Snijders, Tom A.B., and Bosker, Roel J. 2012

Kapitel 17 -

shs

19 1 2017

Contents


```
source("C:/Users/shs/Desktop/pw.R")
```

```
library("AzureML")
ws <- workspace(
  id           = id_shs,
  auth         = auth_shs,
  api_endpoint = "https://studioapi.azureml.net"
)

ds1 <- download.datasets(
  dataset = ws,
  name = "rel_level1.txt"
)

ds2 <- download.datasets(
  dataset = ws,
  name = "rel_level2.txt"
)

ds3 <- download.datasets(
  dataset = ws,
  name = "loop.csv"
)



level1 <- read.table(textConnection(ds1), header=T)
level2 <- read.table(textConnection(ds2), header=T)
loop   <- ds3

###############################################
```

```
start <- Sys.time()

# What have we got?
names(level1)

##  [1] "COUNTRY"               "religiousattendance"   "educationallevel"
##  [4] "income"                "unemployed"            "religiousparents"
##  [7] "loglocalurbanization" "FEMALE"                "SINGLE"
## [10] "DIVORCED"              "widowed"

dim(level1)

## [1] 136611     11

names(level2)

## [1] "COUNTRY"                   "religiousregulation"
## [3] "gini"                      "tertiaryschoolenrollment"
## [5] "urbanization"

dim(level2)

## [1] 60  5
```

Example 17.1

```
# table of religiosity by country
ra.by.c <- table(level1$religiousattendance,by = level1$COUNTRY)
dim(ra.by.c)

## [1]  2 60

# Chi-squared test
(chi2 <- chisq.test(ra.by.c))

##
##  Pearson's Chi-squared test
##
## data:  ra.by.c
## X-squared = 29733, df = 59, p-value < 2.2e-16

# Average
p <- sum(ra.by.c[2,])/sum(ra.by.c)

# Proportions by country
props <- ra.by.c[2,]/(ra.by.c[2,] + ra.by.c[1,])
# The data for Figure 17.1.
props

##          Argentina           Armenia         Australia           Austria
##        0.292292292       0.076000000       0.165200391       0.256229081
##         Azerbaijan        Bangladesh           Belarus           Belgium
##        0.059940060       0.635409836       0.058728099       0.233816014
## BosniaHerzegovina            Brazil          Bulgaria            Canada
##        0.310833333       0.346639372       0.074350649       0.269186047
##              Chile             China          Colombia           Croatia
##        0.266506603       0.005706134       0.455445545       0.265060241
##      CzechRepublic           Denmark DominicanRepublic        EastGermany
##        0.110537504       0.025527737       0.446043165       0.086769231
##            Estonia           Finland            France           Georgia
```

```
##       0.036031589       0.044955045       0.087122660       0.097608025
##             Ghana            Greece           Hungary           Iceland
##       0.789473684       0.133757962       0.142142142       0.032024793
##             India           Ireland             Italy             Japan
##       0.489044289       0.733879222       0.392127554       0.028196403
##            Latvia         Lithuania        Luxembourg         Macedonia
##       0.050547328       0.149481993       0.201550388       0.109547739
##             Malta            Mexico MoldovaRepublicOf       Netherlands
##       0.828000000       0.448878628       0.108739837       0.170953101
##           Nigeria   NorthernIreland            Norway              Peru
##       0.873528422       0.480891720       0.051129101       0.430222956
##            Poland          Portugal           Romania RussianFederation
##       0.577282851       0.382608696       0.217874611       0.025073066
##          Slovakia          Slovenia             Spain            Sweden
##       0.404958678       0.201780415       0.274757908       0.041431262
##       Switzerland            Turkey           Ukraine     UnitedKingdom
##       0.156846473       0.390752493       0.099350974       0.143430291
##      UnitedStates           Uruguay         Venezuela       WestGermany
##       0.440927609       0.132000000       0.309166667       0.160273005
```

```r
# Note that the smallest value
min(props)
```

```
## [1] 0.005706134
```

```r
# is 0.006, obtained for China - overlooked in the book!
```

```r
# To calculate tau-hat, we follow the calculations of p. 292-293
# Total sample size
ntot <- sum(ra.by.c)
# Number of countries
nco  <- dim(ra.by.c)[2]
# Sample sizes by country
nj <- colSums(ra.by.c)
# Their variance
s2nj <- var(nj)
# Formula (3.7), also see p. 292.
(ntilde <- (ntot - sum(nj*nj)/ntot)/(nco-1))
```

```
## [1] 2263.186
```

```r
# A different way of calculating the same
(ntilde <- (ntot/nco) - (var(nj)/ntot))
```

```
## [1] 2263.186
```

```r
# S-squared-between
(s2_b <- p*(1-p)*(chi2$statistic)/(ntilde*(nco-1)))
```

```
##  X-squared
## 0.04043574
```

```r
# S-squared-within
(s2_w <- (sum(ra.by.c[1,]*ra.by.c[2,]/nj))/(ntot-nco))
```

```
## [1] 0.1421351
```

```r
# tau-hat-squared
tau2 <- s2_b - (s2_w/ntilde)
```

```r
# The estimated between-country standard deviation
sqrt(tau2)
```

```
## X-squared
## 0.2009302
```

