

# Midterm

Ted Henson

2/27/2020

```
## -- Attaching packages ----- tidyverse 1.3.0 --
## v ggplot2 3.2.1      v dplyr   0.8.4
## v tibble  2.1.3      v stringr 1.4.0
## v tidyr   1.0.2      v forcats 0.4.0
## v purrr   0.3.3

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

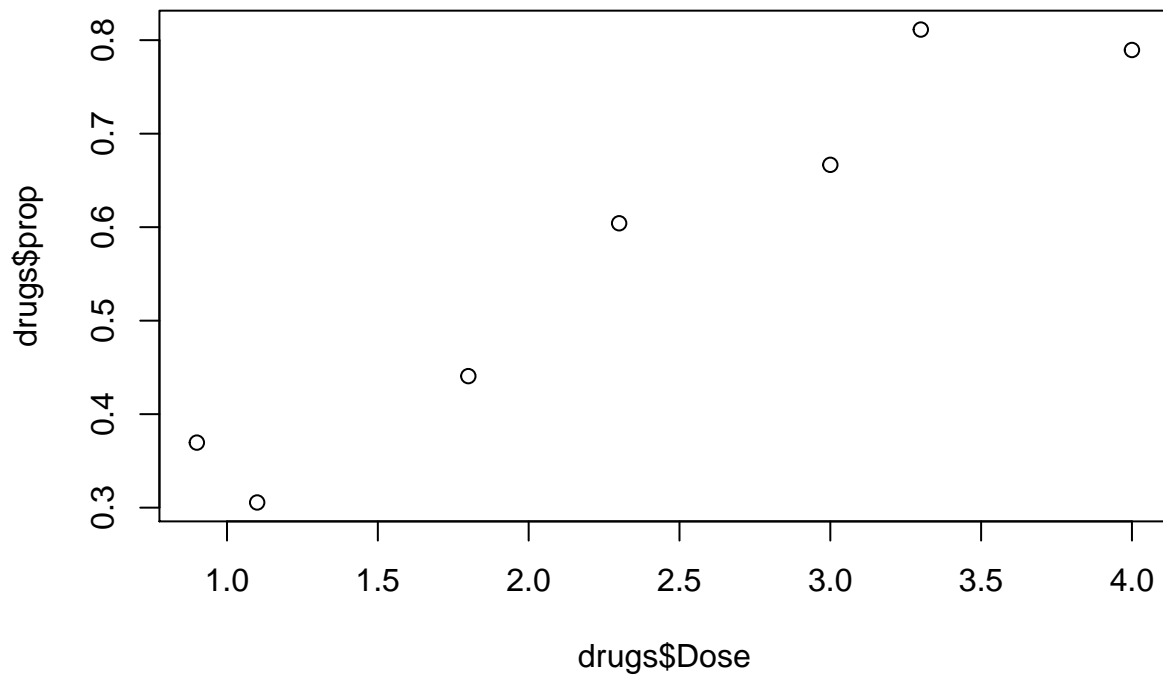
## Parsed with column specification:
## cols(
##   Dose = col_double(),
##   Patients = col_double(),
##   Aftereffects = col_double()
## )

## Parsed with column specification:
## cols(
##   .default = col_double()
## )

## See spec(...) for full column specifications.
```

## Question 1

a)



The plot above shows that the probability of having after-effects is generally higher at higher doses. The highest proportion of after-effects occurred at a dose of 3.3 and the lowest proportion occurred at a dose of 1.1. The shape of this moderately strong positive relationship is roughly linear, but there could be some non linear effects present.

b)

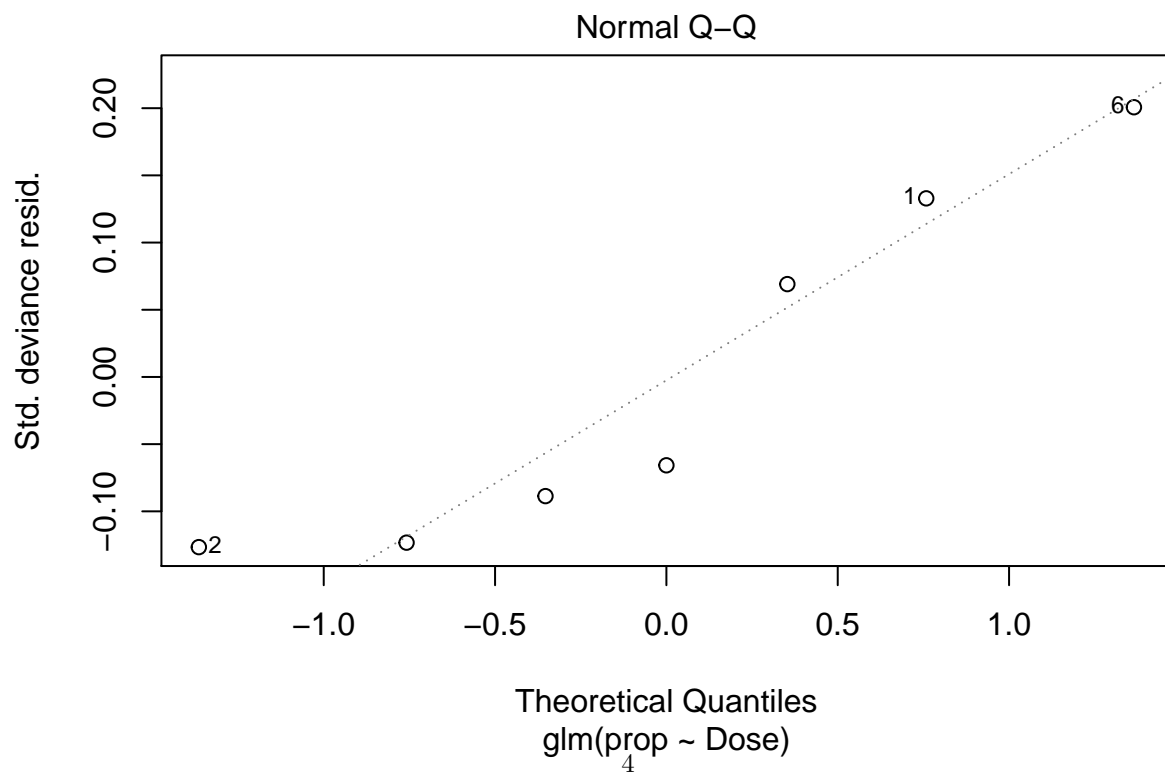
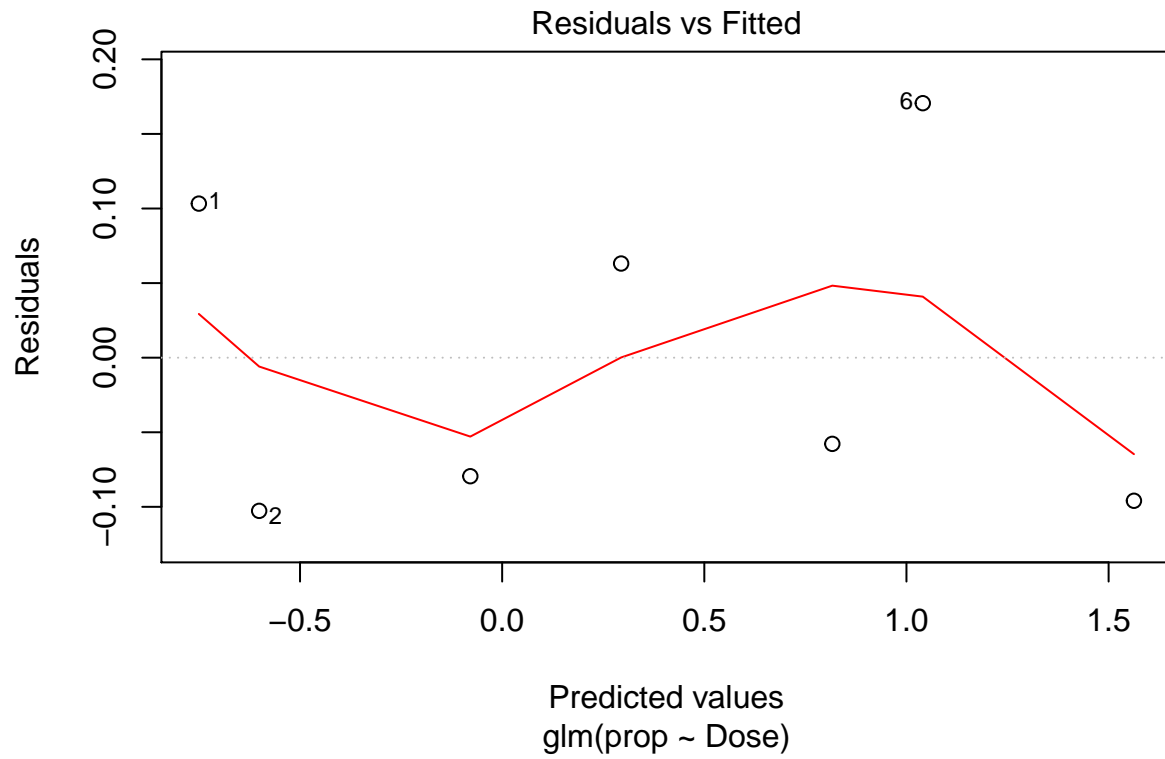
```
## [1] "Summary of Model"
##
## Call:
## glm(formula = prop ~ Dose, family = "binomial", data = drugs)
##
## Deviance Residuals:
##      1       2       3       4       5       6       7
## 0.10329 -0.10269 -0.07947  0.06313 -0.05778  0.17061 -0.09593
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.4218     1.9782  -0.719   0.472
## Dose           0.7461     0.8110   0.920   0.358
##
```

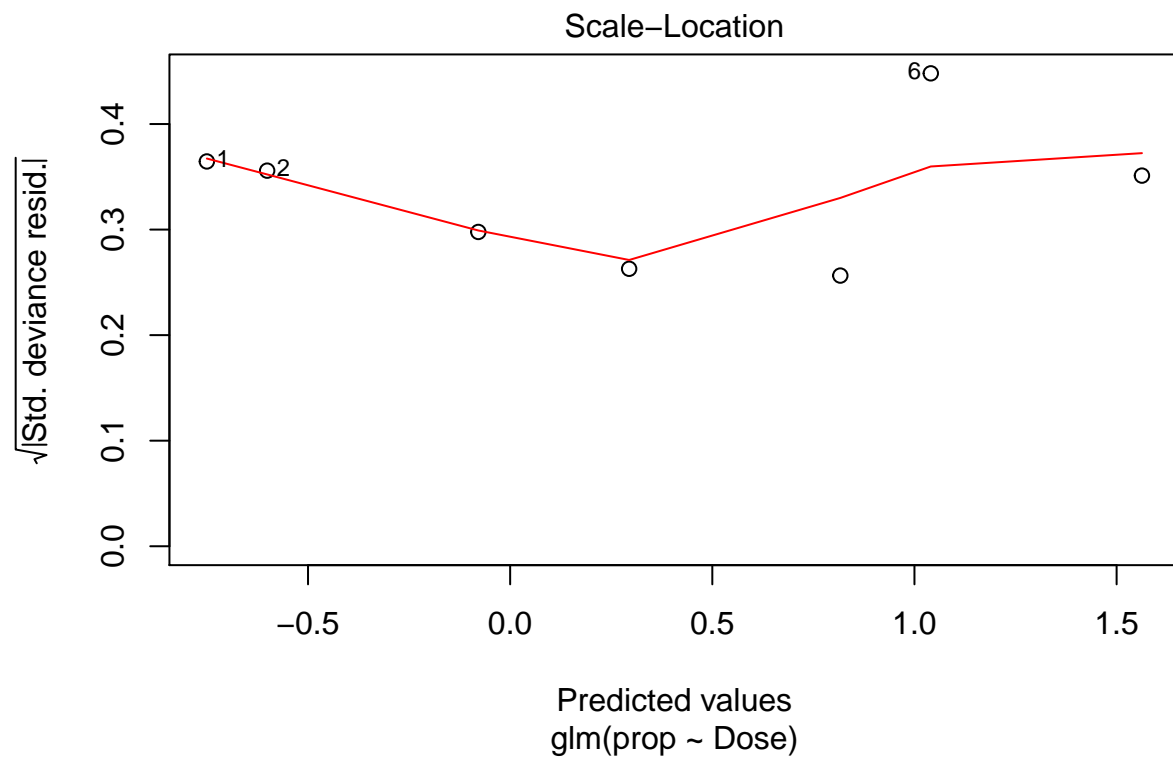
```

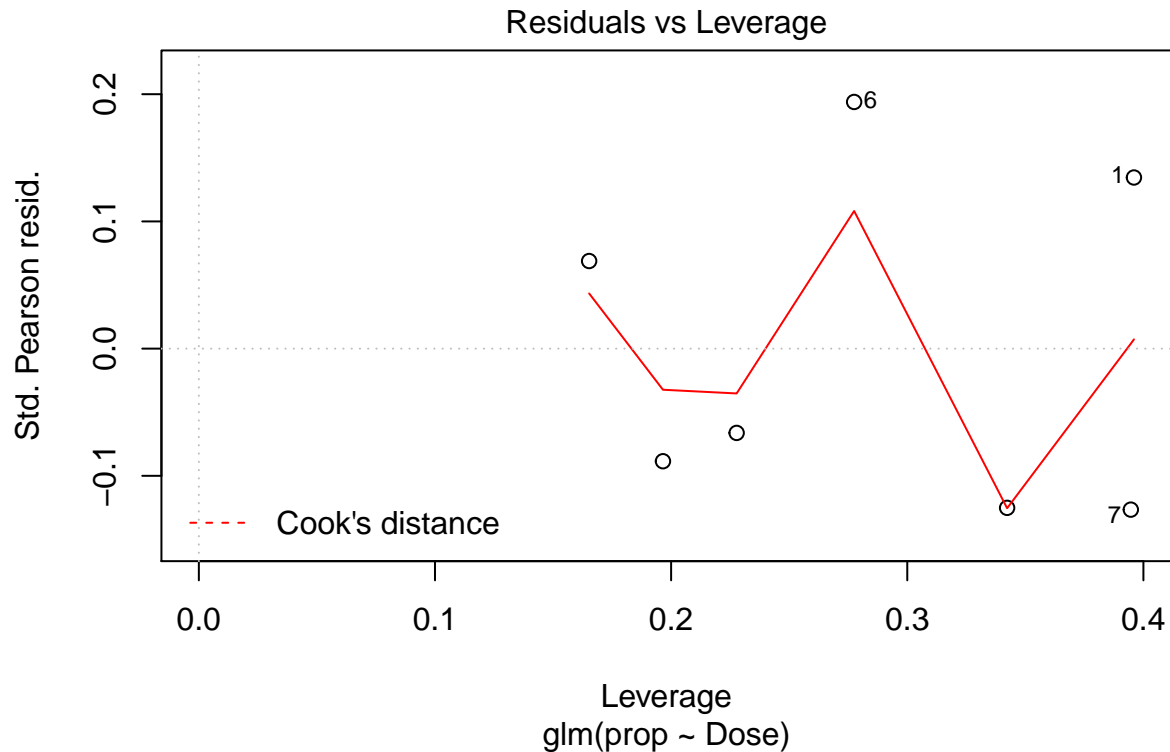
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1.033715  on 6  degrees of freedom
## Residual deviance: 0.073162  on 5  degrees of freedom
## AIC: 9.7881
##
## Number of Fisher Scoring iterations: 4
## [1] "Parameter Estimates:"
## (Intercept)      Dose
## -1.4217741    0.7461213
## [1] "Standard Errors of Parameters:"
## (Intercept)      Dose
##  1.9781923    0.8110293
## [1] "Deviance of the Model:"
## [1] 0.07316179

