

# Toenail data

Emmanuelle

07/04/2020

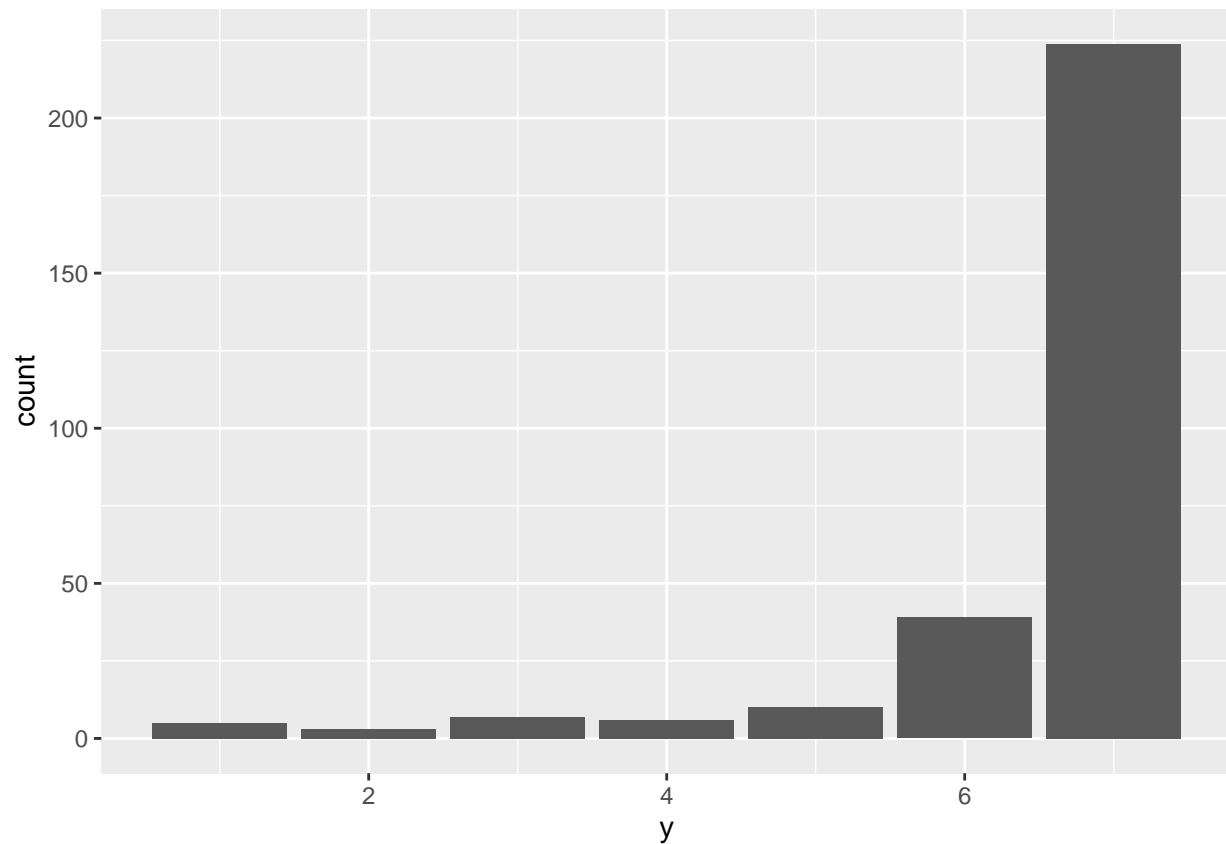
## Toenail data

```
data("Toenail", package = "prLogistic")
```

