21

Bernat Chiva

12/8/2020

```r
library(car)
```

```
## Loading required package: carData
```

```r
library(HH)
```

```
## Loading required package: lattice
## Loading required package: grid
## Loading required package: latticeExtra
## Loading required package: multcomp
## Loading required package: mvtnorm
## Loading required package: survival
## Loading required package: TH.data
## Loading required package: MASS
##
## Attaching package: 'TH.data'
## The following object is masked from 'package:MASS':
##
##     geyser
## Loading required package: gridExtra
##
## Attaching package: 'HH'
## The following objects are masked from 'package:car':
##
##     logit, vif
```

```r
library(RcmdrMisc)
```

```
## Loading required package: sandwich
```

```r
library(doBy)
library(emmeans)
```

```
##
## Attaching package: 'emmeans'

## The following object is masked from 'package:HH':
##
##     as.glht

library(tables)
library(ggplot2)

##
## Attaching package: 'ggplot2'

## The following object is masked from 'package:latticeExtra':
##
##     layer

library(sandwich)
library(survival)
library(estimability)
library(readr)
vitc <- read_delim("C:/Users/berna/OneDrive/Desktop/UPC/S1/5. Models
Lineals/datasets/vitc.csv",
    ";", escape_double = FALSE, locale = locale(decimal_mark = ","),
    trim_ws = TRUE)

##
## -- Column specification --------------------------------------------------
------
## cols(
##   treat = col_character(),
##   week = col_double(),
##   vitc = col_double()
## )

vitc$treat<-as.factor(vitc$treat)
summary(vitc)

##  treat       week            vitc
##  a:24   Min.   : 1.00   Min.   : 3.30
##  b:24   1st Qu.: 3.75   1st Qu.:17.18
##  c:24   Median : 6.50   Median :26.30
##         Mean   : 6.50   Mean   :27.22
##         3rd Qu.: 9.25   3rd Qu.:33.15
##         Max.   :12.00   Max.   :70.30
```

The model we are assuming is:

$$VitC = \alpha_i e^{-\beta_i * week}$$

Is the same as assuming:

$$log(VitC) = log(\alpha_i) - \beta_i * week$$

1. GAMMA

a) Define a generalized linear model with the "gamma" family, useful to check if the treatments loss Vitamin C at the same velocity, that is if $\beta 1 = \beta 2 = \beta 3$ or not, and also to see if the three values of αi are or not statistically equivalent.

```
#Model with both parameters interaction
mod1<-glm(vitc~week*treat, family=Gamma(link="log"), data=vitc)
summary(mod1)
```

```
##
## Call:
## glm(formula = vitc ~ week * treat, family = Gamma(link = "log"),
##     data = vitc)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.02691  -0.25019  -0.06136   0.18828   0.75558
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.84653    0.14949  25.731  < 2e-16 ***
## week        -0.15729    0.02031  -7.744 7.61e-11 ***
## treatb      -0.06931    0.21141  -0.328    0.744
## treatc       0.12314    0.21141   0.582    0.562
## week:treatb  0.12933    0.02872   4.502 2.80e-05 ***
## week:treatc  0.03066    0.02872   1.067    0.290
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 0.1179906)
##
##     Null deviance: 24.790  on 71  degrees of freedom
## Residual deviance:  8.217  on 66  degrees of freedom
## AIC: 514.38
##
## Number of Fisher Scoring iterations: 5
```

We check the intercept for the lines at week 0:

```
#Check intercept
(emm1<-emmeans(mod1,~treat|week, at=list(week=c(0))))
```

```
## week = 0:
##  treat emmean    SE  df asymp.LCL asymp.UCL
##  a       3.85 0.149 Inf      3.55      4.14
##  b       3.78 0.149 Inf      3.48      4.07
##  c       3.97 0.149 Inf      3.68      4.26
##
## Results are given on the log (not the response) scale.
## Confidence level used: 0.95
```

```
(pairs(emm1))
```

```
## week = 0:
##  contrast estimate    SE  df z.ratio p.value
##  a - b      0.0693 0.211 Inf  0.328  0.9425
##  a - c     -0.1231 0.211 Inf -0.582  0.8295
##  b - c     -0.1924 0.211 Inf -0.910  0.6338
##
## Results are given on the log (not the response) scale.
## P value adjustment: tukey method for comparing a family of 3 estimates
```

