Lab 2

Daniel Dema

2025-05-25

```
{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)


library(spida2)   # devtools::install_github('gmonette/spida2')
library(p3d)      # devtools::install_github('gmonette/p3d')
library(car)
library(lattice)
library(latticeExtra)
?Drugs
some( Drugs )

# Which drug seems best to reduce 'neg' symptoms

xyplot( neg ~  year | Subj, Drugs)

dd <- Drugs
gd(3)
dd$id <- reorder( dd$Subj, dd$neg)
gd(3, cex = .9)
xyplot( neg ~ year | id, dd , groups = drug, auto.key =
list(columns=3))

#
# QUESTION 1: ----
#        Can we perform OLS fits on each cluster?
#        Note that the data are balanced with respect to time
#        but not with respect to drugs.
#
#        Also note that Clozapine is more frequently given later in
the study
```

Solution for Question 1:

No, since not all clusters have within-subject variation. For example, F32 is only treated with typical, and M29 is only treated with clozapine. Note that the data is quite unbalanced overall.

```
fit.lm <- lm(neg ~ drug, dd)
summary(fit.lm)
Ld <- Ldiff(fit.lm, 'drug')
wald(fit.lm, Ld)
```

```
Ld <- Ldiff(fit.lm, 'drug', ref = "Atypical")
wald(fit.lm, Ld)

library(nlme)

fit <- lme( neg ~ drug, dd, random = ~ 1 | id )
summary(fit)
wald(fit, -1)

Ld <- Ldiff ( fit, "drug")   # hypothesis matrix to test differences
between drugs
Ld
wald( fit, Ld )

Ld <- Ldiff ( fit,  "drug", ref = "Atypical")
wald( fit, Ld )

fit2 <- update( fit, . ~ . + cvar( drug, id))
summary( fit2 )
wald( fit2, 'cvar')

head( cbind( dd['id'], getX( fit2) ), 18 )

wald( fit2, -1)
wald( fit , -1)

Ld <- Ldiff( fit2, "drug", ref = "Atypical")

wald( fit2, Ld )

fit2l <- update(fit2, . ~ . + year)
summary( fit2l )
ww <-wald( fit2l )
wald( fit2l, 'cvar' )
wald( fit2l, 'drug' )
Ld <- Ldiff( fit2l, "drug", ref = "Atypical")
wald ( fit2l, Ld )

# QUESTION 5: ----
#    How do you explain the differences in the estimation of the
Typical - Clozapine
#    comparison in the 4 analyses:
#
lapply(
  list("pooled" = fit.lm, "no ctx" = fit,"ctx"= fit2,"ctx+year" =
fit2l),
  function( fit ) wald( fit , Ldiff( fit,'drug', ref = "Atypical"))
)
clist <- lapply(
```

```
  list("pooled" = fit.lm, "no ctx" = fit,"ctx"= fit2,"ctx+year" =
fit2l),
  function( fit ) wald( fit , Ldiff( fit,'drug', ref = "Atypical"))
)
clist
do.call( rbind, lapply(clist, function(x) x[[1]][[2]][3,]))
```

Solution for Question 5:

Pooled is the coefficient from regression on the entire pooled data while the other estimates are within-subject (note that we see Simpson's paradox arise here, with the pooled coefficient being negative.). Estimate 3 takes into account between-subject effects while estimate 2 does not (note that we do this since, as mentioned in Question 1, the data is quite unbalanced). Estimate 4 controls for time.

```
## Taking time into account ----
#

fit.my <- update( fit.m, . ~ . + year)

summary( fit.my )   # very significant drop with time

# Previous hypothesis matrix

Lm
wald( fit.my )

# We only need to add year to Lm

Lmy <- Lm

Lmy <- lapply( Lm , function( x ) cbind(x, 0) )

# let's use the average year for predicted values

Lmy[[1]][,6] <- 3.5

Lmy

#
# QUESTION 9:
#
# Should we do the same thing for the second matrix?

> Lm[1]
$predicted
            Clozapine Typical
Atypical  1         0       0 0 0
Clozapine 1         1       0 0 0
```

```
Typical   1         0       1 0 0

> Lm[2]
$differences
                     [,1] [,2] [,3] [,4] [,5]
Clozapine - Atypical    0    1    0    0    0
Typical - Atypical      0    0    1    0    0
Typical - Clozapine     0   -1    1    0    0
```

Solution for Question 9:

No, since Lm[2] gives differences which are time invariant, so we shouldn't try to take year into consideration.