MAST90104: Introduction to Statistical Learning

Assignment 3, 2019 Solutions

Please submit a scanned or other electronic .pdf of your work, **named XabYcd.pdf where XabYcd is your name,** via the Learning Management System - see this link for instructions.

The .pdf must have in **one file**:

- handwritten or typed answers to the questions
- handwritten or typed R code used to produce your answers
- graphics required to answer the questions

If you have more than one file submitted, *only the last LMS .pdf file with your name on it will be marked.*

1. Suppose that there is one factor with 4 levels. Starting with the usual less than full rank model, find the matrices D and E and verify that $I_4 + DE$ is rank 4 for C_4 from the `contr.helmert(4)` matrix in R. Find the resulting reparameterisation and interpret it in terms of the mean responses for each of the 4 levels.

   Solution:

```
contr.helmert(4)

##   [,1] [,2] [,3]
## 1   -1   -1   -1
## 2    1   -1   -1
## 3    0    2   -1
## 4    0    0    3

# The C matrix is obtained from adding a column of
# zeros to contr.sum and then adding a row which
# is the transpose of the first standard unit vector
#
(C <- rbind(t(c(1,0,0,0)),cbind(c(0,0,0,0),contr.helmert(4))))

##   [,1] [,2] [,3] [,4]
##      1    0    0    0
## 1    0   -1   -1   -1
## 2    0    1   -1   -1
## 3    0    0    2   -1
## 4    0    0    0    3

(Cr <- C[-5,])

##   [,1] [,2] [,3] [,4]
##      1    0    0    0
## 1    0   -1   -1   -1
## 2    0    1   -1   -1
## 3    0    0    2   -1

#
# The original design matrix is used
# The column vector d is the same as for contr.treatment
#
(X <- cbind(c(1,1,1,1), diag(4)))
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    1    1    0    0    0
## [2,]    1    0    1    0    0
## [3,]    1    0    0    1    0
## [4,]    1    0    0    0    1

D <- c(1,-1,-1,-1)
E <-  t(c(0,-1,-1,-1))
```

```
#
# Calculate I + DE
# Invert it to check rank
# Calculate the multiplying matrix to reparameterise
#
library(dplyr)

##
## Attaching package:  'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(fractional)
(diag(4) + D%*%E)%*%Cr

##      [,1] [,2] [,3] [,4]
## [1,]    1    0    0    3
## [2,]    0   -1   -1   -4
## [3,]    0    1   -1   -4
## [4,]    0    0    2   -4

mult1 <- solve((diag(4) + D%*%E)%*%Cr) %>% fractional %>% print

##      [,1]  [,2]  [,3]  [,4]
## [1,]     1   1/4   1/4   1/4
## [2,]     .  -1/2   1/2     .
## [3,]     .  -1/6  -1/6   1/3
## [4,]     . -1/12 -1/12 -1/12

mult <- mult1%*%cbind(diag(4),D) %>% fractional %>% print

##                               D
## [1,]     1   1/4   1/4   1/4   1/4
## [2,]     .  -1/2   1/2     .     .
## [3,]     .  -1/6  -1/6   1/3     .
## [4,]     . -1/12 -1/12 -1/12   1/4
```

The original parameterisation $\mu, \tau_1, \tau_2, \tau_3, \tau_4$ is turned to the average of the treatment mean values

$$\gamma_1 = \frac{\sum_{i=1}^{4} \mu + \tau_i}{4}$$

and the differences

$$\gamma_2 = \eta_2 - \eta_1, \gamma_3 = \eta_3 - \eta_2, \gamma_4 = \eta_4 - \eta_3$$

of successive averages

$$\eta_1 = \tau_1, \eta_2 = \frac{\tau_1 + \tau_2}{2}, \eta_3 = \frac{\tau_1 + \tau_2 + \tau_3}{3}, \eta_4 = \frac{\tau_1 + \tau_2 + \tau_3 + \tau_4}{4}$$

of the τ's.

2. **You may not use theR `glm` command for this question.** Fit a binomial regression model to the O-rings data from the Challenger disaster, available in "orings.csv", using a *complementary log-log* link.

   Your solution should include the following:

   (a) parameter estimates

   (b) 95% CIs for the parameter estimates

   (c) a likelihood ratio test for the significance of the temperature coefficient

   (d) an estimate of the probability of damage when the temperature equals 29 Fahrenheit together with a 95% CI

   (e) a plot comparing the fitted c-log-log model to the fitted logit model.

   Solution: For a binomial regression with a c-log-log link we have $y_i \sim \text{bin}(m_i, p_i)$, where $p_i = 1 - \exp(-e^{\eta_i})$ and $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$, so

$$\begin{aligned}
l(\boldsymbol{\beta}) &= c + \sum_i [y_i \log p_i + (m_i - y_i)\log(1 - p_i)] \\
&= c + \sum_i [y_i \log(1 - \exp(-e^{\eta_i})) - (m_i - y_i)e^{\eta_i}] \\
\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_j} &= \sum_i \left[\frac{y_i e^{\eta_i} x_{i,j}}{\exp(e^{\eta_i}) - 1} - (m_i - y_i)e^{\eta_i} x_{i,j}\right] \\
\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_k} &= \sum_i \left[\frac{y_i e^{\eta_i} x_{i,j} x_{i,k}}{\exp(e^{\eta_i}) - 1} - \frac{y_i e^{\eta_i} x_{i,j}}{(\exp(e^{\eta_i}) - 1)^2} \exp(e^{\eta_i}) e^{\eta_i} x_{i,k} - (m_i - y_i)e^{\eta_i} x_{i,j} x_{i,k}\right] \\
-\mathbb{E}\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_k} &= -\sum_i x_{i,j} x_{i,k} e^{\eta_i} \left[\frac{m_i(1 - \exp(-e^{\eta_i}))}{\exp(e^{\eta_i}) - 1}\right. \\
&\qquad \left. - \frac{m_i(1 - \exp(-e^{\eta_i}))\exp(e^{\eta_i})e^{\eta_i}}{(\exp(e^{\eta_i}) - 1)^2} - (m_i - m_i(1 - \exp(-e^{\eta_i})))\right] \\
&= \sum_i x_{i,j} x_{i,k} \frac{m_i e^{2\eta_i}}{\exp(e^{\eta_i}) - 1} \quad = \quad \sum_i x_{i,j} x_{i,k} \frac{m_i(1 - p_i)(\log(1 - p_i))^2}{p_i}
\end{aligned}$$

[2]

   (a) Estimating $\boldsymbol{\beta}$ **Solution:**

```
library(faraway)
data(orings)
logL <- function(beta, orings) {
y <- orings$damage
X <- cbind(1, orings$temp)
zeta <- X %*% beta
p <- 1 - exp(-exp(zeta))
return(sum(y*log(p) + (6 - y)*log(1 - p)))
}
(betahat <- optim(c(10, -.2), logL, orings=orings, control=list(fnscale=-1))$par)

## [1] 10.8622281 -0.2054973
```

[2]

(b) 95% CIs for β_0 and β_1

```
X <- cbind(1, orings$temp)
zetahat <- X %*% betahat
phat <- 1 - exp(-exp(zetahat))
a <- 6*(1 - phat)*(log(1-phat))^2/phat
I11 <- sum(X[,1]^2*a)
I12 <- sum(X[,1]*X[,2]*a)
I22 <- sum(X[,2]^2*a)
Iinv <- solve(matrix(c(I11, I12, I12, I22), 2, 2))
(si_1 <- sqrt(Iinv[1,1]))

## [1] 2.736517

c(betahat[1] - 1.96*si_1, betahat[1] + 1.96*si_1)

## [1]  5.498654 16.225802

(si_2 <- sqrt(Iinv[2,2]))

## [1] 0.04560421

c(betahat[2] - 1.96*si_2, betahat[2] + 1.96*si_2)

## [1] -0.2948815 -0.1161130
```

[2]

Comparing with **glm** output, we see that the estimates and standard errors agree with ours to four significant figures.

```
cloglogmod <- glm(cbind(damage,6-damage) ~ temp, family=binomial(link=cloglog), orings)
summary(cloglogmod)

##
## Call:
## glm(formula = cbind(damage, 6 - damage) ~ temp, family = binomial(link = cloglog),
##     data = orings)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9884  -0.7262  -0.4373  -0.2141   1.9520
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 10.86388    2.73668   3.970 7.20e-05 ***
## temp        -0.20552    0.04561  -4.506 6.59e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 38.898  on 22  degrees of freedom
## Residual deviance: 16.029  on 21  degrees of freedom
## AIC: 32.791
##
## Number of Fisher Scoring iterations: 7
```

(c) Testing $H_0 : \beta_1 = 0$. **Solution:**
First we calculate the deviance for the model including temperature.

```r
y <- orings$damage
m <- rep(6, length(y))
ylogxy <- function(x, y) ifelse(y == 0, 0, y*log(x/y))
(D <- -2*sum(ylogxy(m*phat, y) + ylogxy(m*(1-phat), m - y)))

## [1] 16.02857

(df <- length(y) - length(betahat))

## [1] 21
```

Next we fit the null model and use a likelihood ratio test.

```r
(phatN <- sum(y)/sum(m))

## [1] 0.07971014

(DN <- -2*sum(ylogxy(m*phatN, y) + ylogxy(m*(1-phatN), m - y)))

## [1] 38.89766

(dfN <- length(y) - 1)

## [1] 22

pchisq(DN - D, dfN - df, lower=FALSE) # p-value

## [1] 1.734185e-06
```

We have very strong evidence that $\beta_1 \neq 0$.