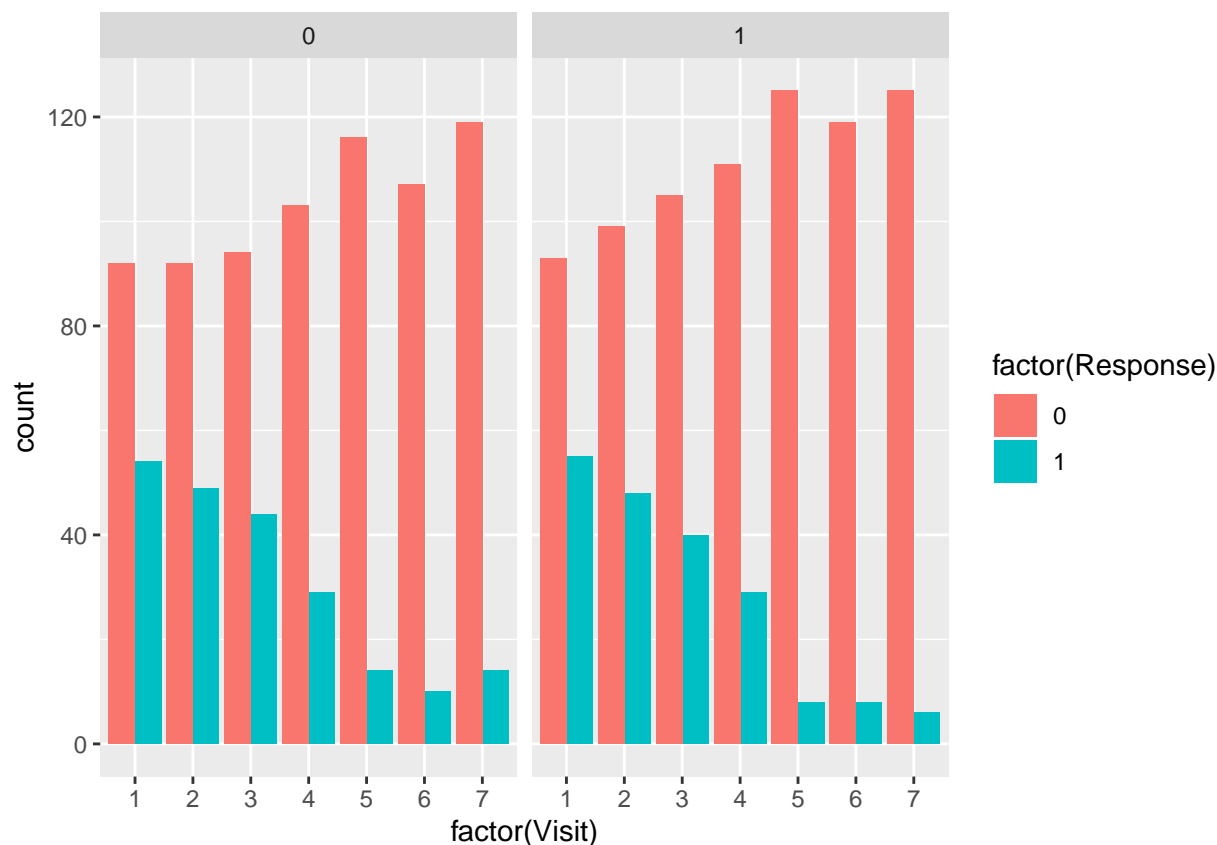
```
ndat<-tapply(Toenail$ID,Toenail$ID,length)  
print(table(ndat))
```

```
## ndat  
##  1  2  3  4  5  6  7  
##  5  3  7  6 10 39 224
```

```
ggplot(data.frame(id=unique(Toenail$ID),y=ndat),aes(x=y)) + geom_bar()
```



```
ggplot(Toenail, aes(x = factor(Visit), fill=factor(Response))) + geom_bar(position='dodge') + facet_wrap
```