Example 17.2

```r
# Calculate odds
ods <- ra.by.c[2,]/ra.by.c[1,]
```

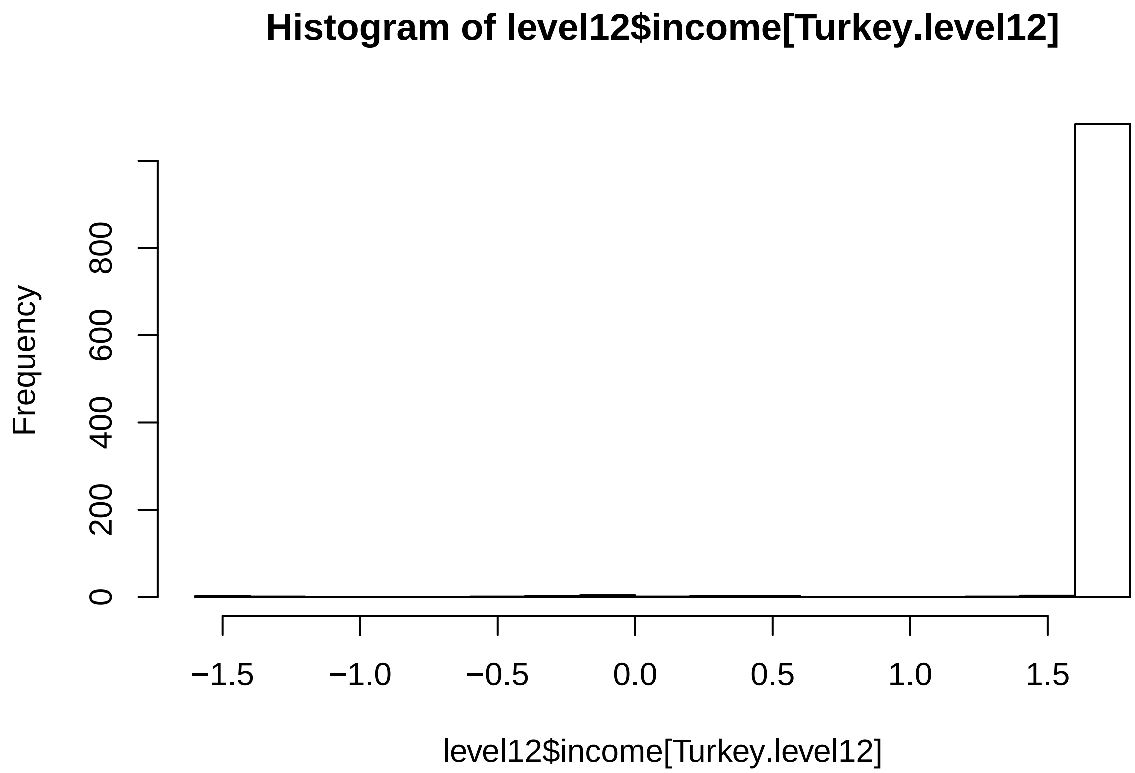
```r
# The log-odds for Figure 17.5.
lods <- log(ra.by.c[2,]/ra.by.c[1,])
# drop the names
names(lods) <- c(1:60)
lods
```

```
##           1           2           3           4           5           6
## -0.88427686 -2.49797873 -1.62003248 -1.06566119 -2.75259858  0.55549632
##           7           8           9          10          11          12
## -2.77431375 -1.18688779 -0.79622628 -0.63384427 -2.52170309 -0.99875615
##          13          14          15          16          17          18
## -1.01241958 -5.16049105 -0.17869179 -1.01983141 -2.08526248 -3.64213047
##          19          20          21          22          23          24
## -0.21667104 -2.35373653 -3.28666249 -3.05609542 -2.34928452 -2.22408928
##          25          26          27          28          29          30
##  1.32175584 -1.86813245 -1.79761086 -3.40869608 -0.04382986  1.01439421
##          31          32          33          34          35          36
## -0.43837789 -3.53995932 -2.93297560 -1.73866965 -1.37663245 -2.09536907
##          37          38          39          40          41          42
##  1.57151868 -0.20520253 -2.10367816 -1.57888747  1.93252307 -0.07647036
##          43          44          45          46          47          48
## -2.92091892 -0.28094156  0.31162908 -0.47849024 -1.27809535 -3.66056834
##          49          50          51          52          53          54
## -0.38484582 -1.37520367 -0.97061514 -3.14140556 -1.68188161 -0.44415025
##          55          56          57          58          59          60
## -2.20445688 -1.78708656 -0.23739826 -1.88338979 -0.80401809 -1.65619819
```

