M3S2 Coursework

Tudor Trita Trita, CID: 01199397

8 March 2019

Part 1, Normal Linear Modelling

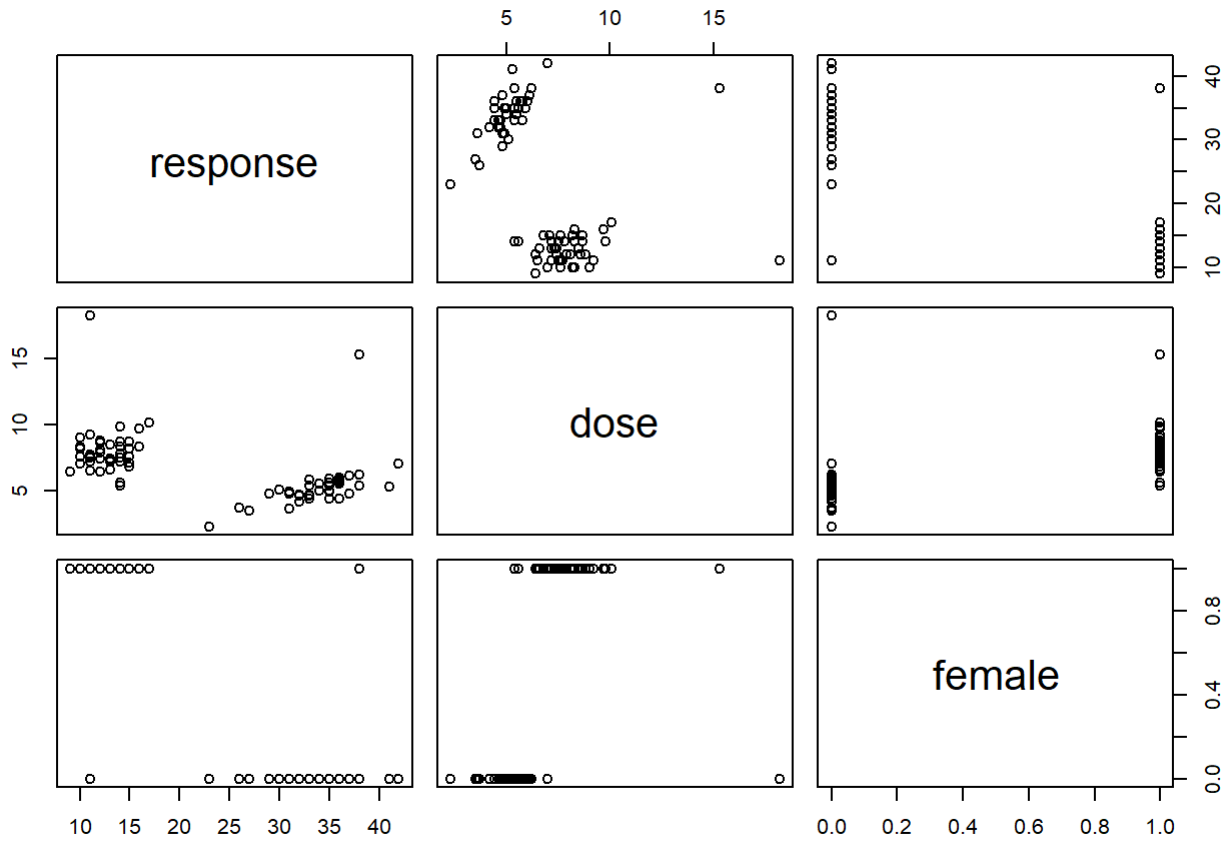
Question 1:

For this question, we are given a set of data to analyse, namely ‘bp’ and we are required to perform a statistical analysis on it. The variables available for analysis are: female, dose and response.

First of all, we import the data from the file ‘1199397.RData’ and we initialise the variables:

```
data = load('1199397.RData')
response = bp$response
dose = bp$dose
female = bp$female
```

As part of our exploratory analysis, let us have a look at the actual data:



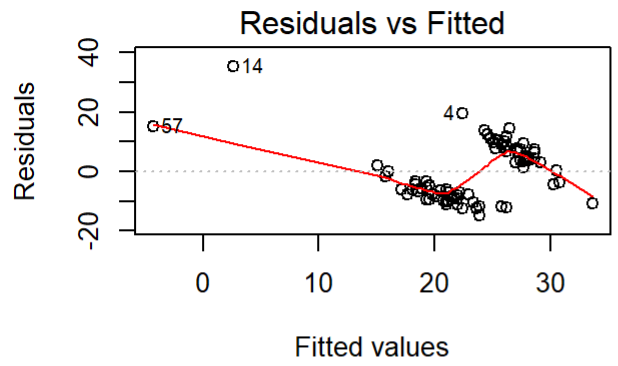
Immediately, we can see that there are two separate clusters of data in every plot. These two clusters are separated by the ‘female’ variable, i.e. each cluster correspond to male or female data. This is an early hint that a linear model would need to take into account of the female variable.

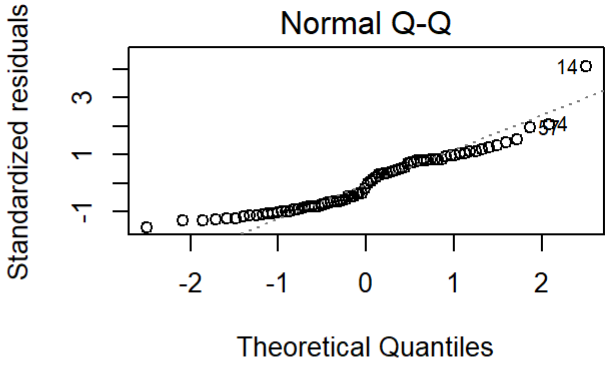
Now, we move on to fit the model as given in the coursework notes. This is a ‘response ~ dose’ model. We then print a summary of this and give the appropriate diagnostics plots.

```
lm1 = lm(response ~ dose)
summary(lm1)
```

