MATH552 Group Project

c.wang35@lancaster.ac.uk & k.molloy@lancaster.ac.uk

```
neomod <- read.table("../data/neomod_isam12.dat", header = T)

#plot(neomod)

summary(neomod)
```

```
##       cns              size             gest             bwt
##  Min.   :0.0000   Min.   :0.0000   Min.   :1.000   Min.   :0.482
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:2.000   1st Qu.:1.159
##  Median :0.0000   Median :1.0000   Median :3.000   Median :1.460
##  Mean   :0.4718   Mean   :0.7103   Mean   :3.197   Mean   :1.699
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:2.041
##  Max.   :1.0000   Max.   :1.0000   Max.   :5.000   Max.   :5.443
##      emp.f            emp.m             edu            re.ad
##  Min.   :0.0000   Min.   :0.0000   Min.   :1.000   Min.   :0.0000
##  1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:2.000   1st Qu.:0.0000
##  Median :1.0000   Median :0.0000   Median :2.000   Median :0.0000
##  Mean   :0.8905   Mean   :0.4832   Mean   :2.421   Mean   :0.4274
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:3.000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :4.000   Max.   :1.0000
##      los              sex             accom
##  Min.   :1.386   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:3.045   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :3.555   Median :1.0000   Median :1.0000
##  Mean   :3.548   Mean   :0.5242   Mean   :0.7298
##  3rd Qu.:4.111   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :5.613   Max.   :1.0000   Max.   :1.0000
```

```
# Attach to allow for better addressing
attach(neomod)
```