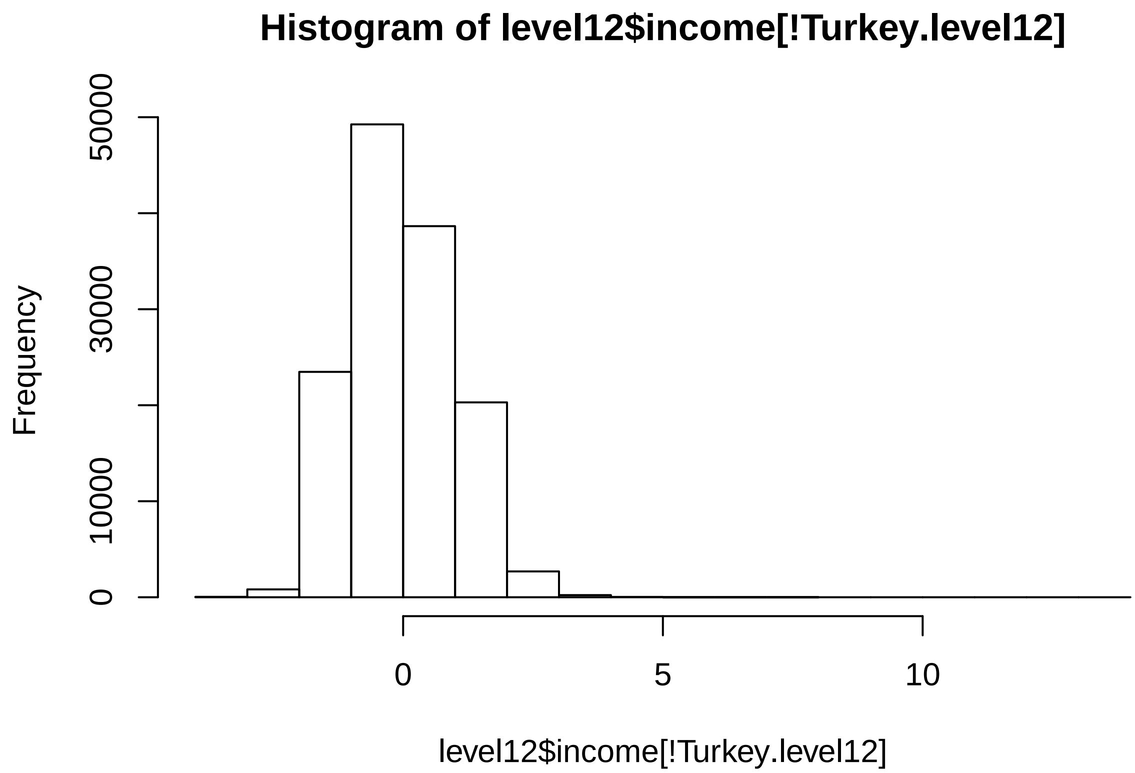
```r
# Data manipulations
level1 <- droplevels(level1)
level2 <- droplevels(level2)
# First merge the two data sets
level12 <- merge(level1,level2)
dim(level12)
```

```
## [1] 136611     15
```

```r
# Something is strange with the data from Turkey
Turkey.level12 <- level12$COUNTRY == "Turkey"
hist(level12$income[Turkey.level12])
```



