A4

Adrien Dinzey, 7842736

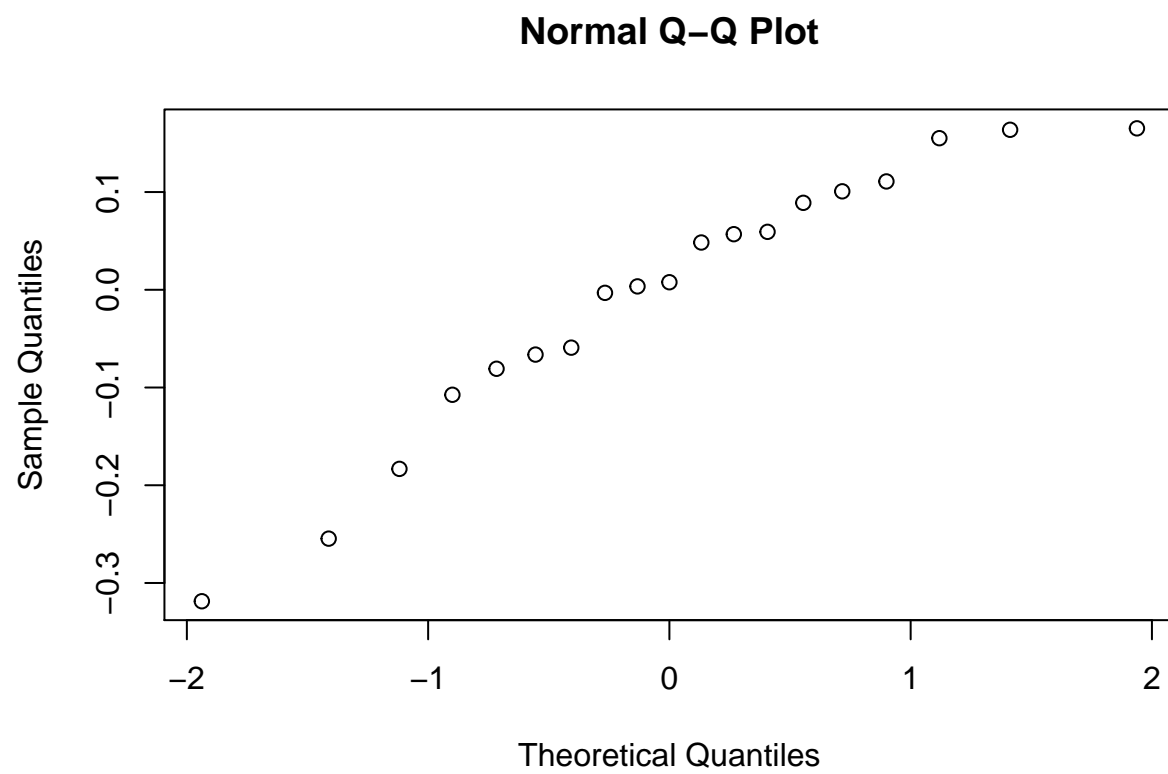23/04/2021

Q1

load in data:

```
library('readxl')
data1<-read_excel("A4-datasets.xlsx")
```

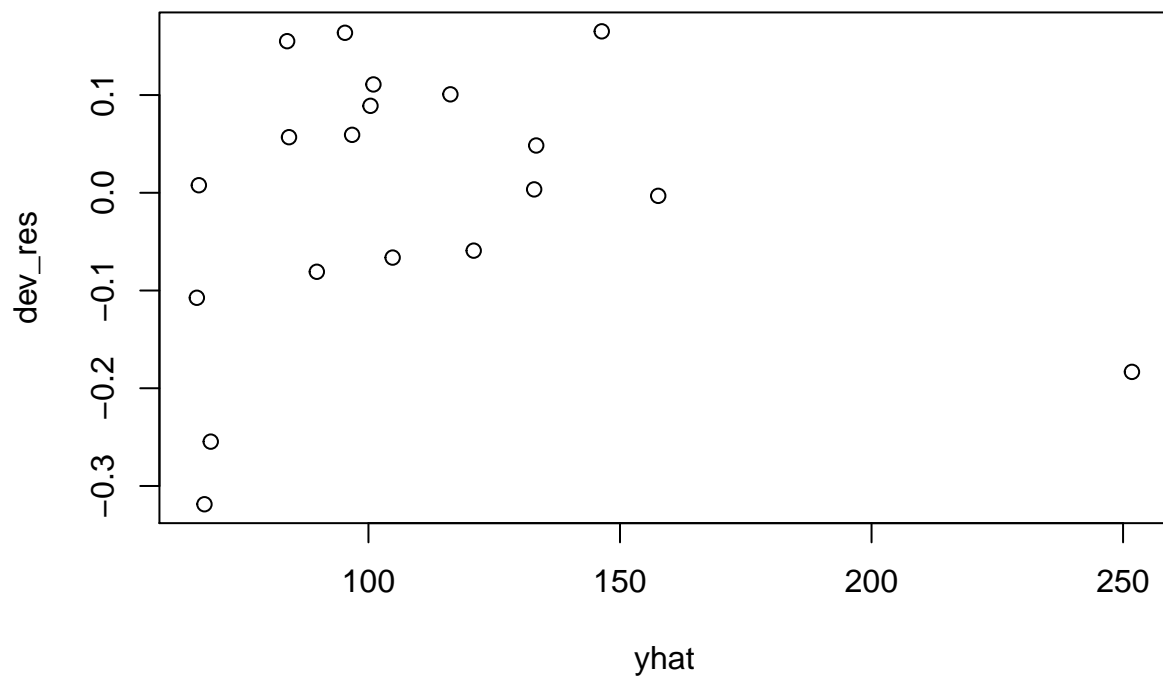Using the cannonical link (that is the recirpocal for Gamma dsitribution)

```
data1.model<-glm(y~x1+x2+x3,family=Gamma(),data=data1)
summary(data1.model)
```

```
##
## Call:
## glm(formula = y ~ x1 + x2 + x3, family = Gamma(), data = data1)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.31869  -0.07357   0.00772   0.09484   0.16517
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.0227448  0.0147363  -1.543  0.14355
## x1           0.0003764  0.0011532   0.326  0.74863
## x2           0.0001708  0.0001021   1.673  0.11496
## x3          -0.0018241  0.0004629  -3.941  0.00131 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 0.02128171)
##
##     Null deviance: 2.65335  on 18  degrees of freedom
## Residual deviance: 0.34362  on 15  degrees of freedom
## AIC: 163.58
##
## Number of Fisher Scoring iterations: 4
```
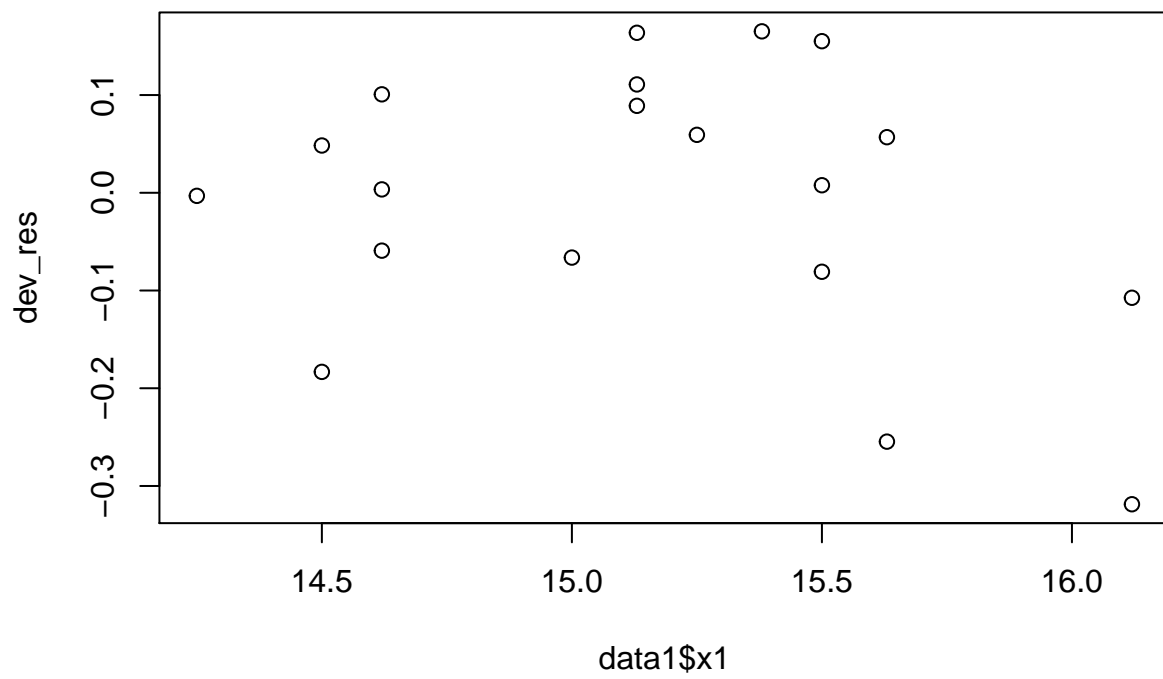
```
yhat<-data1.model$fit
dev_res<-residuals(data1.model,c="deviance")
qqnorm(dev_res)
```
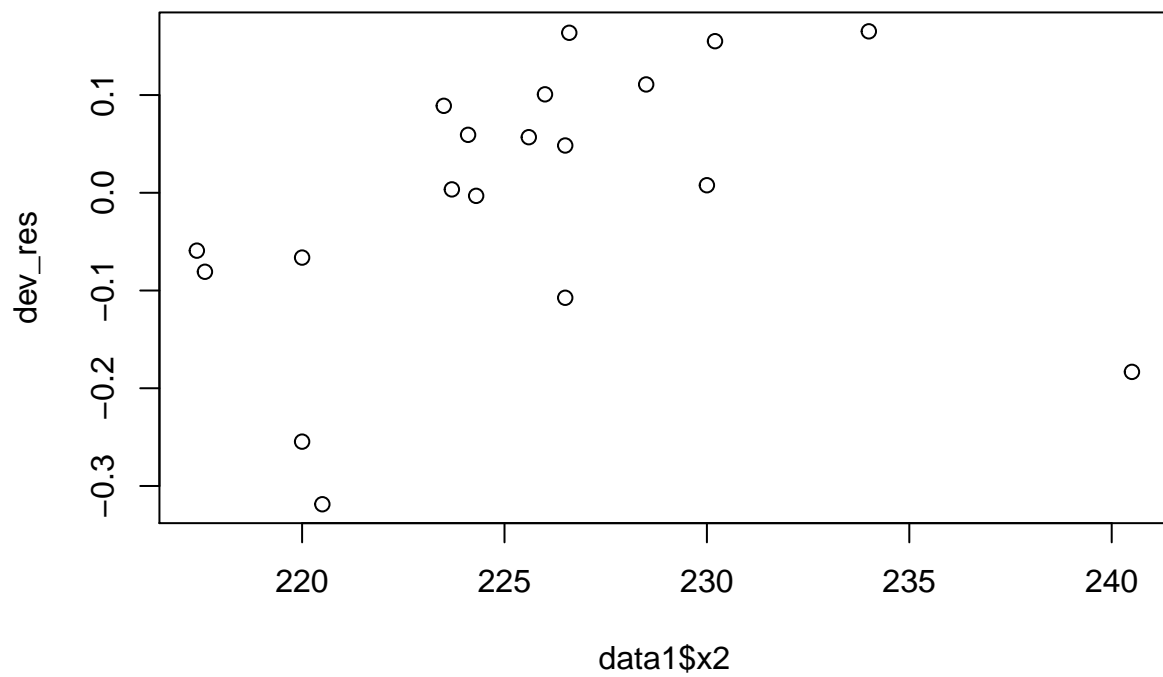
Normal Q–Q Plot