Factorisation

Modify variables that are int or str to factor.

```
neomod$cns    = factor(cns)
neomod$size   = factor(size)
neomod$gest   = factor(gest)
neomod$emp.f  = factor(emp.f)
neomod$emp.m  = factor(emp.m)
neomod$edu    = factor(edu)
```

```r
neomod$re.ad   = factor(re.ad)
neomod$sex     = factor(sex)
neomod$accom   = factor(accom)
```

```r
summary(neomod)
```

```
##  cns       size      gest          bwt          emp.f     emp.m     edu       re.ad
##  0:786   0: 431   1: 44   Min.   :0.482   0: 163   0:769   1: 77   0:852
##  1:702   1:1057   2:384   1st Qu.:1.159   1:1325   1:719   2:871   1:636
##                   3:500   Median :1.460                    3:376
##                   4:355   Mean   :1.699                    4:164
##                   5:205   3rd Qu.:2.041
##                           Max.   :5.443
##        los       sex      accom
##  Min.   :1.386   0:708   0: 402
##  1st Qu.:3.045   1:780   1:1086
##  Median :3.555
##  Mean   :3.548
##  3rd Qu.:4.111
##  Max.   :5.613
```

```r
head(neomod)
```

```
##   cns size gest   bwt emp.f emp.m edu re.ad      los sex accom
## 1   0    1    2 1.040     1     0   4     1 4.510860   1     1
## 2   0    1    3 2.353     1     1   4     0 4.234107   1     1
## 3   1    1    5 3.204     1     1   2     0 1.945910   