```

c)







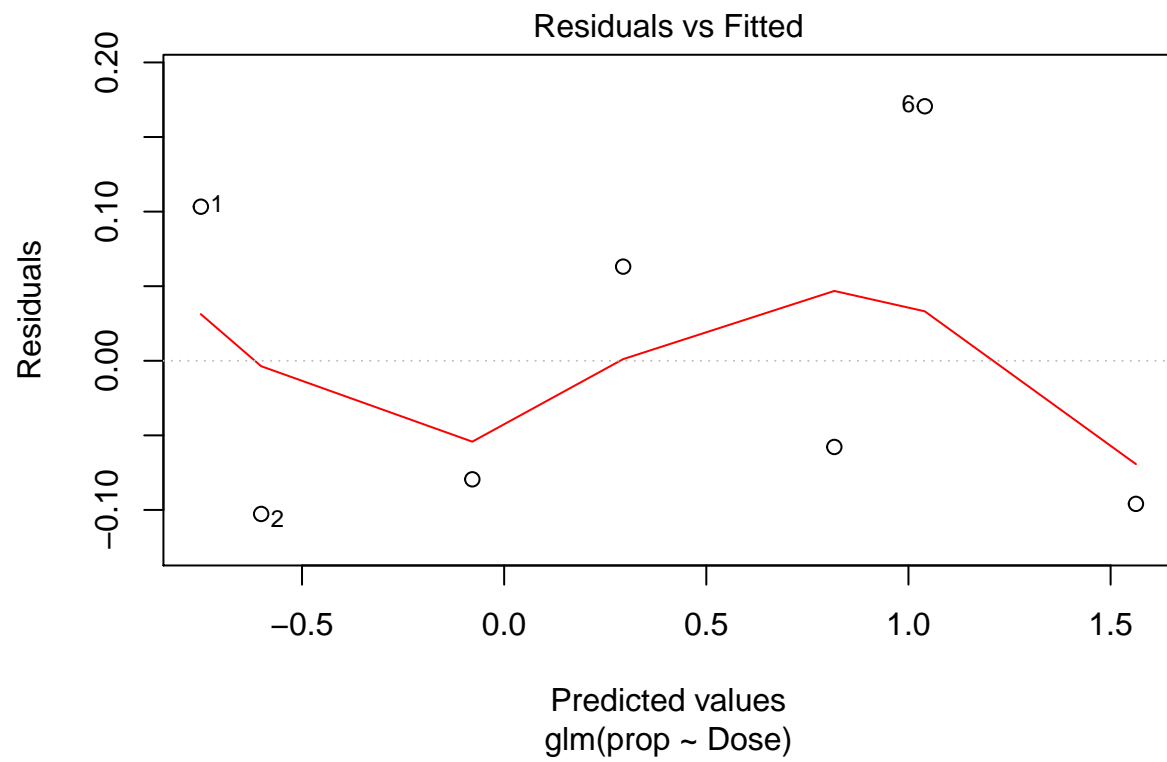
```
## [1] "Chi square test for deviance:"
## [1] 0.999925
## [1] "Chi square test for Pearson Residuals:"
## [1] 0.9999285
## Waiting for profiling to be done...
##           2.5 %   97.5 %
## (Intercept) -6.1714907 2.334529
## Dose        -0.7132819 2.750503
```

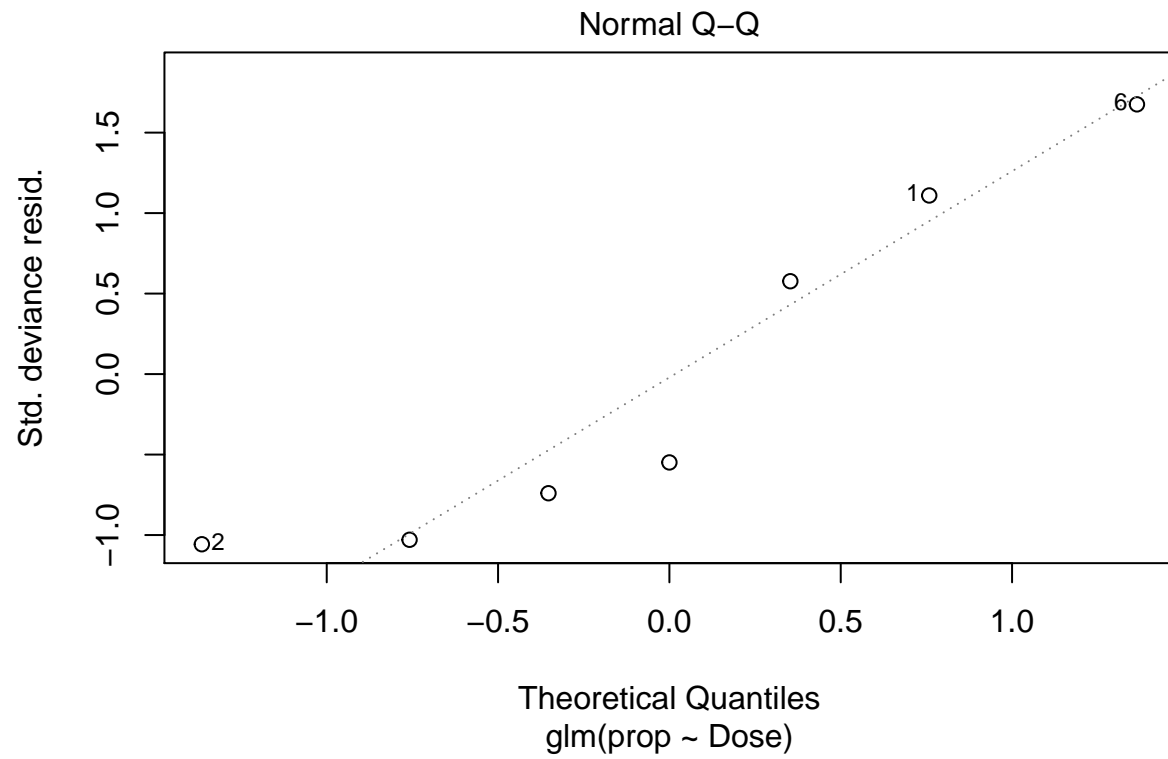
The model appears to fit the data reasonably well despite p values for the coefficients suggesting non significance. The residuals appear to be normally distributed around 0 based on the residuals versus fitted plot. Point 6 appears to be a large residual, but it along with the other points do not appear to be influential based on the leverage versus residuals plot and cooks distance. The chi squared test using the deviance and the pearson residuals both had p values close to 1, indicating no lack of fit.

d)

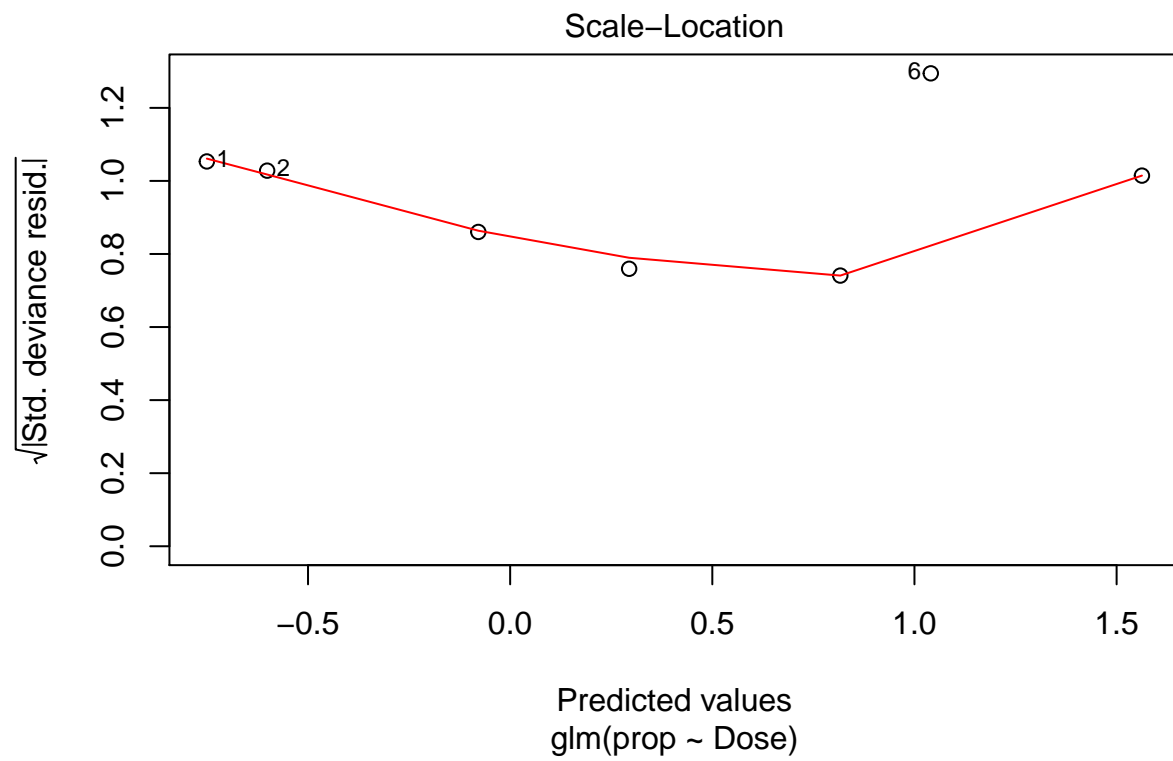
```
## [1] "Chi square test for difference in deviance between base and quadratic model:"
## [1] 0.9761933
## [1] "Chi square test for difference in Pearson residuals between base and quadratic model:"
## [1] 0.9758585
## [1] "Chi square test for difference in deviance between base and quasi model:"
```

```
## [1] 1
## [1] "Chi square test for difference in Pearson residuals between base and quasi model:"
## [1] 1
## [1] "Model Plots of Quasi Model"
```