The intercept is significantly equal for all the lines We assume a new model with the same intercept and without treatment paramater, onlt week and the interaction

```
#Model without treatment
mod12<-glm(vitc~week+treat:week, family=Gamma(link="log"), data=vitc)
summary(mod12)
```

```
##
## Call:
## glm(formula = vitc ~ week + treat:week, family = Gamma(link = "log"),
##     data = vitc)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.02564  -0.24039  -0.03959   0.17580   0.73569
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.86737    0.08537  45.302  < 2e-16 ***
## week        -0.15981    0.01392 -11.481  < 2e-16 ***
## week:treatb  0.12122    0.01333   9.096 2.27e-13 ***
## week:treatc  0.04532    0.01333   3.401  0.00113 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 0.1154367)
##
##     Null deviance: 24.790  on 71  degrees of freedom
## Residual deviance:  8.323  on 68  degrees of freedom
## AIC: 511.32
##
## Number of Fisher Scoring iterations: 4
```

• Estimate αi. Are they statistically different?

To estimate

$$\alpha_1 = exp3.86737 = 47.8165$$

To estimate

$$\alpha_2 = exp3.78 = 43.816$$

To estimate

$$\alpha_3 = exp3.97 = 52.9845$$

These are the estimations of each alpha, although we say that all of them are the same, because the alphas are not significant different from each other.

• **Estimate βi. Are they statistically different?**

To estimate

$$\beta_1 = -0.15981$$

To estimate

$$\beta_2 = 0.12122$$

To estimate

$$\beta_3 = 0.04532$$

The betas are significant different from each other, so we assume different slope for each line.

2. LOGNORMAL

b) Define a LogNormal generalized linear model useful to check if the treatments loss Vitamin C at the same velocity, that is if β1 = β2 = β3 or not, and also to see if the three values of αi are or not statistically equivalent.

```
#Model with both parameters interaction
mod2<-glm(log(vitc)~week*treat, family=gaussian(link="identity"), data=vitc)
summary(mod2)
```

```
##
## Call:
## glm(formula = log(vitc) ~ week * treat, family = gaussian(link =
"identity"),
##     data = vitc)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.15293  -0.18979  -0.01522   0.24540   0.72179
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.82038    0.15788  24.199  < 2e-16 ***
## week        -0.16373    0.02145  -7.632 1.20e-10 ***
## treatb      -0.09785    0.22327  -0.438    0.663
## treatc       0.12472    0.22327   0.559    0.578
## week:treatb  0.13636    0.03034   4.495 2.88e-05 ***
## week:treatc  0.03282    0.03034   1.082    0.283
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.1316045)
##
##     Null deviance: 28.9792  on 71  degrees of freedom
## Residual deviance:  8.6859  on 66  degrees of freedom
## AIC: 66.05
##
## Number of Fisher Scoring iterations: 2
```

To check the intercepts of the three lines we will check the expected values at week 0:

```
(emm2<-emmeans(mod2,~treat|week, at=list(week=c(0))))
```

```
## week = 0:
##  treat emmean    SE  df asymp.LCL asymp.UCL
##  a       3.82 0.158 Inf      3.51      4.13
##  b       3.72 0.158 Inf      3.41      4.03
##  c       3.95 0.158 Inf      3.64      4.25
##
## Results are given on the log (not the response) scale.
## Confidence level used: 0.95
```

```
(pairs(emm2))
```

```
## week = 0:
##  contrast estimate    SE  df z.ratio p.value
##  a - b      0.0978 0.223 Inf  0.438  0.8996
##  a - c     -0.1247 0.223 Inf -0.559  0.8421
##  b - c     -0.2226 0.223 Inf -0.997  0.5789
##
## Results are given on the log (not the response) scale.
## P value adjustment: tukey method for comparing a family of 3 estimates
```

The alphas are not significantly different from each other and the treat parameter is not significant either.

We assume a new model without the treat parameter.

```
#Model with week and interaction
mod21<-glm(log(vitc)~week+treat:week, family=gaussian(link="identity"),
data=vitc)
summary(mod21)
```

```
##
## Call:
## glm(formula = log(vitc) ~ week + treat:week, family = gaussian(link =
"identity"),
##     data = vitc)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.15221  -0.19240   0.00464   0.23452   0.70470
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.82934    0.09048  42.324  < 2e-16 ***
## week        -0.16480    0.01475 -11.171  < 2e-16 ***
## week:treatb  0.12462    0.01412   8.823 7.04e-13 ***
## week:treatc  0.04778    0.01412   3.383  0.00119 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.1296662)
##
##     Null deviance: 28.9792  on 71  degrees of freedom
## Residual deviance:  8.8173  on 68  degrees of freedom
## AIC: 63.131
##
## Number of Fisher Scoring iterations: 2
```

• Estimate αi. Are they statistically different?