1     1
## 4   1    1    3 1.389     1     1   2     0 3.332205   1     1
## 5   1    0    3 0.970     1     0   3     0 3.367296   0     1
## 6   1    0    5 3.033     1     1   4     0 2.833213   1     1
```
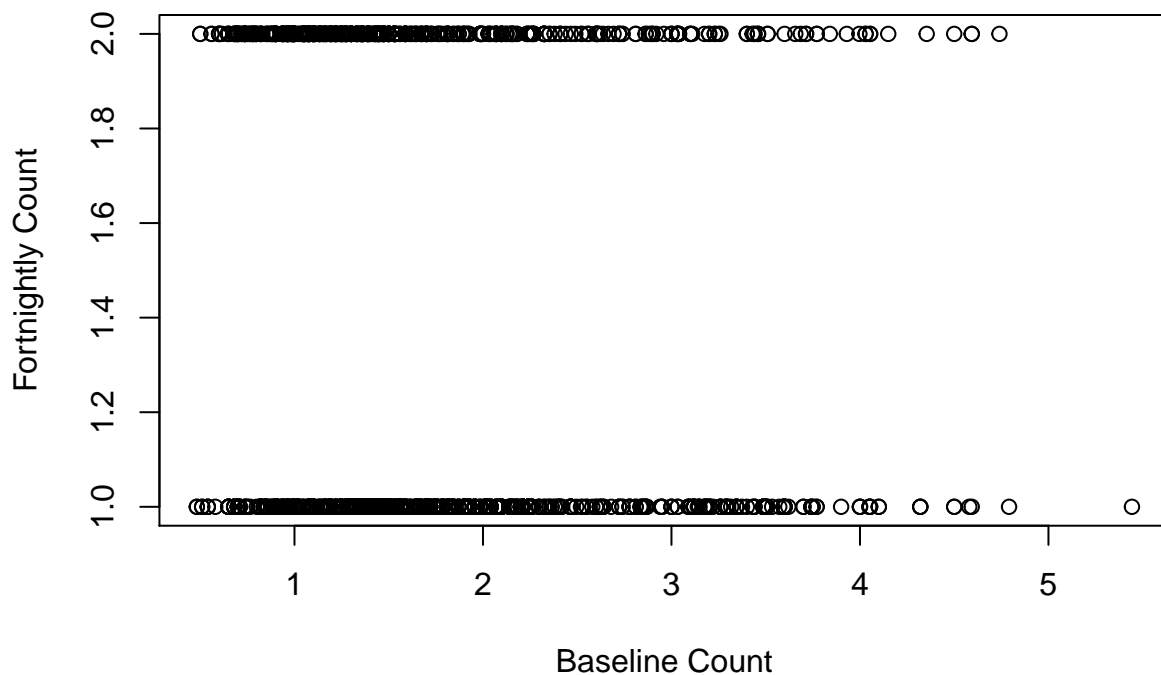
```r
# Representative Model
mod1 = glm(re.ad ~ bwt,
           data=neomod,
           family=binomial)
```
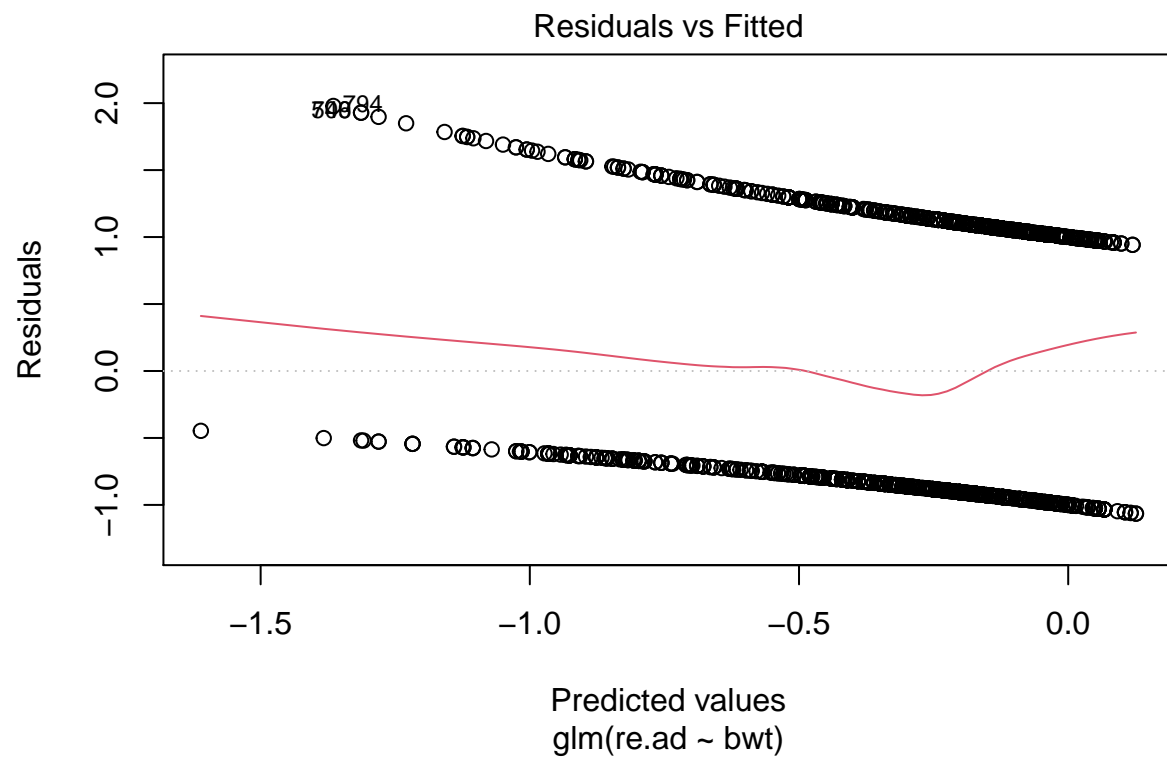
```r
summary(mod1)
```

```
##
## Call:
## glm(formula = re.ad ~ bwt, family = binomial, data = neomod)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2313  -1.0849  -0.8947   1.2369   1.7846
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.29460    0.12571   2.343   0.0191 *
