Multivariate Analysis for the Behavioral Sciences,
Second Edition (Chapman and Hall/CRC, 2019)

Examples of Chapter 10:
Analysis of Longitudinal Data III:
Non-Normal Responses

Kimmo Vehkalahti and Brian S. Everitt

22 November 2018

Contents

Examples

Beat the Blues Revisited

Using the Beat the Blues data introduced in **Chapter 9**:

```r
library("tidyr"); 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
```

```r
BtB <- read.table("data/BtB.txt", header = TRUE, sep = '\t')

# to make sure that the factor levels are logical (esp. Treatment):
BtB <- within(BtB, {
        Drug <- factor(Drug, levels=c("No", "Yes")) # default
      Length <- factor(Length, levels=c("<6m", ">6m")) # default
   Treatment <- factor(Treatment, levels=c("TAU", "BtheB")) # NOT default!
})

glimpse(BtB); head(BtB); tail(BtB)
```

```
## Observations: 100
## Variables: 9
## $ Subject   <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1...
## $ Drug      <fct> No, Yes, Yes, No, Yes, Yes, Yes, No, Yes, Yes, No, Y...
## $ Length    <fct> >6m, >6m, <6m, >6m, >6m, <6m, <6m, >6m, <6m, >6m, >6...
## $ Treatment <fct> TAU, BtheB, TAU, BtheB, BtheB, BtheB, TAU, TAU, Bthe...
## $ BDIpre    <int> 29, 32, 25, 21, 26, 7, 17, 20, 18, 20, 30, 49, 26, 3...
## $ BDI2m     <int> 2, 16, 20, 17, 23, 0, 7, 20, 13, 5, 32, 35, 27, 26, ...
## $ BDI4m     <int> 2, 24, NA, 16, NA, 0, 7, 21, 14, 5, 24, NA, 23, 36, ...
## $ BDI6m     <int> NA, 17, NA, 10, NA, 0, 3, 19, 20, 8, 12, NA, NA, 27,...
## $ BDI8m     <int> NA, 20, NA, 9, NA, 0, 7, 13, 11, 12, 2, NA, NA, 22, ...
```

```
##   Subject Drug Length Treatment BDIpre BDI2m BDI4m BDI6m BDI8m
## 1       1   No    >6m       TAU     29     2     2    NA    NA
## 2       2  Yes    >6m     BtheB     32    16    24    17    20
## 3       3  Yes    <6m       TAU     25    20    NA    NA    NA
## 4       4   No    >6m     BtheB     21    17    16    10     9
## 5       5  Yes    >6m     BtheB     26    23    NA    NA    NA
## 6       6  Yes    <6m     BtheB      7     0     0     0     0
```

```
##     Subject Drug Length Treatment BDIpre BDI2m BDI4m BDI6m BDI8m
## 95       95   No    >6m     BtheB     16    11     4     2     3
## 96       96  Yes    >6m     BtheB     16    16    10    10     8
## 97       97  Yes    <6m       TAU     28    NA    NA    NA    NA
## 98       98   No    >6m     BtheB     11    22     9    11    11
## 99       99   No    <6m       TAU     13     5     5     0     6
```

```
## 100      100  Yes      <6m      TAU      43      NA      NA      NA      NA
# Convert data to long form for the analyses, adding Time:
BtBL <- gather(BtB, key = Visit, value = BDI, BDI2m, BDI4m, BDI6m, BDI8m) %>%
  mutate(Time = as.integer(substr(Visit, 4, 4)))

glimpse(BtBL); head(BtBL); tail(BtBL)
```

```
## Observations: 400
## Variables: 8
## $ Subject   <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1...
## $ Drug      <fct> No, Yes, Yes, No, Yes, Yes, Yes, No, Yes, Yes, No, Y...
## $ Length    <fct> >6m, >6m, <6m, >6m, >6m, <6m, <6m, >6m, <6m, >6m, >6...
## $ Treatment <fct> TAU, BtheB, TAU, BtheB, BtheB, BtheB, TAU, TAU, Bthe...
## $ BDIpre    <int> 29, 32, 25, 21, 26, 7, 17, 20, 18, 20, 30, 49, 26, 3...
## $ Visit     <chr> "BDI2m", "BDI2m", "BDI2m", "BDI2m", "BDI2m", "BDI2m"...
## $ BDI       <int> 2, 16, 20, 17, 23, 0, 7, 20, 13, 5, 32, 35, 27, 26, ...
## $ Time      <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2...

##   Subject Drug Length Treatment BDIpre Visit BDI Time
## 1       1   No    >6m       TAU     29 BDI2m   2    2
## 2       2  Yes    >6m     BtheB     32 BDI2m  16    2
## 3       3  Yes    <6m       TAU     25 BDI2m  20    2
## 4       4   No    >6m     BtheB     21 BDI2m  17    2
## 5       5  Yes    >6m     BtheB     26 BDI2m  23    2
## 6       6  Yes    <6m     BtheB      7 BDI2m   0    2

##     Subject Drug Length Treatment BDIpre Visit BDI Time
## 395      95   No    >6m     BtheB     16 BDI8m   3    8
## 396      96  Yes    >6m     BtheB     16 BDI8m   8    8
## 397      97  Yes    <6m       TAU     28 BDI8m  NA    8
## 398      98   No    >6m     BtheB     11 BDI8m  11    8
## 399      99   No    <6m       TAU     13 BDI8m   6    8
## 400     100  Yes    <6m       TAU     43 BDI8m  NA    8
```

```
# important to order the data with respect to the subjects:
BtBLSORT <- BtBL %>%
  arrange(Subject)

head(BtBLSORT); tail(BtBLSORT)
```

```
##   Subject Drug Length Treatment BDIpre Visit BDI Time
## 1       1   No    >6m       TAU     29 BDI2m   2    2
## 2       1   No    >6m       TAU     29 BDI4m   2    4
## 3       1   No    >6m       TAU     29 BDI6m  NA    6
## 4       1   No    >6m       TAU     29 BDI8m  NA    8
## 5       2  Yes    >6m     BtheB     32 BDI2m  16    2
## 6       2  Yes    >6m     BtheB     32 BDI4m  24    4

##     Subject Drug Length Treatment BDIpre Visit BDI Time
## 395      99   No    <6m       TAU     13 BDI6m   0    6
## 396      99   No    <6m       TAU     13 BDI8m   6    8
## 397     100  Yes    <6m       TAU     43 BDI2m  NA    2
## 398     100  Yes    <6m       TAU     43 BDI4m  NA    4
## 399     100  Yes    <6m       TAU     43 BDI6m  NA    6
## 400     100  Yes    <6m       TAU     43 BDI8m  NA    8
```