Sample Quantiles (y-axis)

Theoretical Quantiles (x-axis)

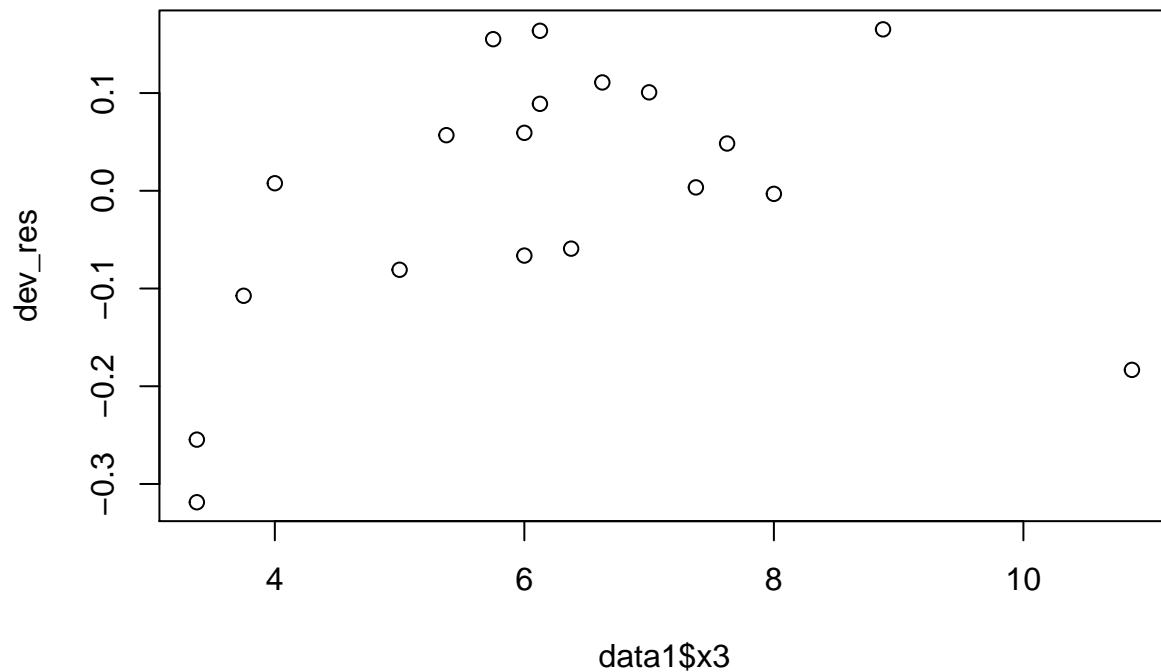```
plot(yhat,dev_res)
```

```r
plot(data1$x1,dev_res)
```

```r
plot(data1$x2,dev_res)
```
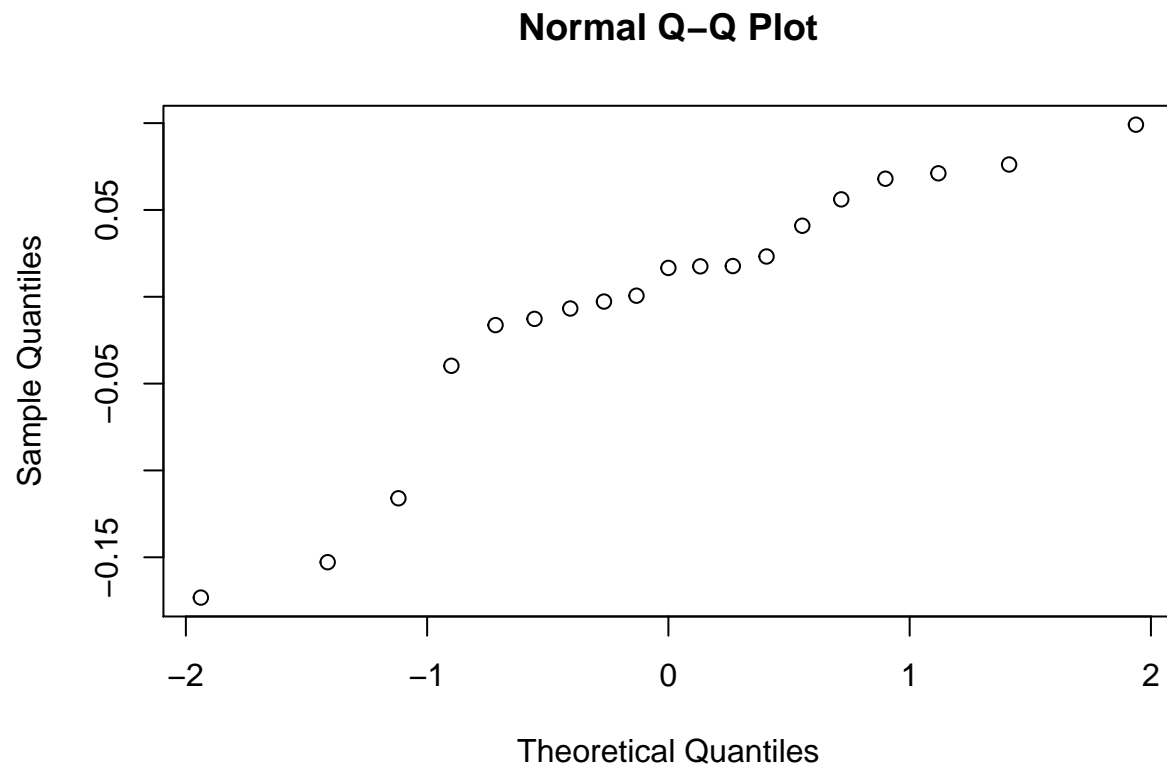
```r
plot(data1$x3,dev_res)
```

Using the log link

```
data1.model<-glm(y~x1+x2+x3,family=Gamma(link=log),data=data1)
summary(data1.model)
```

```
##
## Call:
## glm(formula = y ~ x1 + x2 + x3, family = Gamma(link = log), data = data1)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.17320  -0.01451   0.01663   0.04851   0.09911
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.557725   0.992757   5.598 5.08e-05 ***
## x1          -0.100583   0.068945  -1.459    0.165
## x2          -0.002372   0.005277  -0.449    0.660
## x3           0.182970   0.024159   7.573 1.68e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 0.006358647)
##
##     Null deviance: 2.65335  on 18  degrees of freedom
## Residual deviance: 0.10044  on 15  degrees of freedom
```