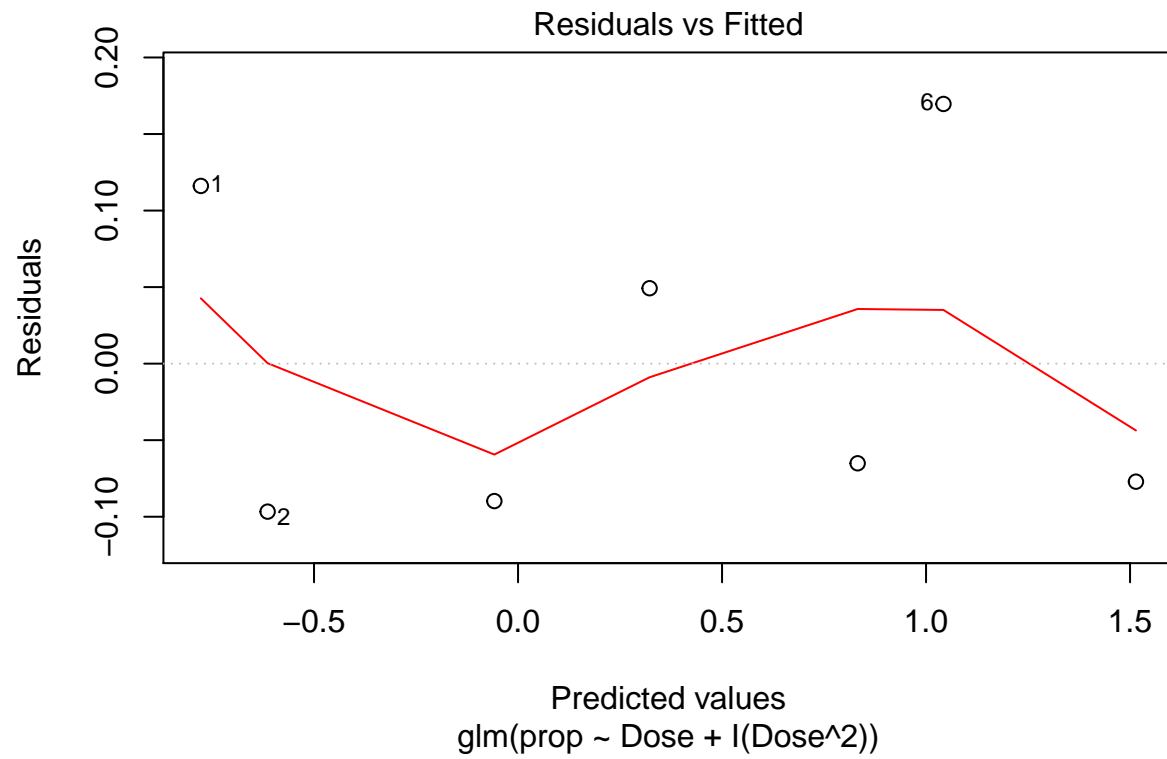


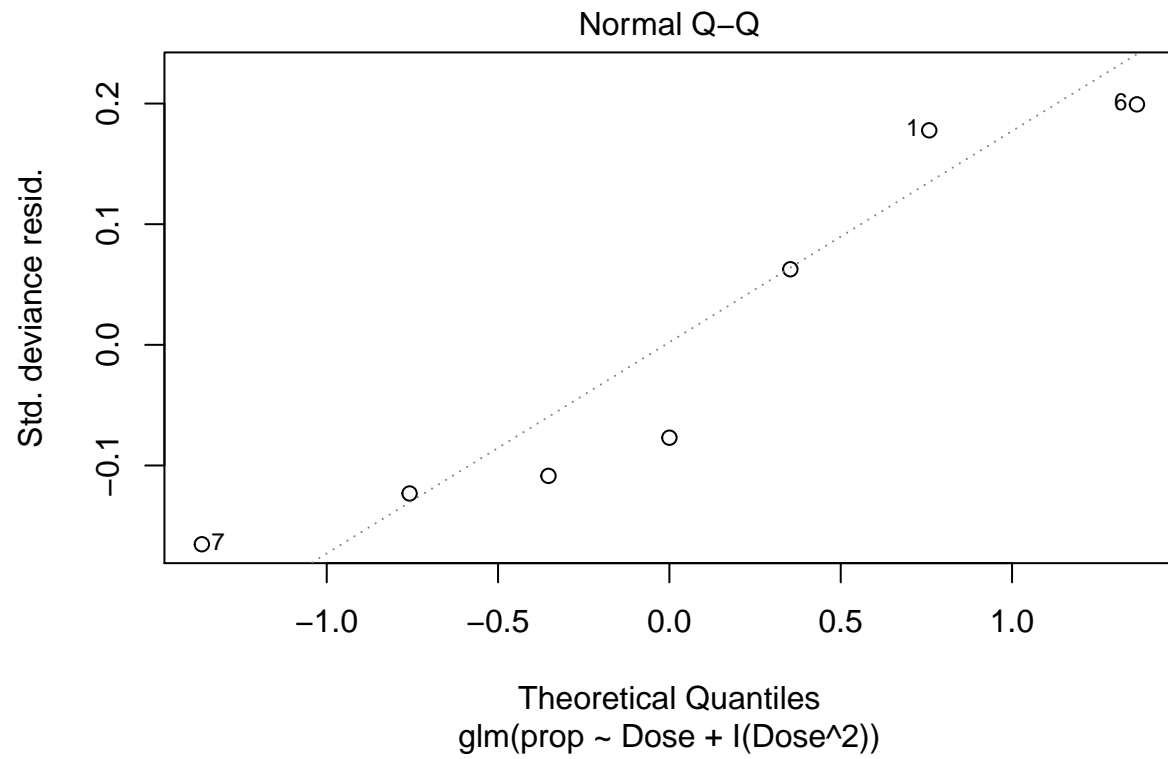


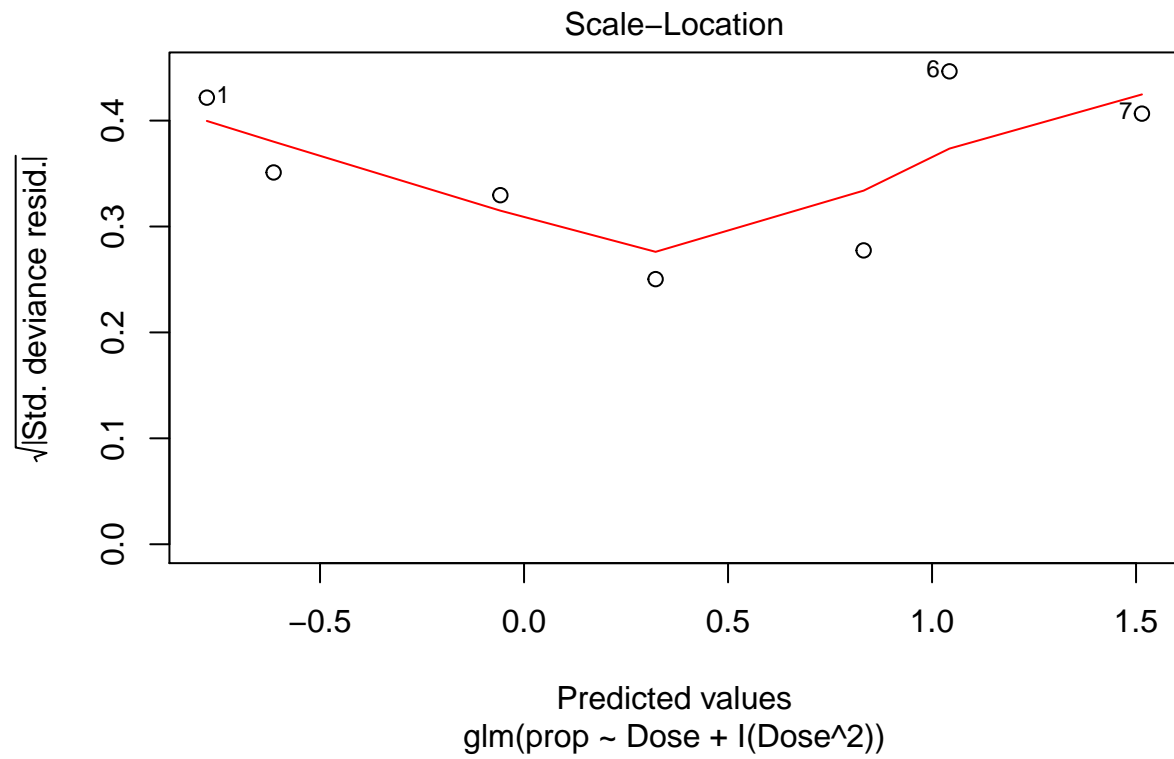


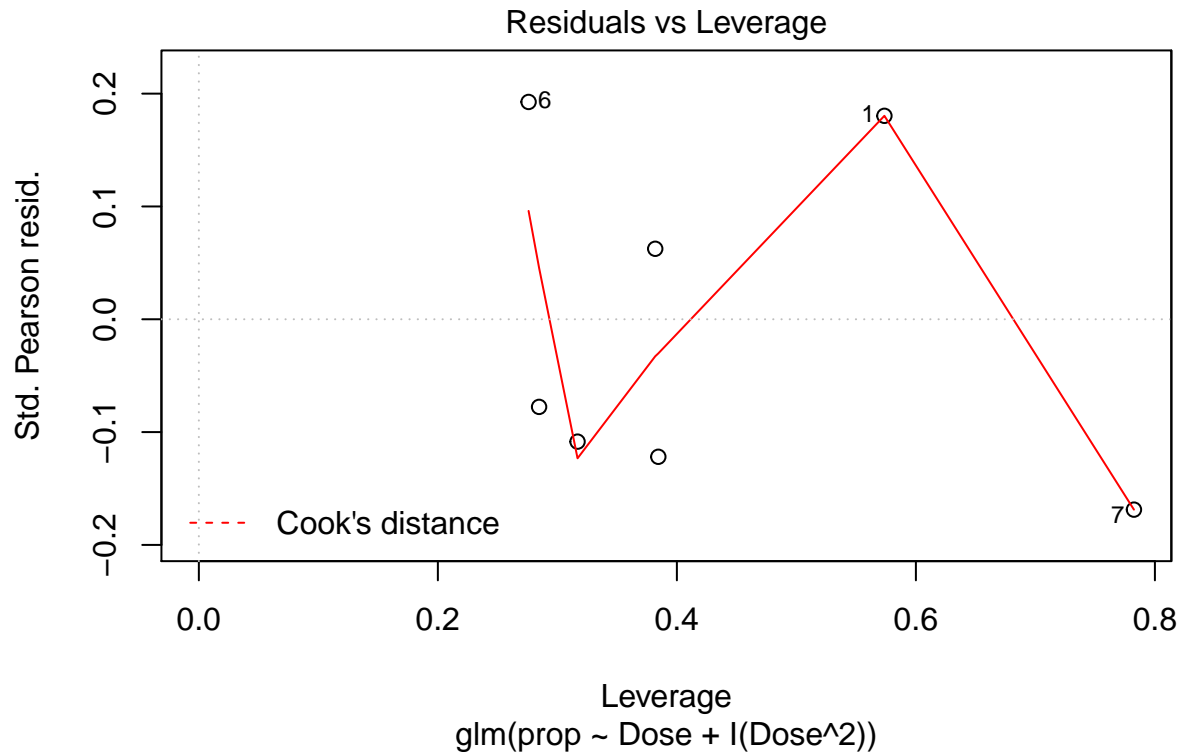


```
## [1] "Model Plots of Quadratic Model"
```









```
## [1] "Summary of Quasi Model:"
##
## Call:
## glm(formula = prop ~ Dose, family = "quasibinomial", data = drugs)
##
## Deviance Residuals:
##      1      2      3      4      5      6      7
## 0.10329 -0.10269 -0.07947  0.06313 -0.05778  0.17061 -0.09593
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.42177    0.23696   -6.00 0.001846 **
## Dose         0.74612    0.09715    7.68 0.000597 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 0.01434819)
##
## Null deviance: 1.033715  on 6  degrees of freedom
## Residual deviance: 0.073162  on 5  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
## [1] "Summary of Quadratic Model:"
##
```

```
## Call:
## glm(formula = prop ~ Dose + I(Dose^2), family = "binomial", data = drugs)
##
## Deviance Residuals:
##      1      2      3      4      5      6      7
## 0.11610 -0.09668 -0.08978  0.04926 -0.06510  0.16964 -0.07710
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.54089    4.45903  -0.346   0.730
## Dose         0.87281    4.32101   0.202   0.840
## I(Dose^2)    -0.02724    0.91156  -0.030   0.976
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1.033715  on 6  degrees of freedom
## Residual deviance: 0.072271  on 4  degrees of freedom
## AIC: 11.764
##
## Number of Fisher Scoring iterations: 4
```

The standard model is better than the quadratic model. The quadratic term in the quadratic model had a p value close to 1 and its deviance and pearson residuals were not reduced enough when compared to the standard model using a chisq test for significance to justify including it. Incorporating an over dispersion parameter does not reduce the deviance or pearson residuals enough to say reduction in deviance is statistically significant according to the chi square test. Plots of the quasibinomial model show there may be influential points according to the cooks distance. The original model had no points close to high influence; however, the standard errors of the coefficients of the quasibinomial model were much lower than both other models, and both the intercept and dose parameter had highly significant p values. This will result in narrower prediction intervals so the quasibinomial model will be used going forward. Treating the dose as a factor would not work in this case because there is only one observation per factored level. Therefore the quasi binomial model with just dose as a predictor is the best.

e)

```
## [1] "Predicted probability for after effects with a dose of 2.6:"
##      1
## 0.6267131
## [1] "Confidence interval for probability of after effects with a dose of 2.6:"
##      1      1
## 0.5786342 0.6747919
```

f)

```
## [1] 1.905553
##      1
## 0.5
##      1      1
## 0.4496334 0.5503666
```

A dose of 1.9055535 would yield a 50% probability of after effects based on the chosen model. A 95% confidence interval for this probability would be 0.4496334, 0.5503666.