Table 10.3

```r
# install.packages("gee")
library("gee")

BtB_gee1 <- gee(BDI ~ BDIpre + Time + Treatment + Drug + Length, id = Subject,
                data = BtBLSORT, family = gaussian, corstr = "independence")
```

```
## Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27

## running glm to get initial regression estimate

##     (Intercept)            BDIpre              Time TreatmentBtheB          DrugYes
##       7.8830747         0.5723729        -0.9608077      -3.3539662       -3.5460058
##       Length>6m
##       1.7530800
```

```r
summary(BtB_gee1)
```

```
##
##  GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
##  gee S-function, version 4.13 modified 98/01/27 (1998)
##
## Model:
##  Link:                      Identity
##  Variance to Mean Relation: Gaussian
##  Correlation Structure:     Independent
##
## Call:
## gee(formula = BDI ~ BDIpre + Time + Treatment + Drug + Length,
##     id = Subject, data = BtBLSORT, family = gaussian, corstr = "independence")
##
## Summary of Residuals:
##         Min           1Q       Median           3Q          Max
## -24.20158432  -5.31202378   0.01101526   5.29503741  27.77789553
##
##
## Coefficients:
##                   Estimate Naive S.E.    Naive z Robust S.E.   Robust z
## (Intercept)      7.8830747 1.78048908   4.427477  2.19973944   3.583640
## BDIpre           0.5723729 0.05486079  10.433188  0.08853253   6.465114
## Time            -0.9608077 0.23263437  -4.130119  0.17688635  -5.431780
## TreatmentBtheB  -3.3539662 1.09831939  -3.053726  1.71390982  -1.956909
## DrugYes         -3.5460058 1.14469086  -3.097785  1.73069664  -2.048889
## Length>6m        1.7530800 1.10849861   1.581490  1.41954159   1.234962
##
## Estimated Scale Parameter:  74.8854
## Number of Iterations:  1
##
## Working Correlation
##      [,1] [,2] [,3] [,4]
## [1,]    1    0    0    0
## [2,]    0    1    0    0
## [3,]    0    0    1    0
## [4,]    0    0    0    1
```

Table 10.4

```
BtB_gee2 <- gee(BDI ~ BDIpre + Time + Treatment + Drug + Length, id = Subject,
                data = BtBLSORT, family = gaussian, corstr = "exchangeable")
```

```
## Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27

## running glm to get initial regression estimate

##    (Intercept)          BDIpre           Time TreatmentBtheB        DrugYes
##      7.8830747       0.5723729     -0.9608077     -3.3539662     -3.5460058
##      Length>6m
##      1.7530800
```

```
summary(BtB_gee2)
```

```
##
##  GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
##  gee S-function, version 4.13 modified 98/01/27 (1998)
##
## Model:
##  Link:                      Identity
##  Variance to Mean Relation: Gaussian
##  Correlation Structure:     Exchangeable
##
## Call:
## gee(formula = BDI ~ BDIpre + Time + Treatment + Drug + Length,
##     id = Subject, data = BtBLSORT, family = gaussian, corstr = "exchangeable")
##
## Summary of Residuals:
##         Min          1Q      Median          3Q         Max
## -25.4478843  -6.3276726  -0.8152833   4.3622258  25.4078115
##
##
## Coefficients:
##                   Estimate Naive S.E.    Naive z Robust S.E.    Robust z
## (Intercept)      5.8855129 2.32380381  2.5327065  2.10712166  2.7931529
## BDIpre           0.6399964 0.08033495  7.9665999  0.07931263  8.0692874
## Time            -0.7078407 0.14254124 -4.9658660  0.15394156 -4.5981134
## TreatmentBtheB  -2.3360241 1.72621751 -1.3532617  1.66217026 -1.4054060
## DrugYes         -2.7742506 1.78397557 -1.5550945  1.64824318 -1.6831561
## Length>6m        0.2084783 1.69179766  0.1232288  1.48052530  0.1408137
##
## Estimated Scale Parameter:  77.14393
## Number of Iterations:  5
##
## Working Correlation
##           [,1]      [,2]      [,3]      [,4]
## [1,] 1.0000000 0.6915241 0.6915241 0.6915241
## [2,] 0.6915241 1.0000000 0.6915241 0.6915241
## [3,] 0.6915241 0.6915241 1.0000000 0.6915241
## [4,] 0.6915241 0.6915241 0.6915241 1.0000000
```