To estimate

$$\alpha_1 = exp3.82 = 45.6042$$

To estimate

$$\alpha_2 = exp3.72 = 41.2644$$

To estimate

$$\alpha_3 = exp3.95 = 51.9354$$

• Estimate βi. Are they statistically different?

To estimate

$$\beta_1 = -0.1648$$

To estimate

$$\beta_2 = 0.12462$$

To estimate

$$\beta_3 = 0.04778$$

3. NORMAL

c) Define a generalized linear model with the "normal" family, useful to check if the treatments loss Vitamin C at the same velocity, that is if β1 = β2 = β3 or not, and also to see if the three values of αi are or not statistically equivalent.

```
mod3<-glm(vitc~week*treat,family= gaussian(link="log"), data=vitc)
summary(mod3)
```

```
##
## Call:
## glm(formula = vitc ~ week * treat, family = gaussian(link = "log"),
##     data = vitc)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -15.328   -5.419   -0.466    5.604   34.472
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.76000    0.14868  25.290  < 2e-16 ***
## week        -0.13944    0.03259  -4.278 6.23e-05 ***
## treatb       0.03691    0.17986   0.205  0.83802
## treatc       0.33305    0.18548   1.796  0.07714 .
## week:treatb  0.10827    0.03593   3.013  0.00366 **
## week:treatc -0.01322    0.04139  -0.319  0.75036
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 86.15934)
##
##     Null deviance: 15538.6  on 71  degrees of freedom
## Residual deviance:  5686.5  on 66  degrees of freedom
## AIC: 532.91
##
## Number of Fisher Scoring iterations: 6
```

```
(emm3<-emmeans(mod3,~treat|week, at=list(week=c(0))))
```

```
## week = 0:
##  treat emmean    SE  df asymp.LCL asymp.UCL
##  a       3.76 0.149 Inf      3.47      4.05
##  b       3.80 0.101 Inf      3.60      4.00
##  c       4.09 0.111 Inf      3.88      4.31
##
## Results are given on the log (not the response) scale.
## Confidence level used: 0.95
```

```
(pairs(emm3))
```

```
## week = 0:
##  contrast estimate    SE  df z.ratio p.value
##  a - b     -0.0369 0.180 Inf -0.205  0.9770
##  a - c     -0.3331 0.185 Inf -1.796  0.1710
##  b - c     -0.2961 0.150 Inf -1.972  0.1191
##
## Results are given on the log (not the response) scale.
## P value adjustment: tukey method for comparing a family of 3 estimates
```

Again, the alphas are not significantly different from each other, we assume the same intercept. As well the paramameter treatment is not significant. We can fit a better model.

```
mod31<-glm(vitc~week+treat:week,family= gaussian(link="log"), data=vitc)
summary(mod31)

##
## Call:
## glm(formula = vitc ~ week + treat:week, family = gaussian(link = "log"),
##     data = vitc)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -16.584   -5.062   -1.577    3.966   34.273
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.87691    0.06813  56.903  < 2e-16 ***
## week        -0.16071    0.02414  -6.659 5.81e-09 ***
## week:treatb  0.11891    0.02223   5.348 1.12e-06 ***
## week:treatc  0.04594    0.02512   1.829   0.0718 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 89.30871)
##
##     Null deviance: 15538.6  on 71  degrees of freedom
## Residual deviance:  6072.9  on 68  degrees of freedom
## AIC: 533.64
##
## Number of Fisher Scoring iterations: 5
```

• Estimate αi. Are they statistically different?