```
## AIC: 140.17
##
## Number of Fisher Scoring iterations: 4

yhat<-data1.model$fit
dev_res<-residuals(data1.model,c="deviance")
qqnorm(dev_res)
```
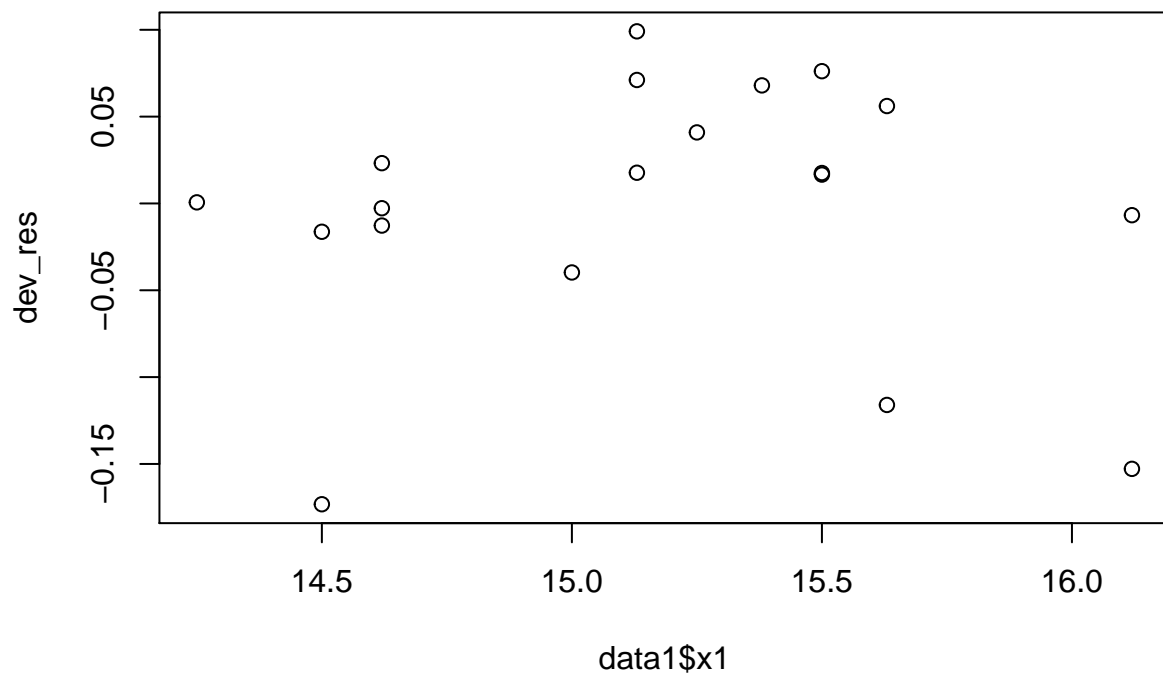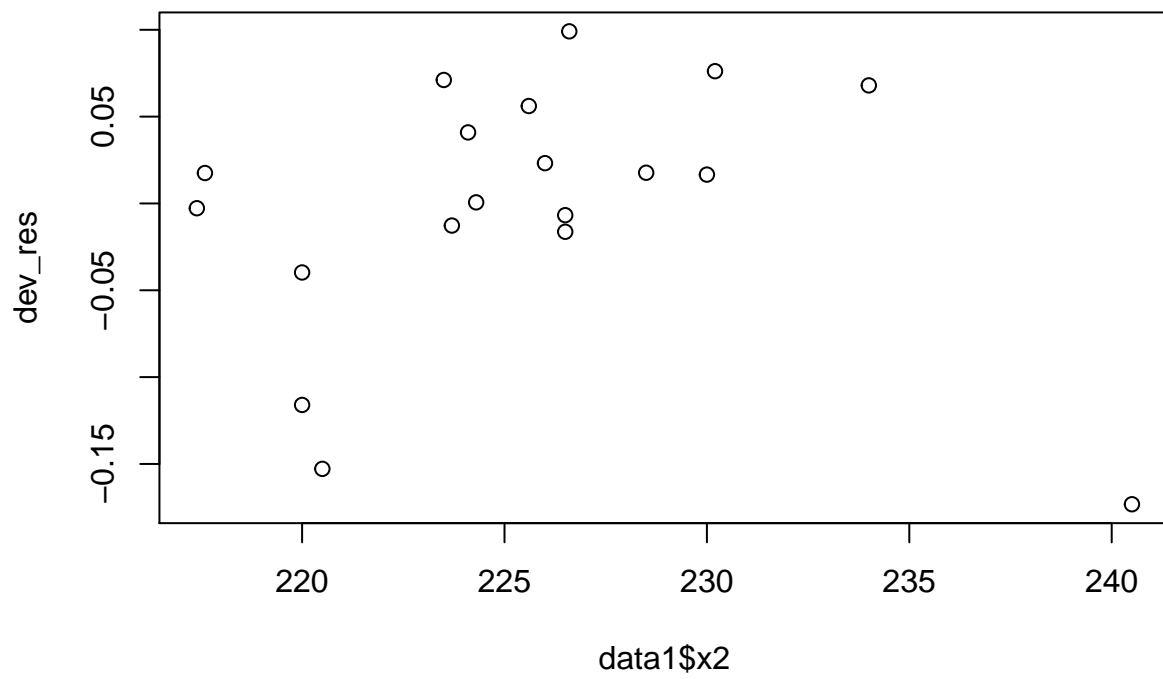
Normal Q–Q Plot
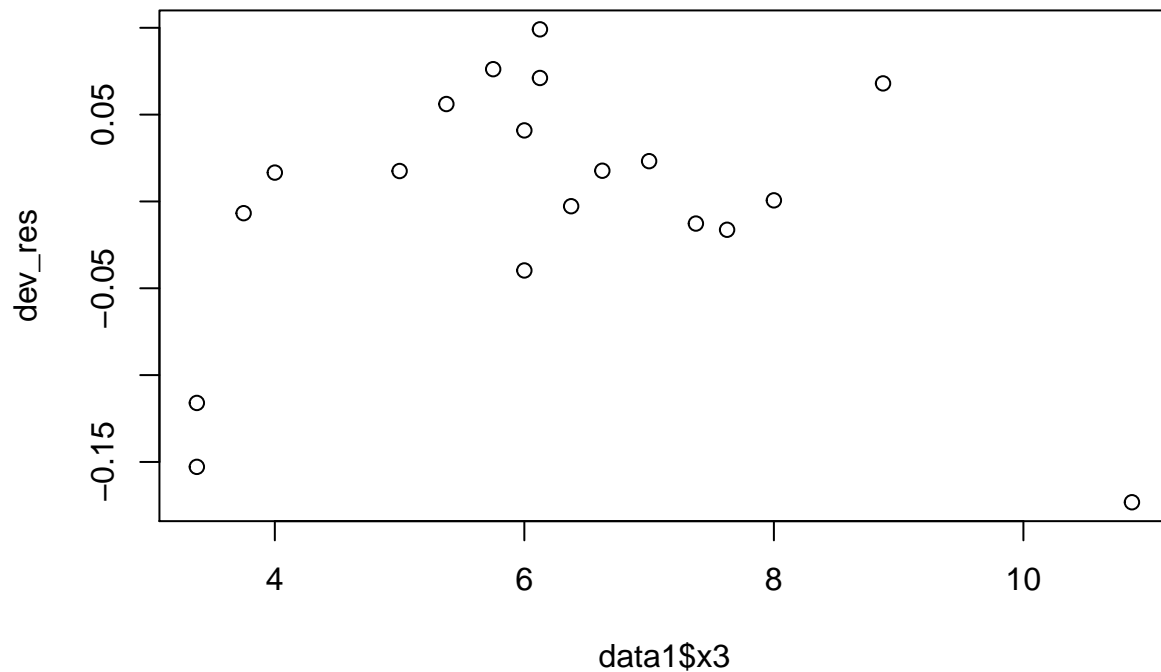


```
plot(yhat,dev_res)
```

```r
plot(data1$x1,dev_res)
```

```r
plot(data1$x2,dev_res)
```
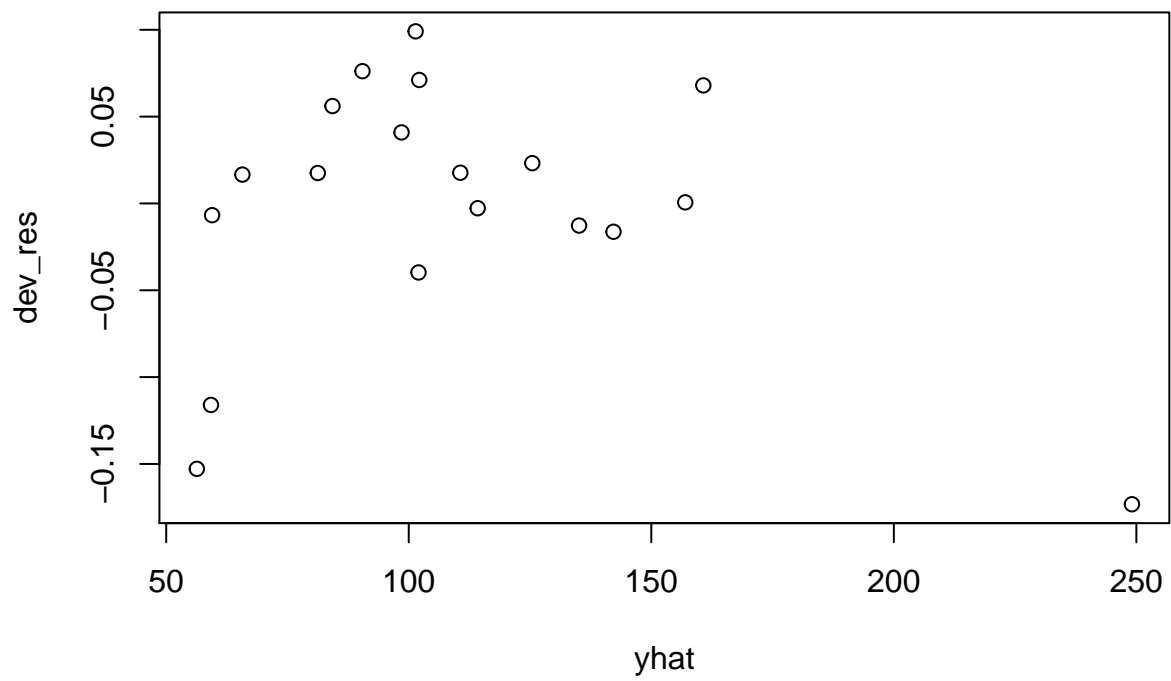
```r
plot(data1$x3,dev_res)
```

We see that the residuals seem to be distributed the same way using either link. The QQ plot of the canonical link seems to look like a normal qq plot while the log link qq plot has a bit of a bend near the middle and is pushed upwards.

Q2