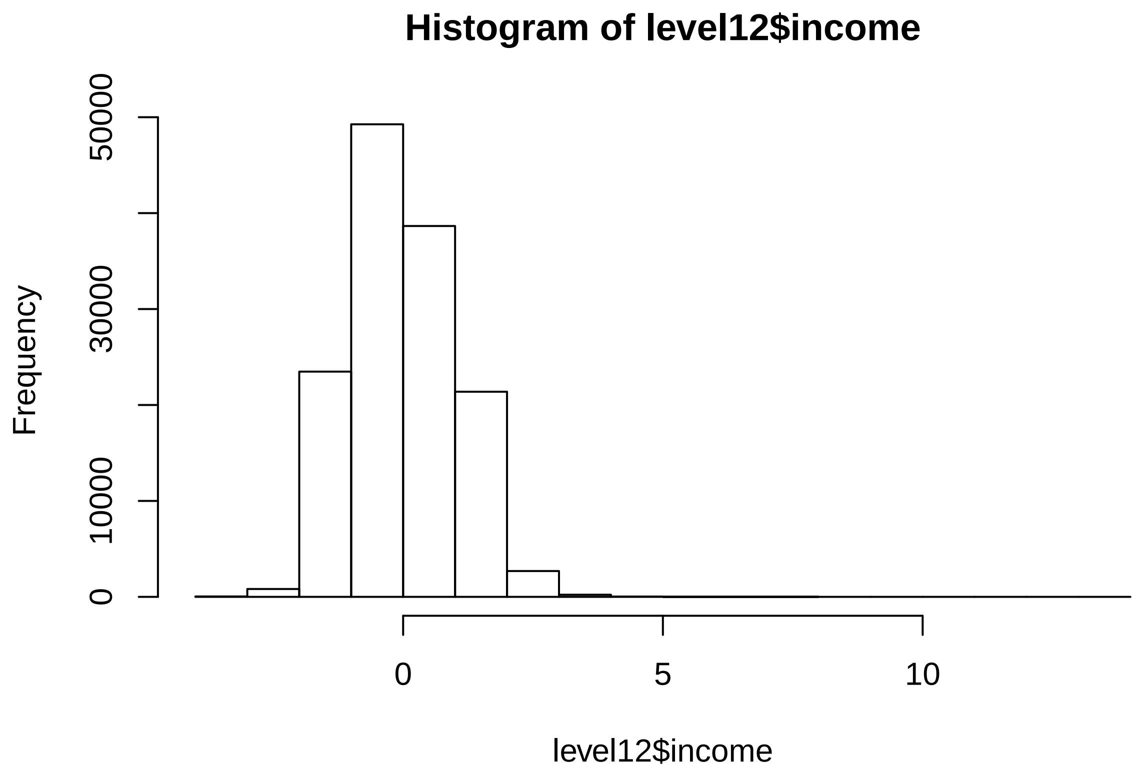

```
hist(level12$income[!Turkey.level12])
```




```
# Therefore we drop this country
level12_nT <- level12[!Turkey.level12,]
dim(level12_nT)
```

```
## [1] 135508     15
```

```
# For income, there are some outliers
hist(level12$income)
```




```
# We truncate income at 3
sum(level12_nT$income > 3)
```

```
## [1] 276
```

```
level12_nT$income <- ifelse(level12_nT$income > 3, 3, level12_nT$income)
sum(level12_nT$income > 3)
```