```
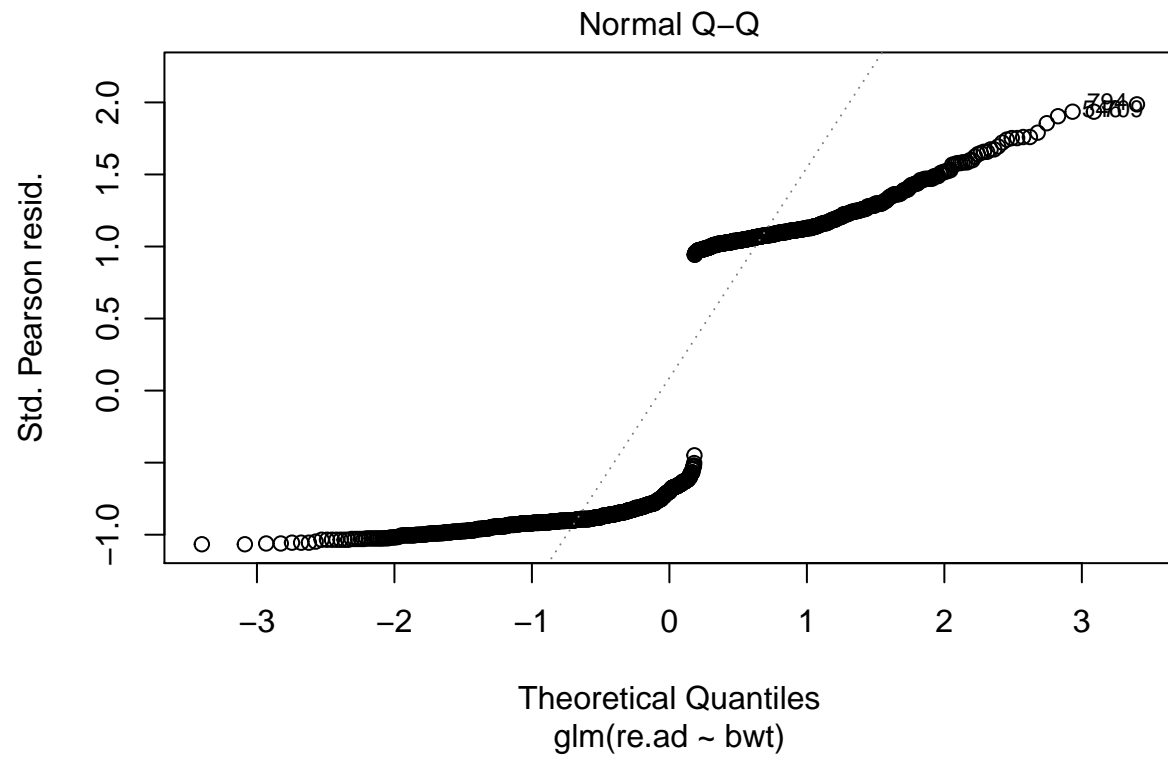
```
## bwt          -0.35011    0.06914  -5.064  4.1e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2031.3  on 1487  degrees of freedom
## Residual deviance: 2004.2  on 1486  degrees of freedom
## AIC: 2008.2
##
## Number of Fisher Scoring iterations: 4
```
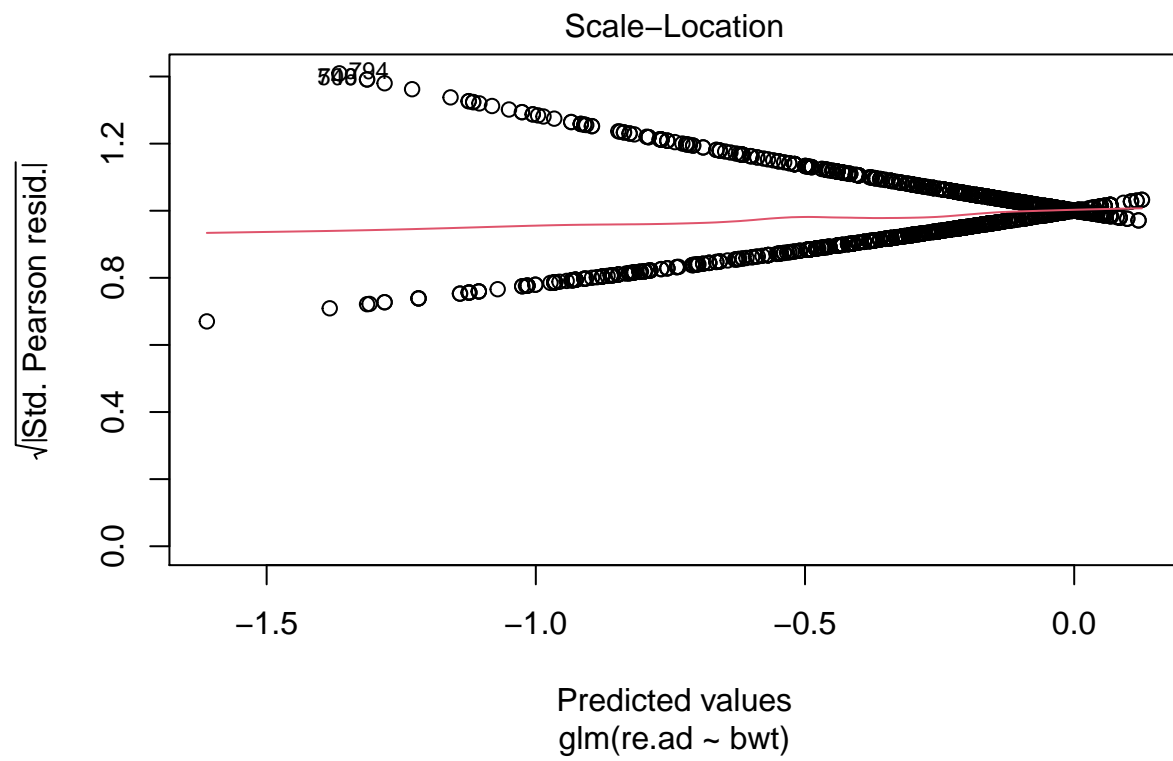
```r
# Representative plot
par(mfrow=c(1,1))
plot(neomod$bwt,
     neomod$re.ad,
     xlab="Baseline Count",
     ylab="Fortnightly Count")
abline(mod1, col=2)
```



```r
# Representative plot
plot(mod1)
```
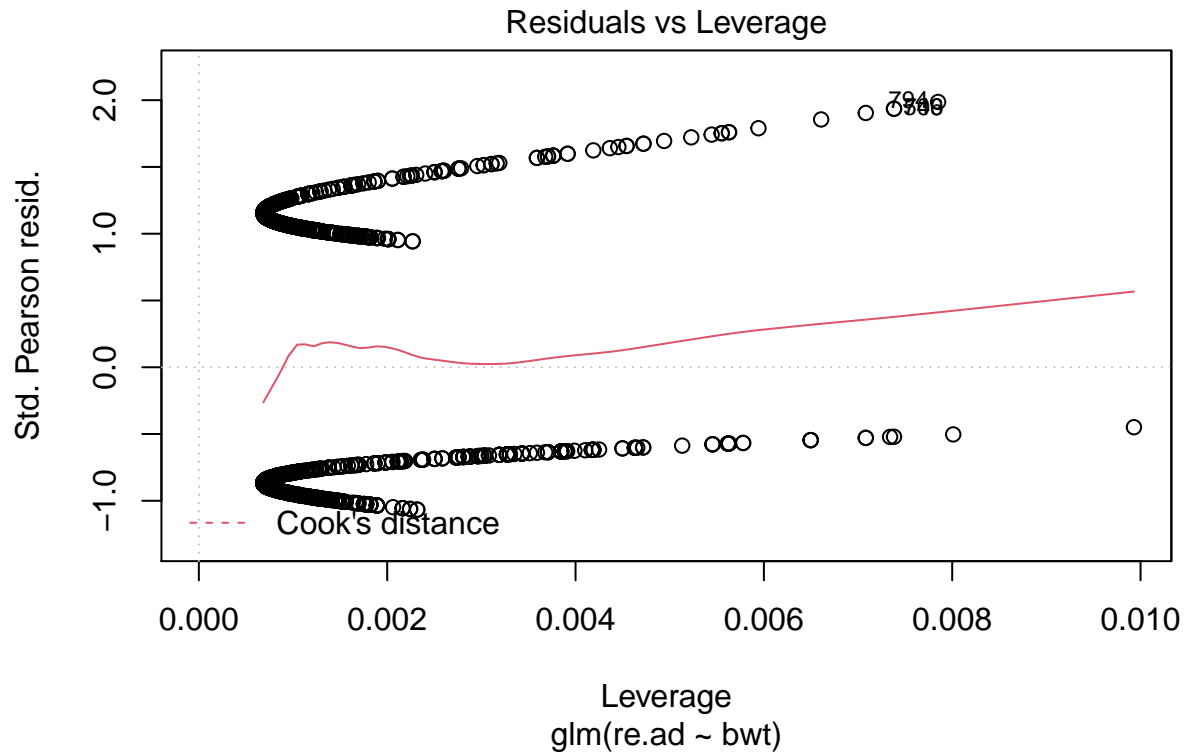
Residuals vs Fitted