Table 10.1: Respiratory Disorder Data

```r
library("tidyr"); library("dplyr")
# install.packages("HSAUR3")
library("HSAUR3")
```

```
## Loading required package: tools
```

```r
data("respiratory", package = "HSAUR3")
RESPL <- respiratory # (data are readily given in long form)
glimpse(RESPL); head(RESPL, n = 5)
```

```
## Observations: 555
## Variables: 7
## $ centre    <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
## $ treatment <fct> placebo, placebo, placebo, placebo, placebo, placebo...
## $ gender    <fct> female, female, female, female, female, female, fema...
## $ age       <dbl> 46, 46, 46, 46, 46, 28, 28, 28, 28, 28, 23, 23, 23, ...
## $ status    <fct> poor, poor, poor, poor, poor, poor, poor, poor, poor...
## $ month     <ord> 0, 1, 2, 3, 4, 0, 1, 2, 3, 4, 0, 1, 2, 3, 4, 0, 1, 2...
## $ subject   <fct> 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4...

##     centre treatment gender age status month subject
## 1        1   placebo female  46   poor     0       1
## 112      1   placebo female  46   poor     1       1
## 223      1   placebo female  46   poor     2       1
## 334      1   placebo female  46   poor     3       1
## 445      1   placebo female  46   poor     4       1
```

```r
RESPL <- within(RESPL, {
    levels(gender) <- c("male", "female") # see the original article by Davis (1991)
    status <- as.numeric(status) - 1 # for converting to wide form below
})
head(RESPL, n = 5)
```

```
##     centre treatment gender age status month subject
## 1        1   placebo   male  46      0     0       1
## 112      1   placebo   male  46      0     1       1
## 223      1   placebo   male  46      0     2       1
## 334      1   placebo   male  46      0     3       1
## 445      1   placebo   male  46      0     4       1
```

```r
# convert data to WIDE form:
# (sep: set valid names month0:month4 here; otherwise just 0:4)
RESP <- RESPL %>%
  spread(key = month, value = status, sep = "") %>%
  arrange(subject)

glimpse(RESP); head(RESP, n = 5)
```

```
## Observations: 111
## Variables: 10
## $ centre    <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
## $ treatment <fct> placebo, placebo, treatment, placebo, placebo, treat...
## $ gender    <fct> male, male, male, male, female, male, male, male, ma...
## $ age       <dbl> 46, 28, 23, 44, 13, 34, 43, 28, 31, 37, 30, 14, 23, ...
```

```
## $ subject   <fct> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1...
## $ month0    <dbl> 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1, 0, 1, 0, 1, 0, 0, 0...
## $ month1    <dbl> 0, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 0, 0...
## $ month2    <dbl> 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 1...
## $ month3    <dbl> 0, 0, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 1...
## $ month4    <dbl> 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 1, 0, 1...

##   centre treatment gender age subject month0 month1 month2 month3 month4
## 1      1   placebo   male  46       1      0      0      0      0      0
## 2      1   placebo   male  28       2      0      0      0      0      0
## 3      1 treatment   male  23       3      1      1      1      1      1
## 4      1   placebo   male  44       4      1      1      1      1      0
## 5      1   placebo female  13       5      1      1      1      1      1
```

```r
# back to long form; leave month0 for baseline:
RESPL <- gather(RESP, key = months, value = status, month1:month4) %>%
  mutate(month = as.integer(substr(months, 6, 6))) %>%
  rename(baseline = month0) %>%
  arrange(subject)

glimpse(RESPL); head(RESPL, n = 5)
```

```
## Observations: 444
## Variables: 9
## $ centre    <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
## $ treatment <fct> placebo, placebo, placebo, placebo, placebo, placebo...
## $ gender    <fct> male, male, male, male, male, male, male, male, male...
## $ age       <dbl> 46, 46, 46, 46, 28, 28, 28, 28, 23, 23, 23, 23, 44, ...
## $ subject   <fct> 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5...
## $ baseline  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
## $ months    <chr> "month1", "month2", "month3", "month4", "month1", "m...
## $ status    <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1...
## $ month     <int> 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2...

##   centre treatment gender age subject baseline months status month
## 1      1   placebo   male  46       1        0 month1      0     1
## 2      1   placebo   male  46       1        0 month2      0     2
## 3      1   placebo   male  46       1        0 month3      0     3
## 4      1   placebo   male  46       1        0 month4      0     4
## 5      1   placebo   male  28       2        0 month1      0     1
```

