Basic SAS and R for HLM

Edps/Psych/Soc 589

Carolyn J. Anderson

Department of Educational Psychology

ILLINOIS

Overview

The following will be demonstrated in class:

1. Data steps.
2. SAS PROC MIXED.
3. R (here and more in Rmarkdown files))

SAS DATA STEPS

1. Creating a SAS data set.
2. Merging files of different lengths (level 1 & level 2).
3. Creating centered variables.
4. Other.

Creating a SAS data set

```
* HSB dat1 : level 1 responses;

LIBNAME sasdata
    'C:\cja\teaching\hlm\lectures\SAS-MIXED';

DATA sasdata.hsb1;
    INPUT id minority female ses mathach;
    LABEL id='school'
        minority='Student ethnicity (1=minority, 0=not)'
        female ='student gener (1=female, 0=male)'
        ses='standardized scale of student ses'
        mathach='Mathematics achievement';
    DATALINES;
```

Check the Log File!

NOTE: The data set SASDATA.HSB1
has 7185 observations and 5 variables.
NOTE: DATA statement used:
real time 0.09 seconds
cpu time 0.09 seconds

Level 2 Data

Need LIBNAME?

```
DATA sasdata.hsb2;
    INPUT id size sector pracad disclim himinty meanses;
    LABEL id='school id'
        size='school enrollment'
        sector='1=Catholic, 0=public'
        pracad='proportion students in academic track'
        disclim='Disciplinary climate'
        himinty='1=40% minority, 0= ¡40% minority'
        meanses='Mean SES' ;
    DATALINES;
```

Log File

NOTE: The data set SASDATA.HSB2
has 160 observations and 7 variables.
NOTE: DATA statement used:
real time 0.03 seconds
cpu time 0.03 seconds

Merging Two Files: Sort then Merge

- * Make sure data files are correctly sorted;

```
PROC SORT    DATA=sasdata.hsb1;    BY id;
PROC SORT    DATA=sasdata.hsb2;    BY id;
```

- *Merge the two files into a new file;

```
DATA hsball;
      MERGE sasdata.hsb1 sasdata.hsb2;
      BY ID;
```

- Check log and data file:
  PROC PRINT DATA= hsball;
  or
  Use explorer and look at the data file hsball under the folder named "work".

Centering Variables

```
DATA hsbcent;
      SET hsball;
      cSES = SES - meanSES;
RUN;
```

If you don't have a variable that contains mean of desired level 1 variables, then you have to compute it. . . .

Computing Group Means

- Make sure that data are sorted by group:

```
PROC SORT   DATA=hsball;
        BY id;
```

- * Create a file that contains means;

```
PROC MEANSDATA=hsball;
        CLASS id;
        VAR SES;
        OUTPUT OUT=grpmeans MEAN=meanSES;
```

- The new file "grpmeans" will contain special variables _TYPE_ and _FREQ_. The descriptive statistics for _TYPE_= $\cdot$ are over all groups and those for _TYPE_=1 are group means.

Alternative code for computing means

```
PROC MEANS DATA=hsball noprint;
     BY id;
     VAR SES;
     OUTPUT OUT=grpmeans MEAN=meanSES;
```

- noprint $\longrightarrow$ nothing is displayed in output window.
- BY $\longrightarrow$ does not produce (include) the overall mean in the sas file "grpmeans".

Finishing Up

- Merge the files "grpmeans" and "hsball":

```
DATA centHSB;
    MERGE grpmeans hsball;
    BY id;
    IF _TYPE_=. THEN DELETE;
    cSES = SES - meanSES;
```

- If you used the alternative code for computing means:

```
DATA centHSB;
    MERGE grpmeans hsball;
    BY id;
    cSES = SES - meanSES;
```

- Check log and data!

SAS PROC MIXED

Basic Syntax:

1. PROC MIXED options
2. CLASS statement
3. MODEL statement
4. RANDOM statement.
5. TITLE statement.

PROC MIXED Options

```
PROC MIXED DATA=sasdata.hsball NOCLPRINT COVTEST
METHOD=ML;
```

1. DATA=<name of sas data set>
2. NOCLPRINT: don't print classification levels/information.
3. COVTEST: hypothesis tests for variances (& covariances).
4. METHOD: estimation method to use.
   - ML = maximum likelihood
   - REML = restricted maximum likelihood (default)
   - MIVQUE0 = Minimum variance quadratic unbiased estimation. (non-iterative).