```
data2<-read_excel("A4-datasets.xlsx",sheet = 2)
```

```
data2.model<-glm(y~x,family = "poisson",data2)
summary(data2.model)
```

```
##
## Call:
## glm(formula = y ~ x, family = "poisson", data = data2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -40.797  -26.045   -4.466   15.657   43.214
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 7.9210837  0.0064719    1224   <2e-16 ***
## x           0.0008271  0.0000044     188   <2e-16 ***
```

11

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 45458  on 14  degrees of freedom
## Residual deviance: 11187  on 13  degrees of freedom
## AIC: 11346
##
## Number of Fisher Scoring iterations: 4
```
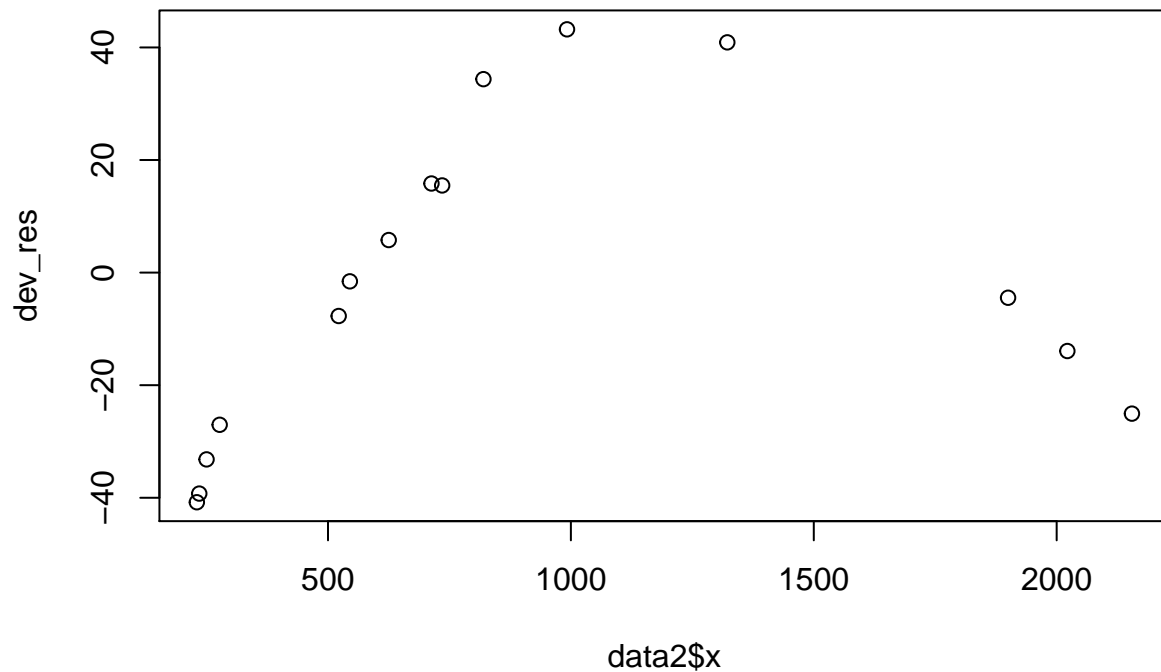
```
yhat <- data2.model$fit
dev_res <- residuals(data2.model, c="deviance")
qqnorm(dev_res)
```

Normal Q–Q Plot


```
plot(yhat,dev_res)
```
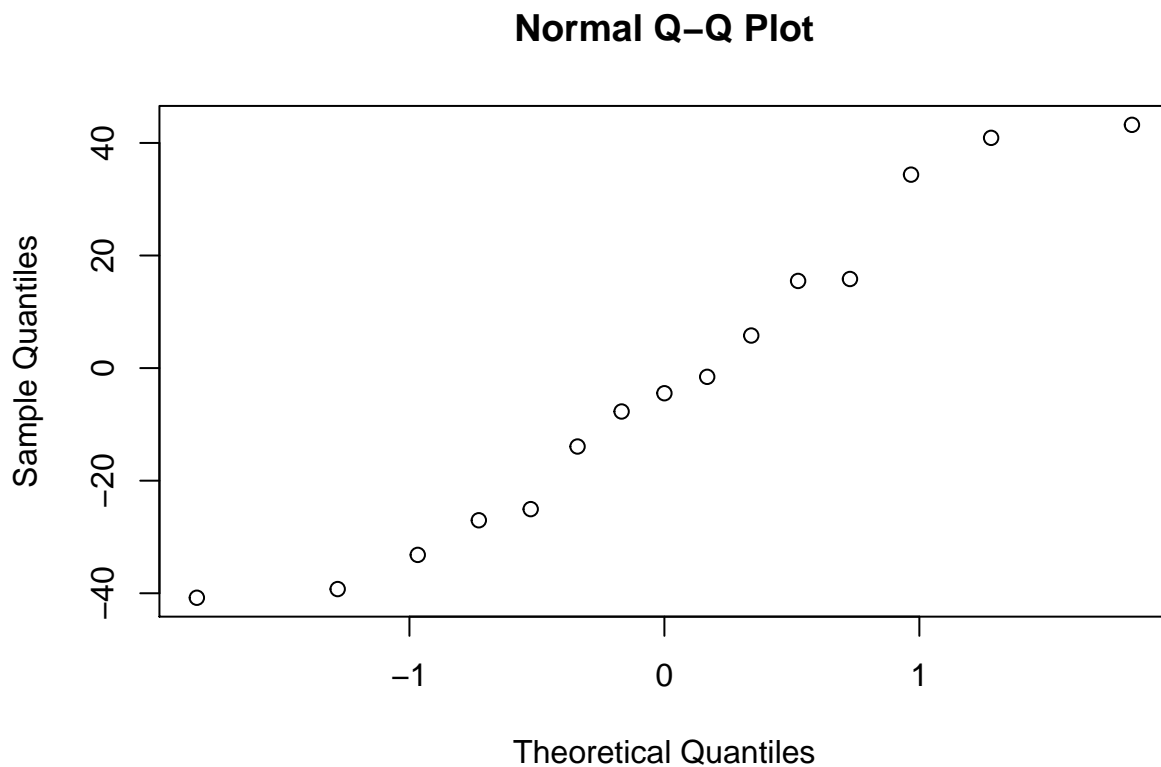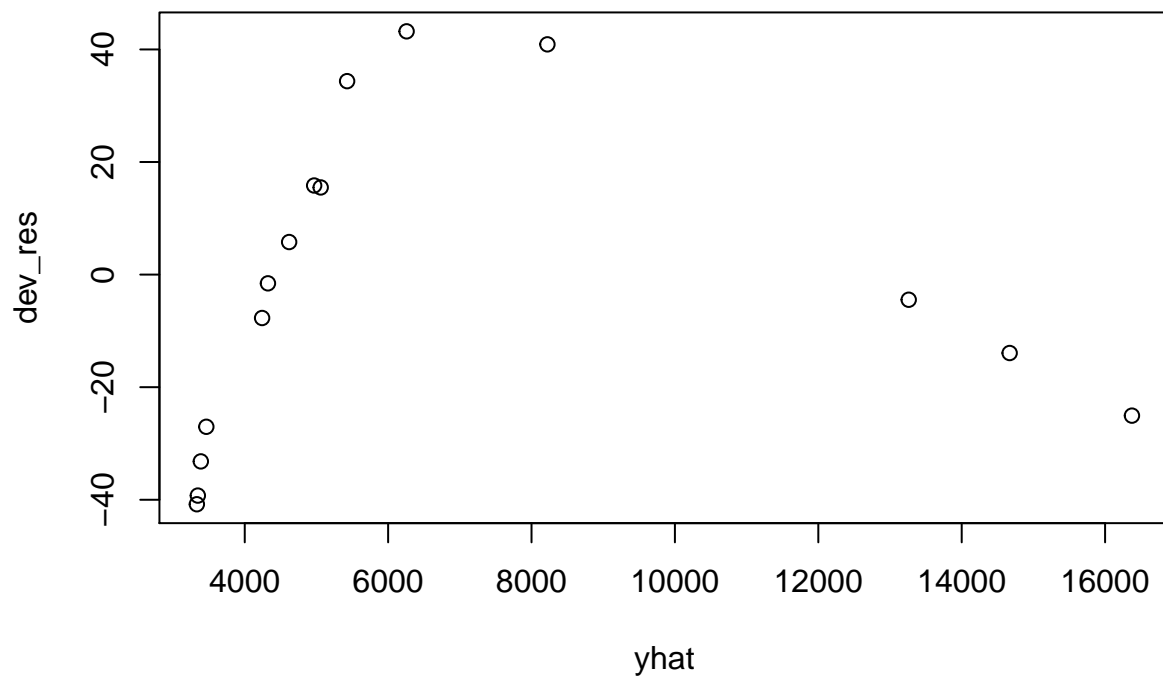
```r
plot(data2$x,dev_res)
```