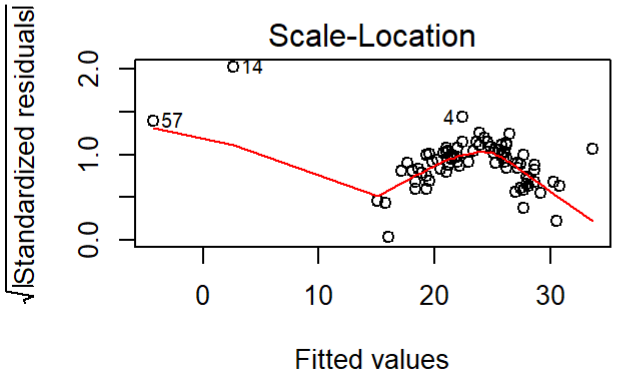
```
##
## Call:
## lm(formula = response ~ dose)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.869  -8.041  -0.872   7.587  35.378
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  39.1472     3.2329  12.109  < 2e-16 ***
## dose         -2.3872     0.4589  -5.202 1.55e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.555 on 78 degrees of freedom
## Multiple R-squared:  0.2576, Adjusted R-squared:  0.248
## F-statistic: 27.06 on 1 and 78 DF,  p-value: 1.55e-06
```

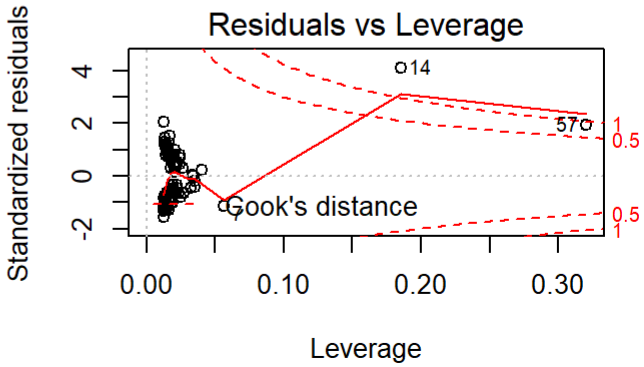
```
par(mfrow=c(2,2))  # Changing panel layout to 2 x 2
plot(lm1)
```





By





looking at the diagnostic plots, we can see that this is not the correct model to fit the data. The two plots in the left column indicate that there is still some signal that needs to be picked up by the model, i.e. 'noise should look like noise' and these two plots do not look like noise. Moving on to the qq plots, we can see that the points do not lie on the y=x axis and they are in fact tilted, indicating that the two sets of quantiles do not come from the same distribution, and thus an indicator of a wrong model.

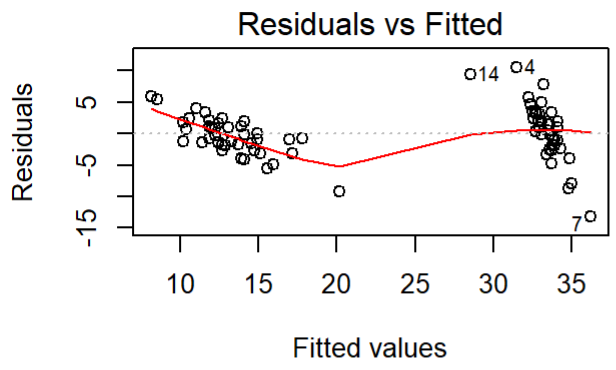
In fact, the estimate for beta is $\hat{\beta} = (39.1472, -2.3872)^T$. This would suggest that increasing dose by an additional unit of the drug has a negative effect, which is counterintuitive. This clearly calls for a better model, and the next one is such one with an interaction term.

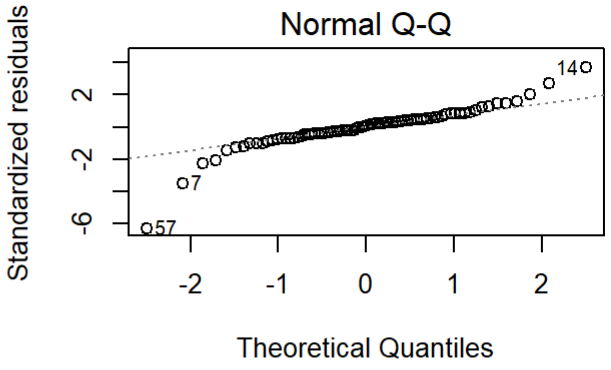
Model 2: response ~ dose + female + dose*female

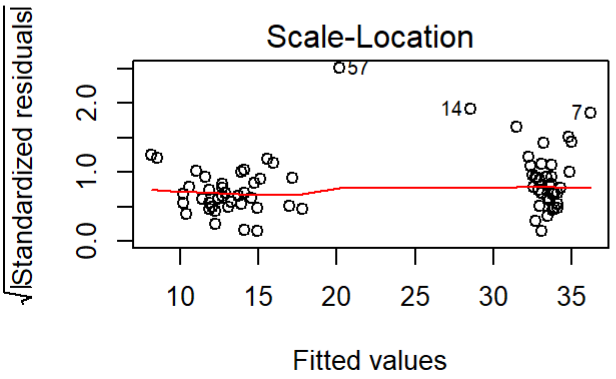
```
lm2 = lm(response ~ dose + female + dose*female)
par(mfrow=c(2,2))  # Changing panel layout to 2 x 2
summary(lm2)
```

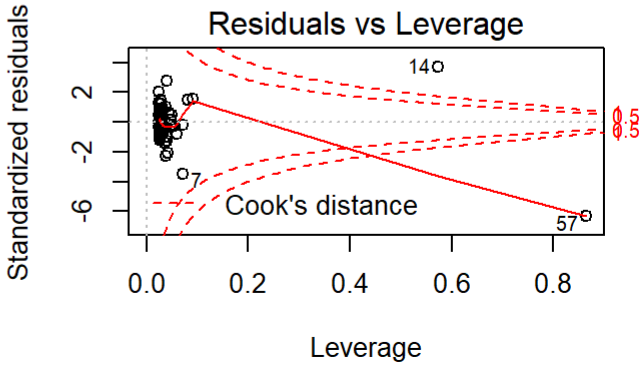
```
##
## Call:
## lm(formula = response ~ dose + female + dose * female)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.2133  -1.8676   0.2004   1.9661  10.5305
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  38.5347     1.6168  23.834  < 2e-16 ***
## dose         -1.0093     0.2798  -3.608 0.000551 ***
## female      -41.5296     3.6021 -11.529  < 2e-16 ***
## dose:female   3.0683     0.4854   6.321 1.62e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.93 on 76 degrees of freedom
## Multiple R-squared:  0.8776, Adjusted R-squared:  0.8728
## F-statistic: 181.7 on 3 and 76 DF,  p-value: < 2.2e-16
```

```
plot(lm2)
```