[2]

Note that our deviance calculations agree with the output from `glm`.

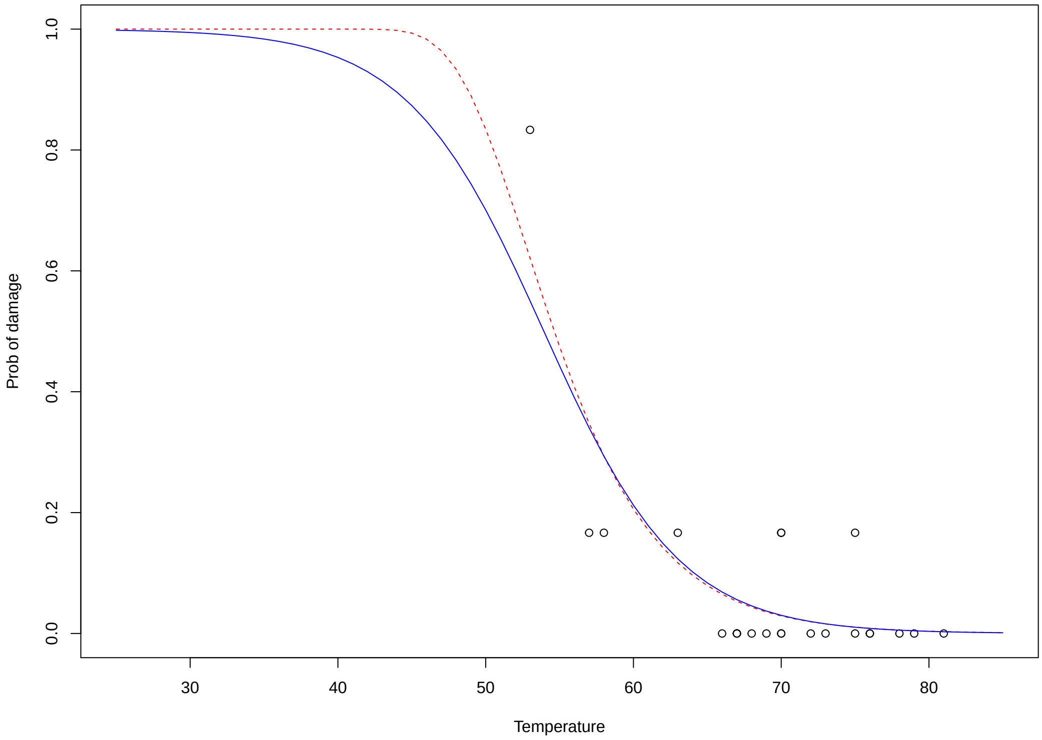
(d) Forecast for the probability of failure when the temperature is 29° Fahrenheit.

```r
options(digits=16)
si2 <- matrix(c(1, 29), 1, 2) %*% Iinv %*% matrix(c(1, 29), 2, 1)
(p29 <- 1 - exp(-exp(betahat[1] + betahat[2]*29)))

## [1] 1

1 - exp(-exp(betahat[1] + betahat[2]*29 - 1.96*sqrt(si2)))[1]

## [1] 0.9997181332034782

1 - exp(-exp(betahat[1] + betahat[2]*29 + 1.96*sqrt(si2)))[1]

## [1] 1
```

[2]

(e) Plot of the fitted c-log-log (dashed line) and logit (solid line) models. They are very close for the observed data points, but the c-log-log model puts much less weight in the left tail, giving a notably larger fit when temperature equals 31°.

```r
plot(damage/6 ~ temp, orings, xlim=c(25,85), ylim=c(0,1),
xlab="Temperature", ylab="Prob of damage")
x <- seq(25,85,1)
lines(x, 1 - exp(-exp(betahat[1] + betahat[2]*x)), col="red", lty=2)
betalogit <- glm(cbind(damage,6-damage) ~ temp, family=binomial, orings)$coefficients
lines(x, ilogit(betalogit[1] + betalogit[2]*x), col="blue")
```



[2]

3. Suppose the Y comes from an exponential family with pdf or pmf f of the form

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

(a) Show that $\mathbb{E}Y = b'(\theta)$

(b) Show that $\operatorname{Var} Y = b''(\theta)a(\phi)$.

Solution:

$$\frac{\partial \log f(y;\theta,\phi)}{\partial\theta} = \frac{y - b'(\theta)}{a(\phi)}$$

But $\mathbb{E}\,\frac{\partial \log f(y;\theta,\phi)}{\partial\theta} = 0$ which gives 3a. *[2]*

Further

$$\frac{\partial^2 \log f(y;\theta,\phi)}{\partial\theta^2} = \frac{-b''(\theta)}{a(\phi)}$$

so

$$\begin{aligned}
\frac{b''(\theta)}{a(\phi)} &= \mathbb{E}\left(-\frac{\partial^2 \log f(y;\theta,\phi)}{\partial\theta^2}\right) \\
&= \operatorname{Var}\frac{\partial \log f(y;\theta,\phi)}{\partial\theta} \\
&= \operatorname{Var}\left(\frac{y - b'(\theta)}{a(\phi)}\right) \\
&= \frac{\operatorname{Var} y}{a^2(\phi)}
\end{aligned}$$