We see that the residuals plotted against yhat seems to look like some sort of F distribution while when plotted against x it looks like an upside down parabola. Its possible that a power link would be better suited for this.
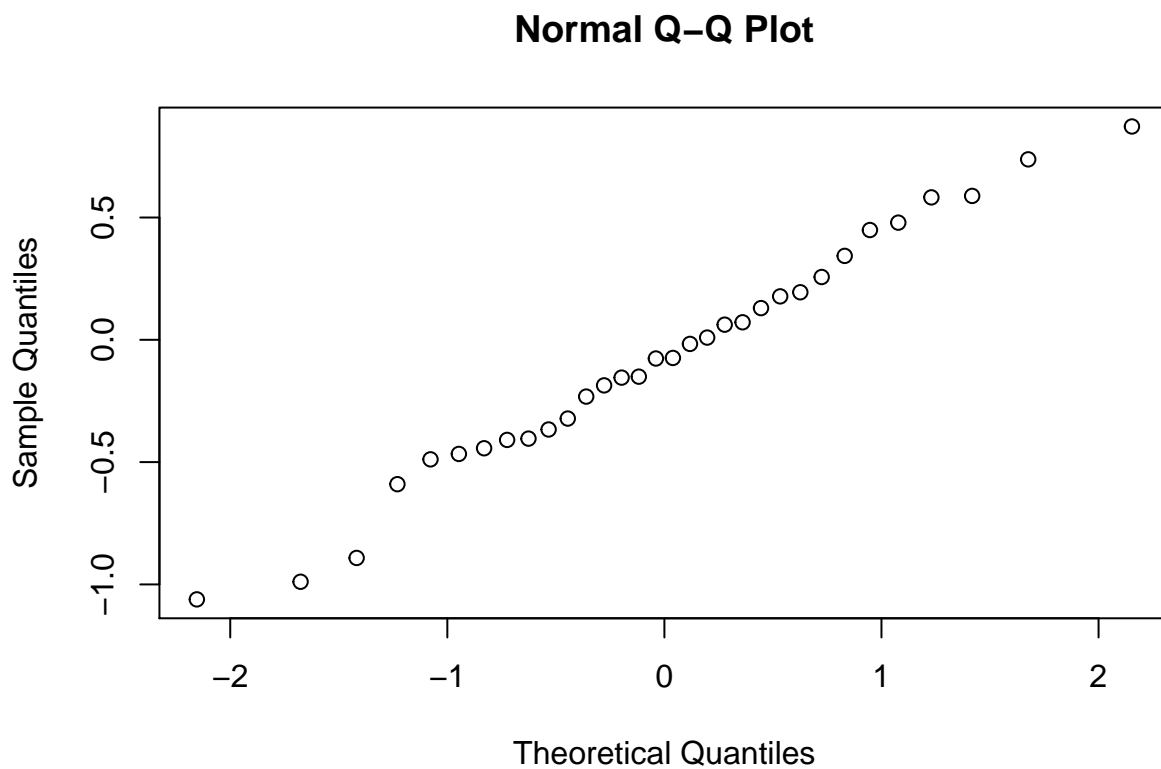
Q3

```
data3<-read_excel("A4-datasets.xlsx",sheet=3)
```

```
data3.model<-glm(y~x1+x2+x3,family=Gamma(link=power(1/2)),data=data3)
summary(data3.model)
```

```
##
## Call:
## glm(formula = y ~ x1 + x2 + x3, family = Gamma(link = power(1/2)),
##     data = data3)
##
## Deviance Residuals:
##      Min         1Q     Median         3Q        Max
## -1.06118   -0.40520   -0.07525    0.21006    0.87221
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.5794452  0.8368281    1.887 0.069510 .
```

```
## x1              0.0154558   0.0040105    3.854 0.000621 ***
## x2              0.0003347   0.0004987    0.671 0.507636
## x3             -0.0392398   0.0425135   -0.923 0.363896
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 0.2417801)
##
##     Null deviance: 10.5283  on 31  degrees of freedom
## Residual deviance:  7.3169  on 28  degrees of freedom
## AIC: 155.34
##
## Number of Fisher Scoring iterations: 19
```
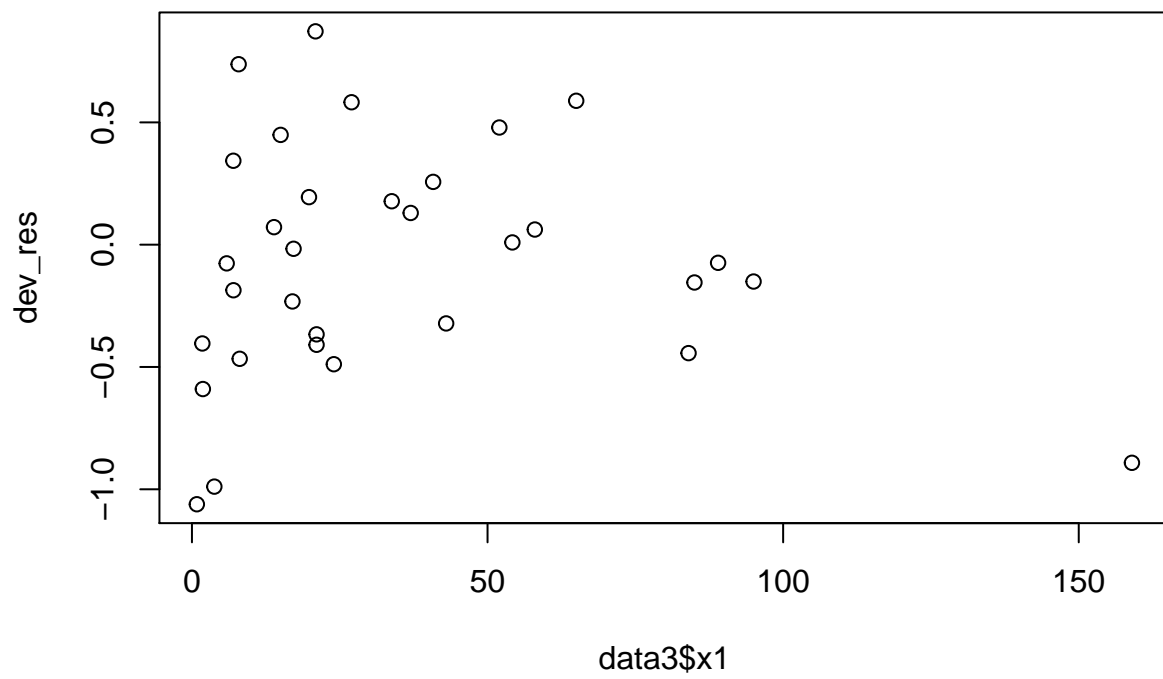
```
yhat<-data3.model$fit
dev_res<-residuals(data3.model,c="deviance")
qqnorm(dev_res)
```
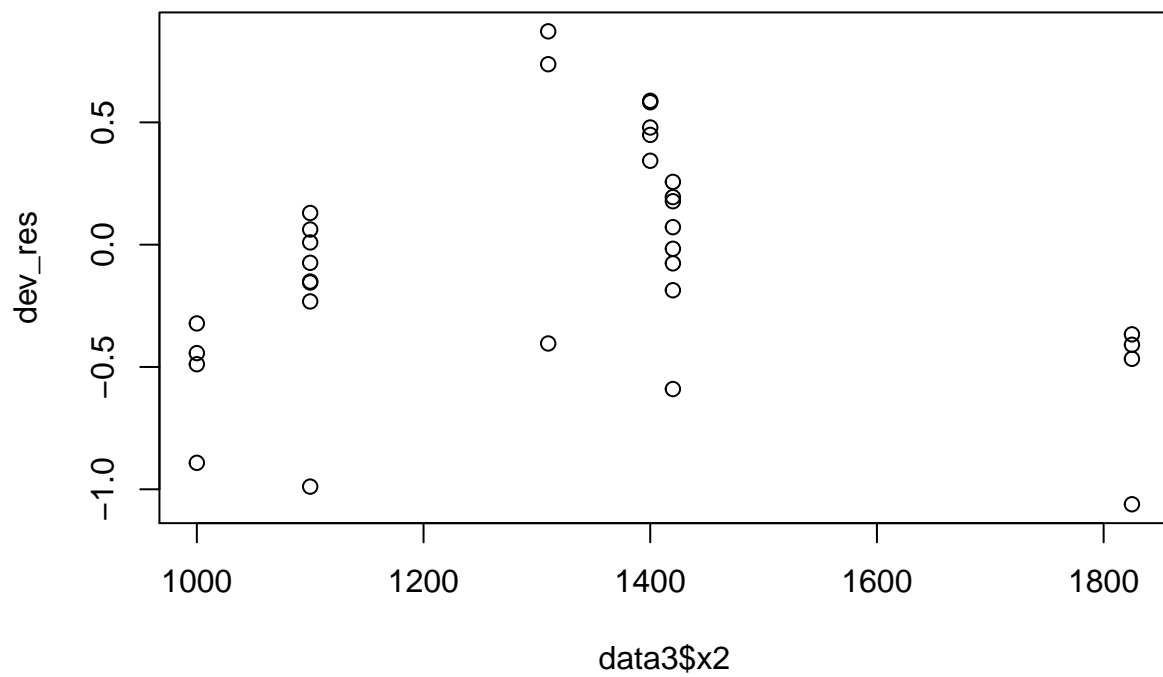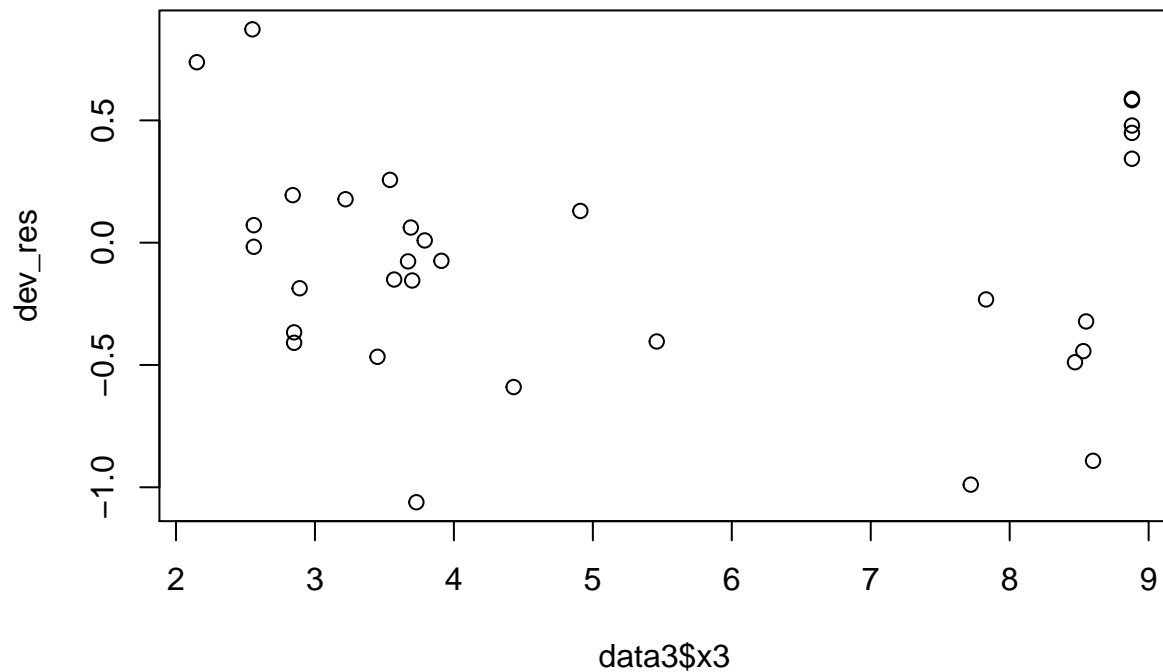
Normal Q–Q Plot


```
plot(yhat,dev_res)
```
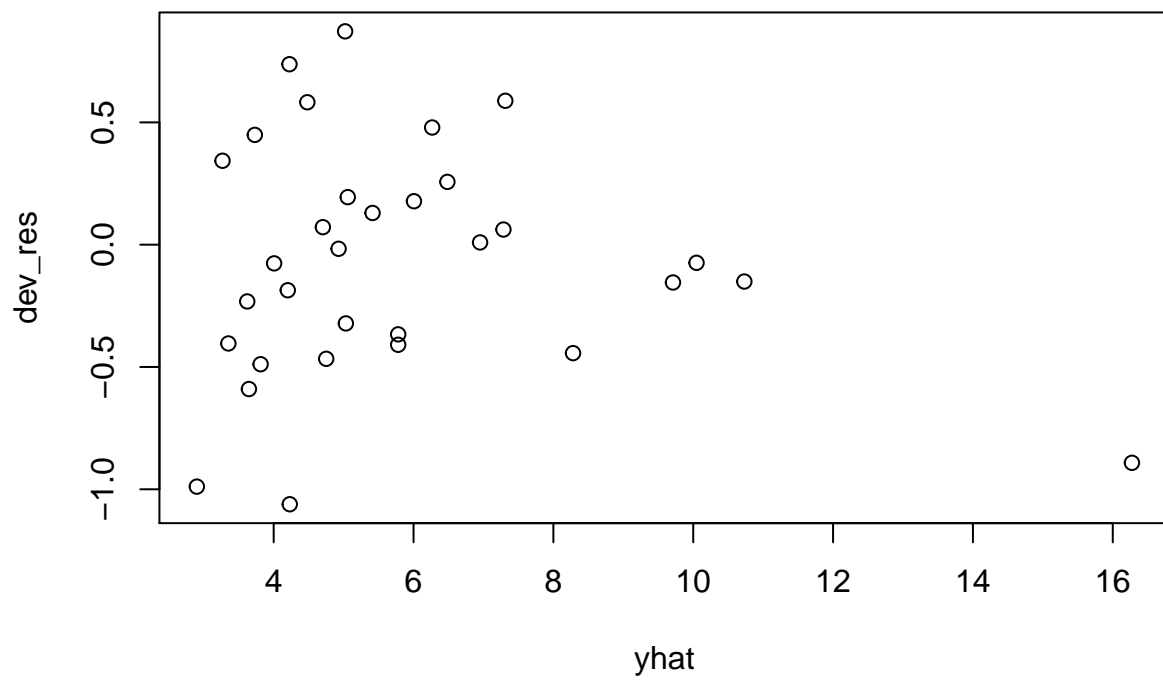
```r
plot(data3$x1,dev_res)
```

```
plot(data3$x2,dev_res)
```

```r
plot(data3$x3,dev_res)
```