Figure 10.2

```r
n = dim(RESP)[1]
RP <- RESP %>%
  group_by(treatment) %>%
  summarise(m0 = 100*sum(month0)/n,
            m1 = 100*sum(month1)/n,
            m2 = 100*sum(month2)/n,
            m3 = 100*sum(month3)/n,
            m4 = 100*sum(month4)/n ) %>%
  select(-treatment) %>%
  ungroup()

RP2 <- cbind(RP[1, ], RP[2, ])
month <- rep(1:5, times = 2)
treatment <- rep(0:1, each = 5)
RPos <- as.data.frame(cbind(treatment, month, t(RP2)))
row.names(RPos) <- 1:10
names(RPos) <- c("treatment", "month", "positive")
RPos$treatment <- factor(RPos$treatment, labels = c("placebo", "treatment"))

library("ggplot2")

p1 <- ggplot(RPos, aes(x = month, y = positive, fill = treatment))
p2 <- p1 + geom_bar(position = "dodge", stat = "identity")
p3 <- p2 + scale_x_continuous(name = "", breaks = 1:5,
               labels = c("Baseline", "Month 1", "Month 2", "Month 3", "Month 4"))
p4 <- p3 + theme_bw() + theme(panel.grid.major = element_blank(),
                              panel.grid.minor = element_blank())
p5 <- p4 + scale_y_continuous(name = "Positive responses (%)")
p6 <- p5 + scale_fill_grey(start = 0.5, end = 0)
p6
```
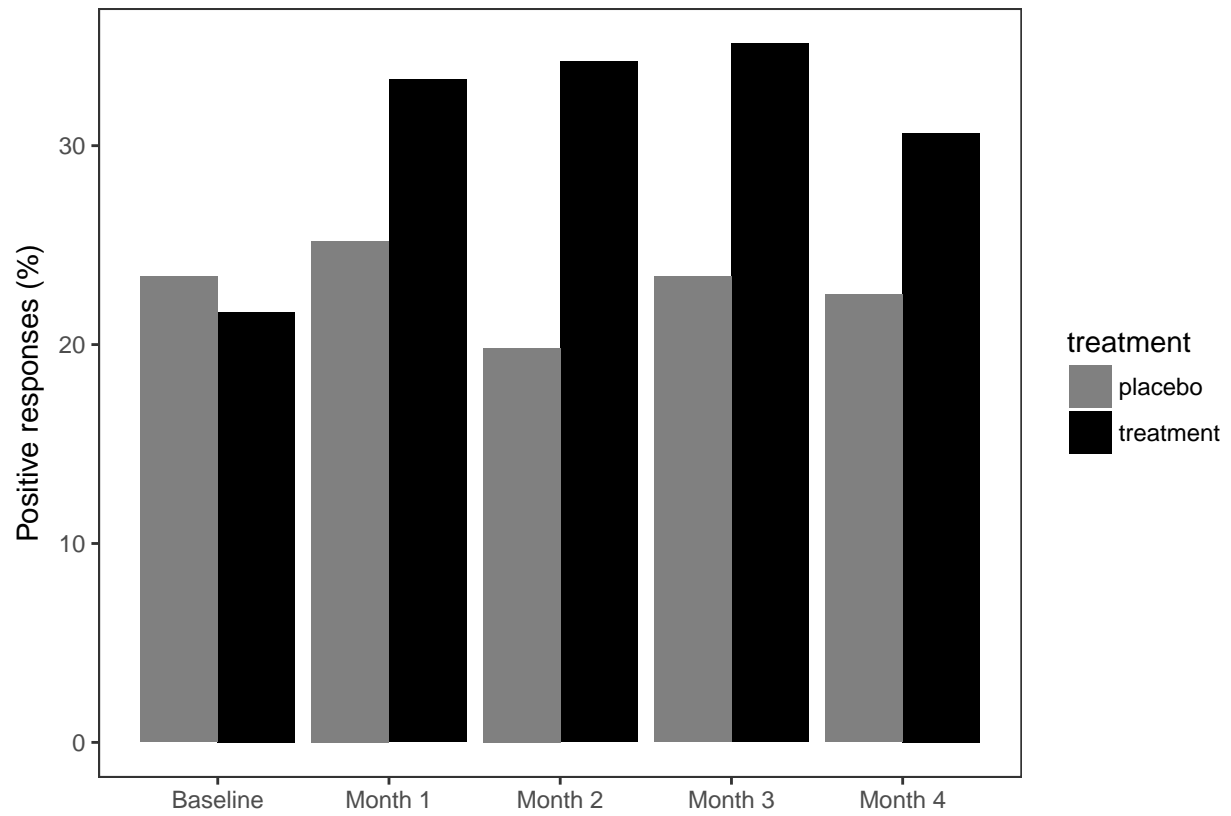
Table 10.5

```r
resp_gee1 <- gee(status ~ age, data = RESPL, family = binomial, id = subject,
                 corstr = "independence", scale.fix = TRUE, scale.value = 1)
```

```
## Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27

## running glm to get initial regression estimate

## (Intercept)         age
##  0.64576265 -0.01201525
```

```r
summary(resp_gee1)
```

```
##
##  GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
##  gee S-function, version 4.13 modified 98/01/27 (1998)
##
## Model:
##  Link:                      Logit
##  Variance to Mean Relation: Binomial
##  Correlation Structure:     Independent
##
## Call:
## gee(formula = status ~ age, id = subject, data = RESPL, family = binomial,
##     corstr = "independence", scale.fix = TRUE, scale.value = 1)
##
## Summary of Residuals:
##        Min         1Q     Median         3Q        Max
## -0.6256488 -0.5508736  0.3828310  0.4320996  0.5427142
##
##
## Coefficients:
##               Estimate  Naive S.E.   Naive z Robust S.E.  Robust z
## (Intercept)  0.64576265 0.255300707  2.529420  0.39567302  1.632061
## age         -0.01201525 0.007061029 -1.701628  0.01173464 -1.023912
##
## Estimated Scale Parameter:  1
## Number of Iterations:  1
##
## Working Correlation
##      [,1] [,2] [,3] [,4]
## [1,]    1    0    0    0
## [2,]    0    1    0    0
## [3,]    0    0    1    0
## [4,]    0    0    0    1
```