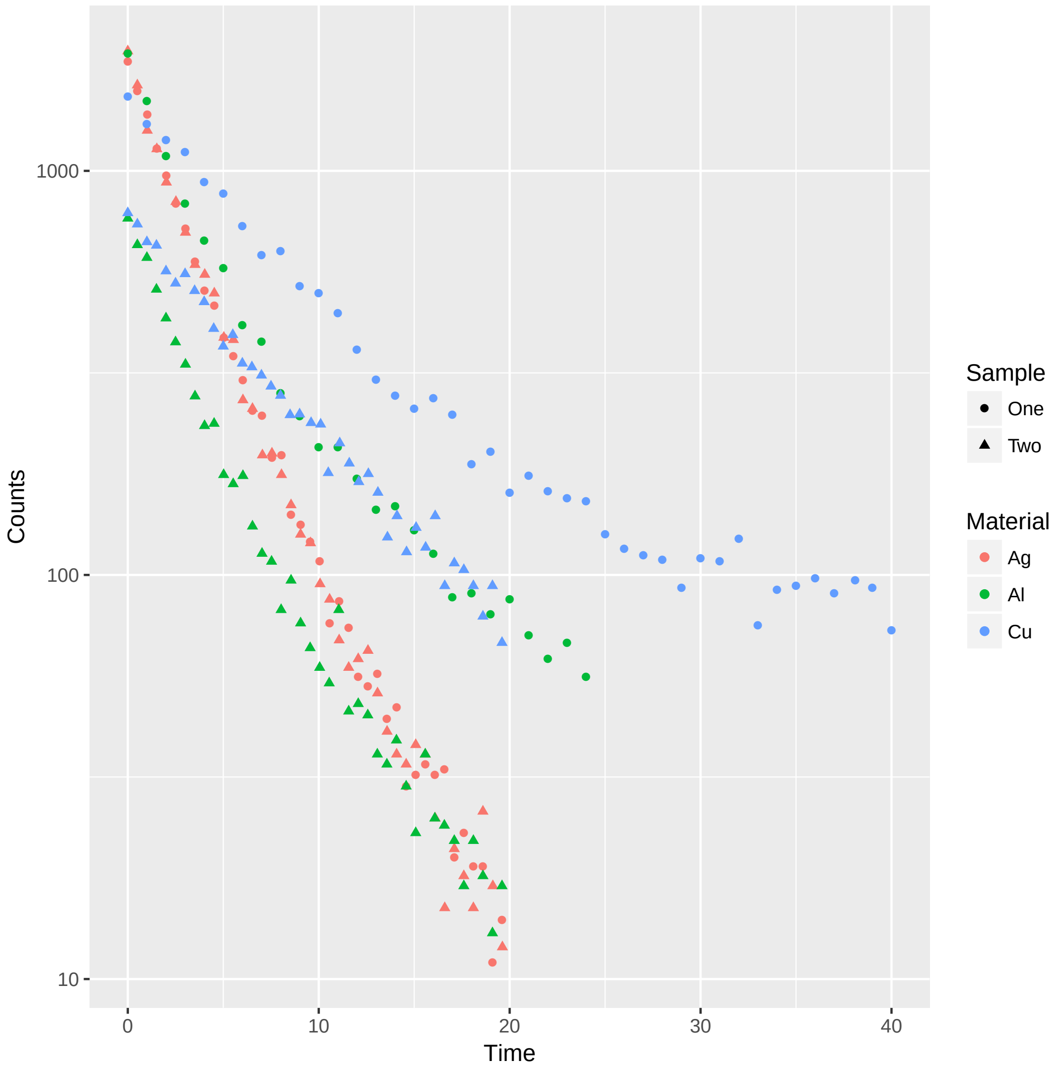
which gives 3b. *[2]*

4. Steve Sahyun of the University of Winsconsin has a website http://sahyun.net/neutron.php in which he explains neutron activation as a means of producing radioactive metals from standard metals. It includes data on samples of Silver (Ag), Aluminium (Al) and Copper (Cu) which have been subject to neutron activation. For each of the elements, two samples have been tested. The data file `Radioactive.csv` contains radioactive counts at various times after the initial measurement (with the square of time also recorded) as well as the name of the element and whether the count is for the first or second sample.

   (a) Use `qplot` (in package `ggplot2`) to plot the log of counts versus time with different colours for the different elements and different shapes for samples one and two. Your answer should include the plot and your command to get it.

   Solution:

```
Radioactive <- read.csv("Radioactive.csv")
library(ggplot2)
qplot(data=Radioactive,x=Time,y=Counts,log = "y",
col = Material,
shape = Sample)
```



[2]

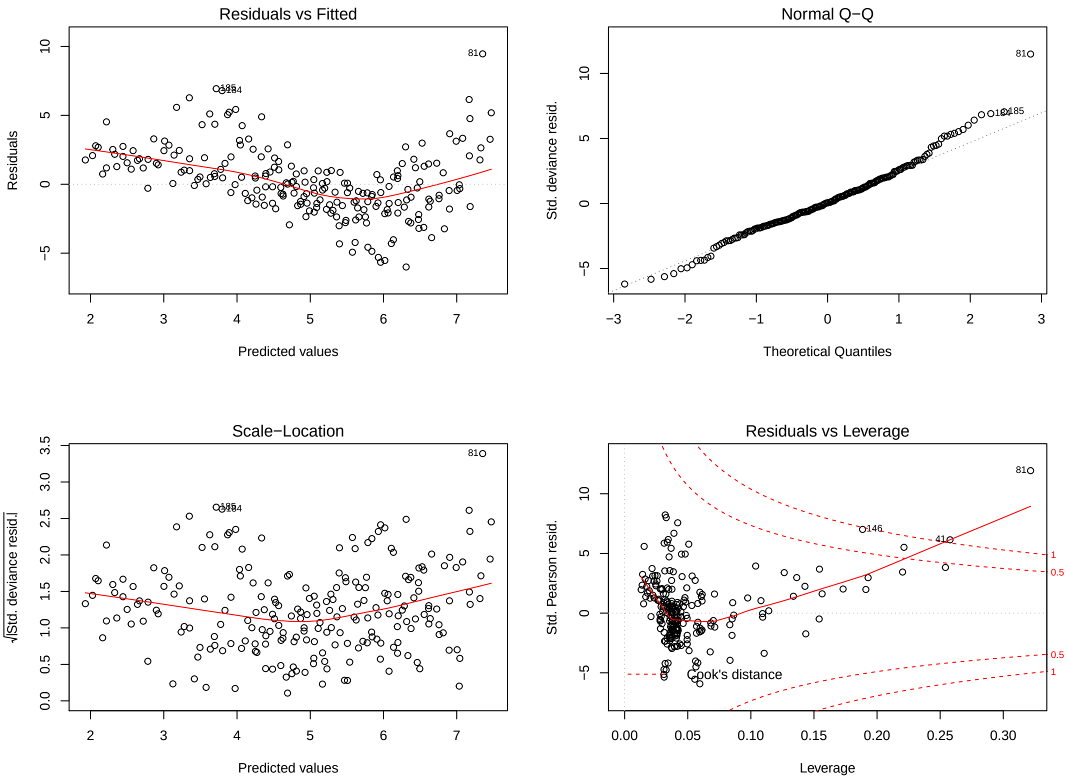
   (b) Standard theory on radioactive emissions suggests that counts of them should be Poisson distributed with a varying rate as the radioactivity decays. This suggests a Poisson model for the counts with a log link. Comment on this in the light of the plot in (a). Fit the Poisson model with the log the mean having a linear dependence on time and including factors for both Material and Sample. Do diagnostic plots using `plot` for the model. Comment and re-fit, if needed, omitting some observations. Use stepwise AIC selection to see if some variables can be omitted. What does the final residual deviance indicate?

Solution:

```
modelfull <- glm(data = Radioactive, Counts ~ Time*Material*Sample,
family = "poisson")
summary(modelfull)

##
## Call:
## glm(formula = Counts ~ Time * Material * Sample, family = "poisson",
##     data = Radioactive)
##
## Deviance Residuals:
##              Min                1Q            Median                3Q
## -6.0044396886934  -1.3666725699640   0.0525317862601   1.6273477341733
##              Max
##  9.4594506366135
##
## Coefficients:
##                                    Estimate         Std. Error    z value
## (Intercept)                   7.455633328488147  0.012123020614535  614.99799
## Time                         -0.276964951310045  0.002566950735178 -107.89648
## MaterialAl                   -0.101620232513983  0.018783805444139   -5.40999
## MaterialCu                   -0.286062569562786  0.017090111983412  -16.73848
## SampleTwo                     0.015857542750063  0.017137267861923    0.92533
## Time:MaterialAl               0.102834354682576  0.003285350079692   31.30088
## Time:MaterialCu               0.188418935644620  0.002724350592823   69.16105
## Time:SampleTwo               -0.005514225110704  0.003659851170003   -1.50668
## MaterialAl:SampleTwo         -0.842550779214367  0.028665555861010  -29.39245
## MaterialCu:SampleTwo         -0.570093309266234  0.025464359731614  -22.38789
## Time:MaterialAl:SampleTwo -0.046016316809879  0.005294003139965   -8.69216
## Time:MaterialCu:SampleTwo -0.025538922499841  0.004189479087120   -6.09597
##                              Pr(>|z|)
## (Intercept)               < 2.22e-16 ***
## Time                      < 2.22e-16 ***
## MaterialAl                6.3028e-08 ***
## MaterialCu                < 2.22e-16 ***
## SampleTwo                    0.35480
## Time:MaterialAl           < 2.22e-16 ***
## Time:MaterialCu           < 2.22e-16 ***
## Time:SampleTwo               0.13189
## MaterialAl:SampleTwo      < 2.22e-16 ***
## MaterialCu:SampleTwo      < 2.22e-16 ***
## Time:MaterialAl:SampleTwo < 2.22e-16 ***
## Time:MaterialCu:SampleTwo 1.0878e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 80585.6018432677  on 225  degrees of freedom
## Residual deviance:  1376.1253954491  on 214  degrees of freedom
## AIC: 2954.0506727316
##
```

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

```
par(mfrow=c(2,2))
plot(modelfull)
```