To estimate

$$\alpha_1 = exp3.76 = 45.6042$$

To estimate

$$\alpha_2 = exp3.8 = 41.2644$$

To estimate

$$\alpha_3 = exp4.09 = 51.9354$$

- **Estimate βi. Are they statistically different?**

To estimate

$$\beta_1 = -0.16071$$

To estimate

$$\beta_2 = 0.11891$$

To estimate

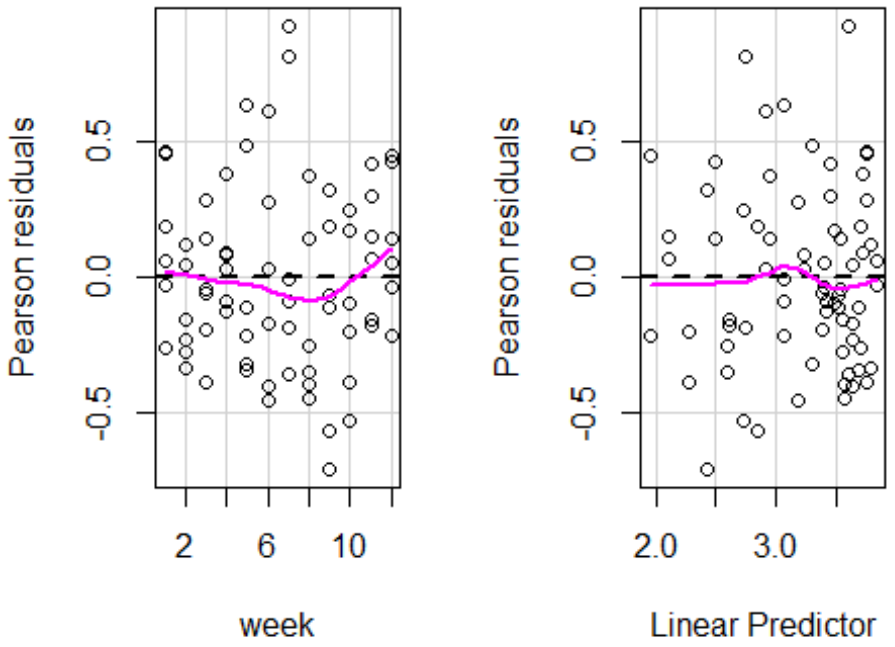
$$\beta_3 = 0.04594$$

The interaction between week and treatment c is not different from the interaction between week and treatment a, so we assumer that are the same

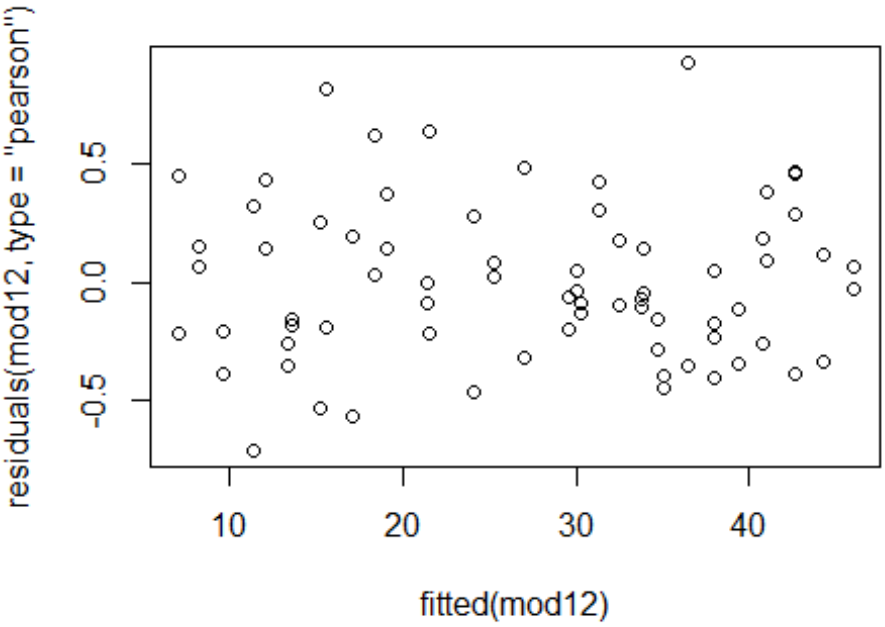
e) Justify which of the four models that you have fitted better verify the model hypothesis and gives place to the bests fits.

GAMMA

```
residualPlots(mod12, test=F)
```




```
plot(fitted(mod12),residuals(mod12,type="pearson"))
```



AIC=511.32

```
sum(residuals(mod12, type="pearson")^2)
```

```
## [1] 7.849694
```

```
qchisq(0.05,68)
```

```
## [1] 50.02023
```