Residuals

Predicted values
glm(re.ad ~ bwt)

Normal Q–Q

Std. Pearson resid.

Theoretical Quantiles
glm(re.ad ~ bwt)

Scale–Location

√|Std. Pearson resid.|

Predicted values
glm(re.ad ~ bwt)

Residuals vs Leverage
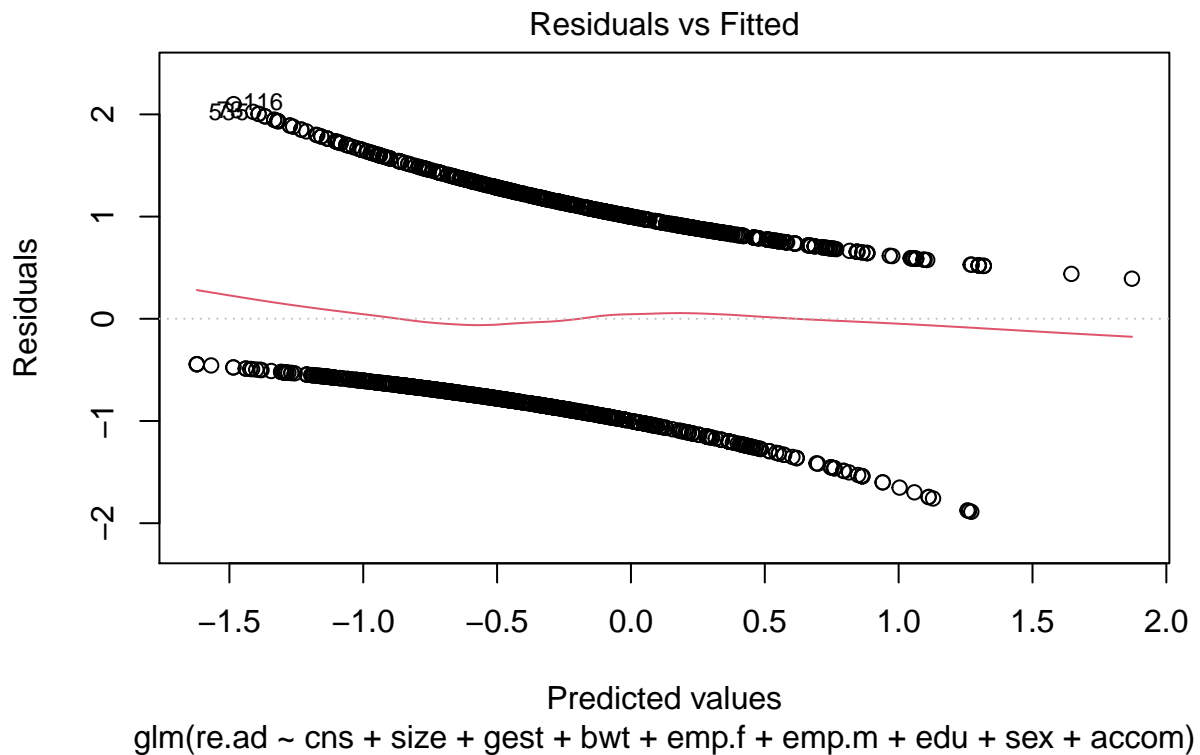

Leverage
glm(re.ad ~ bwt)

```r
# Random model - probably wrong
mod2.re.ad = glm(re.ad ~ cns + size + gest + bwt + emp.f + emp.m + edu + sex + accom,
          data= neomod,
          family=binomial)

summary(mod2.re.ad)
```
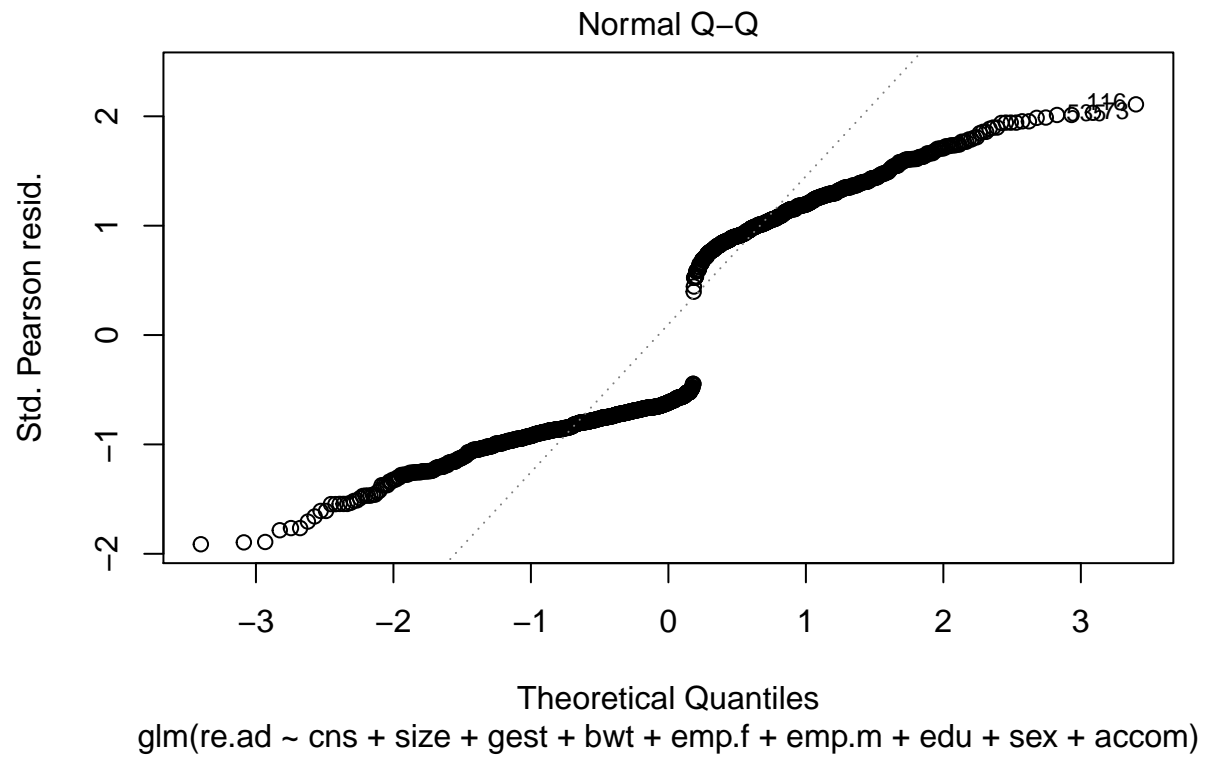
```
##
## Call:
## glm(formula = re.ad ~ cns + size + gest + bwt + emp.f + emp.m +
##     edu + sex + accom, family = binomial, data = neomod)
##
## Deviance Residuals:
##     Min      1Q   Median      3Q      Max
## -1.7427  -1.0038  -0.7997   1.1824   1.8377
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.42839    0.45780   3.120 0.001808 **