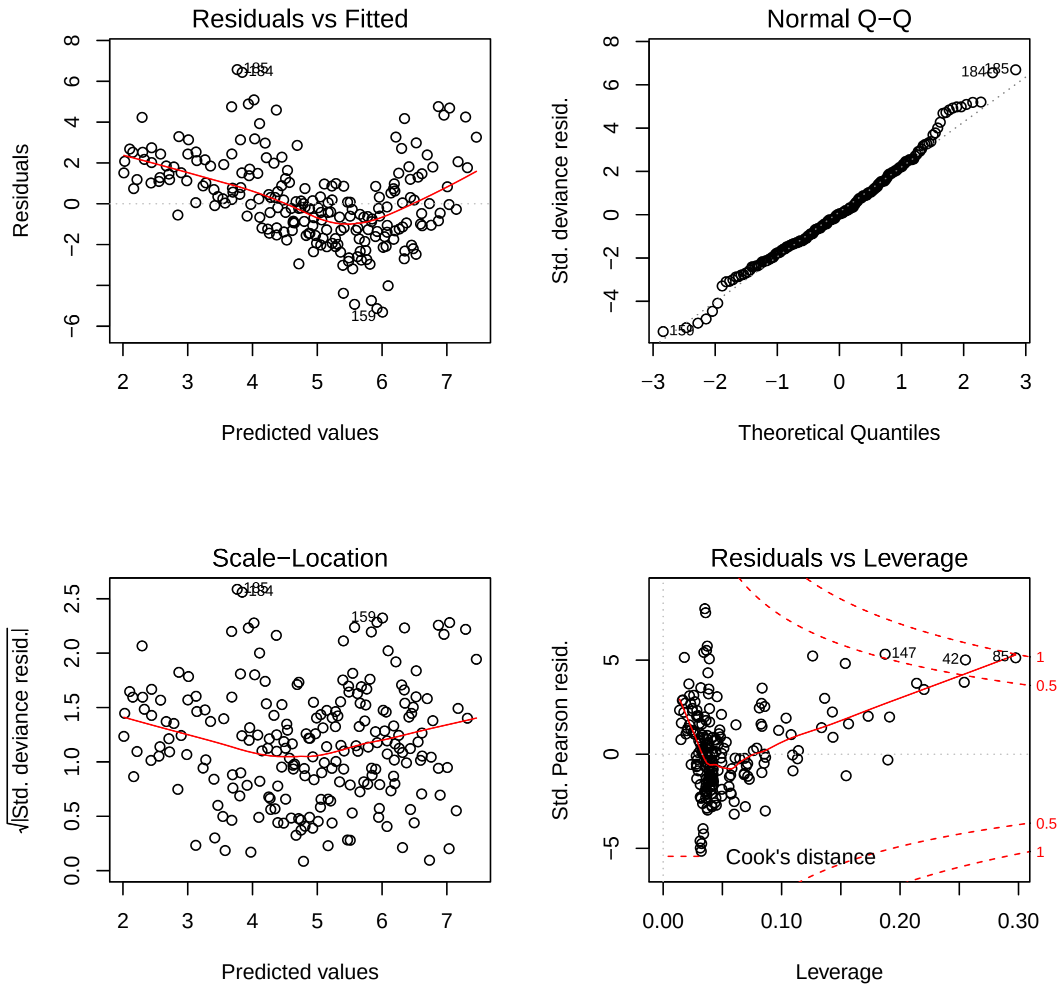


The diagnostic plots show heavily influential outliers which are the counts at the beginning of sample collection. The observations which need to be removed to keep Cook's distance under 1 are numbered 81 to 84, 41 and 146 in the original file.
All of these have been removed (but also accept just removing 81, 41 and 146).
The rerun analysis on plts are:

```
RadioMinus <- Radioactive[-c(81,82,83,84,146,41), ]
modelminus <- glm(data = RadioMinus, Counts ~ Time*Material*Sample,
family = "poisson")
summary(modelminus)

##
## Call:
## glm(formula = Counts ~ Time * Material * Sample, family = "poisson",
##     data = RadioMinus)
##
## Deviance Residuals:
##              Min                 1Q             Median                 3Q
## -5.3097634426731  -1.2480285097480   0.0386464970598   1.4755284983337
##              Max
##  6.5740773466997
##
## Coefficients:
##                                Estimate         Std. Error    z value
## (Intercept)                 7.455633328488158  0.012123020614535  614.99799
## Time                       -0.276964951310046  0.002566950735178 -107.89648
## MaterialAl                 -0.576128791074049  0.034958572274294  -16.48033
```

```
## MaterialCu                -0.327254065450264  0.018165082222128  -18.01556
## SampleTwo                 -0.028462004098048  0.018681555604919   -1.52354
## Time:MaterialAl            0.143681177591306  0.003953247034709   36.34510
## Time:MaterialCu            0.190694226216676  0.002742341707764   69.53700
## Time:SampleTwo             0.001067336709291  0.003796775648721    0.28112
## MaterialAl:SampleTwo      -0.323722673806196  0.041788743596366   -7.74665
## MaterialCu:SampleTwo      -0.484582266530651  0.027233186943884  -17.79381
## Time:MaterialAl:SampleTwo -0.093444701538603  0.005820833355922  -16.05349
## Time:MaterialCu:SampleTwo -0.034395774891891  0.004321004703897   -7.96013
##                             Pr(>|z|)
## (Intercept)               < 2.22e-16 ***
## Time                      < 2.22e-16 ***
## MaterialAl                < 2.22e-16 ***
## MaterialCu                < 2.22e-16 ***
## SampleTwo                    0.12762
## Time:MaterialAl           < 2.22e-16 ***
## Time:MaterialCu           < 2.22e-16 ***
## Time:SampleTwo               0.77862
## MaterialAl:SampleTwo      9.4350e-15 ***
## MaterialCu:SampleTwo      < 2.22e-16 ***
## Time:MaterialAl:SampleTwo < 2.22e-16 ***
## Time:MaterialCu:SampleTwo 1.7185e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 65189.4165245295  on 219  degrees of freedom
## Residual deviance:   952.0770390815  on 208  degrees of freedom
## AIC: 2475.4574225015
##
## Number of Fisher Scoring iterations: 4
```

```
par(mfrow=c(2,2))
plot(modelminus)
```




```
modelminusfinal <- step(modelminus)

## Start:  AIC=2475.46
## Counts ~ Time * Material * Sample
##
##                          Df        Deviance             AIC
## <none>                        952.0770390815 2475.4574225014
## - Time:Material:Sample    2 1212.4834734753 2731.8638568953

summary(modelminusfinal)

##
## Call:
## glm(formula = Counts ~ Time * Material * Sample, family = "poisson",
##     data = RadioMinus)
##
## Deviance Residuals:
##              Min                1Q            Median                3Q
## -5.3097634426731  -1.2480285097480   0.0386464970598   1.4755284983337
##              Max
##  6.5740773466997
##
## Coefficients:
##                                        Estimate         Std. Error    z value
```

```
## (Intercept)                  7.455633328488158  0.012123020614535  614.99799
## Time                        -0.276964951310046  0.002566950735178 -107.89648
## MaterialAl                  -0.576128791074049  0.034958572274294  -16.48033
## MaterialCu                  -0.327254065450264  0.018165082222128  -18.01556
## SampleTwo                   -0.028462004098048  0.018681555604919   -1.52354
## Time:MaterialAl              0.143681177591306  0.003953247034709   36.34510
## Time:MaterialCu              0.190694226216676  0.002742341707764   69.53700
## Time:SampleTwo               0.001067336709291  0.003796775648721    0.28112
## MaterialAl:SampleTwo        -0.323722673806196  0.041788743596366   -7.74665
## MaterialCu:SampleTwo        -0.484582266530651  0.027233186943884  -17.79381
## Time:MaterialAl:SampleTwo -0.093444701538603  0.005820833355922  -16.05349
## Time:MaterialCu:SampleTwo -0.034395774891891  0.004321004703897   -7.96013
##                              Pr(>|z|)
## (Intercept)                 < 2.22e-16 ***
## Time                        < 2.22e-16 ***
## MaterialAl                  < 2.22e-16 ***
## MaterialCu                  < 2.22e-16 ***
## SampleTwo                      0.12762
## Time:MaterialAl             < 2.22e-16 ***
## Time:MaterialCu             < 2.22e-16 ***
## Time:SampleTwo                 0.77862
## MaterialAl:SampleTwo        9.4350e-15 ***
## MaterialCu:SampleTwo        < 2.22e-16 ***
## Time:MaterialAl:SampleTwo < 2.22e-16 ***
## Time:MaterialCu:SampleTwo 1.7185e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 65189.4165245295  on 219  degrees of freedom
## Residual deviance:   952.0770390815  on 208  degrees of freedom
## AIC: 2475.4574225015
##
## Number of Fisher Scoring iterations: 4
```

```
anova(modelminus,modelminusfinal)
```

```
## Analysis of Deviance Table
##
## Model 1: Counts ~ Time * Material * Sample
## Model 2: Counts ~ Time * Material * Sample
##   Resid. Df     Resid. Dev Df Deviance
## 1       208 952.0770390815
## 2       208 952.0770390815  0        0
```

```
pchisq(modelminusfinal$deviance,df=modelminusfinal$df.residual, lower=FALSE)
```

```
## [1] 1.465684513333956e-95
```

No variables can be omitted. The final model has a deviance of 952.1 which is extremely unusual for a chi-square and indicates possible model inadequacy.