CLASS Statement

```
CLASS id gender sector;
```

The CLASS statement

1. Indicates variables that are the "factors" or discrete (nominal), "classification" variables.
2. They may be numeric or character variables.
3. SAS creates dummy codes for them.

Model Statement

```
MODEL math = cSES gender sector / SOLUTION;
```

1. Specify the fixed effects.
2. The response or outcome variable is given to the left of the = sign.
3. Fixed effects are listed on the right side of = sign.
4. The option SOLUTION requests parameter estimates for the fixed effect output.
5. An intercept is included in the model (default).

RANDOM Statement

```
RANDOM intercept / subject= id type=un;
```

- Defines the random effects.
- Must explicitly request random intercept.
- Options:
  - subject= the variable identifying macro units.
  - type=un. Specify the covariance matrix for the random effects (i.e.,) as an "unstructured" general covariance matrix; i.e., square & symmetric.
  - g and gcorr: Requests covariance matrix, , and correlation matrix, respectively, for the random effects (written as a matrix instead of list-wise).

RANDOM Statement (continue)

Options (continued)

- `v` and `vcorr`: Requests covariance matrix, $_j$, and correlation matrix, respectively, for the response variable (i.e., $\boldsymbol{V}_j = \boldsymbol{Z}_j \boldsymbol{T} \boldsymbol{Z}_j' + \sigma^2 \boldsymbol{I}$). Default: SAS gives this for the first marco unit/group.
- `solution`: Requests the empirical Bayes estimates of $_j$.

Output Random Effects to Data File

To output $\hat{U}_j$'s to a SAS DATA file:

```
ODS OUTPUT SolutionR=Ujdata;
```

Obs	Effect	id	Estimate	StdErr Pred	DF	tValue	Probt
1	Intercept	1224	-2.6752	0.8782	7024	-3.05	0.0023
2	Intercept	1288	0.7470	1.1429	7024	0.65	0.5134
3	Intercept	1296	-4.5902	0.8704	7024	-5.27	$< .0001$
4	Intercept	1308	2.9789	1.2502	7024	2.38	0.0172
5	Intercept	1317	0.4963	0.8704	7024	0.57	0.5686

Must include the RANDOM option `SOLUTION`

"Estimate" $= \hat{U}_j$

"StdErr Pred" $=$ standard error of $(\hat{U}_j - U_j)$

Output Fixed Effects to Data File

To output the estimates of the γ's:

```
ODS OUTPUT SolutionF=GammaData;
```

Contents of GammaData:

Obs	Effect	Estimate	StdErr	DF	tValue	Probt
1	Intercept	12.6494	0.2437	159	51.92	$< .0001$
2	cSES	2.1912	0.1086	7024	20.17	$< .0001$

For other statistics that can be output to a SAS file, see documentation for PROC MIXED (Look for table that contains "ODS Tables Produced in PROC MIXED").

Example of Null HLM

```
PROC MIXED DATA=sasdata.hsball NOCLPRINT COVTEST
METHOD=ML;

TITLE 'HSB: null/empty random intercept model';

CLASS id ;

MODEL mathach =     / SOLUTION ;

RANDOM INTERCEPT / SUBJECT = id ;
```

A More Complex Model

```
PROC MIXED DATA=sasdata.hsball
      NOCLPRINT COVTEST METHOD=ML;

TITLE 'HSB: random intercept model, One x';

CLASS id ;

MODEL mathach = cSES     / SOLUTION ;

RANDOM INT / SUBJECT = id ;
```

An Even More Complex Model

```
PROC MIXED DATA=sasdata.hsball NOCLPRINT COVTEST
METHOD=ML;

TITLE 'HSB: random intercept model w/ lots micro and macro';

CLASS id ;

MODEL mathach = cSES minority female meanSES himinty pracad
disclim sector size / SOLUTION ;

RANDOM INTERCEPT / SUBJECT = id ;
```