Q4

```
data4<-read_excel("A4-datasets.xlsx",sheet = 4)
```
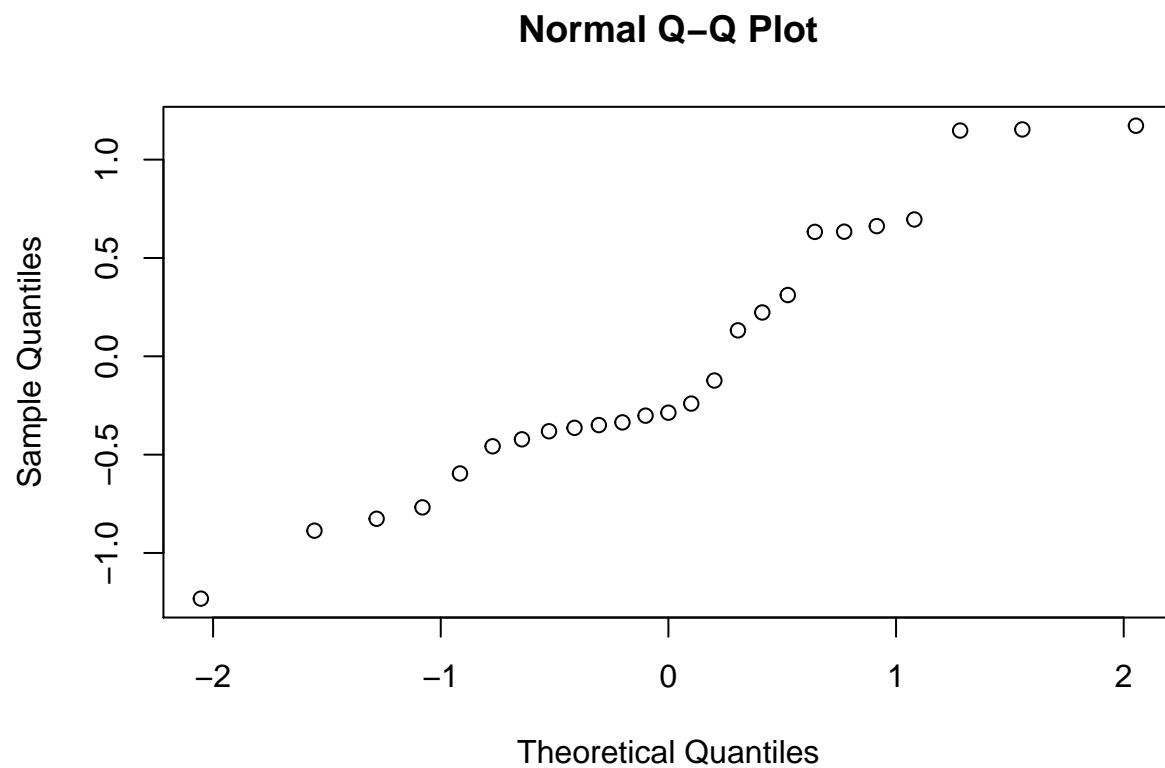
Using canonical link

```
data4.model<-glm(NO~x1+x2+x3,family = poisson,data4)
summary(data4.model)
```
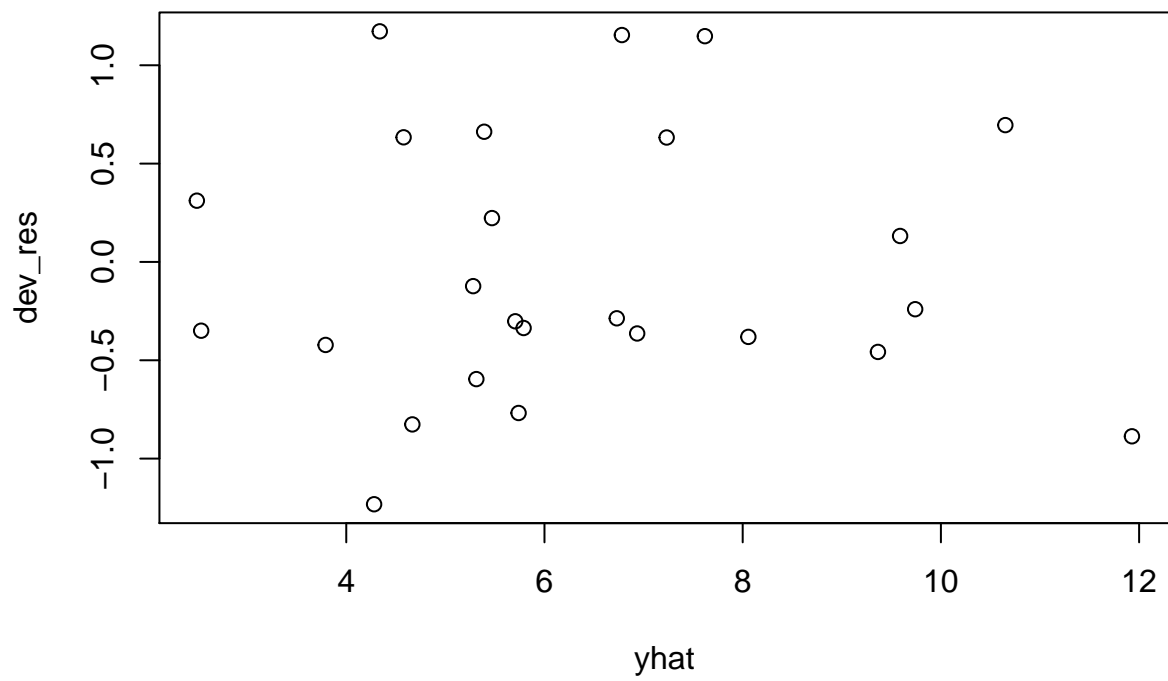
```
##
## Call:
## glm(formula = NO ~ x1 + x2 + x3, family = poisson, data = data4)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2321  -0.4221  -0.2866   0.6327   1.1725
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.8863729  1.7468919   1.080   0.2802
## x1          -0.1047734  0.0521803  -2.008   0.0447 *
## x2           0.0153406  0.0158521   0.968   0.3332
```

```
## x3             -0.0005171  0.0014037  -0.368    0.7126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 32.968  on 24  degrees of freedom
## Residual deviance: 10.989  on 21  degrees of freedom
## AIC: 109.3
##
## Number of Fisher Scoring iterations: 4
```
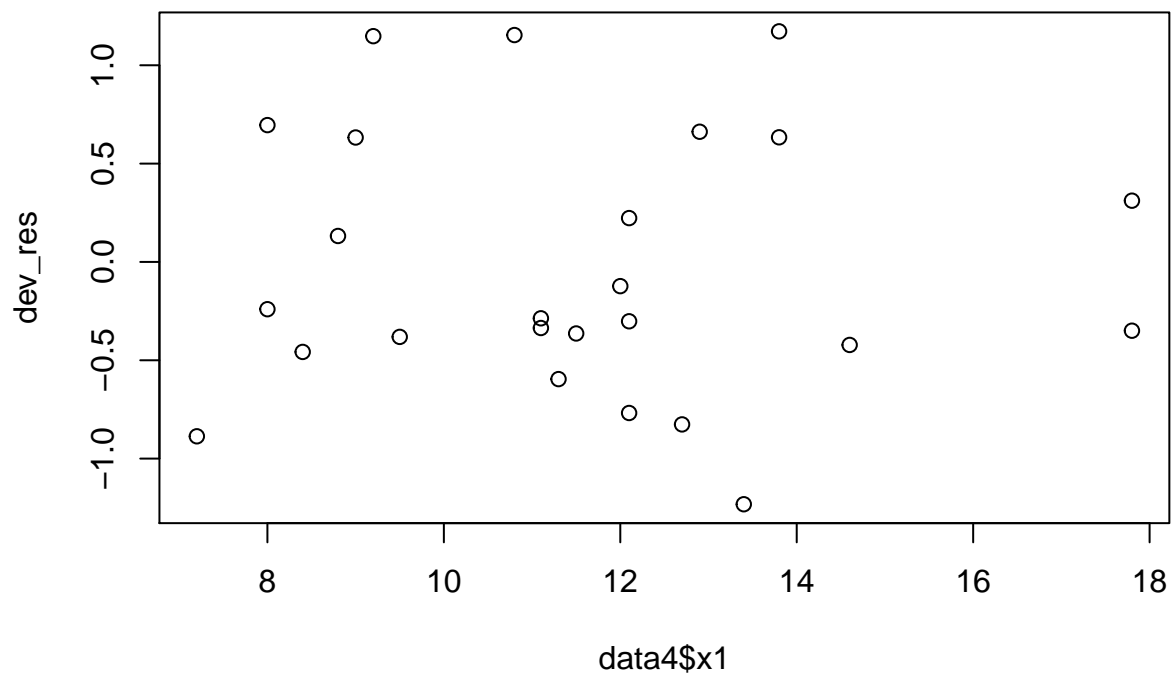
```
yhat <- data4.model$fit
dev_res <- residuals(data4.model, c="deviance")
qqnorm(dev_res)
```
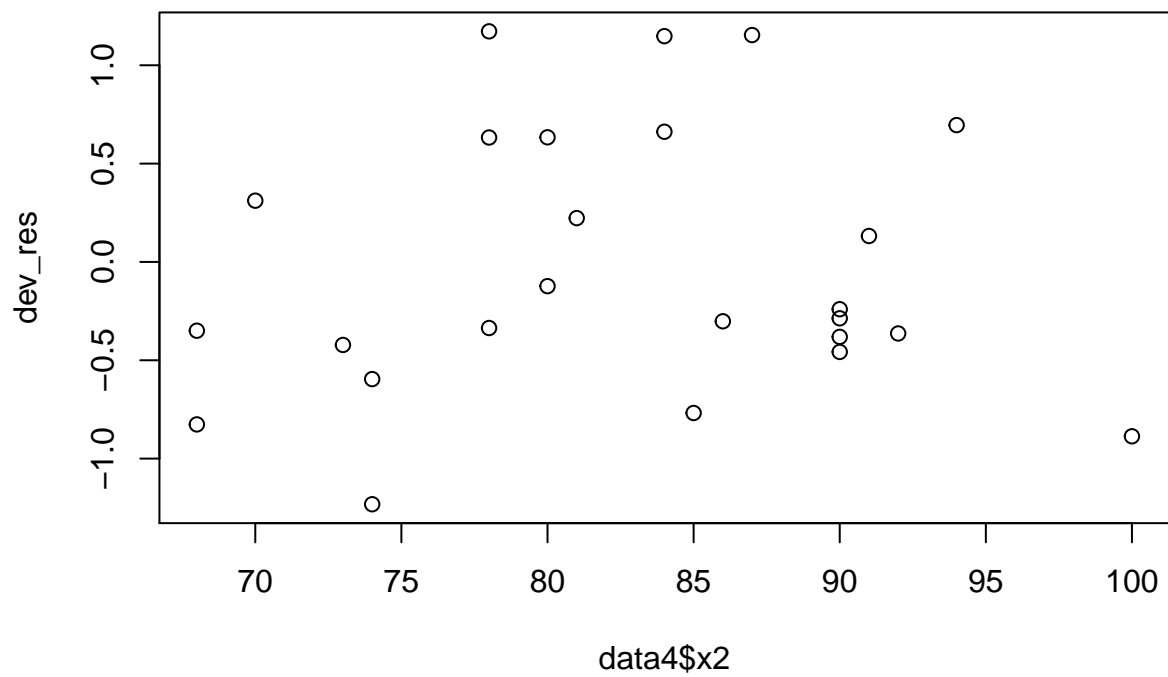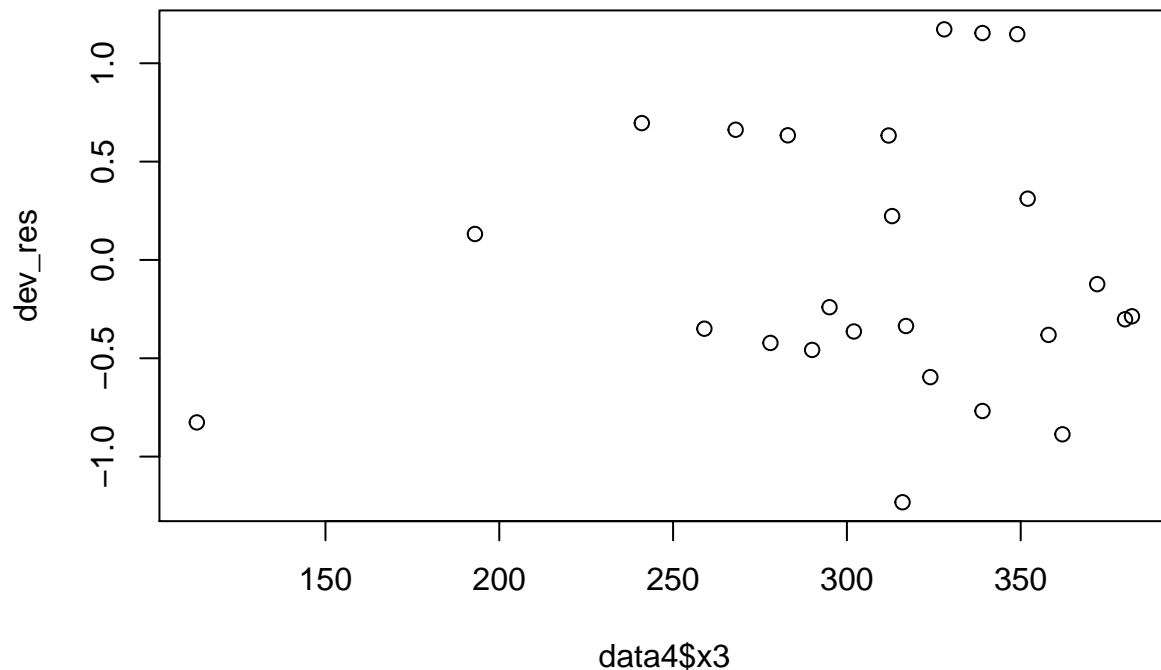