## Question 2

a)

```
backache$y = ifelse(backache$Severity <= 1, 0, 1)
backache$Weightend = backache$Weightend - backache$Weightstart
colnames(backache)[which(colnames(backache) == 'Weightend')] = 'Weightgain'
head(backache$y)
```

```
## [1] 0 1 1 0 0 0
```

```
head(backache$Weightgain)
```

```
## [1] 20.5 9.5 9.5 5.9 4.5 15.0
```

b)

```
##      Severity      Monthstarted      Age      Height
##  Min.   :0.000   Min.   :0.000   Min.   :15.00   Min.   :1.470
## 1st Qu.:1.000   1st Qu.:0.000   1st Qu.:22.00   1st Qu.:1.550
## Median :1.000   Median :4.000   Median :25.00   Median :1.600
## Mean   :1.561   Mean   :3.789   Mean   :26.02   Mean   :1.607
## 3rd Qu.:2.000   3rd Qu.:7.000   3rd Qu.:29.00   3rd Qu.:1.650
## Max.   :3.000   Max.   :9.000   Max.   :42.00   Max.   :1.750
## Weightstart Weightgain Weightbaby PrevKids
##  Min.   :38.20   Min.   : 0.00   Min.   :1.080   Min.   :0.0000
## 1st Qu.:52.30   1st Qu.: 7.30   1st Qu.:2.840   1st Qu.:0.0000
## Median :58.40   Median :10.95   Median :3.240   Median :0.0000
## Mean   :59.64   Mean   :11.04   Mean   :3.217   Mean   :0.7611
## 3rd Qu.:65.90   3rd Qu.:14.03   3rd Qu.:3.587   3rd Qu.:1.0000
## Max.   :95.50   Max.   :39.50   Max.   :6.280   Max.   :7.0000
## PrevBackache Tablets      HotBottle      HotBath
##  Min.   :0.000   Min.   :0.0000   Min.   :0.00   Min.   :0.00000
## 1st Qu.:1.000   1st Qu.:0.0000   1st Qu.:0.00   1st Qu.:0.00000
## Median :1.000   Median :0.0000   Median :0.00   Median :0.00000
## Mean   :1.517   Mean   :0.1278   Mean   :0.05   Mean   :0.08333
## 3rd Qu.:2.000   3rd Qu.:0.0000   3rd Qu.:0.00   3rd Qu.:0.00000
## Max.   :4.000   Max.   :1.0000   Max.   :1.00   Max.   :1.00000
## Cushion      Standing      Sitting      Lyingdown
##  Min.   :0.0000   Min.   :0.00   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:0.0000   1st Qu.:0.00   1st Qu.:0.0000   1st Qu.:0.0000
## Median :0.0000   Median :0.00   Median :0.0000   Median :0.0000
## Mean   :0.3167   Mean   :0.05   Mean   :0.1111   Mean   :0.2111
## 3rd Qu.:1.0000   3rd Qu.:0.00   3rd Qu.:0.0000   3rd Qu.:0.0000
## Max.   :1.0000   Max.   :1.00   Max.   :1.0000   Max.   :1.0000
## Walking      Fatigue      Bending      Lifting
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
## Median :0.00000   Median :0.0000   Median :0.0000   Median :0.0000
## Mean   :0.02778   Mean   :0.2167   Mean   :0.1667   Mean   :0.1333
```



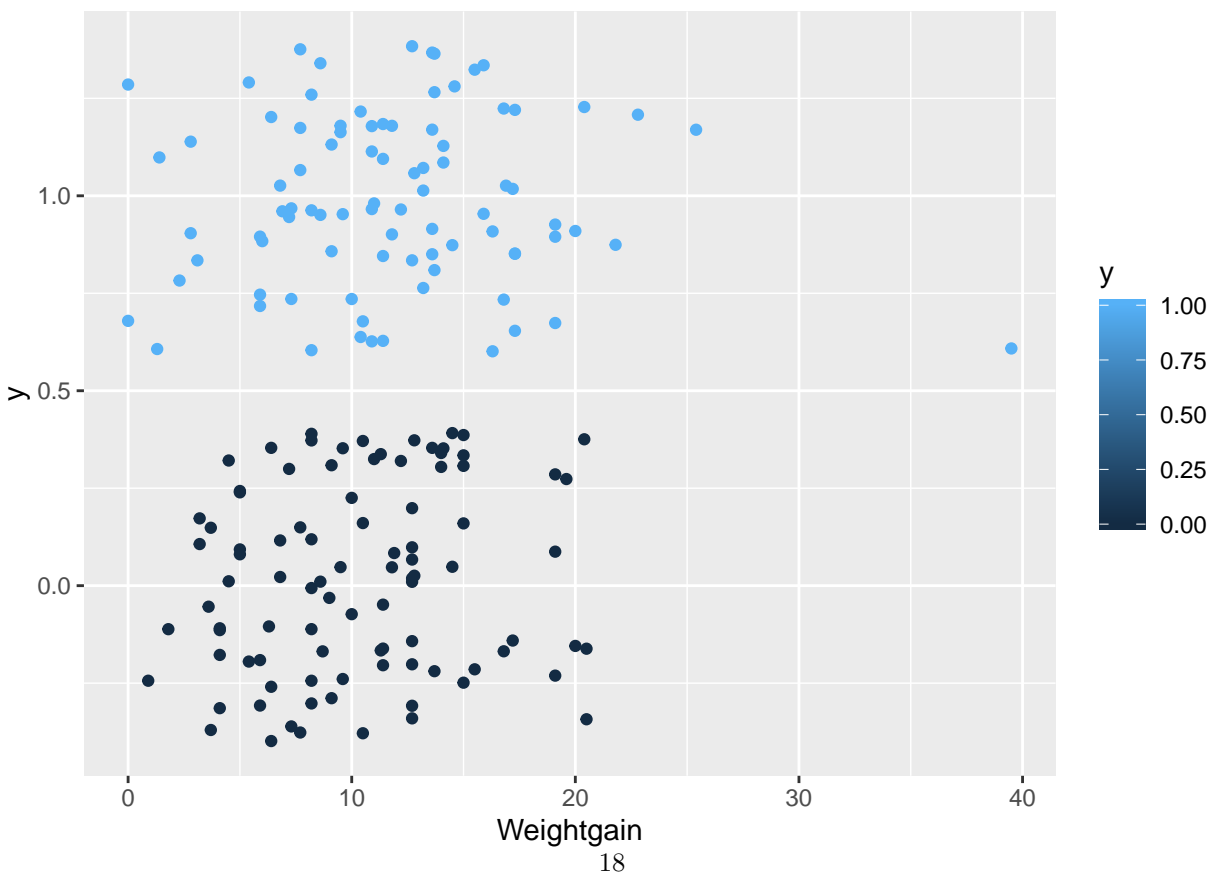
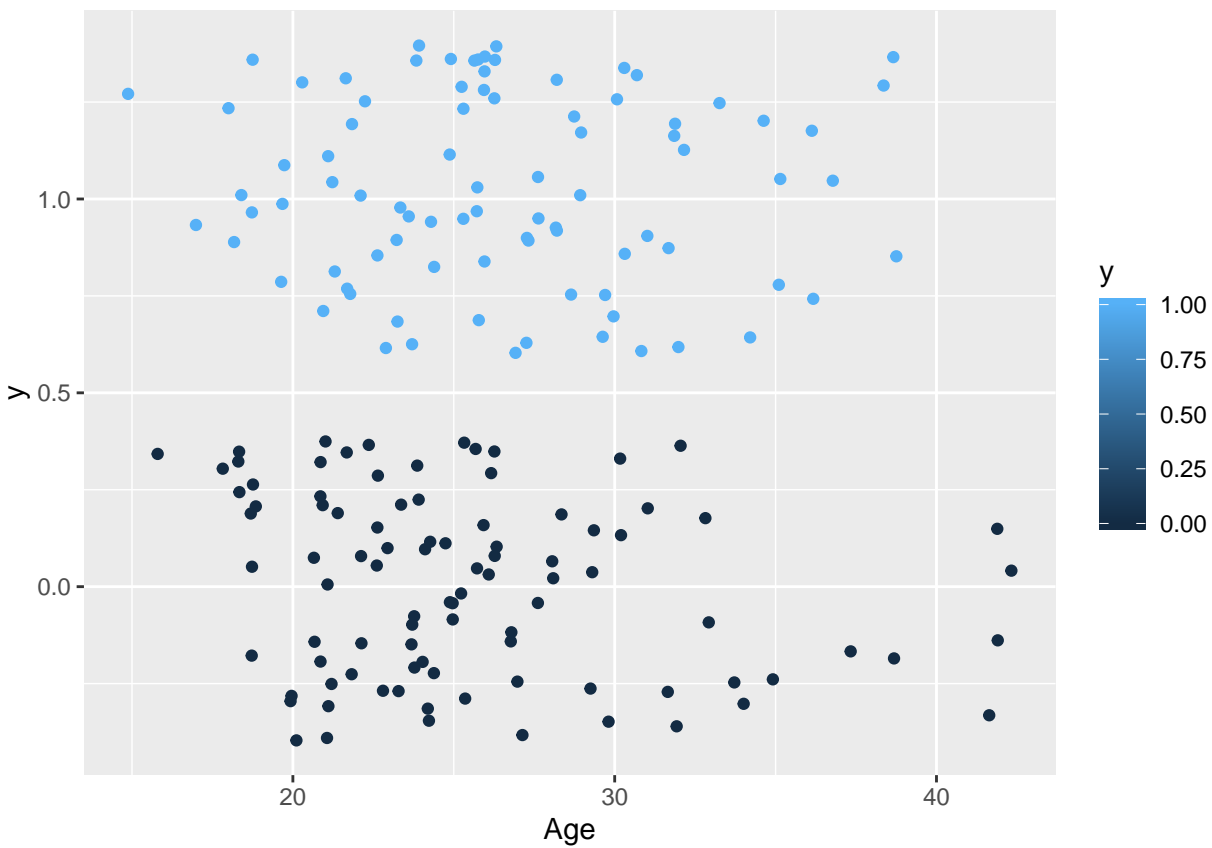
```

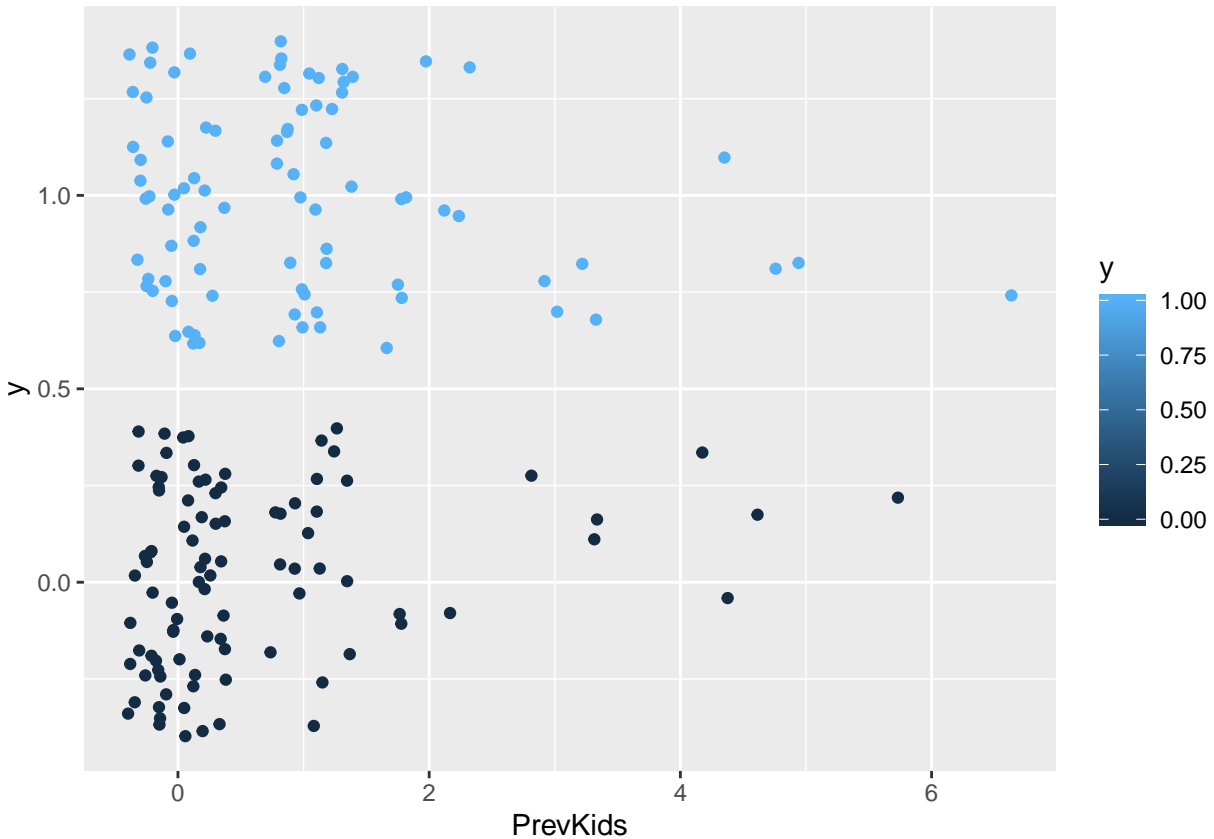
## 3rd Qu.:0.00000 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:0.0000
## Max. :1.00000 Max. :1.0000 Max. :1.0000 Max. :1.0000
## MakingBeds WashingUp Ironing Bowels
## Min. :0.0000 Min. :0.00000 Min. :0.0000 Min. :0.000000
## 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.000000
## Median :0.0000 Median :0.00000 Median :0.0000 Median :0.000000
## Mean :0.1944 Mean :0.07778 Mean :0.2056 Mean :0.005556
## 3rd Qu.:0.0000 3rd Qu.:0.00000 3rd Qu.:0.0000 3rd Qu.:0.000000
## Max. :1.0000 Max. :1.00000 Max. :1.0000 Max. :1.000000
## Intercourse Coughing Sneezing TurningInBed
## Min. :0.00000 Min. :0.00000 Min. :0.00000 Min. :0.00000
## 1st Qu.:0.00000 1st Qu.:0.00000 1st Qu.:0.00000 1st Qu.:0.00000
## Median :0.00000 Median :0.00000 Median :0.00000 Median :0.00000
## Mean :0.02222 Mean :0.02222 Mean :0.01667 Mean :0.05556
## 3rd Qu.:0.00000 3rd Qu.:0.00000 3rd Qu.:0.00000 3rd Qu.:0.00000
## Max. :1.00000 Max. :1.00000 Max. :1.00000 Max. :1.00000
## Standing2 Sitting2 Lyingdown2 Walking2
## Min. :0.0000 Min. :0.0000 Min. :0.00000 Min. :0.0000
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:0.0000
## Median :0.0000 Median :0.0000 Median :0.00000 Median :0.0000
## Mean :0.2611 Mean :0.1667 Mean :0.07778 Mean :0.1389
## 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:0.00000 3rd Qu.:0.0000
## Max. :1.0000 Max. :1.0000 Max. :1.00000 Max. :1.0000
## y
## Min. :0.0000
## 1st Qu.:0.0000
## Median :0.0000
## Mean :0.4833
## 3rd Qu.:1.0000
## Max. :1.0000