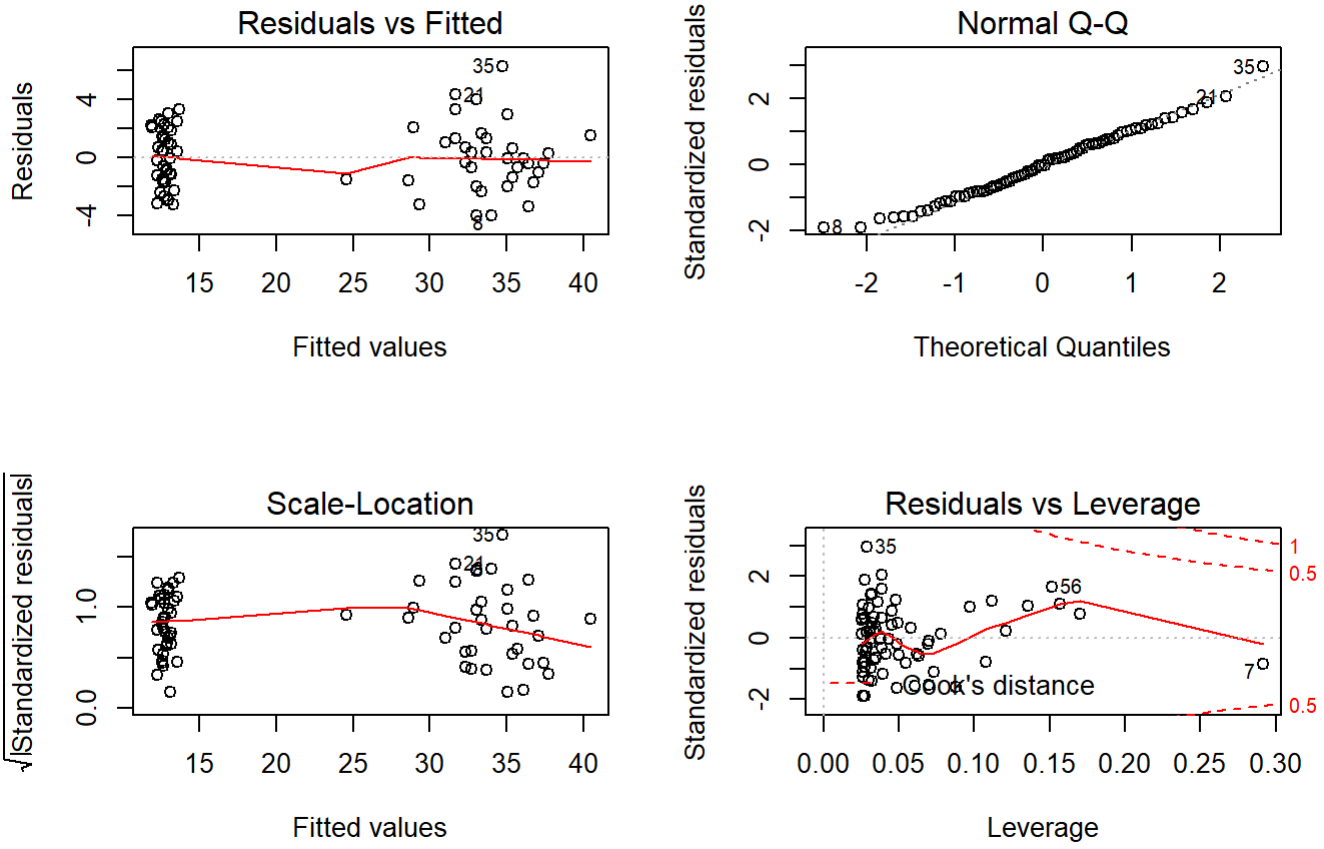




This new interaction model fits better if we look at the QQ-plot. We can see that now most of the quantile-quantile ratios lie on the y=x axis, bar a few points at the extremes. These two points with an unusually large Cook's distance, namely points 14 and 57 are distorting the fit, due to their high leverage. Both of these points have a Cook's distance which is grater than 1. As such, we will consider them outliers, remove them from the dataset and re-fit the model again.

Model 3: Model 2 without outliers

```
response2 = c(response[1:13], response[15:56], response[58:80])
dose2 = c(dose[1:13], dose[15:56], dose[58:80])
female2 = c(female[1:13], female[15:56], female[58:80])
lm3 = lm(response2 ~ dose2 + female2 + dose2*female2)
par(mfrow=c(2,2))  # Changing panel layout to 2 x 2
plot(lm3)
```



This final model seems to be fit the best. Most quantiles fit well on the y=x axis and all of the points have a relatively low Cook's Distance (of < 0.5), implying that outliers are not present in the model anymore. Observation 7 has unusually high leverage, but because it is close to the fit, it has a low Cook's distance, thus not considered an outlier.

```
summary(lm3)
```

```
##
## Call:
## lm(formula = response2 ~ dose2 + female2 + dose2 * female2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0391 -1.6073 -0.0553  1.4281  6.2839
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    16.7747     2.0882   8.033 1.12e-11 ***
## dose2           3.3852     0.4115   8.227 4.82e-12 ***
## female2        -7.0782     3.3225  -2.130   0.0365 *
## dose2:female2  -2.9899     0.5271  -5.673 2.58e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.158 on 74 degrees of freedom
## Multiple R-squared:  0.9626, Adjusted R-squared:  0.9611
## F-statistic: 635.3 on 3 and 74 DF,  p-value: < 2.2e-16
```

When looking at the summary of the model we find that $\hat{\beta} = (16.7747, 3.3852, -7.0782, -2.9899)^T$. This model actually suggests a positive relationship between response and dose, which is what we would expect.

The RSS of lm3 is much lower than that of lm2 and lm1 (2.158, 3.93, 9.555) respectively, showing more evidence that lm3 is the model with the better fit.