Normal Q–Q Plot

```
plot(yhat,dev_res)
```

```
plot(data4$x1,dev_res)
```

```r
plot(data4$x2,dev_res)
```

```
plot(data4$x3,dev_res)
```
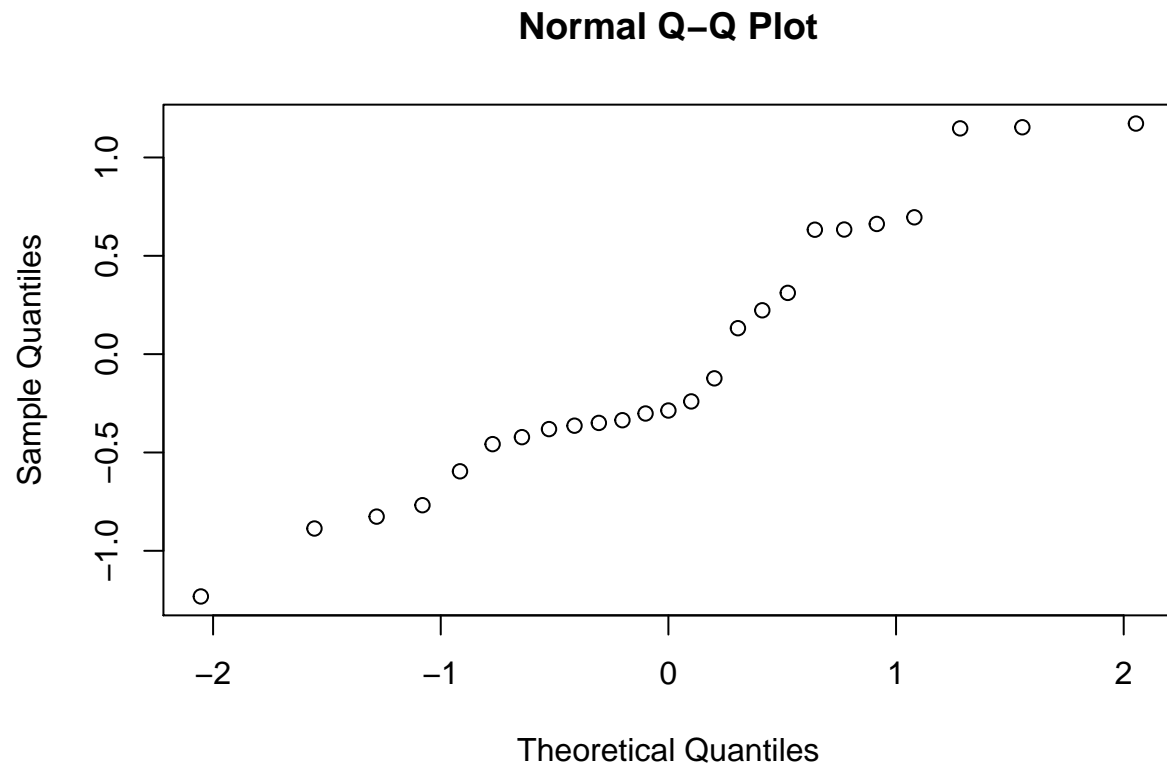
23