## Logistic binary regression with saemix

Fitting a logistic regression model of the form

$$\text{logit}(P(Y_{ij} = 1/\psi_i)) = \mu_1 + \mu_2 \text{Month}_{ij} + \mu_3 \text{Treatment}$$

```
## Using the object called Toenail in this R session as the data.
## Column name(s) do(es) not exist in the dataset, please check
## Remove columns 1 ( )
## No valid name given, attempting automatic recognition
## Automatic recognition of columns Response successful
##
##
## The following SaemixData object was successfully created:
##
## Object of class SaemixData
##   longitudinal data for use with the SAEM algorithm
## Dataset Toenail
##   Structured data: Response ~ Month + Treatment + Response | ID
##   X variable for graphs: Month ( )
##
##
## The following SaemixModel object was successfully created:
##
## Nonlinear mixed-effects model
```

```

## Model function: Binary repeated model Model type: likelihood
## function(psi,id,xidep) {
##   month<-xidep[,1]
##   trt<-xidep[,2]
##   y<-xidep[,3]
##   alp<-psi[id,1]
##   th1<-psi[id,2]
##   th2<-psi[id,3]
##   logit1<-alp+th1*month+th2*trt # model for Logit(P(Y=0))
##   pge0<-exp(logit1)/(1+exp(logit1))
##   P.obs = (y==0)*pge0+(y==1)*(1-pge0)
##   logpdf <- log(P.obs)
##   return(logpdf)
## }
## Nb of parameters: 3
##   parameter names: alp th1 th2
##   distribution:
##   Parameter Distribution Estimated
## [1,] alp      normal      Estimated
## [2,] th1      normal      Estimated
## [3,] th2      normal      Estimated
## Variance-covariance matrix:
##   alp th1 th2
## alp  1  0  0
## th1  0  1  0
## th2  0  0  0
## No covariate in the model.
## Initial values
##   alp th1 th2
## Pop.CondInit 2.5 1.5 -0.5

## Running main SAEM algorithm
## [1] "Tue Apr 7 17:07:06 2020"
## ....
## Minimisation finished
## [1] "Tue Apr 7 17:07:13 2020"

```

## Estimation of the log-likelihood



```
## Nonlinear mixed-effects model fit by the SAEM algorithm
## -----
##      Data
## -----
## Object of class SaemixData
##   longitudinal data for use with the SAEM algorithm
## Dataset Toenail
##   Structured data: Response ~ Month + Treatment + Response | ID
##   X variable for graphs: Month ()
## Dataset characteristics:
##   number of subjects:      294
##   number of observations: 1908
##   average/min/max nb obs: 6.49 / 1 / 7
## First 10 lines of data:
##   ID      Month Treatment Response Response.1 mdv cens occ ytype
## 1  1  0.0000000      1         1           1  0   0   1   1
## 2  1  0.8571429      1         1           1  0   0   1   1
## 3  1  3.5357143      1         1           1  0   0   1   1
## 4  1  4.5357143      1         0           0  0   0   1   1
## 5  1  7.5357143      1         0           0  0   0   1   1
## 6  1 10.0357143      1         0           0  0   0   1   1
## 7  1 13.0714286      1         0           0  0   0   1   1
## 8  2  0.0000000      0         0           0  0   0   1   1
## 9  2  0.9642857      0         0           0  0   0   1   1
## 10 2  2.0000000      0         1           1  0   0   1   1
## -----
##      Model
## -----
## Nonlinear mixed-effects model
```

```

## Model function: Binary repeated model Model type: likelihood
## function(psi,id,xidep) {
##   month<-xidep[,1]
##   trt<-xidep[,2]
##   y<-xidep[,3]
##   alp<-psi[id,1]
##   th1<-psi[id,2]
##   th2<-psi[id,3]
##   logit1<-alp+th1*month+th2*trt # model for Logit(P(Y=0))
##   pge0<-exp(logit1)/(1+exp(logit1))
##   P.obs = (y==0)*pge0+(y==1)*(1-pge0)
##   logpdf <- log(P.obs)
##   return(logpdf)
## }
## <bytecode: 0x558b97100b48>
## Nb of parameters: 3
##   parameter names: alp th1 th2
##   distribution:
##   Parameter Distribution Estimated
## [1,] alp      normal      Estimated
## [2,] th1      normal      Estimated
## [3,] th2      normal      Estimated
## Variance-covariance matrix:
##   alp th1 th2
## alp  1  0  0
## th1  0  1  0
## th2  0  0  0
## No covariate in the model.
## Initial values
##   alp th1 th2
## Pop.CondInit 2.5 1.5 -0.5
## -----
## ---- Key algorithm options ----
## -----
## Estimation of individual parameters (MAP)
## Estimation of log-likelihood by importance sampling
## Number of iterations: K1=300, K2=100
## Number of chains: 1
## Seed: 632545
## Number of MCMC iterations for IS: 5000
## Simulations:
##   nb of simulated datasets used for npde: 1000
##   nb of simulated datasets used for VPC: 100
## Input/output
##   save the results to a file: FALSE
##   save the graphs to files: FALSE
## -----
## ---- Results ----
## -----
## ----- Fixed effects -----
## -----
## Parameter Estimate
## [1,] alp      2.72
## [2,] th1      1.48