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

```
sum(level12_nT$income >= 3)
```

```
## [1] 276
```

mlm0

```
# The multilevel logistic models will be estimated using lme4.
library(lme4)
```

```
## Loading required package: Matrix
```

```
# Table 17.1
summary(mlm0 <- glmer(religiousattendance ~ (1|COUNTRY),
          family = binomial, data=level12_nT))
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: religiousattendance ~ (1 | COUNTRY)
```

```
##    Data: level12_nT
##
##      AIC      BIC   logLik deviance df.resid
## 119290.2 119309.8 -59643.1 119286.2   135506
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.6210 -0.5524 -0.3085 -0.1610 10.9811
##
## Random effects:
##  Groups  Name        Variance Std.Dev.
##  COUNTRY (Intercept) 1.895    1.377
## Number of obs: 135508, groups:  COUNTRY, 59
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.4462     0.1521  -9.508   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Estimated average log-odds is
(b0 <- fixef(mlm0))
```

```
## (Intercept)
##   -1.446213
```

```
# which transformed to a probability is
(p0 <- exp(b0)/(1+exp(b0)))
```

```
## (Intercept)
##   0.1905851
```

```
# The estimated level-2 variance is
(tau00 <- VarCorr(mlm0)$COUNTRY[1,1])
```

```
## [1] 1.895305
```

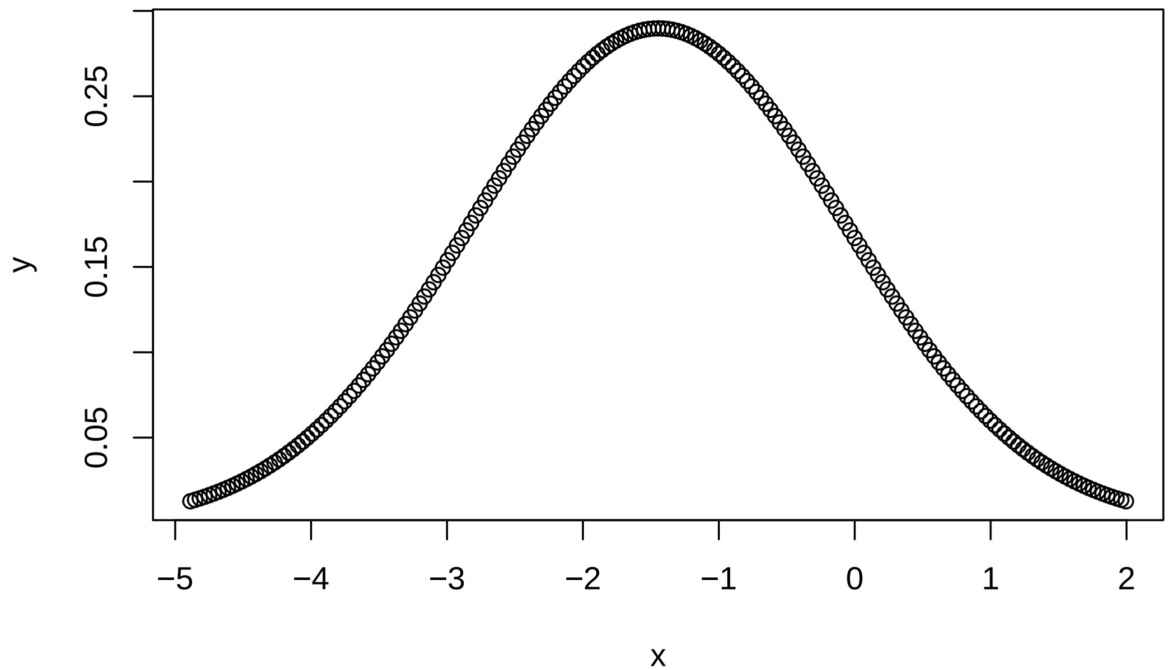
```
# with corresponding standard deviation
(tau0 <- sqrt(tau00))
```

```
## [1] 1.376701
```

```
# Approximation formula (17.13) yields
(var0 <- tau00*p0*(1-p0)*p0*(1-p0))
```

```
## (Intercept)
##  0.04510235
```

```
# The normal density in Figure 17.5 can be obtained by
x <- tau0*c(-100:100)/40+b0
y <- dnorm(x,mean=b0,sd=tau0)
plot(x,y)
```