[2]

(c) Try re-fitting the model including TimeSquare as an extra variable to Time but still including Material and Sample. Carry out the additional steps in part (b). Comment on the contrast and comparison with (b).

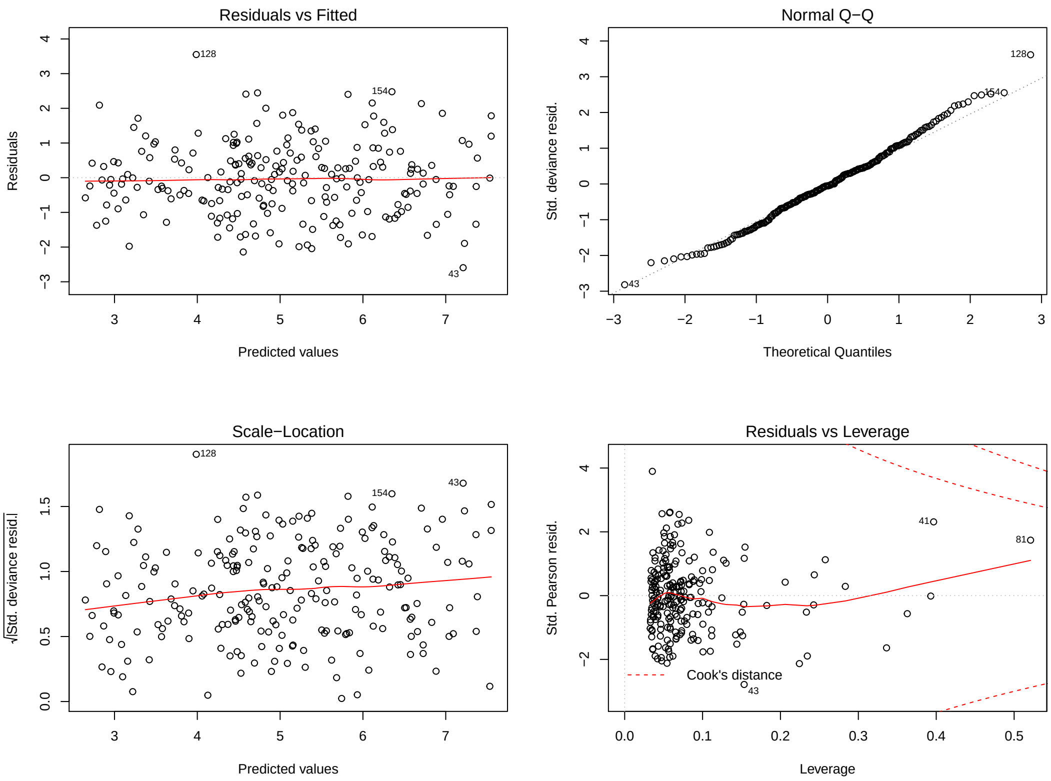
Solution:

```r
modelfull2 <- glm(data = Radioactive, Counts ~ (Time+TimeSquare)*Material*Sample,
family = "poisson")
summary(modelfull2)
```

```
##
## Call:
## glm(formula = Counts ~ (Time + TimeSquare) * Material * Sample,
##     family = "poisson", data = Radioactive)
##
## Deviance Residuals:
##              Min                1Q            Median                3Q
## -2.5939824345665  -0.6981698106619  -0.0515212710755   0.6038569871469
##              Max
##  3.5504964936046
##
## Coefficients:
##                                          Estimate          Std. Error
## (Intercept)                       7.5323357469219019  0.0145061814077899
## Time                             -0.3300310155715374  0.0063181670504271
## TimeSquare                        0.0042652871833848  0.0004542019938688
## MaterialAl                        0.0165231612457993  0.0220207291924353
## MaterialCu                       -0.1671235980349859  0.0205719381934672
## SampleTwo                         0.0176353772990381  0.0204724932413225
## Time:MaterialAl                   0.0562363112303348  0.0082648082517706
## Time:MaterialCu                   0.1898242642186684  0.0068240976861650
## TimeSquare:MaterialAl             0.0014092047975691  0.0005289369150715
## TimeSquare:MaterialCu            -0.0025971318520336  0.0004605635011048
## Time:SampleTwo                   -0.0075411255565626  0.0089649393944596
## TimeSquare:SampleTwo              0.0002057640394388  0.0006465181853376
## MaterialAl:SampleTwo             -0.9128635027220288  0.0340800370730413
## MaterialCu:SampleTwo             -0.7233619798216151  0.0315439704171485
## Time:MaterialAl:SampleTwo        -0.0209656037974567  0.0134592777228336
## Time:MaterialCu:SampleTwo         0.0084199360775660  0.0110461739735809
## TimeSquare:MaterialAl:SampleTwo -0.0003599202954664  0.0008938654021975
## TimeSquare:MaterialCu:SampleTwo -0.0006730240220906  0.0007353004376161
##                                    z value   Pr(>|z|)
## (Intercept)                      519.25007 < 2.22e-16 ***
## Time                             -52.23525 < 2.22e-16 ***
## TimeSquare                         9.39073 < 2.22e-16 ***
## MaterialAl                         0.75035  0.4530465
## MaterialCu                        -8.12386 4.5158e-16 ***
## SampleTwo                          0.86142  0.3890078
## Time:MaterialAl                    6.80431 1.0154e-11 ***
## Time:MaterialCu                   27.81676 < 2.22e-16 ***
## TimeSquare:MaterialAl              2.66422  0.0077167 **
## TimeSquare:MaterialCu             -5.63903 1.7101e-08 ***
## Time:SampleTwo                    -0.84118  0.4002472
## TimeSquare:SampleTwo               0.31826  0.7502840
## MaterialAl:SampleTwo             -26.78587 < 2.22e-16 ***
## MaterialCu:SampleTwo             -22.93186 < 2.22e-16 ***
## Time:MaterialAl:SampleTwo         -1.55771  0.1193028
## Time:MaterialCu:SampleTwo          0.76225  0.4459114
## TimeSquare:MaterialAl:SampleTwo   -0.40266  0.6872013
## TimeSquare:MaterialCu:SampleTwo   -0.91530  0.3600316
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 80585.60184326771  on 225  degrees of freedom
## Residual deviance:   250.55977419795  on 208  degrees of freedom
## AIC: 1840.4850514805
##
## Number of Fisher Scoring iterations: 4
```

```
par(mfrow=c(2,2))
plot(modelfull2)
```




```
modelfinal2 <- step(modelfull2)
```

```
## Start:  AIC=1840.49
## Counts ~ (Time + TimeSquare) * Material * Sample
##
##                                 Df         Deviance               AIC
## - TimeSquare:Material:Sample     2 251.45186979674 1837.3771470793
## <none>                             250.55977419795 1840.4850514805
## - Time:Material:Sample           2 256.61819555912 1842.5434728417
##
## Step:  AIC=1837.38
## Counts ~ Time + TimeSquare + Material + Sample + Time:Material +
##     TimeSquare:Material + Time:Sample + TimeSquare:Sample + Material:Sample +
##     Time:Material:Sample
##
##                           Df         Deviance               AIC
## - TimeSquare:Sample        1 252.50935739784 1836.4346346804
## <none>                       251.45186979674 1837.3771470793
## - Time:Material:Sample     2 287.96294513063 1869.8882224132
## - TimeSquare:Material      2 525.65729145982 2107.5825687423
```

```
##
## Step:  AIC=1836.43
## Counts ~ Time + TimeSquare + Material + Sample + Time:Material +
##     TimeSquare:Material + Time:Sample + Material:Sample + Time:Material:Sample
##
##                         Df        Deviance             AIC
## <none>                     252.50935739784 1836.4346346804
## - Time:Material:Sample  2 288.85086205080 1868.7761393333
## - TimeSquare:Material   2 538.66132824805 2118.5866055306
```

```
summary(modelfinal2)
```

```
##
## Call:
## glm(formula = Counts ~ Time + TimeSquare + Material + Sample +
##     Time:Material + TimeSquare:Material + Time:Sample + Material:Sample +
##     Time:Material:Sample, family = "poisson", data = Radioactive)
##
## Deviance Residuals:
##              Min                1Q            Median                3Q
## -2.5703433142756  -0.7202328515142  -0.0642142036212   0.6205451515729
##              Max
##  3.5635716762162
##
## Coefficients:
##                                      Estimate          Std. Error
## (Intercept)                 7.5342205890020715  0.0132350878180143
## Time                       -0.3313336210670742  0.0048114049288053
## TimeSquare                  0.0043665965102361  0.0003232399650526
## MaterialAl                  0.0135888225649133  0.0207885638722534
## MaterialCu                 -0.1716491712719468  0.0196051075534495
## SampleTwo                   0.0138512984380801  0.0166658604401971
## Time:MaterialAl             0.0580895856762715  0.0068323997267187
## Time:MaterialCu             0.1918406634839754  0.0054329895285709
## TimeSquare:MaterialAl       0.0012781007257425  0.0004047216606247
## TimeSquare:MaterialCu      -0.0027207465590010  0.0003317177717333
## Time:SampleTwo             -0.0049075778879757  0.0034490987335881
## MaterialAl:SampleTwo       -0.9051102351489030  0.0277005593959261
## MaterialCu:SampleTwo       -0.7006113901977439  0.0255021274804134
## Time:MaterialAl:SampleTwo -0.0259378215949388  0.0050057320624506
## Time:MaterialCu:SampleTwo -0.0022756988878154  0.0041334077445747
##                             z value   Pr(>|z|)
## (Intercept)               569.26110 < 2.22e-16 ***
## Time                      -68.86421 < 2.22e-16 ***
## TimeSquare                 13.50884 < 2.22e-16 ***
## MaterialAl                  0.65367  0.5133256
## MaterialCu                 -8.75533 < 2.22e-16 ***
## SampleTwo                   0.83112  0.4059069
## Time:MaterialAl             8.50208 < 2.22e-16 ***
## Time:MaterialCu            35.31033 < 2.22e-16 ***
## TimeSquare:MaterialAl       3.15797  0.0015887 **
## TimeSquare:MaterialCu      -8.20199 2.3644e-16 ***
## Time:SampleTwo             -1.42286  0.1547773
## MaterialAl:SampleTwo      -32.67480 < 2.22e-16 ***
## MaterialCu:SampleTwo      -27.47266 < 2.22e-16 ***
## Time:MaterialAl:SampleTwo  -5.18162 2.1996e-07 ***
## Time:MaterialCu:SampleTwo  -0.55056  0.5819337
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 80585.60184326771  on 225  degrees of freedom
## Residual deviance:   252.50935739784  on 211  degrees of freedom
## AIC: 1836.4346346804
##
## Number of Fisher Scoring iterations: 4
```

```
anova(modelfull2,modelfinal2)
```

```
## Analysis of Deviance Table
##
## Model 1: Counts ~ (Time + TimeSquare) * Material * Sample
## Model 2: Counts ~ Time + TimeSquare + Material + Sample + Time:Material +
##     TimeSquare:Material + Time:Sample + Material:Sample + Time:Material:Sample
##   Resid. Df      Resid. Dev Df         Deviance
## 1       208 250.55977419795
## 2       211 252.50935739784 -3 -1.9495831998937
```

```
pchisq(modelfinal2$deviance,df=modelfinal2$df.residual, lower=FALSE)
```

```
## [1] 0.02658778522608837
```