```

```
## [3,] th2      -0.49
## -----
## ----- Variance of random effects -----
## -----
##      Parameter Estimate
## alp omega2.alp 65.9
## th1 omega2.th1  1.2
## -----
## ----- Correlation matrix of random effects -----
## -----
##      omega2.alp omega2.th1
## omega2.alp 1      0
## omega2.th1 0      1
## -----
## ----- Statistical criteria -----
## -----
##
## Likelihood computed by importance sampling
##      -2LL= 1121.365
##      AIC = 1133.365
##      BIC = 1155.467
## -----
```

## Dropout

Assumptions:

- missing visits are considered to be missing at random (MAR), in which case maximum likelihood theory states that we don't need to account for missing data
- on the other hand, some patients leave the study before the last visit and this dropout has been posited in some publications to be linked to study outcome (in particular, patients feeling improvement in their condition were thought to drop out of the study more easily)
  - what we have is the last visit

Models for dropout:

- event
  - 1 at last visit if patient drops out
  - 0 (censored) if the patient remains until the last visit
  - Monolix needs to start with an event=0, so for the subjects with only one visit, we assume the dropout took place after 2 weeks
- dropout
  - non informative dropout: baseline hazard modelled as a Weibull function
  - informative dropout: baseline hazard multiplied by an exponential function of the logit
  - could also consider perhaps the difference in observed events between the last two visits

$$h(t) = \beta \lambda t^{\beta-1} e^{\gamma \text{logit}}$$

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0         0         0         0         0         0
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.6786  1.0000  1.0000  1.0321  1.0357  1.6071
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 1.500   2.000   2.000   2.059   2.071   3.536
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
##      2.500    3.000    3.000    3.134    3.250    4.750
##      Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##      4.000    6.000    6.107    6.283    6.500    8.750
##      Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##      5.250    9.000    9.107    9.311    9.580   12.536
##      Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##      10.50   12.00   12.20   12.48   12.75   18.50

##      ID Response Treatment Month Visit ytype deltaR
## 207 45      0      0  0.0  1.0    2      0
## 2071 45      1      0  0.0  1.0    1      0
## 2072 45      1      0  0.5  1.5    2      0

##      ID Response Treatment Month Visit ytype deltaR
## 208 48      0      1  0.0  1.0    2      0
## 2081 48      0      1  0.0  1.0    1      0
## 2082 48      1      1  0.5  1.5    2      0
```

## Monolix estimates

Convergence profiles are not fantastic (evidence of fluctuations after the initial iteration phase, during convergence)

- model with non-informative dropout
  - very similar estimates for the logistic model as saemix
    - \* good pd plot for the binary data (slight overestimation of 90% percentile but lower bound and median very good)
  - dropout model
    - \* Weibull model: estimates of  $\beta=0.5$  and  $\lambda=0.03$ ,  $-2*LL=1457.32$ ,  $BIC=1497.1$
    - \* constant hazard model ( $h=1/\lambda$ ): estimate of  $\lambda=113$  ( $1/\lambda=0.009$ ),  $-2LL=1466.45$ ,  $BIC=1500.55$ 
      - but very poor convergence plots and  $th2$  (treatment effect) now very close to 0 indicating some convergence problem (we didn't change the model for the binary data...)
- model with informative dropout (Weibull+exponential term)
  - again  $th2$  (treatment effect) estimated to a very small value (-0.05, with an SE of 0.7)
  - the value of the link is estimated to be  $6.10^{-5}$  with large SE suggesting an absence of link, with values of  $\beta$  and  $\lambda$  close to the previous estimates, and the likelihood is the same as without a link:  $-2*LL=1457.8$
- model with informative dropout linked to  $\delta R$ 
  - estimate of  $th2$  positive this time (0.765) but again with high SE, estimates of  $th1$  and  $\alpha$  same as previously, and  $-2LL=1458.07$  for this model
  - estimates of  $\beta$  and  $\lambda$  similar to previous values
  - again estimate of the link is not significantly different from 0

## Conclusion

Maybe a MAR assumption for this data, without considering dropouts, could be an acceptable way to analyse the binary longitudinal response, as there doesn't seem to be much link between dropout and either the observed change in response or the predicted logit.