Example 17.3

```
# First calculation of some country-level averages

edu.ave <- ave(level12_nT$educationallevel,level12_nT$COUNTRY)
inc.ave <- ave(level12_nT$income,level12_nT$COUNTRY)
unempl.ave <- ave(level12_nT$unemployed,level12_nT$COUNTRY)
single.ave <- ave(level12_nT$SINGLE,level12_nT$COUNTRY)
div.ave <- ave(level12_nT$DIVORCED,level12_nT$COUNTRY)
wid.ave <- ave(level12_nT$widowed,level12_nT$COUNTRY)
urb.ave <- ave(level12_nT$loglocalurbanization,level12_nT$COUNTRY)
```

```
# Deviation scores; use 17 as provisional centering constant
edumin <- level12_nT$educationallevel - 17
edumin.ave <- edu.ave - 17
eduminmin <- edumin - edumin.ave
level12_nT$gini <- level12_nT$gini - 35
level12_nT$loglocalurbanization <- level12_nT$loglocalurbanization - 10
```

```
# Make a function for easy display of mean and s.d.
d <- function(x){(c(mean(x),sqrt(var(x))))}
# Means and s.d.s of dependent and explanatory variables
d(level12_nT$religiousattendance)
```

```
## [1] 0.2372185 0.4253789
```

```
d(eduminmin)
```

```
## [1] -1.578521e-16  2.479466e+00
d(level12_nT$income)
## [1] -0.03415624  0.99021284
d(level12_nT$unemployed)
## [1] 0.1887707 0.3857990
d(level12_nT$FEMALE)
## [1] 0.524345 0.499287
d(level12_nT$SINGLE)
## [1] 0.2253711 0.4170631
d(level12_nT$DIVORCED)
## [1] 0.06492623 0.24536712
d(level12_nT$widowed)
## [1] 0.07915857 0.26904537
d(level12_nT$loglocalurbanization)
## [1] 0.09015702 2.18033215
d(level12_nT$gini)
## [1] 0.1016302 9.5839313
d(edumin.ave)
## [1] 0.8157685 0.9382561
d(unempl.ave)
## [1] 0.1887707 0.0846463
d(div.ave)
## [1] 0.06492623 0.03454115
```