```

The patient variable should be eliminated from analysis since it is the unique identifier. There are also a few patients ages 15-17 so that could potentially be an error in the data set. There are also no observations with months November or December so that could also be an error. Summary tables and plots of the continuous variables showed no obvious errors. There are a few binary variables with only a few values of 1 so that could be erroneous or a problem in the analysis. Significance from these variables should be made cautiously.

c)





The age does not show a clear positive or negative relationship to the response, but most older women did not have severe back aches. A high weight gain may increase the risk of the response based on the plot, but statistical tests would need to confirm this. The more previous kids the mother had also may increase the risk of back aches.

d)

```
##
## Call:
## glm(formula = y ~ Age + Height + Weightstart + Weightgain + Weightbaby +
##       PrevKids + PrevBackache, family = "binomial", data = backache)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1683  -0.9978  -0.6009   1.0444   2.0874
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   6.0760572  4.1534418   1.463   0.1435
## Age          -0.0129749  0.0362921  -0.358   0.7207
## Height        -6.4468910  2.8618026  -2.253   0.0243 *
## Weightstart    0.0409242  0.0183875   2.226   0.0260 *
## Weightgain     0.0854778  0.0350612   2.438   0.0148 *
## Weightbaby    -0.0185621  0.2978181  -0.062   0.9503
## PrevKids       0.0002787  0.1795266   0.002   0.9988
## PrevBackache   0.8203859  0.2048231   4.005 6.19e-05 ***
```

```

## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 249.33  on 179  degrees of freedom
## Residual deviance: 216.07  on 172  degrees of freedom
## AIC: 232.07
##
## Number of Fisher Scoring iterations: 4
##
## Call:  glm(formula = y ~ Height + Weightstart + Weightgain + PrevBackache,
##      family = "binomial", data = backache)
##
## Coefficients:
##      (Intercept)      Height  Weightstart      Weightgain  PrevBackache
##           5.93391      -6.56023         0.04053         0.08482         0.79393
##
## Degrees of Freedom: 179 Total (i.e. Null);  175 Residual
## Null Deviance:      249.3
## Residual Deviance: 216.2      AIC: 226.2
##
## Waiting for profiling to be done...
##
##              2.5 %      97.5 %
## (Intercept) -1.97331473 14.3863932
## Age         -0.08530090  0.0578633
## Height      -12.23186073 -0.9531226
## Weightstart  0.00550310  0.0780286
## Weightgain   0.01934338  0.1571395
## Weightbaby  -0.60192148  0.5723205
## PrevKids    -0.35268819  0.3629317
## PrevBackache 0.43807159  1.2460929
##
## Call:
## glm(formula = y ~ Height + Weightgain + Weightstart + PrevBackache,
##      family = "binomial", data = backache)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2254  -0.9906  -0.6076   1.0612   2.1118
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   5.93391    4.13045   1.437  0.15082
## Height       -6.56023    2.84956  -2.302  0.02132 *
## Weightgain     0.08482    0.03256   2.605  0.00919 **
## Weightstart   0.04053    0.01822   2.225  0.02611 *
## PrevBackache  0.79393    0.17318   4.584 4.55e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##

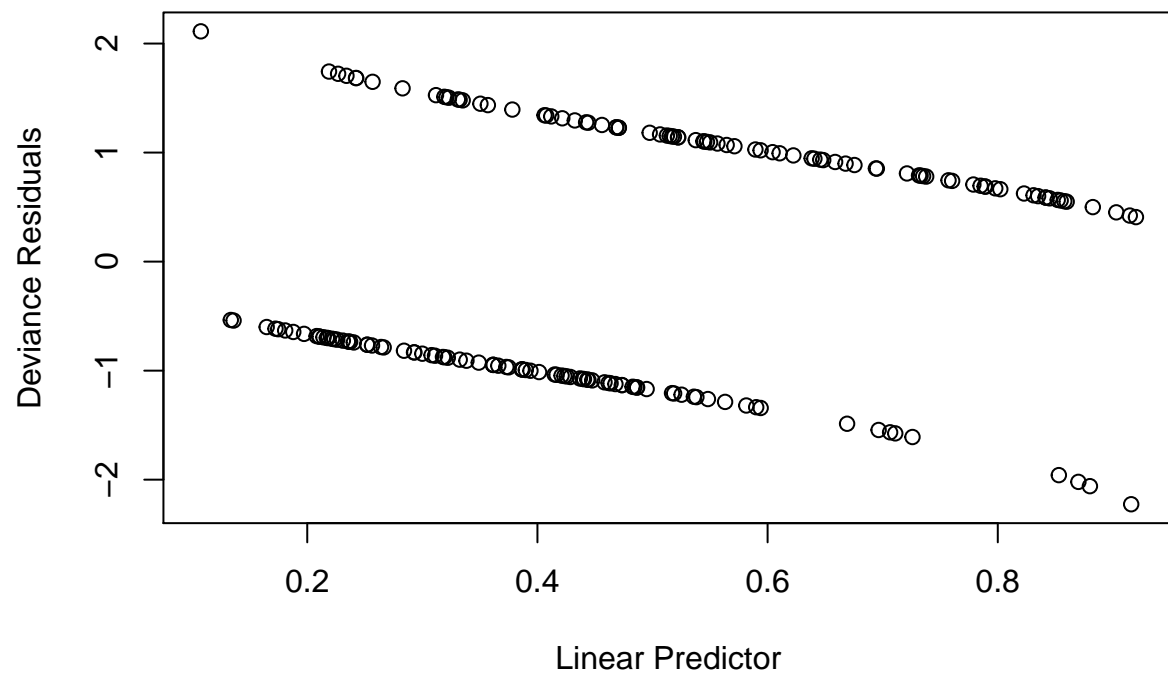
```

```
## Null deviance: 249.33 on 179 degrees of freedom
## Residual deviance: 216.23 on 175 degrees of freedom
## AIC: 226.23
##
## Number of Fisher Scoring iterations: 4
```

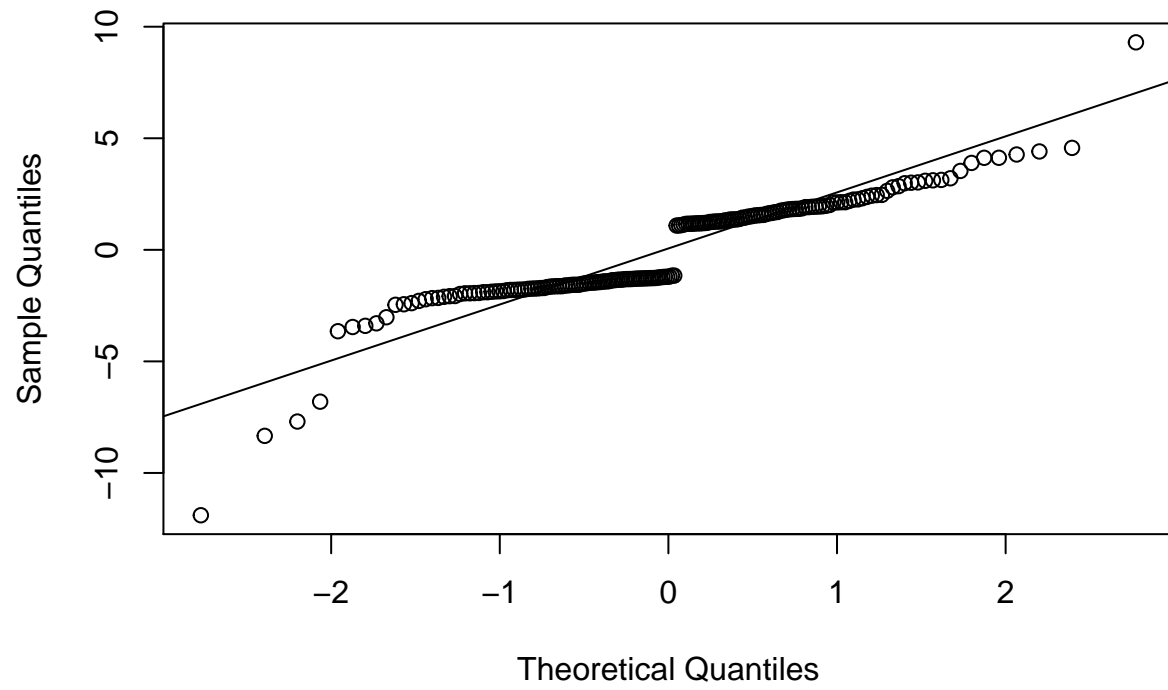
The significant variables in predicting backaches appear to be the mother's height, weight at the start of the pregnancy, weight gain during pregnancy, and the number of previous backaches. These variables had small p values and had confidence intervals not including 0. All other variables had non significant p values and had 95% confidence intervals containing.

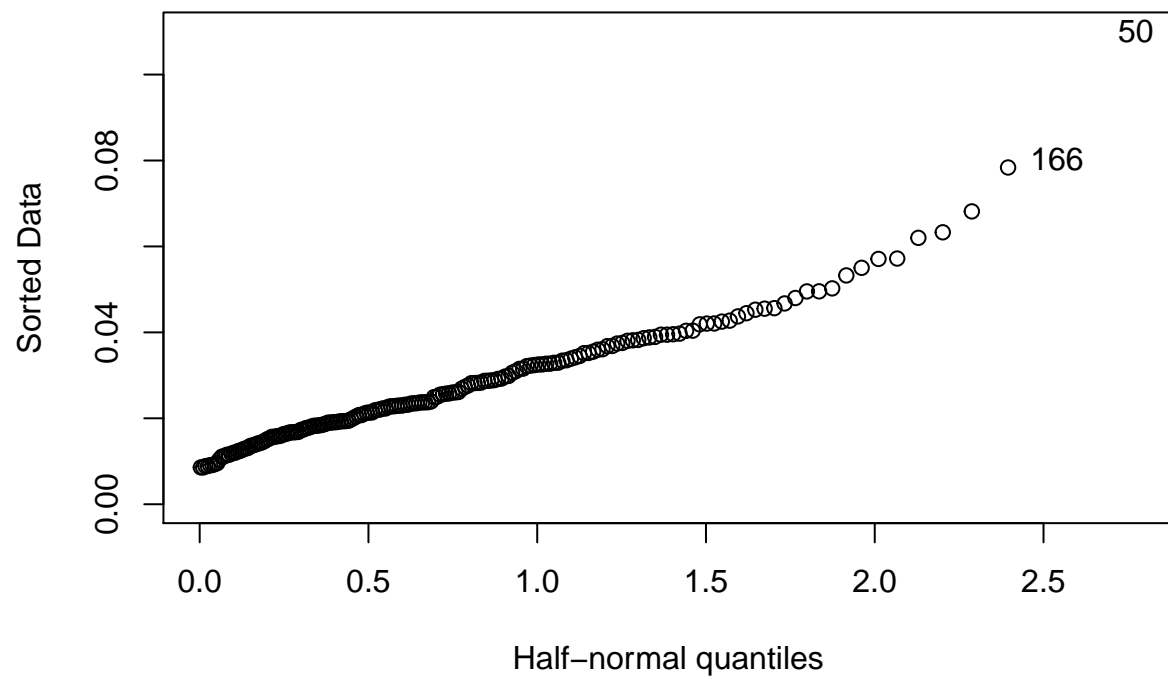
e)

```
## Loading required package: reshape
##
## Attaching package: 'reshape'
## The following object is masked from 'package:dplyr':
##
## rename
## The following objects are masked from 'package:tidyr':
##
## expand, smiths
## Loading required package: MASS
##
## Attaching package: 'MASS'
## The following object is masked from 'package:dplyr':
##
## select
##
## Hosmer and Lemeshow test (binary model)
##
## data: backache$y, reduced.mod$fitted.values
## X-squared = 0.53832, df = 0, p-value < 2.2e-16
```