Table 10.6

```
resp_gee2 <- gee(status ~ age, data = RESPL, family = binomial, id = subject,
                 corstr = "exchangeable", scale.fix = TRUE, scale.value = 1)
```

```
## Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27

## running glm to get initial regression estimate

## (Intercept)          age
##  0.64576265 -0.01201525
```

```
summary(resp_gee2)
```

```
##
##  GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
##  gee S-function, version 4.13 modified 98/01/27 (1998)
##
## Model:
##  Link:                    Logit
##  Variance to Mean Relation: Binomial
##  Correlation Structure:    Exchangeable
##
## Call:
## gee(formula = status ~ age, id = subject, data = RESPL, family = binomial,
##     corstr = "exchangeable", scale.fix = TRUE, scale.value = 1)
##
## Summary of Residuals:
##        Min          1Q      Median          3Q         Max
## -0.6256488 -0.5508736   0.3828310   0.4320996   0.5427142
##
##
## Coefficients:
##              Estimate Naive S.E.    Naive z Robust S.E.   Robust z
## (Intercept)  0.64576265 0.41037540  1.573590  0.39567302  1.632061
## age         -0.01201525 0.01135004 -1.058609  0.01173464 -1.023912
##
## Estimated Scale Parameter:  1
## Number of Iterations:  1
##
## Working Correlation
##            [,1]      [,2]      [,3]      [,4]
## [1,] 1.0000000 0.5279328 0.5279328 0.5279328
## [2,] 0.5279328 1.0000000 0.5279328 0.5279328
## [3,] 0.5279328 0.5279328 1.0000000 0.5279328
## [4,] 0.5279328 0.5279328 0.5279328 1.0000000
```

Table 10.7

```
formula3 <- status ~ age + month + baseline + centre + treatment + gender
resp_gee3 <- gee(formula3, data = RESPL, family = binomial, id = subject,
                 corstr = "exchangeable", scale.fix = TRUE, scale.value = 1)
```

```
## Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27

## running glm to get initial regression estimate

##       (Intercept)                age             month
##       -0.74049186         -0.01818538        -0.06425079
##          baseline             centre2 treatmenttreatment
##        1.88406115          0.67230658         1.30055737
##      genderfemale
##        0.11936743
```

```
summary(resp_gee3)
```

```
##
##  GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
##  gee S-function, version 4.13 modified 98/01/27 (1998)
##
## Model:
##  Link:                      Logit
##  Variance to Mean Relation: Binomial
##  Correlation Structure:     Exchangeable
##
## Call:
## gee(formula = formula3, id = subject, data = RESPL, family = binomial,
##     corstr = "exchangeable", scale.fix = TRUE, scale.value = 1)
##
## Summary of Residuals:
##         Min          1Q      Median          3Q         Max
## -0.93336064 -0.29731249  0.08709546  0.32904495  0.85591042
##
##
## Coefficients:
##                      Estimate Naive S.E.    Naive z Robust S.E.
## (Intercept)       -0.73323775 0.51777213 -1.4161399  0.51901959
## age               -0.01841147 0.01256999 -1.4647157  0.01301183
## month             -0.06419992 0.08100426 -0.7925499  0.08150945
## baseline           1.87781449 0.34189674  5.4923439  0.35009816
## centre2            0.68091583 0.33960534  2.0050210  0.35676012
## treatmenttreatment 1.29222142 0.33562959  3.8501415  0.35053559
## genderfemale       0.13076523 0.41791342  0.3129003  0.44409802
##                     Robust z
## (Intercept)       -1.4127362
## age               -1.4149795
## month             -0.7876377
## baseline           5.3636799
## centre2            1.9086097
## treatmenttreatment 3.6864200
## genderfemale       0.2944513
##
```

```
## Estimated Scale Parameter:  1
## Number of Iterations:  2
##
## Working Correlation
##           [,1]     [,2]     [,3]     [,4]
## [1,] 1.000000 0.336918 0.336918 0.336918
## [2,] 0.336918 1.000000 0.336918 0.336918
## [3,] 0.336918 0.336918 1.000000 0.336918
## [4,] 0.336918 0.336918 0.336918 1.000000
```

Table 10.10

```r
# install.packages("lme4")
library("lme4")
```

```
## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following object is masked from 'package:tidyr':
##
##     expand
```

```r
formula4 <- status ~ age + month + baseline + centre + treatment + gender + (1 | subject)
resp_glmer <- glmer(formula4, data = RESPL, family = binomial)
summary(resp_glmer)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: status ~ age + month + baseline + centre + treatment + gender +
##     (1 | subject)
##    Data: RESPL
##
##      AIC      BIC   logLik deviance df.resid
##    444.3    477.1   -214.2    428.3      436
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.8574 -0.3590  0.1427  0.3693  2.2393
##
## Random effects:
##  Groups  Name        Variance Std.Dev.
##  subject (Intercept) 3.89     1.972
## Number of obs: 444, groups:  subject, 111
##
## Fixed effects:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -1.39580    0.83828  -1.665 0.095895 .
## age                 -0.02546    0.02014  -1.264 0.206131
## month               -0.10133    0.12518  -0.810 0.418221
## baseline             3.07833    0.60271   5.107 3.26e-07 ***
## centre2              1.04669    0.54784   1.911 0.056059 .
## treatmenttreatment   2.16324    0.55643   3.888 0.000101 ***
## genderfemale         0.20250    0.67270   0.301 0.763394
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) age    month  baseln centr2 trtmnt
## age        -0.670
## month      -0.356  0.009
## baseline   -0.349  0.015 -0.041
## centre2    -0.078 -0.223 -0.015 -0.150
```

```
## trtmnttrtmn -0.395 -0.050 -0.031  0.301  0.058
## genderfemal -0.008 -0.263 -0.003  0.102 -0.147  0.219
```