Finally, we present the data with the two lines, each representing the model for male and female. The top line represents females and the bottom one represents males:

```
plot(dose2, response2)
abline(16.7747, 3.3852)
abline(16.7747-7.0782, 3.3852 - 2.9899)
title(main="Plot of interaction model")
```

We perform an ANOVA test for lm3 and lm1 and we then make a judgement on which model to reject:

```
anova(lm1, lm3)  # Smaller, Bigger
```

```
## Warning in anova.lmlist(object, ...): models with response '"response2"'
## removed because response differs from model 1
```

```
## Analysis of Variance Table
##
## Response: response
##           Df Sum Sq Mean Sq F value   Pr(>F)
## dose       1 2470.5  2470.5  27.059 1.55e-06 ***
## Residuals 78 7121.5    91.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can see that the F-statistic is 27.059, and the p-value is 1.55e-06. Since the p-value is less than 0.01 we make the following conclusion: There is sufficient evidence to reject the null hypothesis at the 1% level, (corresponding to the smaller model), and therefore accept the alternative. This means we should proceed to use the larger model, which in this case is lm3.

Conclusion of findings:

In the analysis above, we have found that the model fit previously, namely ‘response ~ dose’ fitted poorly and an interaction term was needed. This means that we need to take into account how the sex of the subject affects the results. After fitting a new model ’response ~ dose + female + female*dose’, and taking points 14 and 57 as outliers, we found that this new model fits the data a lot better. In particular, the two results found are as follows: For female subjets, a unit increase in dose (in mg/kg) should result in a decrease in systolic blood pressure of approx. 3.4 mmHg, and for male subjects, we found that a unit increase in dose (in mg/kg) should result in a decrease in systolic blood pressure of approx. 0.4 mmHg according to our new model.

It is important to note that other variables may be at play which not have been accounted for, such as the weight of a patient and the tolerance to the drug in question.

Question 2:

Let X be the design matrix, $X_{(i)}$ be the design matrix when ith observation is removed, and let $y_{(i)} = y - y_i$, and all other variables as in the question.

We first note that $\hat{\beta} = (X^T X)^{-1} X^T \mathbf{y}$ and that $\hat{\beta}_{(i)} = (X_{(i)}^T X_{(i)})^{-1} X_{(i)}^T \mathbf{y}_{(i)}$.

Also, note that $(X_{(i)}^T X_{(i)})^{-1} = X^T X - x_i^T x_i$, which allows us to use the formula given in the notes, thus:

$$(X_{(i)}^T X_{(i)})^{-1} = (X^T X)^{-1} + \frac{(X^T X)^{-1} x_i^T x_i (X^T X)^{-1}}{1 - h_{ii}}.$$

And we use the fact that:

$$X_{(i)}^T y_{(i)} = X^T y - x_i y_i,$$

Using the definition of $\hat{\beta}_{(i)}$ and after some tedious algebra:

$$\begin{aligned}
& \hat{\beta}_{(i)} = (X_{(i)}^T X_{(i)})^{-1} X_{(i)}^T y_{(i)} \\
\Rightarrow \quad & \hat{\beta}_{(i)} = \left((X^T X)^{-1} + \frac{(X^T X)^{-1} x_i^T x_i (X^T X)^{-1}}{1 - h_{ii}} \right) \left(X^T y - x_i y_i \right) \\
\Rightarrow \quad & \hat{\beta}_{(i)} = \hat{\beta} - \frac{(X^T X)^{-1} x_i^T}{1 - h_{ii}} \cdot \left(y_i (1 - h_{ii}) - x_i \hat{\beta} + h_{ii} y_i \right) \\
\Rightarrow \quad & \hat{\beta}_{(i)} = \hat{\beta} - \frac{(X^T X)^{-1} x_i^T}{1 - h_{ii}} \cdot (y_i - \hat{y}_i)
\end{aligned}$$

Therefore:

$$\hat{\beta}_{(i)} - \hat{\beta} = \frac{(X^T X)^{-1} x_i^T (y_i - \hat{y}_i)}{1 - h_{ii}}$$

as required.

Taking the top of Cook’s equation and substituting the result we just found together with $e_i = y_i - \hat{y}_i$:

$$
\begin{aligned}
&(\hat{\beta}_{(i)} - \hat{\beta})^T (X^T X)(\hat{\beta}_{(i)} - \hat{\beta}) \\
=&\frac{e_i^2}{(1-h_{ii})^2} \cdot x_i (X^T X)^{-1} (X^T X)(X^T X)^{-1} x_i^T \\
=&\frac{e_i^2}{(1-h_{ii})^2} \cdot x_i (X^T X)^{-1} x_i^T \\
=&\frac{e_i^2}{(1-h_{ii})^2} \cdot h_{ii}
\end{aligned}
$$

Now, working with the standardised residuals:

$$
r_i = \frac{e_i}{\sqrt{\hat{\sigma}^2(1-h_{ii})}} \quad \Rightarrow \quad r_i^2 = \frac{e_i^2}{\hat{\sigma}^2(1-h_{ii})}
$$

Substituting () into the formula for Cook's distance:

$$
\begin{aligned}
C_i &= \frac{1}{p\hat{\sigma}^2} \cdot \frac{e_i^2}{(1-h_{ii})^2} \cdot h_{ii} \\
&= \frac{h_{ii}}{p(1-h_{ii})} \cdot \left[\frac{e_i^2}{\hat{\sigma}^2(1-h_{ii})} \right] \\
&= \frac{r_i^2}{p} \cdot \frac{h_{ii}}{1-h_{ii}},
\end{aligned}
$$

as required.

Question 3:

We know that

$$
\hat{y} = Py,
$$

where

$$
P = X(X^T X)^{-1} X^T
$$

is the 'hat matrix' and is itself a projection matrix.

Now taking

$$
e = y - \hat{y} = y - Py = (I - P)y,
$$

where I is the identity matrix.

It turns out that (I - P) projects to the vector space perpendicular to the row-space of X. After equating:

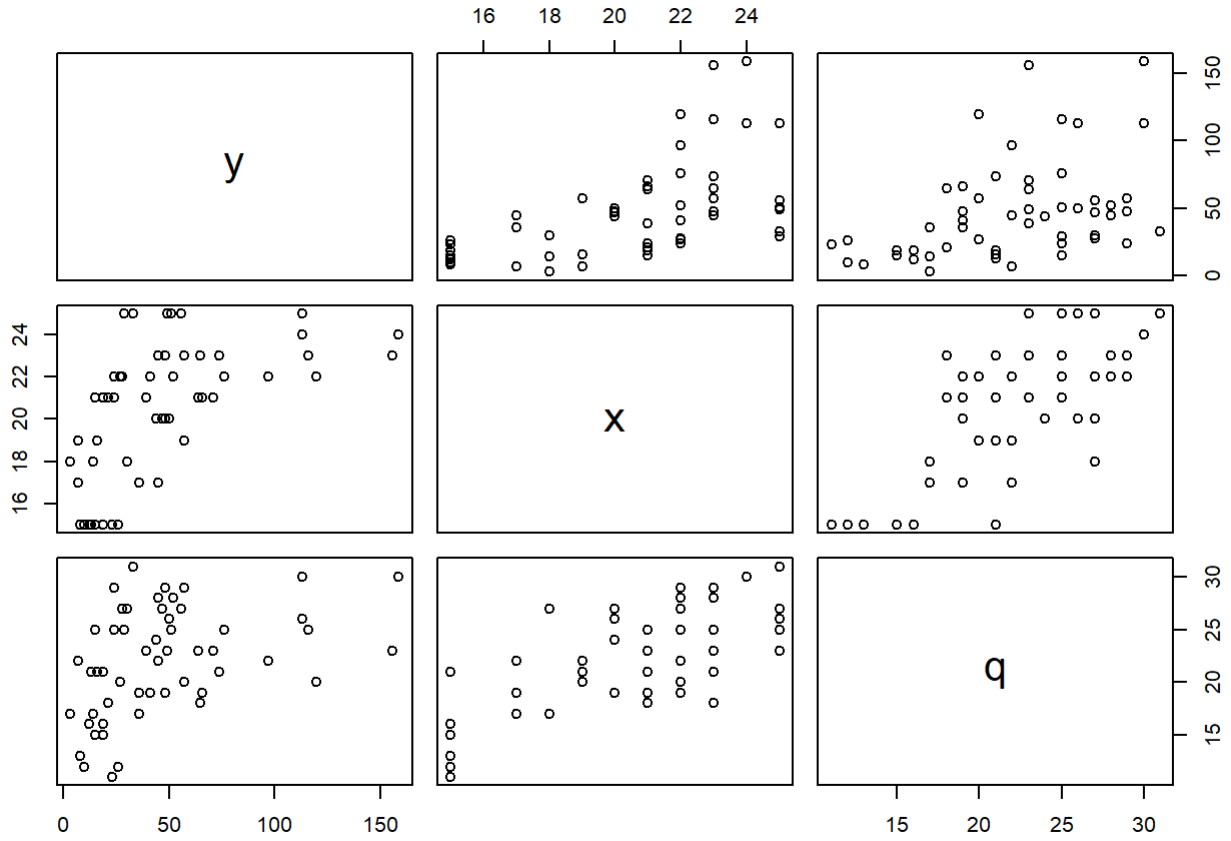
$$
e = (I - P)y = X \cdot (\alpha, \beta)^T,
$$

we can see that the RHS of this equation is a linear combination of the rows of X, and it is given in the question, but now is in matrix form. But equating two perpendicular matrices can only happen if they are equal to 0. Therefore, the only way this equation is satisfied is if it alpha and beta are precisely 0. Thus result is shown.

Part 2: Generalized Linear Modelling

For this part, we are concerned with reading accuracy in children. We begin by plotting the data ‘read’.