## cns1        -0.01603    0.11128  -0.144 0.885433
## size1        0.11750    0.12115   0.970 0.332117
## gest2       -0.76823    0.35558  -2.160 0.030735 *
## gest3       -1.31581    0.36022  -3.653 0.000259 ***
## gest4       -1.44663    0.38183  -3.789 0.000151 ***
## gest5       -1.20174    0.45826  -2.622 0.008731 **
## bwt         -0.21167    0.11790  -1.795 0.072601 .
```
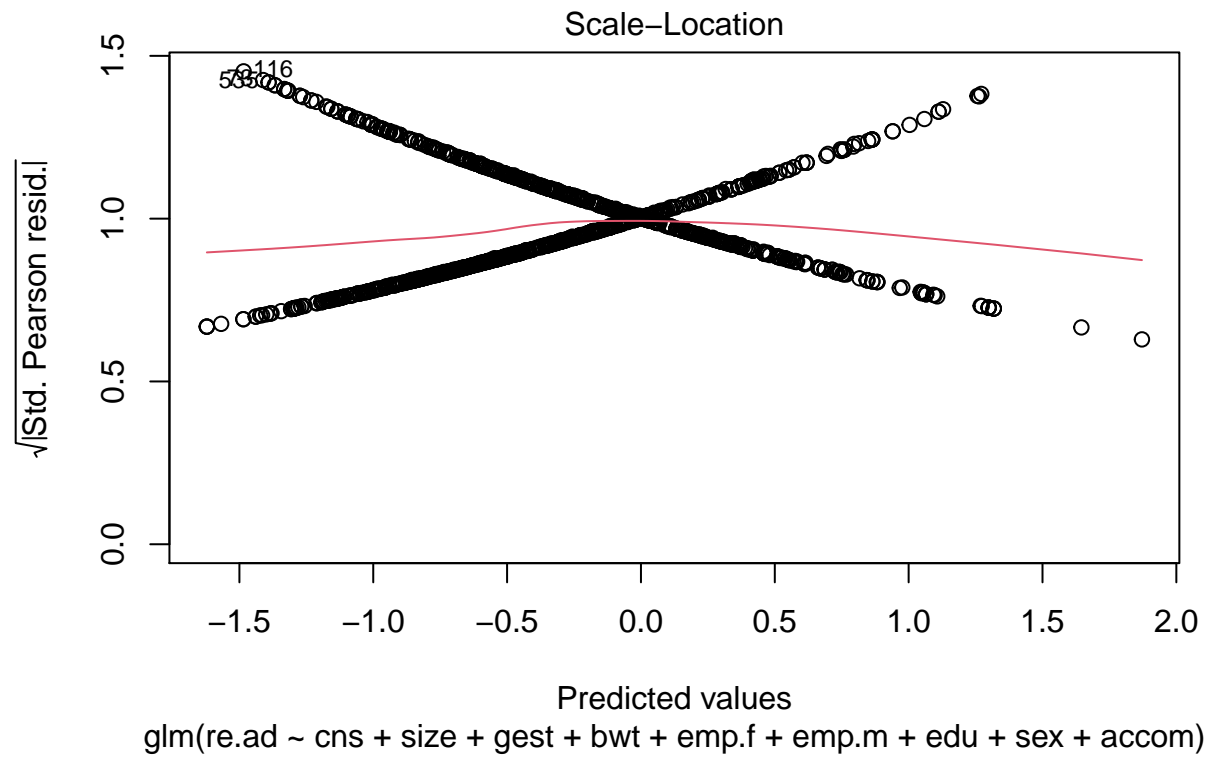
```
## emp.f1       -0.21245     0.18752  -1.133 0.257238
## emp.m1       -0.30657     0.11281  -2.717 0.006578 **
## edu2         -0.06275     0.25075  -0.250 0.802405
## edu3          0.07672     0.26961   0.285 0.775986
## edu4          0.17870     0.29607   0.604 0.546116
## sex1          0.55258     0.11275   4.901 9.54e-07 ***
## accom1       -0.35208     0.13795  -2.552 0.010704 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2031.3  on 1487  degrees of freedom
## Residual deviance: 1926.1  on 1473  degrees of freedom
## AIC: 1956.1
##
## Number of Fisher Scoring iterations: 4
```

```r
plot(mod2.re.ad)
```