The fit is much better with TimeSquare included, with no very influential outliers. The residual deviance is much closer to the degrees of freedom although it is nominally significant.

[2]

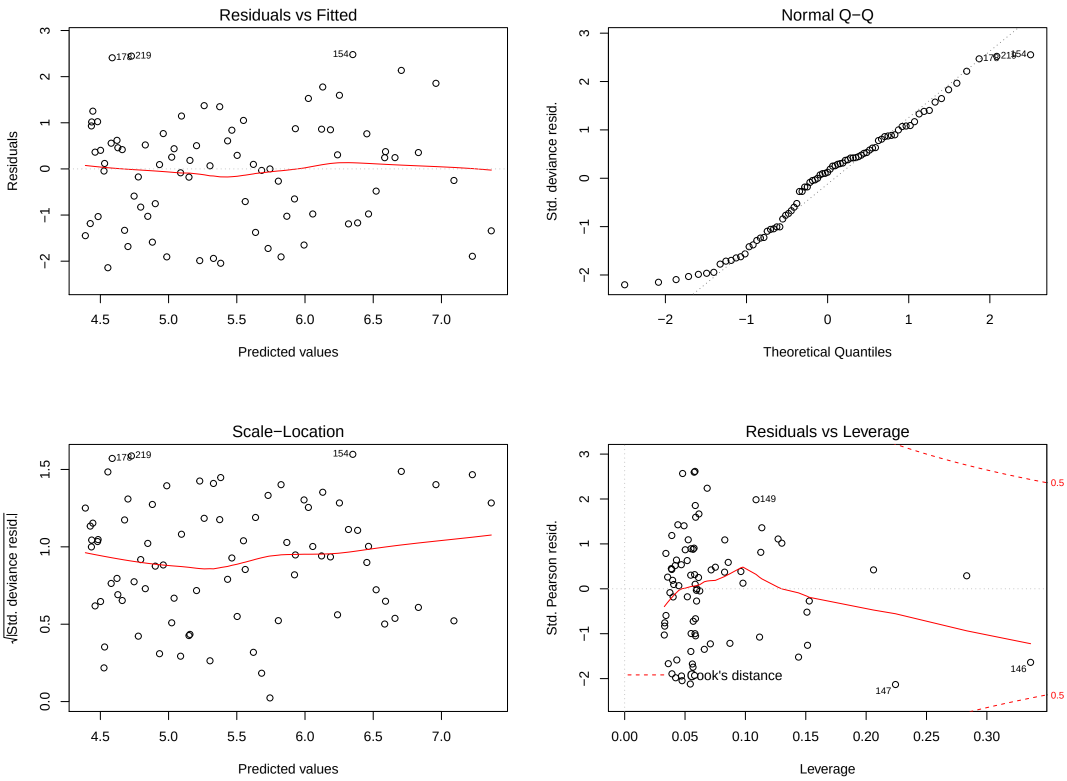
(d) An alternative approach would be to run the same analyses just using the data for one material at a time. Do this. Comment on the comparison with (b) and (c).

```
RadioCu <- Radioactive[Radioactive$Material=="Cu",]
modelfull2Cu <- glm(data = RadioCu, Counts ~ (Time+TimeSquare)*Sample,
family = "poisson")
summary(modelfull2Cu)
```

```
##
## Call:
## glm(formula = Counts ~ (Time + TimeSquare) * Sample, family = "poisson",
##     data = RadioCu)
##
## Deviance Residuals:
##               Min                1Q            Median                3Q
## -2.1417633294710  -1.0258012929768   0.1182330092738   0.7659038704425
##               Max
##  2.4783657642320
##
## Coefficients:
##                                Estimate          Std. Error   z value
## (Intercept)           7.365212148887e+00 1.458682083252e-02 504.92237
## Time                 -1.402067513529e-01 2.578579910187e-03 -54.37363
## TimeSquare            1.668155331351e-03 7.628425339189e-05  21.86762
## SampleTwo            -7.057266025226e-01 2.399789761962e-02 -29.40785
## Time:SampleTwo        8.788105210036e-04 6.453512307905e-03   0.13618
## TimeSquare:SampleTwo -4.672599826519e-04 3.502584325697e-04  -1.33404
##                      Pr(>|z|)
## (Intercept)           < 2e-16 ***
## Time                  < 2e-16 ***
## TimeSquare            < 2e-16 ***
```

```
## SampleTwo                < 2e-16 ***
## Time:SampleTwo           0.89168
## TimeSquare:SampleTwo     0.18219
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 18202.06221164008  on 80  degrees of freedom
## Residual deviance:   110.03743654979  on 75  degrees of freedom
## AIC: 712.55592083833
##
## Number of Fisher Scoring iterations: 4
```

```
par(mfrow=c(2,2))
plot(modelfull2Cu)
```




```
modelCu <- step(modelfull2Cu)
```

```
## Start:  AIC=712.5599999999999
## Counts ~ (Time + TimeSquare) * Sample
##
##                     Df        Deviance             AIC
## - Time:Sample        1 110.05598283896 710.57446712750
## - TimeSquare:Sample  1 111.82329397605 712.34177826460
## <none>                 110.03743654979 712.55592083833
##
## Step:  AIC=710.5700000000001
## Counts ~ Time + TimeSquare + Sample + TimeSquare:Sample
##
##                     Df        Deviance             AIC
```

```
## <none>                   110.0559828390  710.5744671275
## - TimeSquare:Sample  1  121.8272694175  720.3457537060
## - Time               1 3446.9311495449 4045.4496338334
```

```
summary(modelCu)
```

```
##
## Call:
## glm(formula = Counts ~ Time + TimeSquare + Sample + TimeSquare:Sample,
##     family = "poisson", data = RadioCu)
##
## Deviance Residuals:
##              Min                1Q            Median                3Q
## -2.1393533243836  -1.0060594149061   0.1215788825808   0.7598958455108
##              Max
##  2.4719038169136
##
## Coefficients:
##                               Estimate           Std. Error   z value
## (Intercept)          7.364606596899e+00  1.389505789400e-02 530.01626
## Time                -1.400664149973e-01  2.363787427098e-03 -59.25508
## TimeSquare           1.664223045628e-03  7.061720768664e-05  23.56682
## SampleTwo           -7.032145257598e-01  1.534591890615e-02 -45.82420
## TimeSquare:SampleTwo -4.226656522997e-04  1.241424380039e-04  -3.40468
##                       Pr(>|z|)
## (Intercept)          < 2.22e-16 ***
## Time                 < 2.22e-16 ***
## TimeSquare           < 2.22e-16 ***
## SampleTwo            < 2.22e-16 ***
## TimeSquare:SampleTwo 0.00066241 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 18202.06221164008  on 80  degrees of freedom
## Residual deviance:   110.05598283896  on 76  degrees of freedom
## AIC: 710.5744671275
##
## Number of Fisher Scoring iterations: 4
```

```
anova(modelfull2Cu,modelCu)
```

```
## Analysis of Deviance Table
##
## Model 1: Counts ~ (Time + TimeSquare) * Sample
## Model 2: Counts ~ Time + TimeSquare + Sample + TimeSquare:Sample
##   Resid. Df      Resid. Dev Df           Deviance
## 1        75 110.03743654979
## 2        76 110.05598283896 -1 -0.018546289170061
```