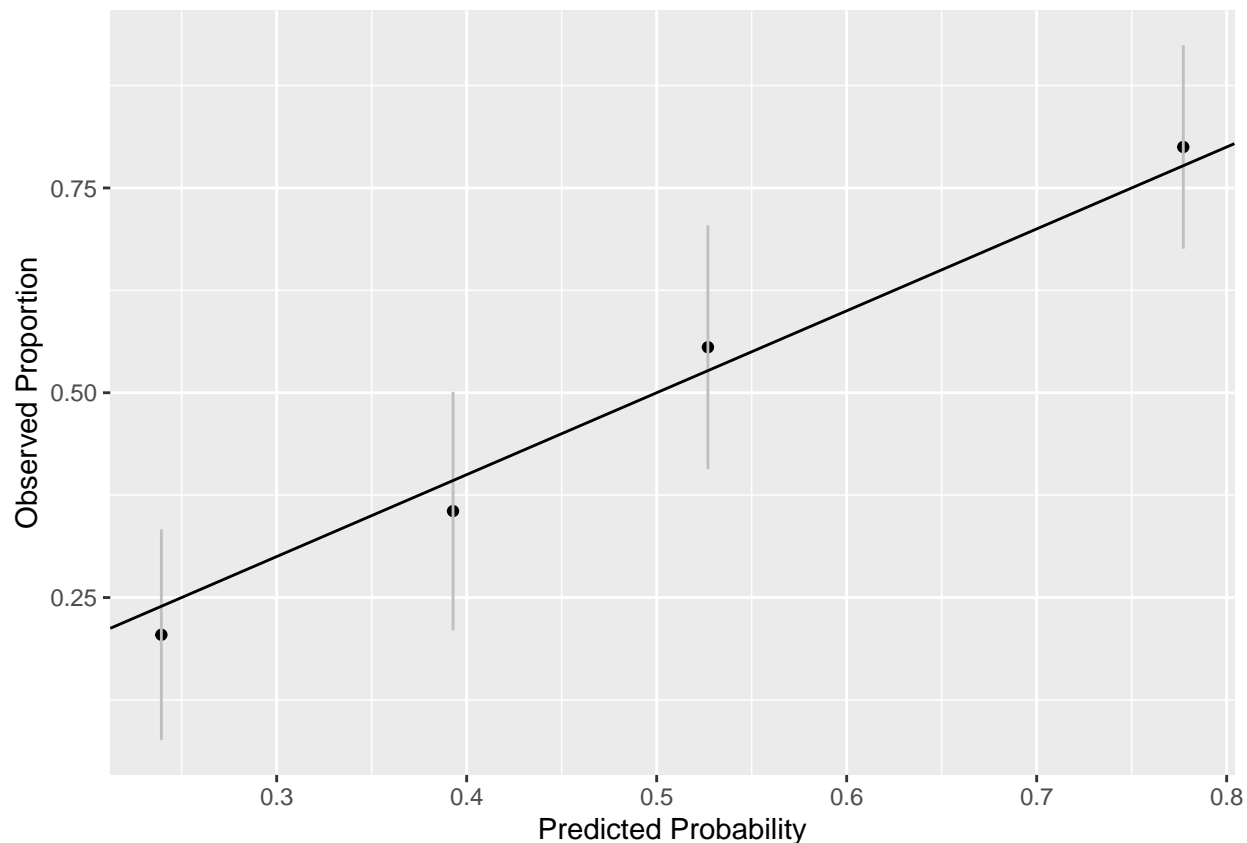


Normal Q-Q Plot









The Hosmer-Lemeshow test of the reduced model shows that this model is a good fit of the data with a p value close to zero. There are no points with extremely large residuals or leverage. The plot of the observed proportion by the expected proportion confirms that the observed proportions are well within the standard errors of the fitted model.

f)

```
##
## Call:
## glm(formula = y ~ Tablets + HotBottle + HotBath + Cushion + Standing +
##       Sitting + Lyingdown + Walking, family = "binomial", data = backache)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2693  -0.6241  -0.6241   0.9534   1.8610
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.5369     0.2862  -5.370 7.86e-08 ***
## Tablets       1.6074     0.5797   2.773  0.00556 **
## HotBottle     1.7074     0.9143   1.867  0.06184 .
## HotBath       1.6769     0.7395   2.268  0.02334 *
## Cushion       1.7859     0.4036   4.425 9.64e-06 ***
## Standing      0.7758     0.8844   0.877  0.38039
## Sitting       2.0896     0.7048   2.965  0.00303 **
## Lyingdown     1.3165     0.4514   2.917  0.00354 **
```

```

## Walking      0.5698      1.1338      0.503  0.61528
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 249.33  on 179  degrees of freedom
## Residual deviance: 187.25  on 171  degrees of freedom
## AIC: 205.25
##
## Number of Fisher Scoring iterations: 5
## Waiting for profiling to be done...

##              2.5 %      97.5 %
## (Intercept) -2.13053926 -1.002437
## Tablets      0.50979586  2.807900
## HotBottle     0.01348778  3.759854
## HotBath       0.31994176  3.297335
## Cushion       1.01628507  2.605518
## Standing     -0.95654800  2.596849
## Sitting       0.81782177  3.659881
## Lyingdown     0.45045689  2.230417
## Walking      -1.68627427  2.905335

## Single term deletions
##
## Model:
## y ~ Tablets + HotBottle + HotBath + Cushion + Standing + Sitting +
##      Lyingdown + Walking
##      Df Deviance    AIC F value    Pr(>F)
## <none>          187.25 205.25
## Tablets      1    195.65 211.65   7.6760 0.006215 **
## HotBottle    1    191.15 207.15   3.5658 0.060676 .
## HotBath      1    193.24 209.24   5.4716 0.020483 *
## Cushion      1    208.95 224.95  19.8186 1.532e-05 ***
## Standing     1    188.03 204.03   0.7100 0.400628
## Sitting      1    198.43 214.43  10.2110 0.001663 **
## Lyingdown    1    196.24 212.24   8.2152 0.004676 **
## Walking      1    187.50 203.50   0.2309 0.631478
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Based on the p values using the dropped F test, the variables Tablets, HotBath, Cushion, Sitting, and Lyingdown are all significant under the .05 threshold at least. They also had 95% confidence intervals for the parameter estimates not containing 0. Hotbottle was almost significant. Standing and Walking had fairly high p values.

g)

```

##
## Call:
## glm(formula = y ~ ., family = "binomial", data = dat)
##
## Deviance Residuals:

```

```

##      Min      1Q   Median      3Q      Max
## -3.3523 -0.6680 -0.6583   0.8454   1.7941
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.3863    0.2789  -4.971 6.67e-07 ***
## Fatigue        0.8727    0.4865   1.794 0.072849 .
## Bending        0.3998    0.5785   0.691 0.489471
## Lifting        1.5674    0.7235   2.167 0.030268 *
## MakingBeds     0.3597    0.6001   0.599 0.548881
## WashingUp      1.2847    1.0140   1.267 0.205158
## Ironing        0.9996    0.5895   1.696 0.089915 .
## Bowels        18.9524  3956.1803   0.005 0.996178
## Intercourse     0.3446    1.5362   0.224 0.822529
## Coughing       34.0889  2422.3426   0.014 0.988772
## Sneezing      -17.8991  1771.4397  -0.010 0.991938
## TurningInBed   -0.2182    0.8589  -0.254 0.799484
## Standing2      1.6418    0.4787   3.429 0.000605 ***
## Sitting2       0.7815    0.4963   1.575 0.115348
## Lyingdown2     0.8705    0.7176   1.213 0.225111
## Walking2      -0.0328    0.6205  -0.053 0.957837
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 249.33  on 179  degrees of freedom
## Residual deviance: 182.77  on 164  degrees of freedom
## AIC: 214.77
##
## Number of Fisher Scoring iterations: 16
## Waiting for profiling to be done...
##
##              2.5 %      97.5 %
## (Intercept)  -1.9616070 -0.8630597
## Fatigue      -0.0761620  1.8459952
## Bending      -0.7515475  1.5389539
## Lifting       0.2377333  3.1612078
## MakingBeds   -0.8335061  1.5393212
## WashingUp    -0.6679099  3.4776954
## Ironing      -0.1339810  2.2043165
## Bowels       -787.1259564      NA
## Intercourse  -2.9074039  3.7534441
## Coughing     -216.7607452      NA
## Sneezing      NA 341.3891964
## TurningInBed -1.9698337  1.4860455
## Standing2    0.7309770  2.6244267
## Sitting2     -0.1919398  1.7680393
## Lyingdown2   -0.5071578  2.3622975
## Walking2     -1.2663536  1.1957862
##
## Single term deletions
##
## Model:

```

```
## y ~ Fatigue + Bending + Lifting + MakingBeds + WashingUp + Ironing +
##     Bowels + Intercourse + Coughing + Sneezing + TurningInBed +
##     Standing2 + Sitting2 + Lyingdown2 + Walking2
##           Df Deviance    AIC F value   Pr(>F)
## <none>           182.77 214.77
## Fatigue         1   186.02 216.02   2.9170 0.089541 .
## Bending         1   183.24 213.24   0.4259 0.514924
## Lifting         1   188.18 218.18   4.8590 0.028894 *
## MakingBeds      1   183.12 213.12   0.3207 0.571975
## WashingUp       1   184.43 214.43   1.4948 0.223220
## Ironing         1   185.75 215.75   2.6743 0.103898
## Bowels          1   185.94 215.94   2.8447 0.093575 .
## Intercourse     1   182.82 212.82   0.0451 0.832006
## Coughing        1   190.46 220.46   6.9030 0.009422 **
## Sneezing        1   186.06 216.06   2.9575 0.087366 .
## TurningInBed    1   182.83 212.83   0.0581 0.809897
## Standing2       1   195.58 225.58  11.4958 0.000874 ***
## Sitting2        1   185.25 215.25   2.2278 0.137471
## Lyingdown2      1   184.29 214.29   1.3657 0.244256
## Walking2        1   182.77 212.77   0.0025 0.960116
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

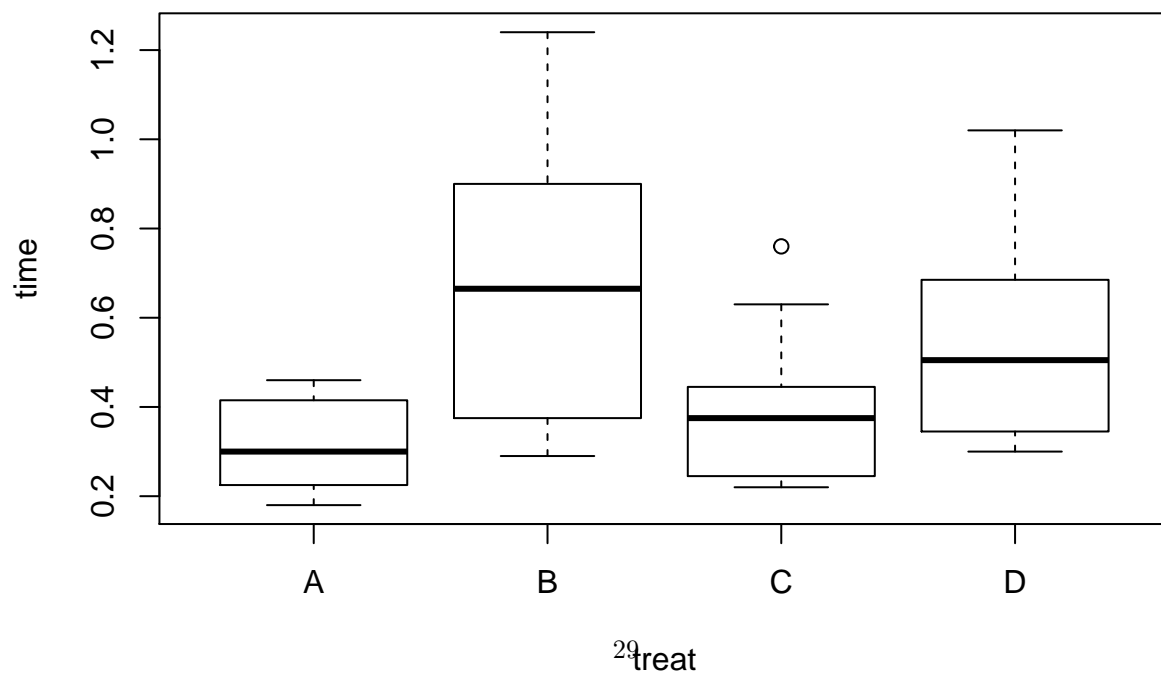
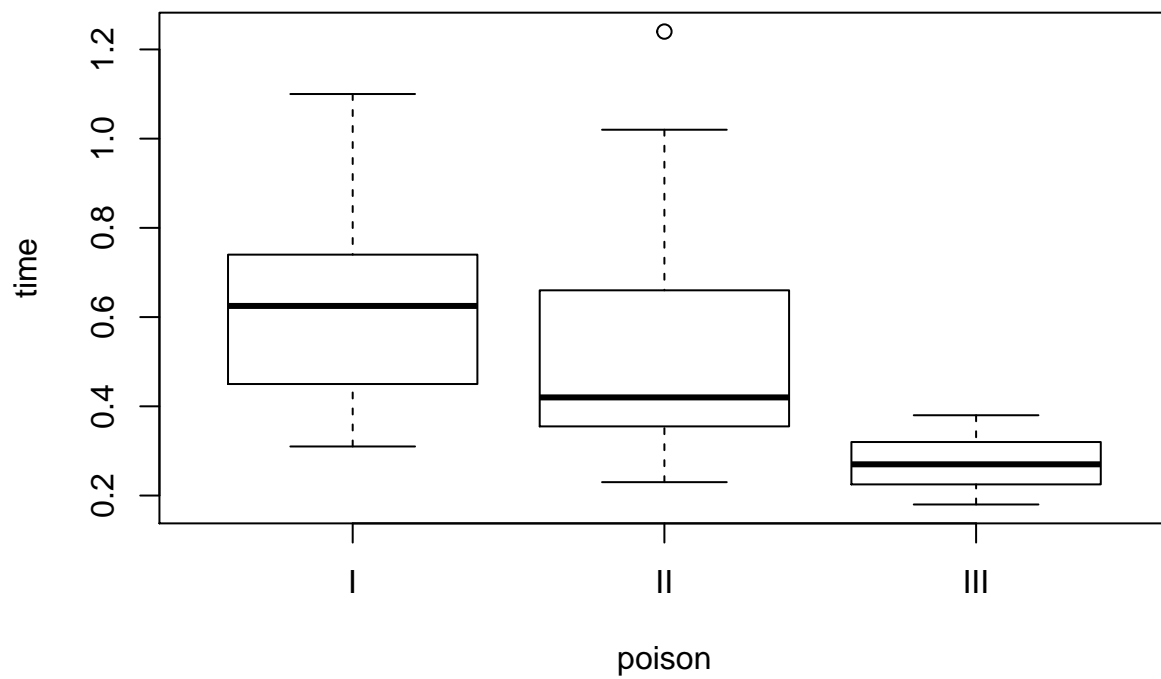
The dropped F test showed that Coughing, Standing2, and Lifting were all significant below the .05 threshold. Standing2 and Lifting had significant p values from the t distribution (summary output) whereas coughing was close to 1 and had a very high standard error. There were only 4 responses as 1 for coughing so that explains the high standard error. Standing2 and lifting are most likely significant, but Coughing has too few responses as 1 to conclude whether it is significant in conjunction with the high p value and standard error.