Table 10.2: Data from a Clinical Trial of Patients Suffering from Epilepsy

```
library("HSAUR3")
EPIL <- epilepsy # (data are readily given in long form)
glimpse(EPIL); head(EPIL)
```

```
## Observations: 236
## Variables: 6
## $ treatment    <fct> placebo, placebo, placebo, placebo, placebo, plac...
## $ base         <int> 11, 11, 11, 11, 11, 11, 11, 11, 6, 6, 6, 6, 8, 8,...
## $ age          <int> 31, 31, 31, 31, 30, 30, 30, 30, 25, 25, 25, 25, 3...
## $ seizure.rate <int> 5, 3, 3, 3, 3, 5, 3, 3, 2, 4, 0, 5, 4, 4, 1, 4, 7...
## $ period       <ord> 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1...
## $ subject      <fct> 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 5...

##     treatment base age seizure.rate period subject
## 1     placebo   11  31            5      1       1
## 110   placebo   11  31            3      2       1
## 112   placebo   11  31            3      3       1
## 114   placebo   11  31            3      4       1
## 2     placebo   11  30            3      1       2
## 210   placebo   11  30            5      2       2
```

```
# convert data to WIDE form:
# (sep: set valid names month0:month4 here! otherwise just 0:4)

EPI <- EPIL %>%
  spread(key = period, value = seizure.rate, sep = "") %>%
  arrange(subject)

glimpse(EPI); head(EPI)
```

```
## Observations: 59
## Variables: 8
## $ treatment <fct> placebo, placebo, placebo, placebo, placebo, placebo...
## $ base      <int> 11, 11, 6, 8, 66, 27, 12, 52, 23, 10, 52, 33, 18, 42...
## $ age       <int> 31, 30, 25, 36, 22, 29, 31, 42, 37, 28, 36, 24, 23, ...
## $ subject   <fct> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1...
## $ period1   <int> 5, 3, 2, 4, 7, 5, 6, 40, 5, 14, 26, 12, 4, 7, 16, 11...
## $ period2   <int> 3, 5, 4, 4, 18, 2, 4, 20, 6, 13, 12, 6, 4, 9, 24, 0,...
## $ period3   <int> 3, 3, 0, 1, 9, 8, 0, 23, 6, 6, 6, 8, 6, 12, 10, 0, 3...
## $ period4   <int> 3, 3, 5, 4, 21, 7, 2, 12, 5, 0, 22, 4, 2, 14, 9, 5, ...

##   treatment base age subject period1 period2 period3 period4
## 1   placebo   11  31       1       5       3       3       3
## 2   placebo   11  30       2       3       5       3       3
## 3   placebo    6  25       3       2       4       0       5
## 4   placebo    8  36       4       4       4       1       4
## 5   placebo   66  22       5       7      18       9      21
## 6   placebo   27  29       6       5       2       8       7
```

```
# correct the baseline when converting to long form:

EPIL <- gather(EPI, key = periods, value = seizure.rate, period1:period4) %>%
```

```
  mutate( base2 = base/4, week = as.integer(substr(periods, 7, 7)) ) %>%
  arrange(subject)

glimpse(EPIL); head(EPIL)
```

```
## Observations: 236
## Variables: 8
## $ treatment    <fct> placebo, placebo, placebo, placebo, placebo, plac...
## $ base         <int> 11, 11, 11, 11, 11, 11, 11, 11, 6, 6, 6, 6, 8, 8,...
## $ age          <int> 31, 31, 31, 31, 30, 30, 30, 30, 25, 25, 25, 25, 3...
## $ subject      <fct> 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 5...
## $ periods      <chr> "period1", "period2", "period3", "period4", "peri...
## $ seizure.rate <int> 5, 3, 3, 3, 3, 5, 3, 3, 2, 4, 0, 5, 4, 4, 1, 4, 7...
## $ base2        <dbl> 2.75, 2.75, 2.75, 2.75, 2.75, 2.75, 2.75, 2.75, 1...
## $ week         <int> 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1...

##   treatment base age subject periods seizure.rate base2 week
## 1   placebo   11  31       1 period1            5  2.75    1
## 2   placebo   11  31       1 period2            3  2.75    2
## 3   placebo   11  31       1 period3            3  2.75    3
## 4   placebo   11  31       1 period4            3  2.75    4
## 5   placebo   11  30       2 period1            3  2.75    1
## 6   placebo   11  30       2 period2            5  2.75    2
```