mlm1

```
# Table 17.2
summary(mlm1 <- glmer(religiousattendance ~  eduminmin + income
    + unemployed + FEMALE + SINGLE + DIVORCED + widowed
    + loglocalurbanization
    + gini +  edumin.ave +  unempl.ave
   + div.ave +  (1 |COUNTRY) ,
          family = binomial, data=level12_nT))
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge with max|grad| = 0.00179243 (tol =
## 0.001, component 1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: religiousattendance ~ eduminmin + income + unemployed + FEMALE +
##     SINGLE + DIVORCED + widowed + loglocalurbanization + gini +
```

```
##     edumin.ave + unempl.ave + div.ave + (1 | COUNTRY)
##    Data: level12_nT
##
##      AIC      BIC   logLik deviance df.resid
## 115997.8 116135.2 -57984.9 115969.8   135494
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.1870 -0.5147 -0.2871 -0.1201 13.3946
##
## Random effects:
##  Groups  Name        Variance Std.Dev.
##  COUNTRY (Intercept) 1.177    1.085
## Number of obs: 135508, groups:  COUNTRY, 59
##
## Fixed effects:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -2.068961   0.382261   -5.41 6.22e-08 ***
## eduminmin            -0.029032   0.003208   -9.05  < 2e-16 ***
## income               -0.063819   0.008236   -7.75 9.28e-15 ***
## unemployed            0.018059   0.019510    0.93  0.35462
## FEMALE                0.508078   0.016020   31.72  < 2e-16 ***
## SINGLE               -0.269316   0.019090  -14.11  < 2e-16 ***
## DIVORCED             -0.489450   0.035799  -13.67  < 2e-16 ***
## widowed               0.517557   0.026708   19.38  < 2e-16 ***
## loglocalurbanization -0.066492   0.003929  -16.92  < 2e-16 ***
## gini                  0.035117   0.016404    2.14  0.03230 *
## edumin.ave           -0.358993   0.134557   -2.67  0.00763 **
## unempl.ave            6.031720   1.559317    3.87  0.00011 ***
## div.ave              -7.110607   1.796962   -3.96 7.59e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation matrix not shown by default, as p = 13 > 12.
## Use print(x, correlation=TRUE)  or
##   vcov(x)     if you need it
## convergence code: 0
## Model failed to converge with max|grad| = 0.00179243 (tol = 0.001, component 1)
# The coefficient for edumin.ave is different from the Table in the book.
# I (T.S.) do not know what happened.
# Let us assume it was a transcription error.
```

mlm2

```
#Table 17.3
# (the heading of the table should be "Logistic random slopes model" (etc.))
########## summary(mlm2 <- lmer(religiousattendance ~  eduminmin + income
##########    + unemployed + FEMALE + SINGLE + DIVORCED + widowed
##########    + loglocalurbanization
##########    + gini +  edumin.ave +  unempl.ave
##########   + div.ave +  (income + eduminmin |COUNTRY) ,
##########          family = binomial, data=level12_nT))
```

```
##########
# This gives different numerical results than Table 17.3;
# the intercept parameter is quite different, because
# a different centering is applied.
# The other parameter estimates and standard errors are very similar.
# This model is a quite complicated model,
# and estimating it is hard, and sensitive to minor details
# of the specification.
# The centering choices applied in this r script
# gives somewhat better stability than those that produced
# Table 17.3 in the book.
```

Example 17.5

```
# First we have to compute the linear predictor.
# The matrix of explanatory variables (the "design matrix")
# is available for mer objects produced by lme4 as
X1 <- getME(mlm1,"X")
# The parameter estimates for the fixed effects are available as
(beta1 <- fixef(mlm1))
```

```
##          (Intercept)             eduminmin                income
##          -2.06896092           -0.02903164           -0.06381936
##           unemployed                FEMALE                SINGLE
##           0.01805938            0.50807824           -0.26931567
##             DIVORCED               widowed loglocalurbanization
##          -0.48944975            0.51755696           -0.06649177
##                 gini            edumin.ave            unempl.ave
##           0.03511653           -0.35899266            6.03172032
##              div.ave
##          -7.11060716
```

```
# The linear predictor, i.e., linear combination of the rows of X1
# with weights being the estimated fixed effect parameters, is
pred1 <- X1 %*% beta1
# and has variance
(sigma2_F <- var(pred1))
```

```
##          [,1]
## [1,] 1.048072
```

```
# The explained variance according to formula (17.22) is
sigma2_F/(sigma2_F + VarCorr(mlm1)$COUNTRY[1,1] + pi^2/3)
```

```
##           [,1]
## [1,] 0.1900545
```

```
stop <- Sys.time()
stop-start
```

```
## Time difference of 25.4258 mins
```

Example 17.6

```
# Read data
#loop <- read.table("LOOPDIC.DAT",header=FALSE)
```

```r
# What do we have?
dim(loop)
```

```
## [1] 3432   10
```

```r
colSums(loop)
```

```
##            V1           V2           V3           V4           V5
## 337045581.000     3432.000     2696.000        1.363     2000.000
##            V6           V7           V8           V9          V10
##       171.000     -122.119       98.000     -183.467     -106.175
```

```r
names(loop)[1:3] <- c("school","cons","scisub")
names(loop)[5:6] <- c("gender","minority")
```