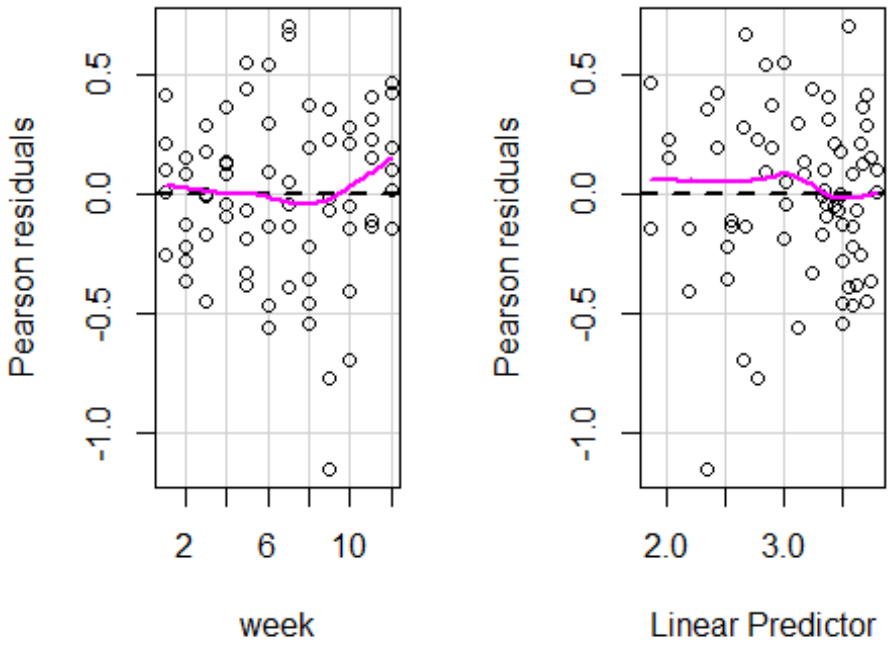
```
pchisq(8.323,68,log=FALSE, lower.tail=TRUE)
```

```
## [1] 6.788502e-20
```

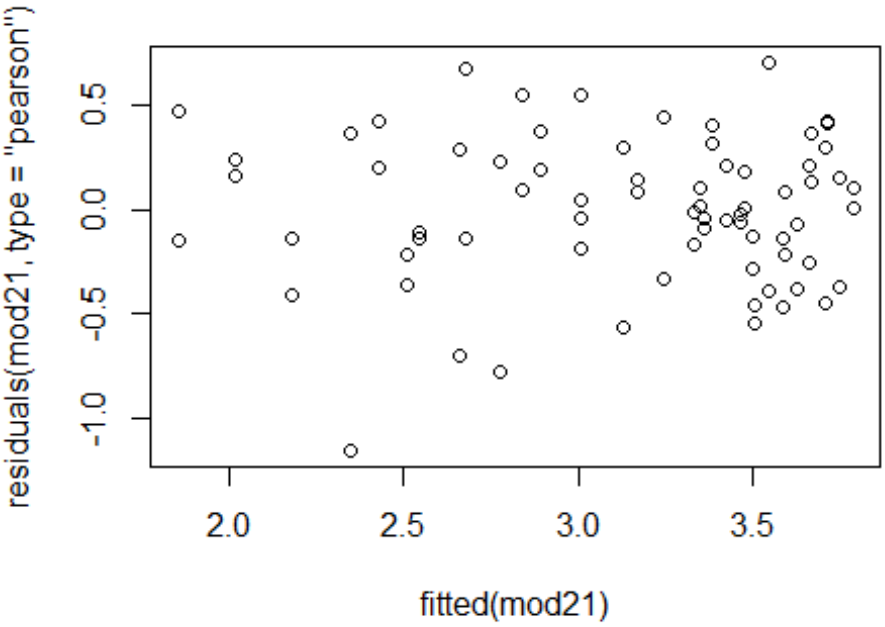
We accept the null hypothesis of the deviance test, and canc conclude that out model is better than the null mmodel and we are explaining something. The AIC is 511.32, in order to compare with the other models. As well we assume normality for the residuals.

LOGNORMAL

```
residualPlots(mod21, test=F)
```




```
plot(fitted(mod21),residuals(mod21,type="pearson"))
```



AIC=63.131 The AIC is the lowest among the models

```
sum(residuals(mod21, type="pearson")^2)
```

```
## [1] 8.817304
```

```
qchisq(0.05,68)
```

```
## [1] 50.02023
```