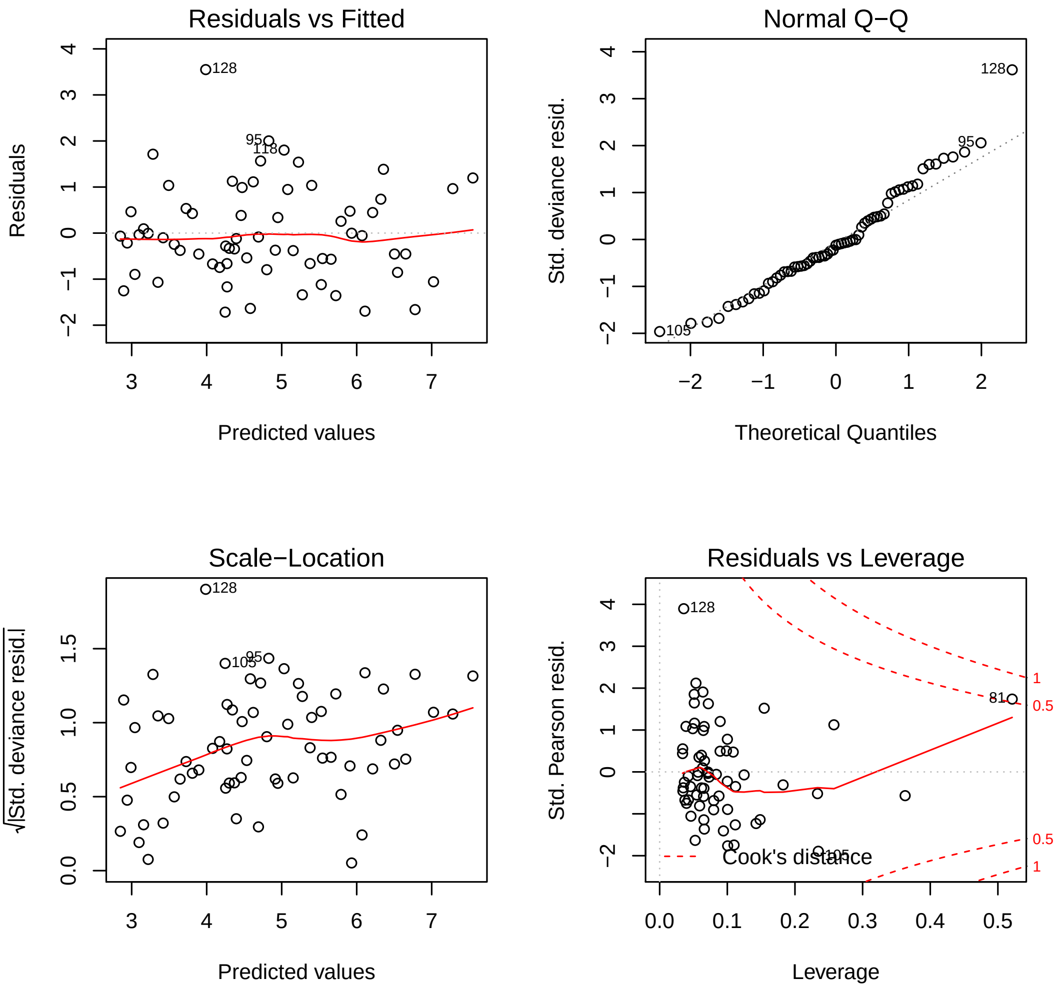
```
pchisq(modelCu$deviance,modelCu$df.residual)
```

```
## [1] 0.9935173649796516
```

```
RadioAl <- Radioactive[Radioactive$Material=="Al",]
modelfull2Al <- glm(data = RadioAl, Counts ~ (Time+TimeSquare)*Sample,
family = "poisson")
summary(modelfull2Al)
```

```
##
## Call:
## glm(formula = Counts ~ (Time + TimeSquare) * Sample, family = "poisson",
##     data = RadioAl)
##
## Deviance Residuals:
##              Min                 1Q            Median                 3Q
## -1.7165690719926  -0.6625926714441  -0.1185447715317   0.5343025289861
##              Max
##  3.5504964936045
##
## Coefficients:
##                                 Estimate          Std. Error   z value
## (Intercept)           7.5488589081677064  0.0165675349744874 455.64165
## Time                 -0.2737947043412010  0.0053280221997880 -51.38768
## TimeSquare            0.0056744919809538  0.0002710623708502  20.93427
## SampleTwo            -0.8952281254229862  0.0272456592392967 -32.85764
## Time:SampleTwo       -0.0285067293540204  0.0100390247770404  -2.83959
## TimeSquare:SampleTwo -0.0001541562560275  0.0006172597453856  -0.24974
##                        Pr(>|z|)
## (Intercept)          < 2.22e-16 ***
## Time                 < 2.22e-16 ***
## TimeSquare           < 2.22e-16 ***
## SampleTwo            < 2.22e-16 ***
## Time:SampleTwo        0.0045171 **
## TimeSquare:SampleTwo  0.8027862
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 21337.425532206551  on 64  degrees of freedom
## Residual deviance:    68.222210852186  on 59  degrees of freedom
## AIC: 510.22909249968
##
## Number of Fisher Scoring iterations: 4
```

```
plot(modelfull2Al)
```




```
modelAl <- step(modelfull2Al)

## Start:  AIC=510.23
## Counts ~ (Time + TimeSquare) * Sample
##
##                        Df        Deviance            AIC
## - TimeSquare:Sample  1 68.284650139319 508.29153178682
## <none>                 68.222210852186 510.22909249968
## - Time:Sample        1 76.267740195921 516.27462184342
##
## Step:  AIC=508.29
## Counts ~ Time + TimeSquare + Sample + Time:Sample
##
##                  Df        Deviance             AIC
## <none>              68.28465013932  508.29153178682
## - Time:Sample   1 143.55272336671  581.55960501420
## - TimeSquare    1 562.17838183544 1000.18526348294

summary(modelAl)

##
## Call:
## glm(formula = Counts ~ Time + TimeSquare + Sample + Time:Sample,
##     family = "poisson", data = RadioAl)
```

```
##
## Deviance Residuals:
##              Min               1Q           Median               3Q
## -1.7203326859399  -0.6539579070119  -0.1311928805266   0.5336869745698
##              Max
##  3.5635716762162
##
## Coefficients:
##                          Estimate          Std. Error   z value
## (Intercept)     7.5478094115669991  0.0160311209314948 470.82231
## Time           -0.2732440353908038  0.0048509863570963 -56.32752
## TimeSquare      0.0056446972359786  0.0002435478342577  23.17696
## SampleTwo      -0.8912589367108276  0.0221262307371837 -40.28065
## Time:SampleTwo -0.0308453994829138  0.0036278190978888  -8.50246
##                  Pr(>|z|)
## (Intercept)    < 2.22e-16 ***
## Time           < 2.22e-16 ***
## TimeSquare     < 2.22e-16 ***
## SampleTwo      < 2.22e-16 ***
## Time:SampleTwo < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 21337.425532206551  on 64  degrees of freedom
## Residual deviance:    68.284650139319  on 60  degrees of freedom
## AIC: 508.29153178682
##
## Number of Fisher Scoring iterations: 4
```

```
anova(modelfull2Al,modelAl)
```

```
## Analysis of Deviance Table
##
## Model 1: Counts ~ (Time + TimeSquare) * Sample
## Model 2: Counts ~ Time + TimeSquare + Sample + Time:Sample
##   Resid. Df       Resid. Dev Df           Deviance
## 1        59 68.222210852186
## 2        60 68.284650139319 -1 -0.062439287132705
```

```
pchisq(modelAl$deviance,modelAl$df.residual)
```