R Set Up and Data Steps

- Load packages
  - lme4
  - lmerTest
- Set working directory either using tool bar or command

```
setwd('D:< path to where data live>')
```

- Read in level 1 data

```
hsb1 <- read.table(file="HSB1data.txt", header=TRUE)
```

- Read in level 2 data

```
hsb2 <- read.table(file="HSB2data.txt", header=TRUE)
```

R Data Steps

- Merge (don't need to sort because already sorted)

```
hsb <- merge(hsb1,hsb2, by=c('id'))
```

- Take a look at data: `head(hsb)`
- Create any transformation of variables that you want,e.g.,

```
meanfemale <- aggregate(female ~ id, data=hsb,
FUN='mean')
names(meanfemale) <- c('id', 'meanfemale')
hsb <- merge(hsb,meanfemale, by =c('id'))
```

R Data Steps: Centering & Saving

- To center a variable `meanses` is already in the dataset:
  `hsb$ses.centered <- ses - meanses`
- If you want to save the data as a txt file (so that you don't have to go through all steps again)
  `write.table(hsb, 'hsb.txt', row.names=F, na='.')`
- Make variables easier to use:
  `attach(hsb)`

Fitting Model

- Fit a null model
  `model.null ← lmer(mathach ∼ 1 + (1 | id), data=hsb, REML=FALSE)`
- To see the results:
  `summary(model.null)`
- summary from lme4 is returned
  some computational error has occurred in lmerTest Problem?
  Linear mixed model fit by maximum likelihood ['lmerMod']
  Formula: mathach $\sim 1 + (1 \mid \text{id})$
  Data: hsb

AIC	BIC	logLik	deviance	df.resid
47121.8	47142.4	-23557.9	47115.8	7182

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.06262	-0.75365	0.02676	0.76070	2.74184

Random effects:

Groups	Name	Variance	Std.Dev.	
id	(Intercept)	8.553	2.925	$\leftarrow \tau_{00}$
Residual		39.148	6.257	$\leftarrow \sigma^2$

Number of obs: 7185, groups: id, 160

Fixed effects:

	Estimate	Std. Error	t value	
(Intercept)	12.6371	0.2436	51.87	$\leftarrow \gamma_{00}$

ICC

- ICC,
  - cut-and-paste: `icc <- 8.552/(8.553+39.148)`
  - or Use function I wrote:

```
icc.lmer <- function(modl) {
    vars <- as.data.frame(VarCorr(modl))[4]
    total <- sum(vars)
    tau00 <- vars[1,1]
    icc <- tau00/total
    return(icc)
}
```

  - To use it: `icc.lmer(model.null)`
- Yields 0.1793108

Fitting Model

- To get t-tests for fixed effects: `require(lmerTest)`
- A simple model

`model.simple ← lmer(mathach ∼ 1 + ses.centered + (1 | id), data=hsb, REML=FALSE)`

To view the results: `summary(model.simple)`

If you want icc: `icc(model.simple)`

Linear mixed model fit by maximum likelihood t-tests use Satterthwaite approximations to degrees of freedom [lmerMod]

Formula: mathach $1 + \text{ses} + (1 \mid \text{id})$

Data: hsb

AIC	BIC	logLik	deviance	df.resid
46649.0	46676.5	-23320.5	46641.0	7181

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.1263	-0.7277	0.0220	0.7578	2.9186

Random effects:

Groups	Name	Variance	Std.Dev.
id	(Intercept)	4.729	2.175 ← τ_{00}
Residual		37.030	6.085 ← σ^2

Number of obs: 7185, groups: id, 160

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> |t|)
(Intercept)	12.6576	0.1873	149.0000	67.57	¡2e-16 *** ← γ_{00}
ses	2.3915	0.1057	6837.0000	22.63	¡2e-16 *** ← γ_{10}

—

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Correlation of Fixed Effects:

	(Intr)
ses	0.003

Fitting Model

. . . and icc.lmer(model.simple)
0.1132352

A more complex model

```
model.complex ← lmer(mathach ∼ 1 + hsb$ses.centered +
minority + female + meanses + himinty + pracad + disclim +
sector + size + (1 | id), data=hsb, REML=FALSE)
```

- To view results: `summary(model.complex)`
- If you want an icc: `icc(model.complex)`