## h)

There are many factors that might be associated with whether or not a pregnant woman will experience severe or non severe back aches. These factors may or may not cause back aches, they could simply be related. A patients individual traits such as a history of backaches, their weight at the start of the pregnancy, and weight gain during pregnancy could increase the chance of back aches with pregnancy. A shorter women has much less of a chance of having back aches. Additionally, if the patient lifts objects or stands frequently the chances of back aches increases. Statistical tests showed that perceiving activities as pain relieving such as a hot bath, lying down, sitting, or using tablets would increase the likelihood of having back aches, but this is evidence of confounding: if a women perceives that resting alleviates back pain then she most likely has back pain.

### Question 3

a)



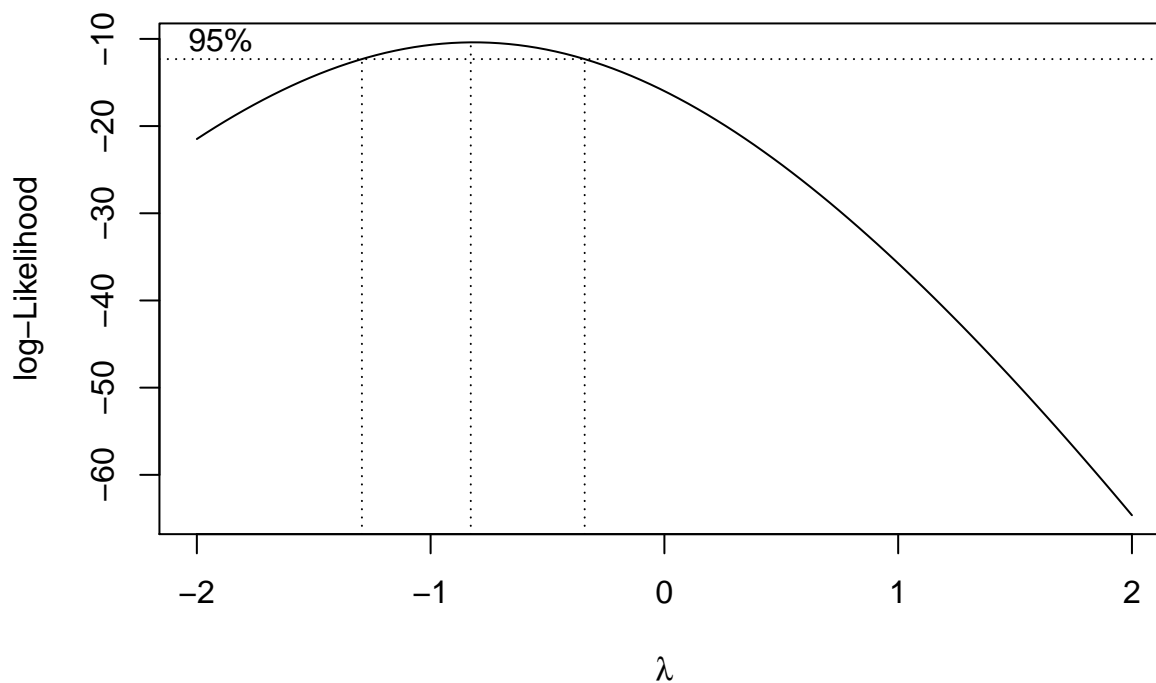
Survival time appears to be longer for poison I then poison II. Poison III appears to clearly yield the shortest survival time. Treatments B and D appear to yield similar survival times, and they both yield longer times than A and C. A and C are about the same, with A yielding slightly shorter survival times.

b)

```
##
## Call:
## lm(formula = time ~ poison + treat + I(poison:treat), data = rats)
##
## Residuals:
```

|  | Min      | 1Q       | Median  | 3Q      | Max     |
|--|----------|----------|---------|---------|---------|
|  | -0.32500 | -0.04875 | 0.00500 | 0.04312 | 0.42500 |

```
##
## Coefficients: (5 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.41250    0.07457   5.532 2.94e-06 ***
## poisonII         0.14000    0.14914   0.939  0.3541
## poisonIII        -0.20250    0.10546  -1.920  0.0628 .
## treatB           0.12500    0.10546   1.185  0.2437
## treatC           0.02500    0.10546   0.237  0.8140
## treatD           0.11500    0.10546   1.090  0.2827
## I(poison:treat)I:B  0.34250    0.14914   2.297  0.0276 *
## I(poison:treat)I:C  0.13000    0.14914   0.872  0.3892
## I(poison:treat)I:D  0.08250    0.14914   0.553  0.5836
## I(poison:treat)II:A -0.23250    0.14914  -1.559  0.1278
## I(poison:treat)II:B  0.13750    0.14914   0.922  0.3627
## I(poison:treat)II:C -0.20250    0.14914  -1.358  0.1830
## I(poison:treat)II:D      NA         NA      NA      NA
## I(poison:treat)III:A     NA         NA      NA      NA
## I(poison:treat)III:B     NA         NA      NA      NA
## I(poison:treat)III:C     NA         NA      NA      NA
## I(poison:treat)III:D     NA         NA      NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1491 on 36 degrees of freedom
## Multiple R-squared:  0.7335, Adjusted R-squared:  0.6521
## F-statistic: 9.01 on 11 and 36 DF, p-value: 1.986e-07
```



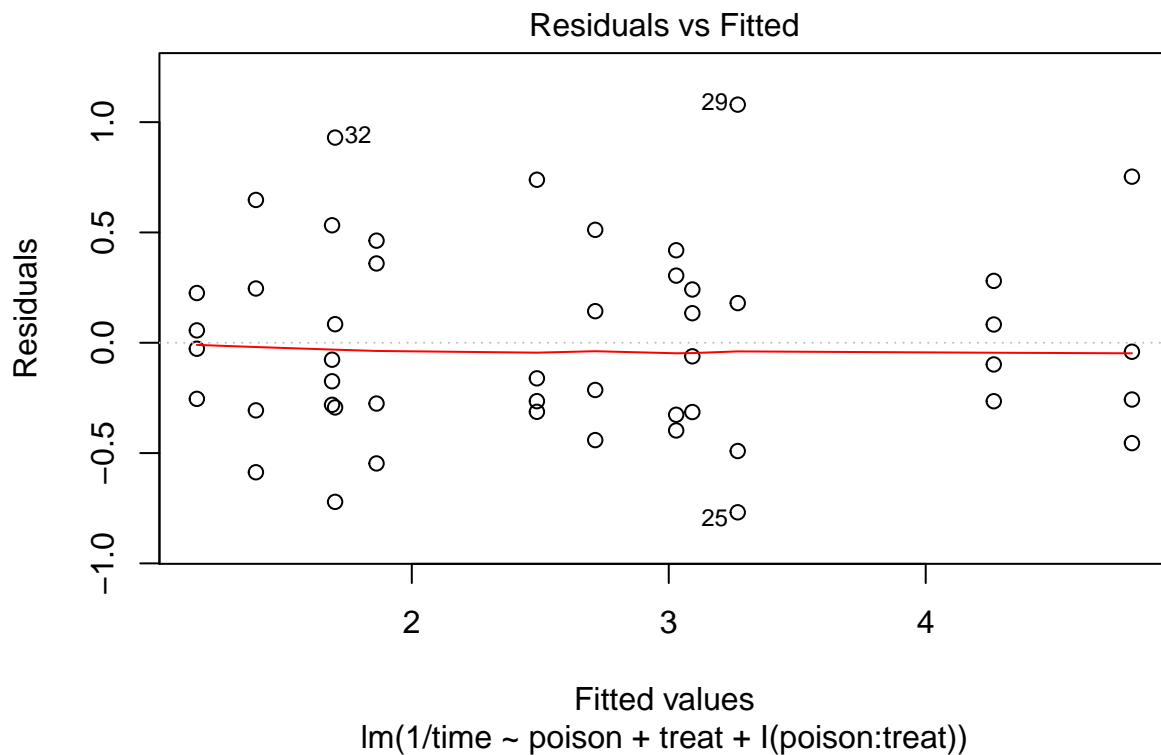
```
## [1] -0.8282828
##
## Shapiro-Wilk normality test
##
## data: 1/rats$time
## W = 0.9586, p-value = 0.08865
```

A box cox transformation would yield an optimal lambda value of -.83. This could be rounded to -1 to yield a more interpretable response: 1/time. A Shapiro Wilks test for normality yielded a p value of about .09 after transforming the response, so the distribution of the transformed response follows a normal distribution reasonably well.

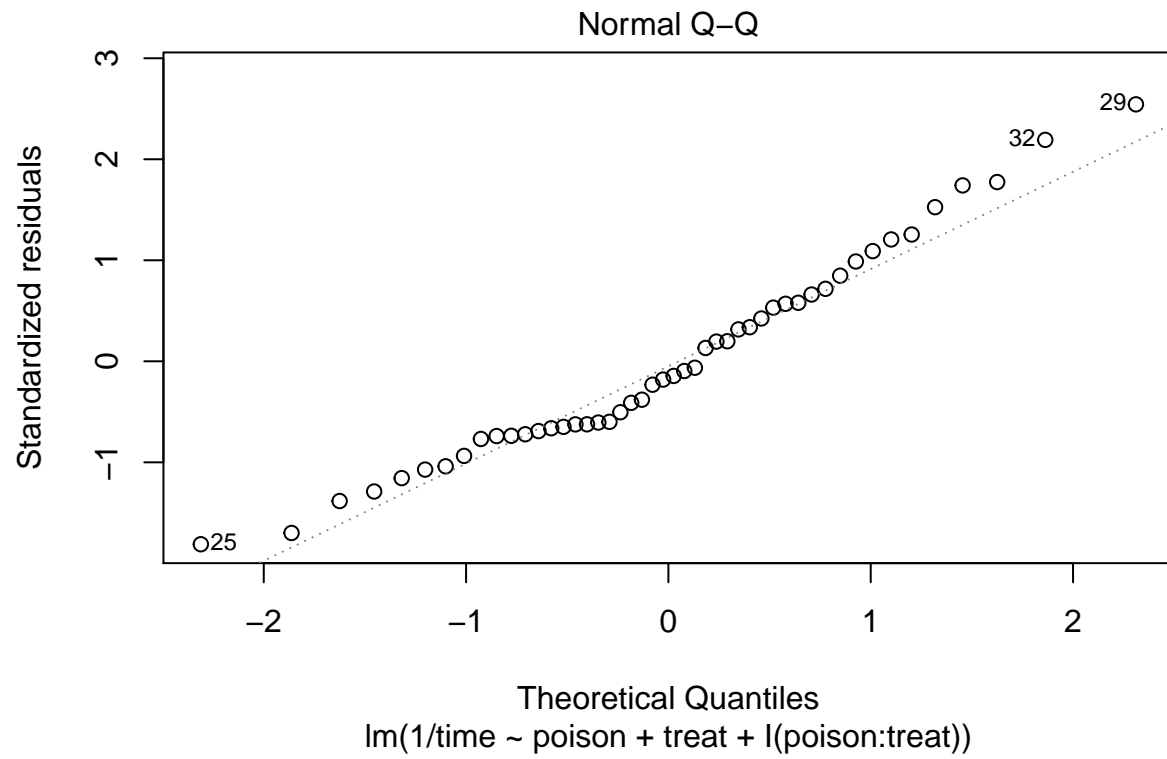
c)