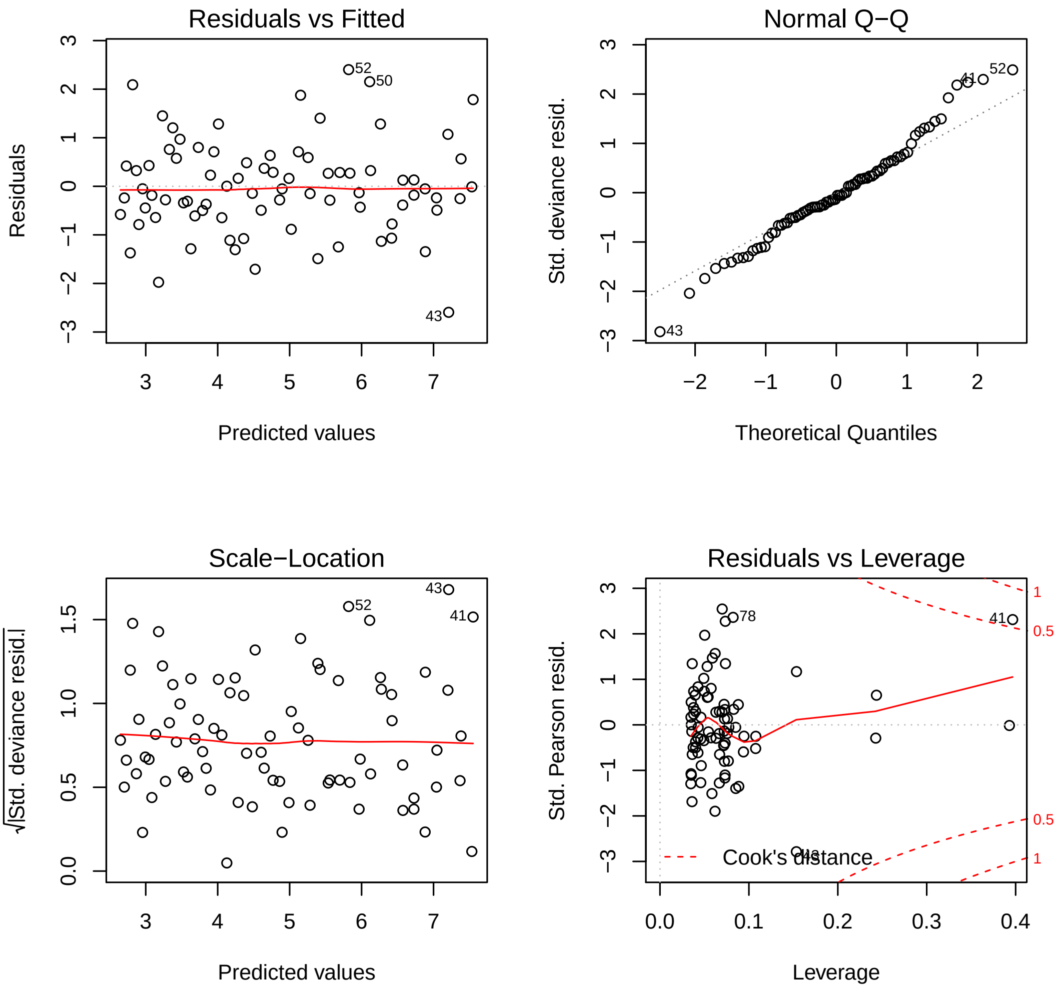
```
## [1] 0.7835590390146442
```

```
RadioAg <- Radioactive[Radioactive$Material=="Ag",]
modelfull2Ag <- glm(data = RadioAg, Counts ~ (Time+TimeSquare)*Sample,
family = "poisson")
summary(modelfull2Ag)
```

```
##
## Call:
## glm(formula = Counts ~ (Time + TimeSquare) * Sample, family = "poisson",
##     data = RadioAg)
##
## Deviance Residuals:
##               Min                1Q            Median
## -2.59398243456703  -0.51823985699471  -0.09200015020171
##                3Q               Max
```

```
##  0.50510117792126   2.40252416234687
##
## Coefficients:
##                                Estimate          Std. Error   z value
## (Intercept)           7.5323357469219463  0.0145061814077901 519.25007
## Time                 -0.3300310155715474  0.0063181670504271 -52.23525
## TimeSquare            0.0042652871833853  0.0004542019938688   9.39073
## SampleTwo             0.0176353772990058  0.0204724932413227   0.86142
## Time:SampleTwo       -0.0075411255565521  0.0089649393944596  -0.84118
## TimeSquare:SampleTwo  0.0002057640394382  0.0006465181853376   0.31826
##                      Pr(>|z|)
## (Intercept)           < 2e-16 ***
## Time                  < 2e-16 ***
## TimeSquare            < 2e-16 ***
## SampleTwo             0.38901
## Time:SampleTwo        0.40025
## TimeSquare:SampleTwo  0.75028
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 39933.420938691626  on 79  degrees of freedom
## Residual deviance:    72.300126795976  on 74  degrees of freedom
## AIC: 617.70003814246
##
## Number of Fisher Scoring iterations: 4
```

```
plot(modelfull2Ag)
```




```
modelAg <- step(modelfull2Ag)

## Start:  AIC=617.7
## Counts ~ (Time + TimeSquare) * Sample
##
##                        Df        Deviance             AIC
## - TimeSquare:Sample     1 72.401413282470 615.80132462895
## - Time:Sample           1 73.007694146566 616.40760549305
## <none>                    72.300126795976 617.70003814246
##
## Step:  AIC=615.8
## Counts ~ Time + TimeSquare + Sample + Time:Sample
##
##                  Df         Deviance             AIC
## <none>                72.401413282470 615.80132462896
## - Time:Sample     1  74.426445209067 615.82635655555
## - TimeSquare     1 242.797097417761 784.19700876425

summary(modelAg)

##
## Call:
## glm(formula = Counts ~ Time + TimeSquare + Sample + Time:Sample,
##     family = "poisson", data = RadioAg)
```

```
##
## Deviance Residuals:
##               Min                 1Q             Median                 3Q
## -2.5703433142758  -0.5068229497676  -0.1105699060263   0.5000769879591
##               Max
##  2.3600052905389
##
## Coefficients:
##                           Estimate          Std. Error   z value Pr(>|z|)
## (Intercept)     7.5342205890020777  0.0132350878180140 569.26110  < 2e-16
## Time           -0.3313336210670748  0.0048114049288052 -68.86421  < 2e-16
## TimeSquare      0.0043665965102361  0.0003232399650526  13.50884  < 2e-16
## SampleTwo       0.0138512984380779  0.0166658604401969   0.83112  0.40591
## Time:SampleTwo -0.0049075778879754  0.0034490987335880  -1.42286  0.15478
##
## (Intercept)    ***
## Time           ***
## TimeSquare     ***
## SampleTwo
## Time:SampleTwo
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 39933.42093869163  on 79  degrees of freedom
## Residual deviance:    72.40141328247  on 75  degrees of freedom
## AIC: 615.80132462896
##
## Number of Fisher Scoring iterations: 4
```

```
anova(modelfull2Ag,modelAg)
```

```
## Analysis of Deviance Table
##
## Model 1: Counts ~ (Time + TimeSquare) * Sample
## Model 2: Counts ~ Time + TimeSquare + Sample + Time:Sample
##   Resid. Df      Resid. Dev Df         Deviance
## 1        74 72.300126795976
## 2        75 72.401413282470 -1 -0.1012864864939
```

```
pchisq(modelAg$deviance,modelAg$df.residual)
```

```
## [1] 0.4364212737244908
```

The importance of TimeSquare in the previous analysis indicates that this should be included in the separate analyses for the separate materials. The degrees of freedom are now less. The selected models are similar only varying in whether Time or TimeSquare needs a sample adjustment. The p-values for the selected models are all not-significant - perhaps to do with degrees of freedom, perhaps with the alternate selected data. Probably the separate analyses are a better summary, as the original picture appears to indicate.

[2]