```
plot(read)
```



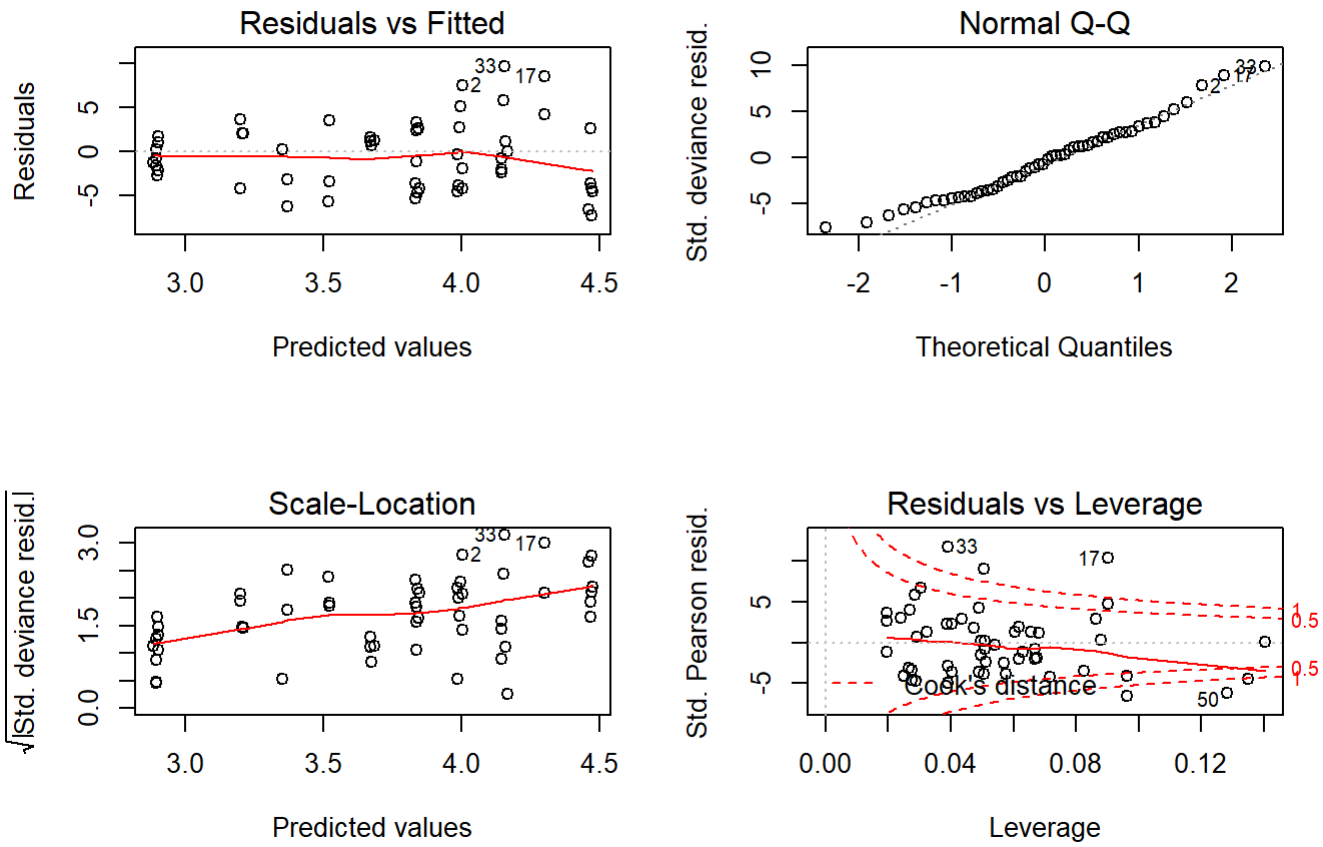
Question 4:

To repeat the initial analysis, we fit a Poisson GLM:

```
x = read$x  # Score out of 25 on a standardised inventory for attention (higher is bet.)
q = read$q  # Standardised measure of verbal fleuncy (higher is better)
y = read$y  # count of words correctly pronounced before the third error
glm1 = glm(y ~ x + q, family=poisson(link='log'))  # GLM without interaction term
summary(glm1)
```

```
##
## Call:
## glm(formula = y ~ x + q, family = poisson(link = "log"))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -7.2815  -3.5994  -0.5083   2.1170   9.7113
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.532923   0.157538   3.383 0.000717 ***
## x            0.159579   0.009631  16.569  < 2e-16 ***
## q           -0.002005   0.005940  -0.338 0.735657
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1353.08  on 53  degrees of freedom
## Residual deviance:  813.58  on 51  degrees of freedom
## AIC: 1110.4
##
## Number of Fisher Scoring iterations: 5
```

```
par(mfrow=c(2,2))  # Changing panel layout to 2 x 2
plot(glm1)
```




```
par(mfrow=c(1,1))  # Reverting panel layout back to 1 x 1
```

The estimated beta is $\hat{\beta} = (0.532923, 0.159579, -0.002005)^T$. This suggests a weak positive correlation between y and x, and almost no relation between y and q.

Upon inspecting the diagnostic plots, we can see that the QQ plot indicates a misfit at the tails of the distribution.

There are a lot of points with a Cook's distance which is close to 1. This is undesirable, as it indicates a sub-optimal fit for this model.

We also take note of the AIC value, which is equal to 1110.4 for this model. The AIC will be used in Question 8.

Question 5:

For a distribution to be a member of the exponential family it must be of the form

$$\exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right]$$

If we take the PMF of the negative binomial, and we find that

$$
\begin{aligned}
\binom{y+r-1}{y}(1-p)^r p^y &= \exp\left[\log\left(\binom{y+r-1}{y}(1-p)^r p^y\right)\right] \\
&= \exp\left[\log\binom{y+r-1}{y} + \log((1-p)^r) + \log(p^y)\right] \\
&= \exp\left[\log\binom{y+r-1}{y} + r\log(1-p) + y\log(p)\right] \\
&= \exp\left[\frac{y\log(p) - (-r\log(1-p))}{1} + \log\binom{y+r-1}{y}\right] \\
&= \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right]
\end{aligned}
$$

if we take $\theta = \log(p)$, then $p = \exp(\theta)$.

Furthermore, if we let $a(\phi) = 1, b(\theta) = -r\log(1-p) = -r\log(1-\exp(\theta))$ and

$$
c(y,\phi) = \log\binom{y+r-1}{y},
$$

this yields the required form.

Question 6:

To derive the IWLS algorithm, we do some calculations similar to the ones on Pages 61, 62 and 63 of the course lecture notes.

We have:

$$
\begin{aligned}
\mu_i &= \mathbb{E}(Y_i) = \frac{re^{\theta_i}}{1-e^{\theta_i}} \\
\theta_i &= \log(p) \\
a(\phi) &= 1 \\
b(\theta_i) &= -r\log(1-e^{\theta_i}) \\
b'(\theta_i) &= \frac{d}{d\theta_i}\left(-r\log(1-e^{\theta_i})\right) = \frac{re^{\theta_i}}{1-e^{\theta_i}} \\
b''(\theta_i) &= \frac{re^{\theta_i}}{(1-e^{\theta_i})^2} \\
V(\mu_i) &= \frac{Var(Y_i)}{a(\phi)} = \frac{re^{\theta_i}}{(1-e^{\theta_i})^2} \\
\eta_i &= \log(\mu_i) \\
\frac{d\eta_i}{d\mu_i} &= \frac{1}{\mu_i} = \frac{1-e^{\theta_i}}{re^{\theta_i}} \\
z_i &= \hat{\eta}_i + (y_i - \hat{\mu}_i)\frac{d\eta_i}{d\mu_i}\Big|_{\mu_i=\hat{\mu}_i} \\
&= \hat{\eta}_i + (y_i - \hat{\mu}_i)\frac{1}{\mu_i} \\
w_{ii}^{-1} &= Var(Y_i)\left(\frac{d\eta_i}{d\mu_i}\right)^2 \\
&= \frac{re^{\theta_i}}{(1-e^{\theta_i})^2}\left(-\frac{1}{\mu_i^2}\right) \\
&= \frac{1}{re_i^{\theta}} = \frac{1}{rp} \\
w_{ii} &= rp \\
p &= \frac{\mu}{\mu + r}
\end{aligned}
$$

With the response variable z and the weight w computed we simply follow $\mathbf{Algorithm\ 3.1}$ from the notes with our z and w.

Question 7:

In this question, I will implement the algorithm derived in question 6:

We begin by having an estimate of beta, and in our case:

```
beta = c(-0.7, 0.25, 0.02)  # Sensible estimate
```

Next, we use our design matrix and pick $r = 3$, as we are aiming for 3 pronunciation errors:

```
X = cbind(1, x, q)
r = 3  # Pick r
```

Following this, we generate a deviance function for the negative binomial. The formulas for the deviance function is given by

$$D = 2\phi\{\ell(y, \phi; y) - \ell(\hat{\mu}, \phi; y)\}$$

.

In our case, the likelihood is

$$\prod_{y=0}^{n} \binom{y_i + r - 1}{y_i} (1-p)^r p^{y_i}$$

and the deviance, after some algebra simplifies to:

$$D = 2\sum_{y=0}^{n} \left[y_i \log\left(\frac{y_i(r+\hat{\mu})}{\hat{\mu}(r+y_i)} \right) + r \log\left(\frac{\hat{\mu}+r}{y_i+r} \right) \right].$$

This can be coded as:

```
# Neg. bin. Deviance function
D = function(y, r, mu){
  a = y*log((y*(r + mu))/(mu*(r + y)))  # Simplified
  b = r*log((mu + r)/(y + r))  # Simplified
  a[y==0] = 0  # Check for log(0/0)
  2*sum(a + b)
}
```

Next we initialise first D, jj and a counter:

```
oldD = D(y, r, exp(X %*% beta))
jj = 0
counter = 0  # Keeping track of iterations
```

Finally the iterations:

```
while(jj==0){
  eta = X %*% beta  # Estimated linear predictor
  mu = exp(eta)  # Estimated mean response
  deta_dmu = 1/mu
  z = eta + (y - mu)*deta_dmu  # Adjusted dependent variable
  w =  r*mu/(mu + r)  # = r*p
  lmod = lm(z ~ x + q, weights=w, data = read)  # Regress y on x + q with weights w
  beta = as.vector(lmod$coeff)  # New beta
  newD = D(y, r, exp(X %*% beta))
  control = abs(newD - oldD)/(abs(newD) + 0.1)
  counter = counter  + 1
  if(control < 1e-8)
    jj = 1
  oldD = newD
}
```

Displaying results:

```
counter # No. of iterations taken
```

```
## [1] 6
```

```
beta  # Final estimate
```

```
## [1]  0.273277461  0.174806561 -0.004742558
```

Question 8

We now check a few quantities for our negative binomial model:

```
# Computing standard errors:
J = t(X) %*% diag(as.vector(w)) %*% X
invJ = solve(J)
beta.sd = sqrt(as.vector(diag(invJ)))
beta.sd
```

```
## [1] 0.53530948 0.03830815 0.02448842
```

We now calculate the deviance residuals:

```
p = as.vector(exp(X %*% beta))
a = y*log((y*(r + mu))/(mu*(r + y)))  # Simplified
b = r*log((mu + r)/(y + r))  # Simplified
a[y==0] = 0
d = sign(y - mu)*sqrt(2*(a + b))
summary(d)
```

```
##        V1
##  Min.   :-2.5177
##  1st Qu.:-0.9918
##  Median :-0.1200
##  Mean   :-0.1783
##  3rd Qu.: 0.5872
##  Max.   : 1.7192
```

Caculating p-values for individual parameters:

```
# Testing individual parameters and corresponding p-values are:
z = beta/beta.sd
z
```

```
## [1]  0.5105037  4.5631690 -0.1936653
```

```
# P-Values:
2*(1 - pnorm(abs(z), lower.tail=TRUE))
```

```
## [1] 6.096986e-01 5.038722e-06 8.464380e-01
```

Attempting to plot negative binomial function from estimates of beta:

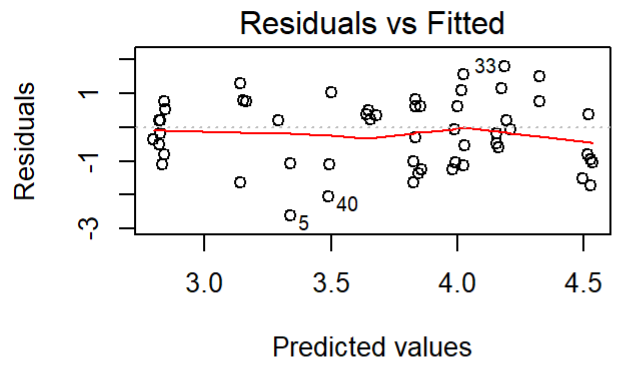
```
library(MASS)
lm = MASS::glm.nb(y ~ x + q, data = read)
summary(lm)
```

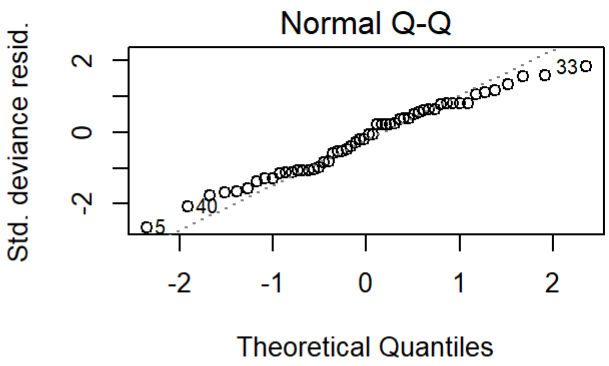
```
##
## Call:
## MASS::glm.nb(formula = y ~ x + q, data = read, init.theta = 3.289395476,
##     link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6110  -1.0355  -0.1250   0.6135   1.7974
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.274117   0.513859   0.533    0.594
## x            0.174761   0.036720   4.759 1.94e-06 ***
## q           -0.004737   0.023468  -0.202    0.840
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(3.2894) family taken to be 1)
##
##     Null deviance: 98.367  on 53  degrees of freedom
## Residual deviance: 56.496  on 51  degrees of freedom
## AIC: 491.38
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  3.289
##           Std. Err.:  0.669
##
##  2 x log-likelihood:  -483.385
```

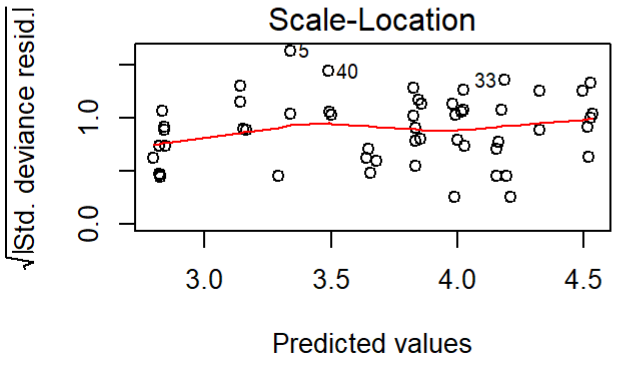
The summary agrees approx. with our results from IWLS.

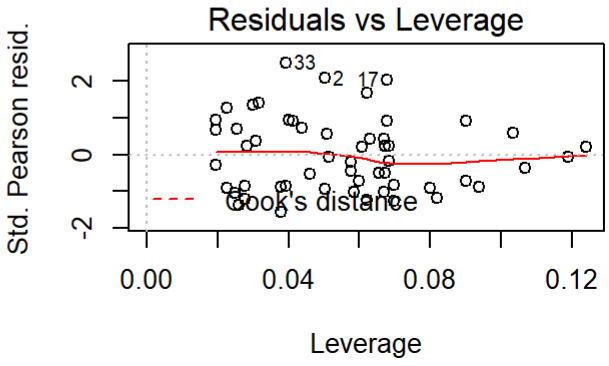
We now plot diagnostics plots for the neg. binomial model:

```
par(mfrow=c(2,2))  # Changing panel layout to 2 x 2
plot(lm)
```



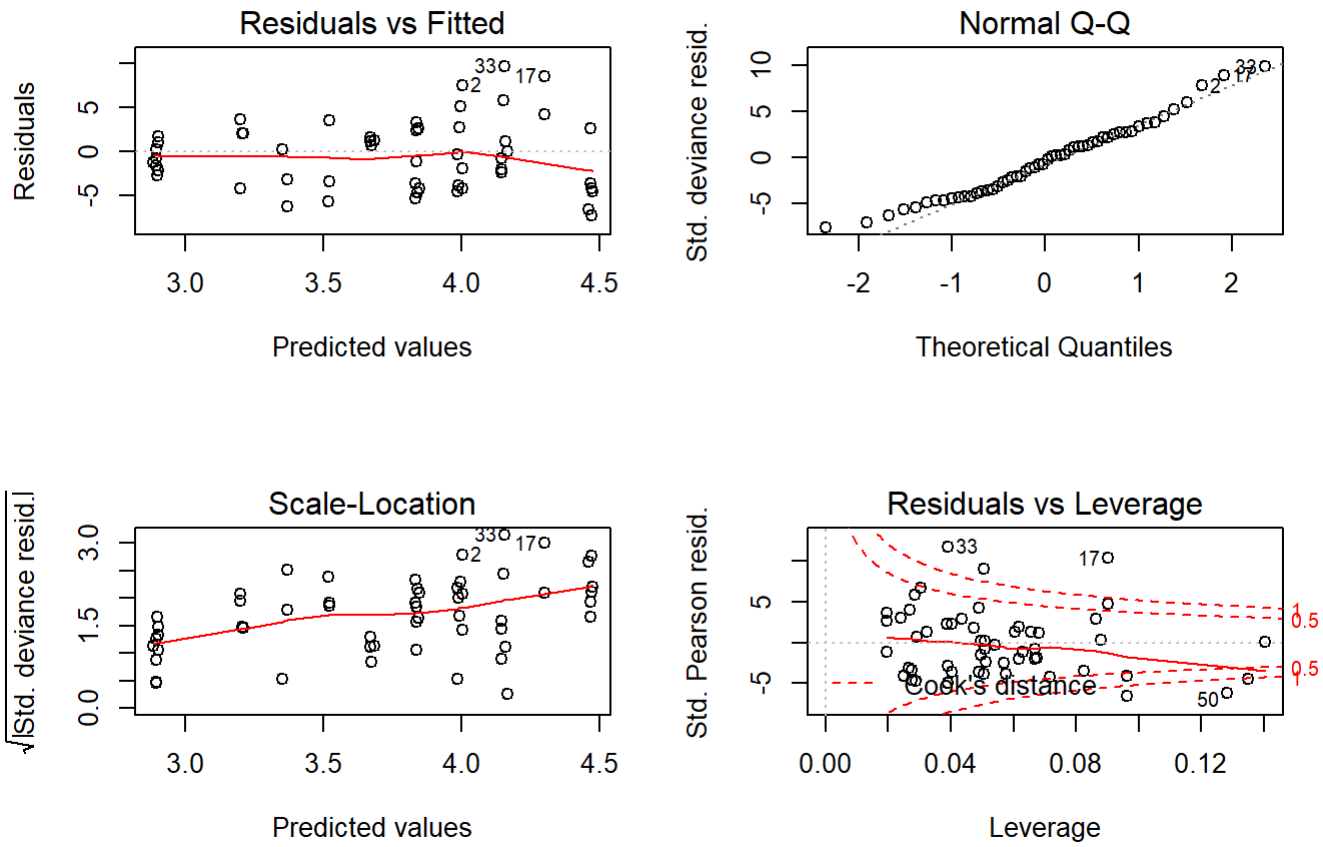






And diagnostics plots for the Poisson model again:

```
par(mfrow=c(2,2))  # Changing panel layout to 2 x 2
plot(glm1)
```



When comparing the Diagnostics plots:

Residuals vs Fitted plots: There does not seem to be a loss of signal from these plots, and we can see that the residuals are smaller for the Binomial model, compared to the Poisson model, so this plot supports the Binomial model as the better fit.

Scale-Location: The std. deviance residual is smaller for the Binomial model compared to the Poisson model, suggesting that the Binomial is a better fit here too.

Normal Q-Q Plot: The two distributions produce similar Q-Q plots, and both have some distortions at the tails, but the middle quantiles lie approx. on the y=x line suggesting a good fit for both.

Residual vs Leverage: Here, the Binomial model is a clear winner. The leverage for both plots is of a similar amount, but the std. pearson residuals are different. This is greatly reflected in the Cook's distance; in the binomial model, every point has a low Cook's distance, and the 0.5 contour line is not even present, whereas the Poisson model has lots of points with high Cook's distances. According to this last plot, the Binomial model is the better fit.

When looking at the AIC, the binomial model has an AIC of 491.38, whereas the Poisson model has an AIC of 1110.4, and a lower AIC suggests a better fit, thus the Binomial model comes on top here too.

Extracting coefficients from models:

```
betanbin = exp(coef(glm1))
betanbin
```

```
## (Intercept)           x           q
##   1.7039048   1.1730165   0.9979967
```

```
betapoiss = exp(coef(lm))
betapoiss
```

```
## (Intercept)           x           q
##    1.315368    1.190962    0.995274
```

One of the assumptions made was that the children were of the same age.

Brief summary of the trends found:

After the statistical analysis trying to decide which model is best between the Poisson and the Negative Binomial model, we have decided to pick the Binomial model for the reasons outlined above. With this model, the results are as follows: A unit increase in either x or q will result in approx. 2 extra words correctly pronounced before the third error. This is to be expected, as one would expect a more attentive child or a verbally more fluent child to pronounce words more correctly.