Using log link

```
data4.model<-glm(NO~x1+x2+x3,family = poisson(link=log),data4)
summary(data4.model)
```

```
##
## Call:
## glm(formula = NO ~ x1 + x2 + x3, family = poisson(link = log),
##     data = data4)
##
## Deviance Residuals:
##     Min       1Q    Median       3Q       Max
## -1.2321   -0.4221   -0.2866    0.6327    1.1725
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.8863729  1.7468919    1.080    0.2802
## x1          -0.1047734  0.0521803   -2.008    0.0447 *
## x2           0.0153406  0.0158521    0.968    0.3332
## x3          -0.0005171  0.0014037   -0.368    0.7126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 32.968  on 24  degrees of freedom
```
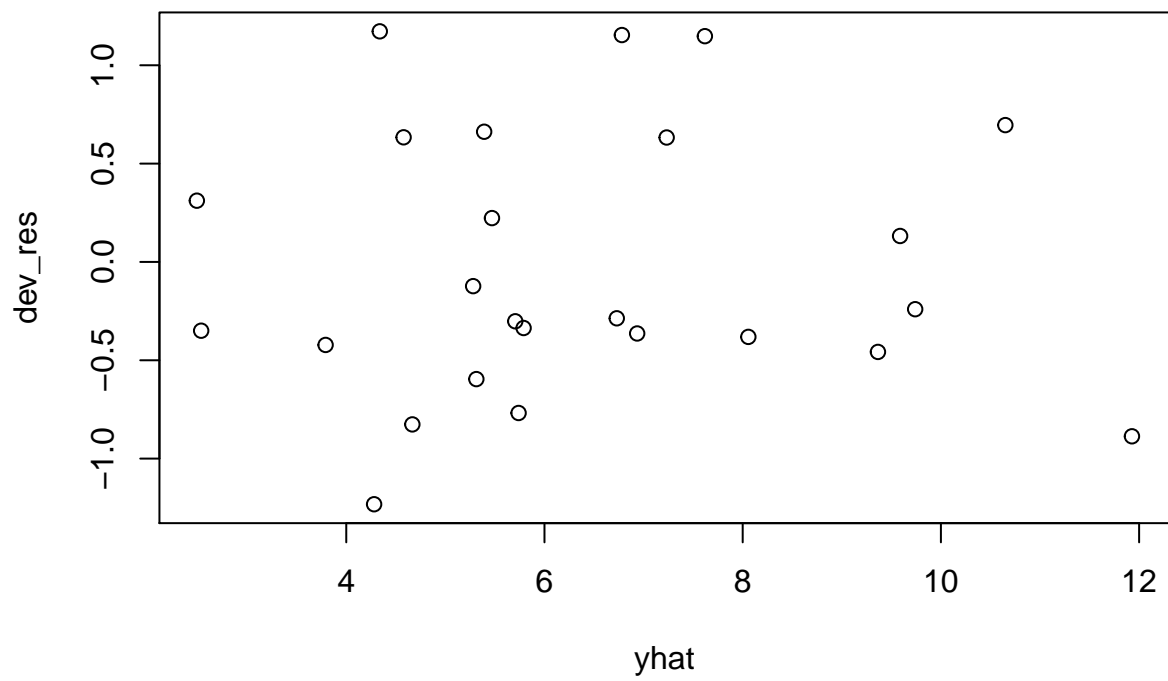
```
## Residual deviance: 10.989  on 21  degrees of freedom
## AIC: 109.3
##
## Number of Fisher Scoring iterations: 4
```
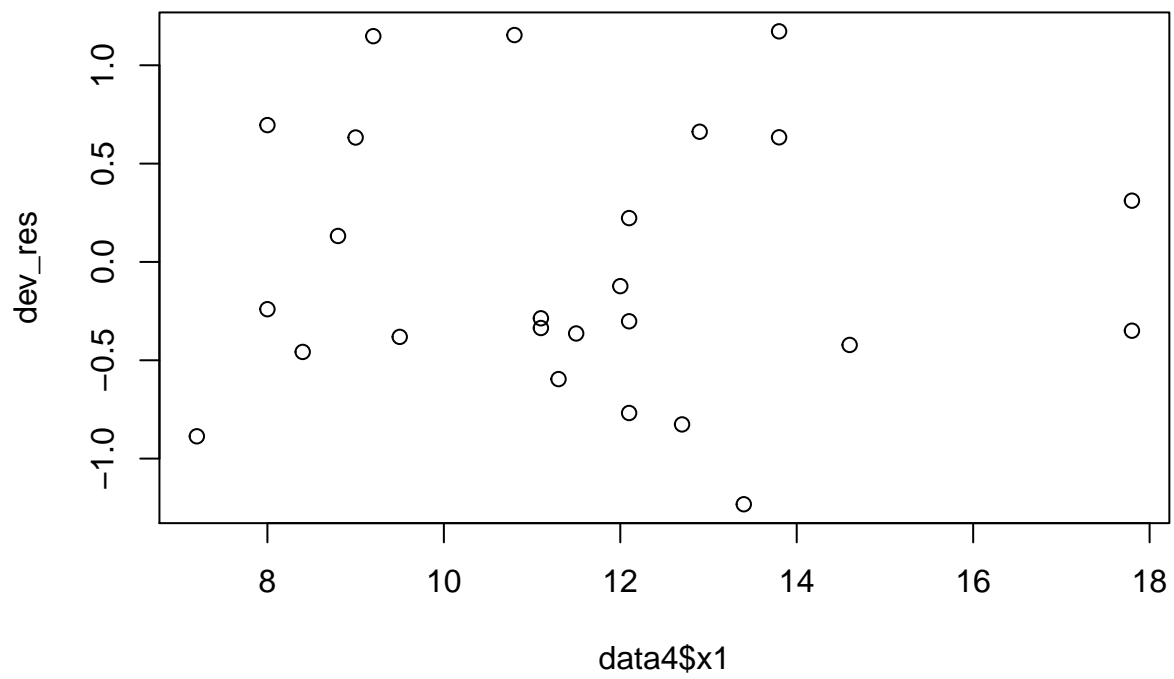
```
yhat <- data4.model$fit
dev_res <- residuals(data4.model, c="deviance")
qqnorm(dev_res)
```
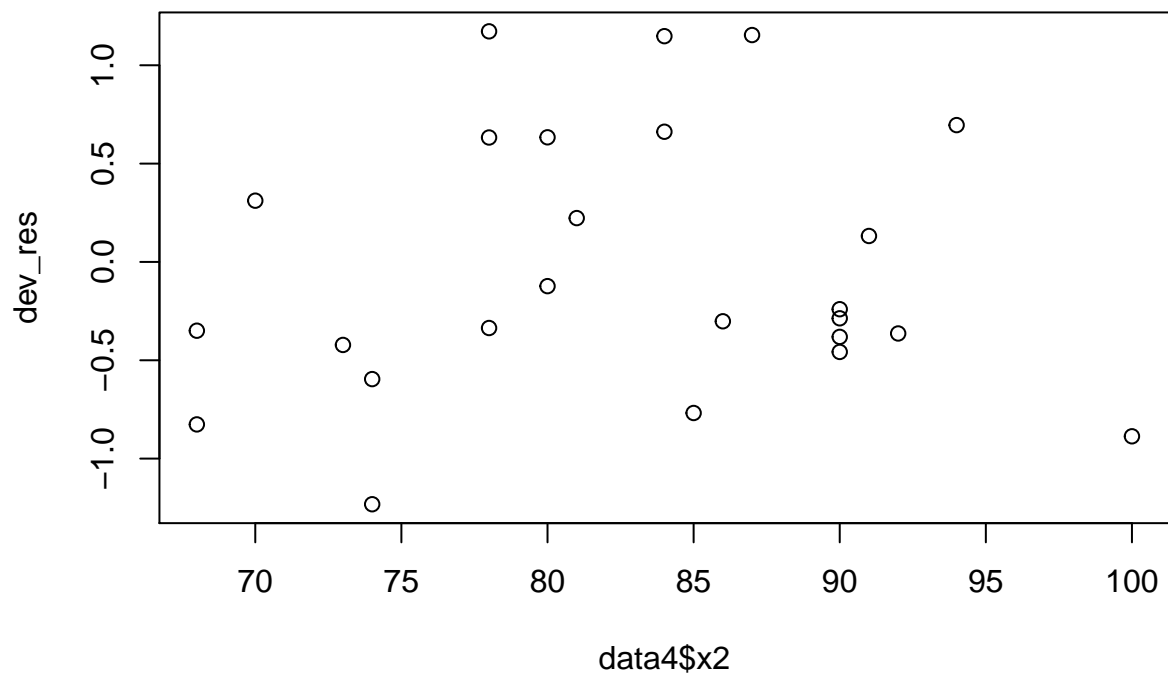
Normal Q–Q Plot
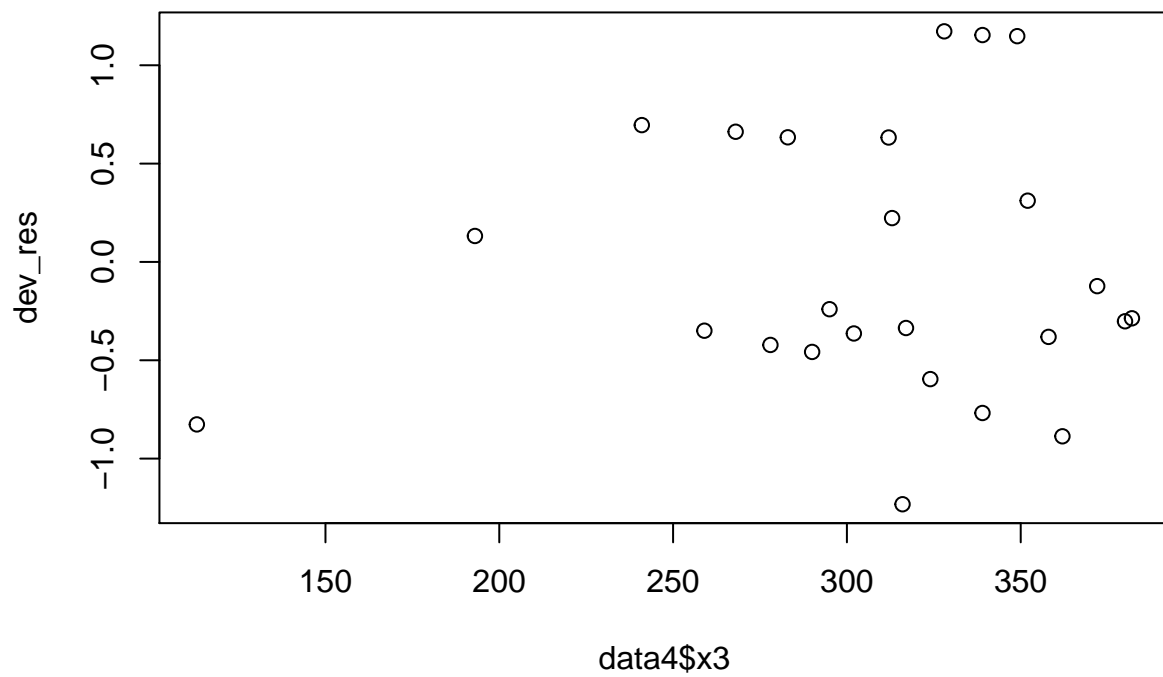


```
plot(yhat,dev_res)
```

```r
plot(data4$x1,dev_res)
```

```
plot(data4$x2,dev_res)
```

```r
plot(data4$x3,dev_res)
```

We see we get the exact same answer using log link or canonical link here because the canonical link for poisson distribution is the log link.