Residuals vs Fitted

glm(re.ad ~ cns + size + gest + bwt + emp.f + emp.m + edu + sex + accom)

8

Normal Q–Q

Theoretical Quantiles
glm(re.ad ~ cns + size + gest + bwt + emp.f + emp.m + edu + sex + accom)

Scale–Location

√|Std. Pearson resid.|

Predicted values
glm(re.ad ~ cns + size + gest + bwt + emp.f + emp.m + edu + sex + accom)

Residuals vs Leverage
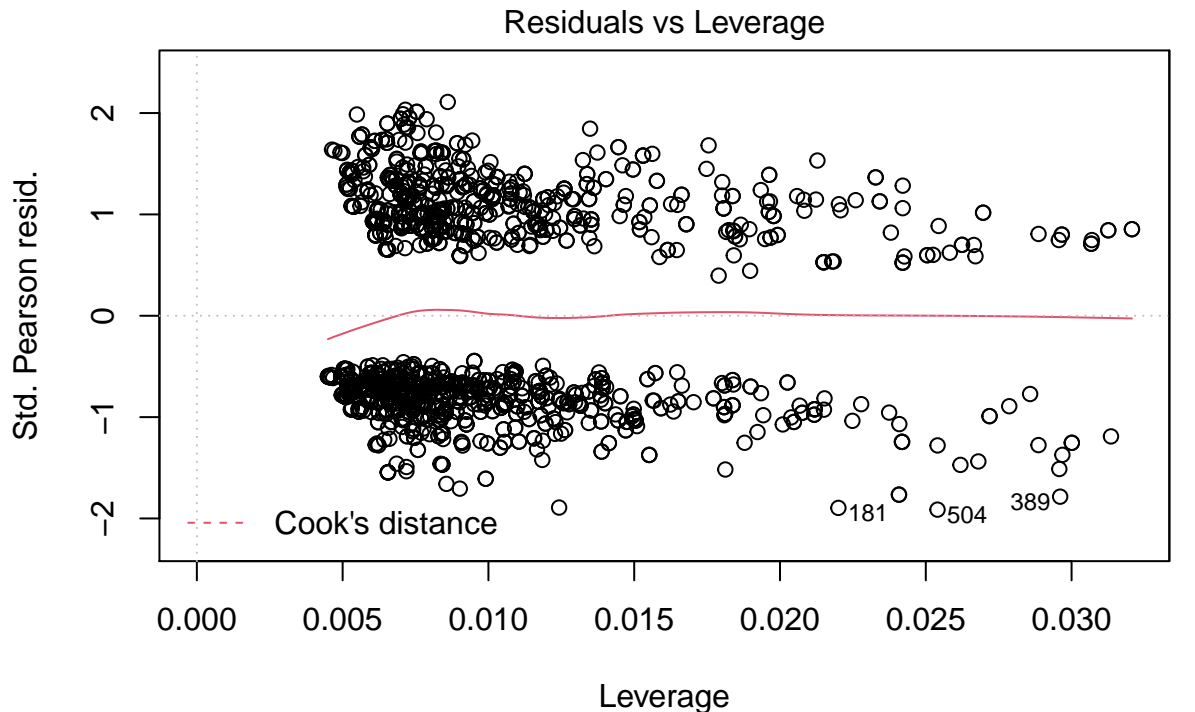

glm(re.ad ~ cns + size + gest + bwt + emp.f + emp.m + edu + sex + accom)

```
# Random model - probably wrong
mod2.los = glm(los ~ cns + size + gest + bwt + emp.f + emp.m + edu + sex + accom,
          data= neomod,
          family=Gamma(link=inverse))

summary(mod2.los)
```

```
##
## Call:
## glm(formula = los ~ cns + size + gest + bwt + emp.f + emp.m +
##     edu + sex + accom, family = Gamma(link = inverse), data = neomod)
##
## Deviance Residuals:
##      Min       1Q    Median       3Q       Max
## -0.62858  -0.07900  -0.00551   0.07554   0.61557
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.167692   0.007882  21.276  < 2e-16 ***
## cns1        0.012280   0.002216   5.540 3.57e-08 ***
## size1       0.003584   0.002393   1.498 0.134469
## gest2       0.019846   0.005224   3.799 0.000151 ***
## gest3       0.050687   0.005485   9.241  < 2e-16 ***
## gest4       0.080400   0.006195  12.978  < 2e-16 ***
## gest5       0.082450   0.008470   9.734  < 2e-16 ***
## bwt         0.032725   0.002614  12.519  < 2e-16 ***
```

```
## emp.f1         0.006436    0.003642    1.767 0.077369 .
## emp.m1         0.004156    0.002248    1.848 0.064734 .
## edu2           0.009237    0.005032    1.836 0.066624 .
## edu3           0.001732    0.005381    0.322 0.747680
## edu4          -0.003761    0.005877   -0.640 0.522338
## sex1          -0.011305    0.002212   -5.111 3.62e-07 ***
## accom1        -0.002905    0.002726   -1.065 0.286849
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 0.02225456)
##
##     Null deviance: 72.645  on 1487  degrees of freedom
## Residual deviance: 33.360  on 1473  degrees of freedom
## AIC: 2305.1
##
## Number of Fisher Scoring iterations: 4
```

```
plot(mod2.los)
```


Residuals vs Fitted

glm(los ~ cns + size + gest + bwt + emp.f + emp.m + edu + sex + accom)

12

Normal Q–Q

Std. Pearson resid.

Theoretical Quantiles
glm(los ~ cns + size + gest + bwt + emp.f + emp.m + edu + sex + accom)

Scale–Location

√|Std. Pearson resid.|

Predicted values
glm(los ~ cns + size + gest + bwt + emp.f + emp.m + edu + sex + accom)

Residuals vs Leverage

Std. Pearson resid.

Leverage
glm(los ~ cns + size + gest + bwt + emp.f + emp.m + edu + sex + accom)