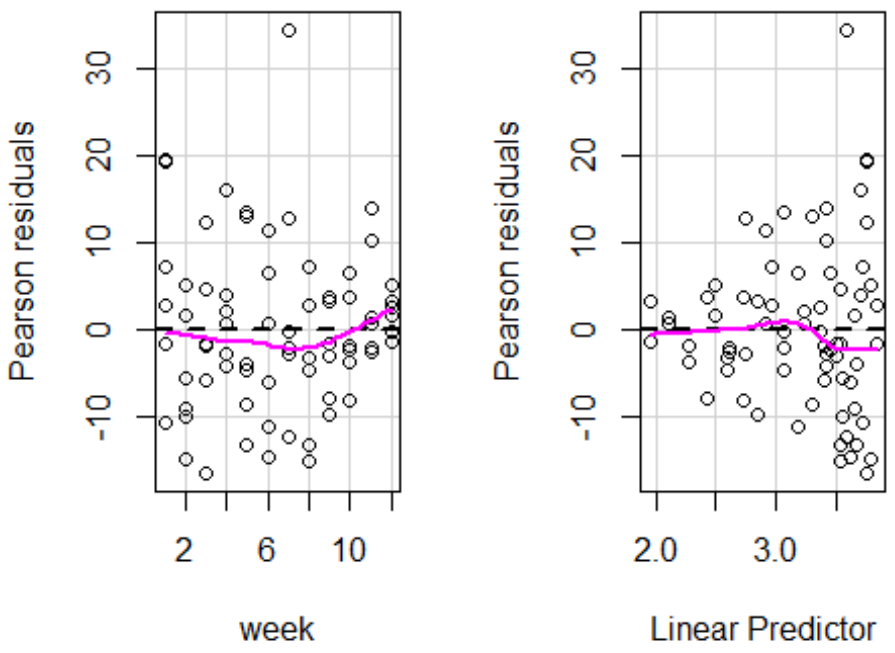
```
pchisq(8.8173,68,log=FALSE, lower.tail=TRUE)
```

```
## [1] 3.800119e-19
```

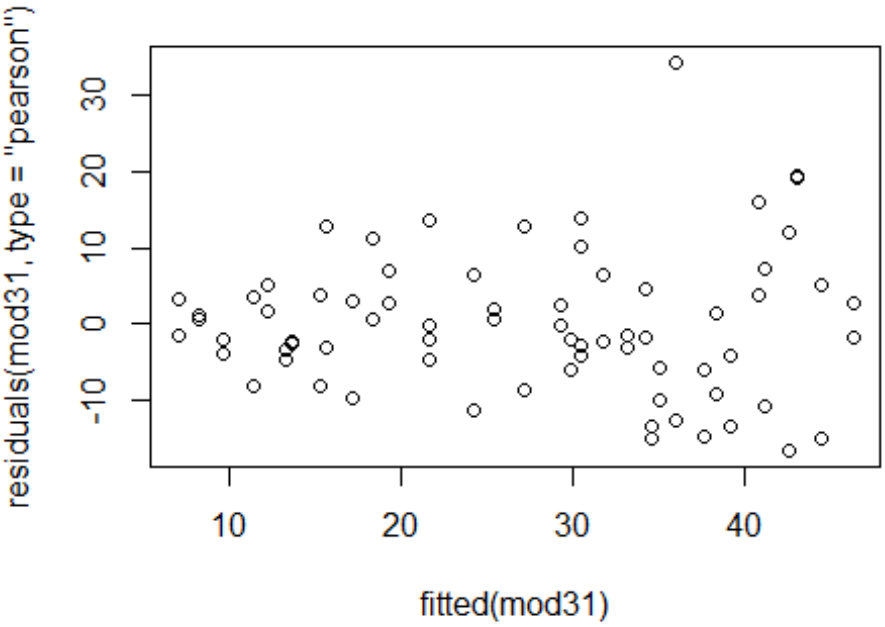
We reject the null hypothesis and can state that our model is better than the null model, as well the AIC is the lowest until now. The residuals are as well verified that follow a normal distribution.

NORMAL

```
residualPlots(mod31,test=F)
```




```
plot(fitted(mod31),residuals(mod31,type="pearson"))
```



AIC=533.64

```
sum(residuals(mod31, type="pearson")^2)
```

```
## [1] 6072.931
```

```
qchisq(0.05,68)
```

```
## [1] 50.02023
```

```
pchisq(0.05,68,log=FALSE, lower.tail=TRUE)
```

```
## [1] 1.120079e-93
```

The residuals are clearly not ollowing a normal distribution, hence this third model is no appropiate for our data.

Conclusions: We have fitted three differetn models for our non-normal data. We have tried three different transformations to get to the linearity. Overall for out Vitamin C data tha model that fits better and the one we should stick to is the LogNormal transformation with the logarithm link. The AIC is the lowest amongs our models and the normality of the residuals is verified.