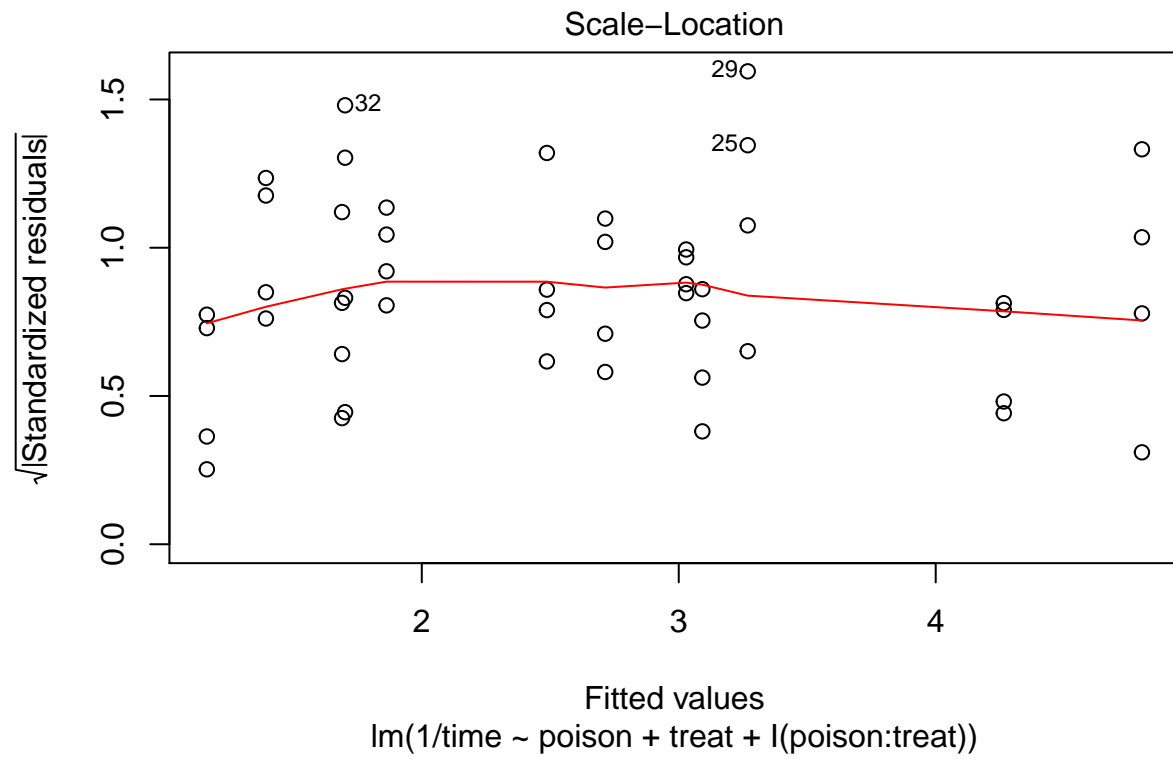
```
##
## Call:
## lm(formula = 1/time ~ poison + treat + I(poison:treat), data = rats)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.76847 -0.29642 -0.06914  0.25458  1.07936
##
## Coefficients: (5 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.48688    0.24499  10.151 4.16e-12 ***
## poisonII       0.92553    0.48999   1.889  0.0670 .
```

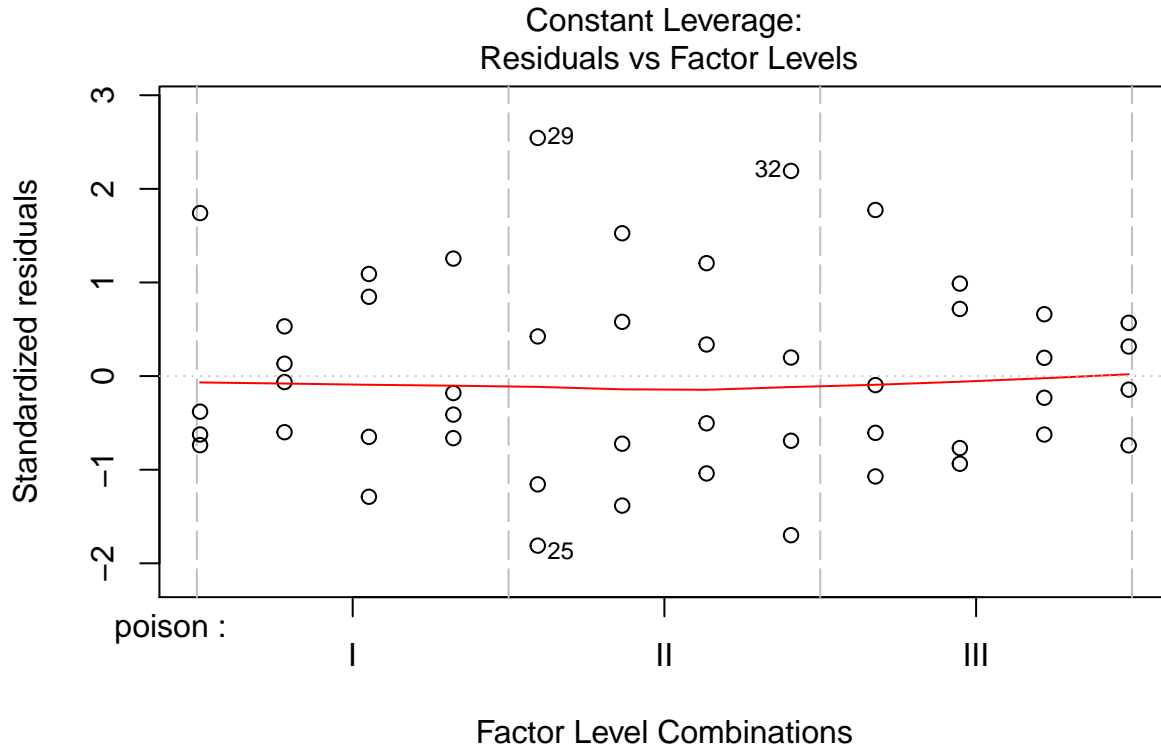
```
## poisonIII      2.31580    0.34647    6.684 8.56e-08 ***
## treatB        -1.77371    0.34647   -5.119 1.04e-05 ***
## treatC        -0.53770    0.34647   -1.552  0.1294
## treatD        -1.71088    0.34647   -4.938 1.82e-05 ***
## I(poison:treat)I:B    0.45030    0.48999    0.919  0.3642
## I(poison:treat)I:C   -0.08646    0.48999   -0.176  0.8609
## I(poison:treat)I:D    0.91368    0.48999    1.865  0.0704 .
## I(poison:treat)II:A   -0.14394    0.48999   -0.294  0.7706
## I(poison:treat)II:B   -0.24531    0.48999   -0.501  0.6197
## I(poison:treat)II:C   -0.16080    0.48999   -0.328  0.7447
## I(poison:treat)II:D      NA         NA         NA     NA
## I(poison:treat)III:A   NA         NA         NA     NA
## I(poison:treat)III:B   NA         NA         NA     NA
## I(poison:treat)III:C   NA         NA         NA     NA
## I(poison:treat)III:D   NA         NA         NA     NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.49 on 36 degrees of freedom
## Multiple R-squared:  0.8681, Adjusted R-squared:  0.8277
## F-statistic: 21.53 on 11 and 36 DF, p-value: 1.289e-12
```











The model appears to fit the data ok. There are do not appear to be any huge outliers in the residuals, but they do appear to be larger for Poison level II compared to I and III. There do not appear to be any points of influence as the values have constant leverage.

d)

```
## Analysis of Variance Table
##
## Response: 1/time
##           Df Sum Sq Mean Sq F value    Pr(>F)
## poison      2 34.877  17.4386  72.6347 2.310e-13 ***
## treat       3 20.414   6.8048  28.3431 1.376e-09 ***
## I(poison:treat) 6  1.571   0.2618   1.0904  0.3867
## Residuals   36  8.643   0.2401
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Call:
## lm(formula = 1/time ~ poison + treat, data = rats)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.82757 -0.37619  0.02116  0.27568  1.18153
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.6977     0.1744  15.473 < 2e-16 ***
## poisonII       0.4686     0.1744   2.688 0.01026 *
## poisonIII      1.9964     0.1744  11.451 1.69e-14 ***
## treatB        -1.6574     0.2013  -8.233 2.66e-10 ***
## treatC        -0.5721     0.2013  -2.842 0.00689 **
## treatD        -1.3583     0.2013  -6.747 3.35e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4931 on 42 degrees of freedom
## Multiple R-squared:  0.8441, Adjusted R-squared:  0.8255
## F-statistic: 45.47 on 5 and 42 DF,  p-value: 6.974e-16
```

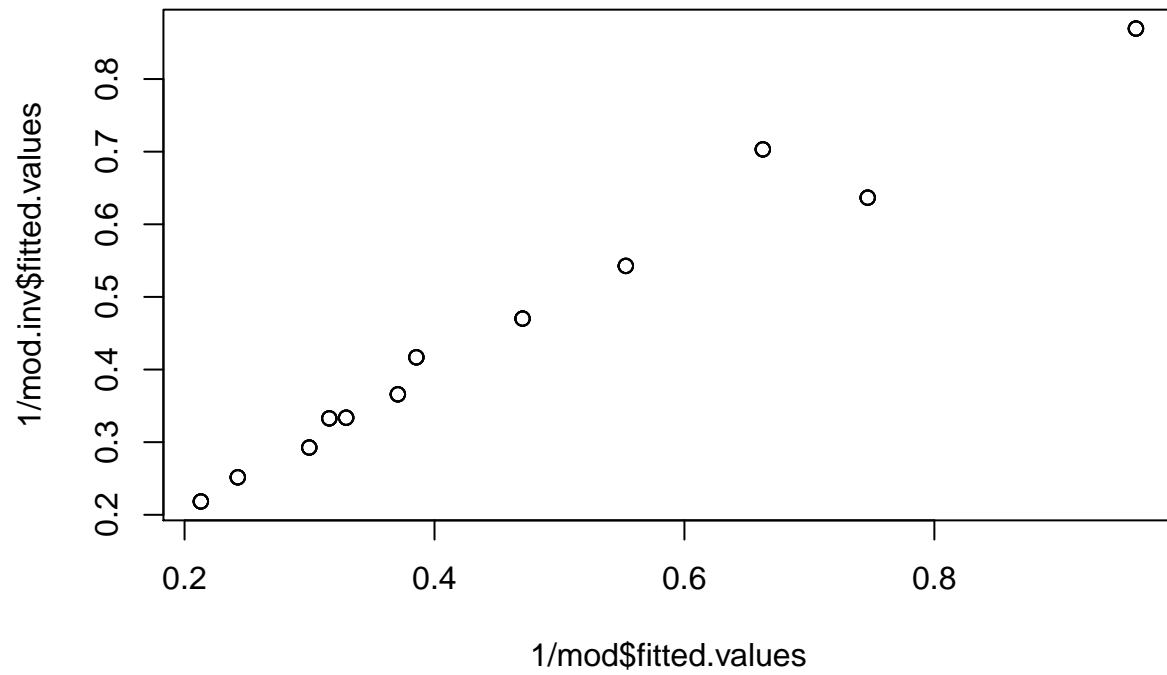
Based on the anova test and corresponding p value, and the p values from the t distribution and standard errors from the model, the interaction term does not appear to be significant. After creating the model with no interaction term, the shortest survival time would result from having poison III and treatment A.

e)

```
##
## Call:
## glm(formula = 1/time ~ poison + treat, family = inverse.gaussian(link = "identity"),
##      data = rats)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.48207  -0.08969  -0.00734   0.07049   0.30456
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.7327     0.2710  10.082 8.79e-13 ***
## poisonII       0.2720     0.1309   2.077 0.043947 *
## poisonIII      1.8468     0.2875   6.424 9.74e-08 ***
## treatB        -1.5826     0.2764  -5.726 9.83e-07 ***
## treatC        -0.6058     0.3229  -1.876 0.067624 .
## treatD        -1.1621     0.2912  -3.991 0.000259 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for inverse.gaussian family taken to be 0.02674546)
##
##      Null deviance: 4.7060  on 47  degrees of freedom
## Residual deviance: 1.2735  on 42  degrees of freedom
## AIC: 99.241
##
## Number of Fisher Scoring iterations: 7
```

Creating an inverse gaussian model with an identity link, the shortest survival time would result from having poison III and treatment A agreeing with previous model.

f)



Yes the models produce a very similar fit. They disagree slightly on a couple of points, but on the whole, they produce similar fits.