5
```

```
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"))
```