Model 1 in Table 17.4

```r
(summary(m1 <- glmer(scisub ~ gender + minority + (1 | school),
                     family = binomial, data=loop)))
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: scisub ~ gender + minority + (1 | school)
##    Data: loop
##
##      AIC      BIC   logLik deviance df.resid
##   3247.6   3272.2  -1619.8   3239.6     3428
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.6588  0.2189  0.3338  0.5483  2.0061
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  school (Intercept) 0.4551   0.6746
## Number of obs: 3432, groups:  school, 240
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   2.4833     0.1101  22.561  < 2e-16 ***
## gender       -1.5151     0.1084 -13.983  < 2e-16 ***
## minority     -0.7291     0.1912  -3.814 0.000137 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) gender
## gender   -0.776
## minority -0.147  0.045
```

```r
# For the explained variation,
# first we have to compute the linear predictor.
# The matrix of explanatory variables (the "design matrix")
# is available for mer objects produced by lme4 as
X1 <- getME(m1,"X")
```

```r
# The parameter estimates for the fixed effects are available as
(beta1 <- fixef(m1))
```

```
## (Intercept)      gender    minority
##   2.4832963  -1.5151412  -0.7291387
```

```r
# The linear predictor, i.e., linear combination of the rows of X1
# with weights being the estimated fixed effect parameters, is
pred1 <- X1 %*% beta1
# and has variance
(sigma2_F <- var(pred1))
```

```
##         [,1]
## [1,] 0.58247
```

```r
# The explained variance according to formula (17.22) is
sigma2_F/(sigma2_F + VarCorr(m1)$school[1,1] + pi^2/3)
```

```
##           [,1]
## [1,] 0.1345982
```

Model 2 in Table 17.5

```r
(summary(m2 <- glm(scisub ~ gender,
                   family = binomial, data=loop)))
```

```
##
## Call:
## glm(formula = scisub ~ gender, family = binomial, data = loop)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1665   0.4485   0.4485   0.8438   0.8438
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.24629    0.08983   25.01   <2e-16 ***
## gender      -1.39661    0.10224  -13.66   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3567.9  on 3431  degrees of freedom
## Residual deviance: 3345.3  on 3430  degrees of freedom
## AIC: 3349.3
##
## Number of Fisher Scoring iterations: 4
```

Model 3 in Table 17.5

```r
(summary(m3 <- glmer(scisub ~ gender + (1 | school),
                     family = binomial, data=loop)))
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
```

```
##  Family: binomial  ( logit )
## Formula: scisub ~ gender + (1 | school)
##    Data: loop
##
##      AIC      BIC   logLik deviance df.resid
##   3259.4   3277.8  -1626.7   3253.4     3429
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.6429  0.2204  0.3323  0.5484  2.1763
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  school (Intercept) 0.4852   0.6966
## Number of obs: 3432, groups:  school, 240
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   2.4361     0.1093   22.29   <2e-16 ***
## gender       -1.5072     0.1080  -13.96   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##        (Intr)
## gender -0.773
```

```
deviance(m3)
```

```
## [1] 2994.569
```

Model 1 in Table 17.4

```
(summary(m4 <- glmer(scisub ~ minority + (1 | school),
                     family = binomial, data=loop)))
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: scisub ~ minority + (1 | school)
##    Data: loop
##
##      AIC      BIC   logLik deviance df.resid
##   3477.9   3496.3  -1736.0   3471.9     3429
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9705  0.3610  0.4344  0.5120  1.8789
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  school (Intercept) 0.3973   0.6303
## Number of obs: 3432, groups:  school, 240
##
## Fixed effects:
```

```
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.44245    0.06626  21.769  < 2e-16 ***
## minority    -0.65453    0.18218  -3.593 0.000327 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr)
## minority -0.179
```

```
deviance(m4)
```

```
## [1] 3232.393
```

```
# The parameter estimates are slightly different compared to
# Tables 17.4 and 17.5, because of the differences
# of the algorithms used by lme4 and MIXOR.
```