Figure 10.3

```r
# Convert data to long form, including the baseline;
# correct the baseline first:
EPI0 <- EPI %>%
  mutate(base2 = base/4)

EPIL0 <- gather(EPI0, key = periods, value = seizure.rate, base2, period1:period4) %>%
  arrange(subject)

glimpse(EPIL0); head(EPIL0)
```

```
## Observations: 295
## Variables: 6
## $ treatment    <fct> placebo, placebo, placebo, placebo, placebo, plac...
## $ base         <int> 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 6, 6, 6, ...
## $ age          <int> 31, 31, 31, 31, 31, 30, 30, 30, 30, 30, 25, 25, 2...
## $ subject      <fct> 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4...
## $ periods      <chr> "base2", "period1", "period2", "period3", "period...
## $ seizure.rate <dbl> 2.75, 5.00, 3.00, 3.00, 3.00, 2.75, 3.00, 5.00, 3...

##   treatment base age subject periods seizure.rate
## 1   placebo   11  31       1   base2         2.75
## 2   placebo   11  31       1 period1         5.00
## 3   placebo   11  31       1 period2         3.00
## 4   placebo   11  31       1 period3         3.00
## 5   placebo   11  31       1 period4         3.00
## 6   placebo   11  30       2   base2         2.75
```

```r
# Mark some of the outliers:
# https://stackoverflow.com/questions/33524669/labeling-outliers-of-boxplots-in-r
is_outlier <- function(x) {
  # changed the upper coeff from 1.5 to 3.5 (only more extreme observations highlighted)
  return(x < quantile(x, 0.25) - 1.5 * IQR(x) | x > quantile(x, 0.75) + 3.5 * IQR(x))
}
# use subject id (not seizure.rate itself!) for identification:
EPIL0 <- EPIL0 %>%
  mutate(outlier = ifelse(is_outlier(seizure.rate), subject, as.numeric(NA)))

p1 <- ggplot(EPIL0, aes(x = factor(periods), y = seizure.rate, fill = treatment))
p2 <- p1 + geom_boxplot() + geom_text(aes(label = outlier), na.rm = TRUE, hjust = -0.6)
p3 <- p2 + facet_grid(treatment ~ ., labeller = label_parsed)
p4 <- p3 + theme_bw() + theme(panel.grid.major = element_blank(),
                              panel.grid.minor = element_blank())
p5 <- p4 + theme(legend.position = "none")
p6 <- p5 + scale_x_discrete(name = "",
                            labels = c("Baseline", "Time 1", "Time 2", "Time 3", "Time 4"))
p7 <- p6 + scale_y_continuous(name = "Seizure rate (count)")
p8 <- p7 + scale_fill_grey(start = 1, end = 1) # B&W for the book
p7; p8
```
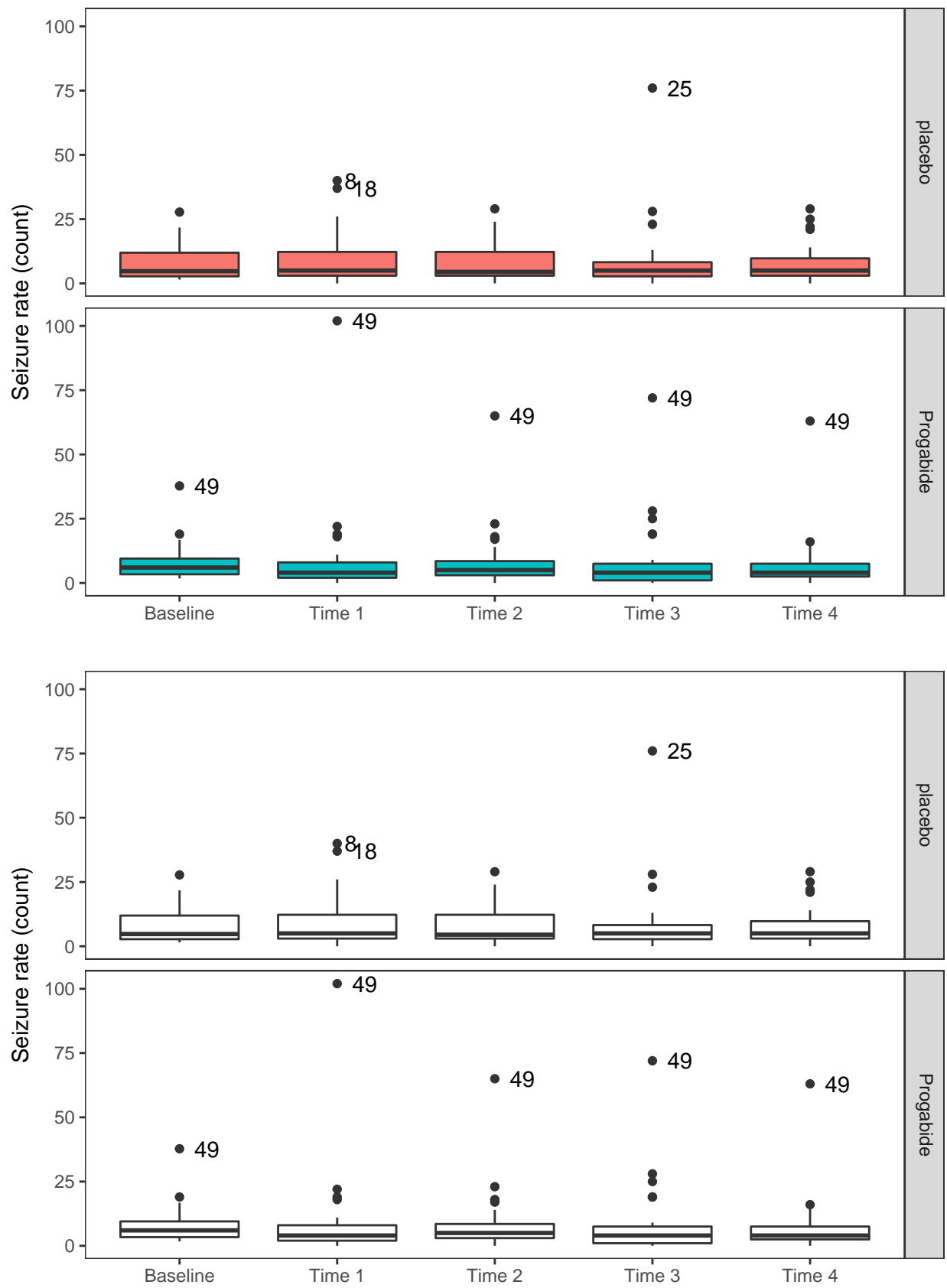
Table 10.8

```
t1 <- tapply(EPIL0$seizure.rate, EPIL0$period, mean)
t2 <- tapply(EPIL0$seizure.rate, EPIL0$period, var)
t1 <- as.data.frame(t1)
t2 <- as.data.frame(t2)

tab108 <- cbind(t1, t2)
names(tab108) = c("mean", "variance")
tab108
```

```
##              mean   variance
## base2    7.805085   45.14885
## period1 8.949153 220.08358
## period2 8.355932 103.78492
## period3 8.440678 200.18177
## period4 7.305085  93.11222
```

Table 10.9

```
formula5 <- seizure.rate ~ week + age + log(base2) + treatment
epil_gee1 <- gee(formula5, data = EPIL, family = "poisson", id = subject,
                 corstr = "exchangeable", scale.fix = FALSE)
```

```
## Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27

## running glm to get initial regression estimate

##        (Intercept)                week                 age
##       -0.84405973          -0.05872397          0.01932016
##         log(base2) treatmentProgabide
##         1.22082626          -0.03005351
```

```
summary(epil_gee1)
```

```
##
##  GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
##  gee S-function, version 4.13 modified 98/01/27 (1998)
##
## Model:
##  Link:                     Logarithm
##  Variance to Mean Relation: Poisson
##  Correlation Structure:     Exchangeable
##
## Call:
## gee(formula = formula5, id = subject, data = EPIL, family = "poisson",
##     corstr = "exchangeable", scale.fix = FALSE)
##
## Summary of Residuals:
##        Min          1Q      Median          3Q         Max
## -15.5346018  -3.2421580  -0.3594229   1.4768610  60.2253586
##
##
## Coefficients:
##                       Estimate Naive S.E.    Naive z Robust S.E.
## (Intercept)        -0.87530135 0.52946461 -1.6531820 0.473365597
## week               -0.05872335 0.03459127 -1.6976352 0.034997602
## age                 0.01999638 0.01250451  1.5991331 0.009792718
## log(base2)          1.22469907 0.10554688 11.6033662 0.155673659
## treatmentProgabide -0.02421150 0.15532727 -0.1558741 0.191087900
##                      Robust z
## (Intercept)        -1.8491022
## week               -1.6779251
## age                 2.0419638
## log(base2)          7.8670925
## treatmentProgabide -0.1267035
##
## Estimated Scale Parameter:  4.822177
## Number of Iterations:  2
##
## Working Correlation
##            [,1]      [,2]      [,3]      [,4]
## [1,] 1.0000000 0.3986404 0.3986404 0.3986404
```

```
## [2,] 0.3986404 1.0000000 0.3986404 0.3986404
## [3,] 0.3986404 0.3986404 1.0000000 0.3986404
## [4,] 0.3986404 0.3986404 0.3986404 1.0000000
```

Table 10.11

```
formula6 <- seizure.rate ~ week + age + log(base2) + treatment + (1 | subject)
epil_glmm <- glmer(formula6, data = EPIL, family = "poisson")
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge with max|grad| = 0.00241629 (tol =
## 0.001, component 1)
```

```
summary(epil_glmm)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##    Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula:
## seizure.rate ~ week + age + log(base2) + treatment + (1 | subject)
##    Data: EPIL
##
##      AIC      BIC   logLik deviance df.resid
##   1346.2   1367.0   -667.1   1334.2      230
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2981 -0.8756 -0.0889  0.6043  7.2421
##
## Random effects:
##  Groups  Name        Variance Std.Dev.
##  subject (Intercept) 0.268    0.5177
## Number of obs: 236, groups:  subject, 59
##
## Fixed effects:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -0.17533    0.43865  -0.400  0.68937
## week              -0.05872    0.02018  -2.910  0.00362 **
## age                0.01073    0.01217   0.882  0.37800
## log(base2)         1.02527    0.10115  10.136  < 2e-16 ***
## treatmentProgabide -0.31998    0.15060  -2.125  0.03362 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) week   age    lg(b2)
## week        -0.112
## age         -0.872  0.000
## log(base2)  -0.545  0.000  0.161
## trtmntPrgbd -0.243  0.000  0.110 -0.043
## convergence code: 0
## Model failed to converge with max|grad| = 0.00241629 (tol = 0.001, component 1)
```

Table 10.12

```
formula7 <- seizure.rate ~ week + age + log(base2) + treatment + (1 + week | subject)
epil_glmm2 <- glmer(formula7, data = EPIL, family = "poisson")
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge with max|grad| = 0.0012166 (tol =
## 0.001, component 1)
```

```
summary(epil_glmm2)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula: seizure.rate ~ week + age + log(base2) + treatment + (1 + week |
##     subject)
##    Data: EPIL
##
##      AIC      BIC   logLik deviance df.resid
##   1330.4   1358.1   -657.2   1314.4      228
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9840 -0.7807 -0.1003  0.5385  6.3781
##
## Random effects:
##  Groups  Name        Variance Std.Dev. Corr
##  subject (Intercept) 0.4134   0.6429
##          week        0.0213   0.1459   -0.60
## Number of obs: 236, groups:  subject, 59
##
## Fixed effects:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -0.18761    0.44351  -0.423   0.6723
## week               -0.05135    0.03264  -1.573   0.1157
## age                 0.01001    0.01233   0.812   0.4166
## log(base2)          1.02520    0.10085  10.165   <2e-16 ***
## treatmentProgabide -0.31674    0.15018  -2.109   0.0349 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) week   age    lg(b2)
## week       -0.136
## age        -0.864 -0.051
## log(base2) -0.537 -0.019  0.162
## trtmntPrgbd -0.243 -0.004  0.113 -0.042
## convergence code: 0
## Model failed to converge with max|grad| = 0.0012166 (tol = 0.001, component 1)
```