EdPsy/Psych/Stat 587
Spring 2019
C.J. Anderson

SAS Computer Lab Session 3
Thursday March 7, 2019

In this lab you will

- Fit random intercept and slopes models.
- Test parameters from the marginal model.
- Robust estimation.
- Information criteria.
- Compute R_1^2 and R_2^2.
- Use of the contrast statement.
- Computing the harmonic mean (of sample sizes n_j's).
- Model refinement

For this computer lab, we'll continue to use the USA TIMSS data where science scores are the outcome measure.

1. Re-run your SAS program(s) from Computer labs 1 and 2 (homeworks 1 and 2) except
   - Add the option "IC" to the PROC MIXED statement to produce various information criteria.
   - Add "DDFM= SATTERTH" as an option to the MODEL
   - Add code that you will need for computing R_1^2 and R_2^2 using the sas macro hlmsqr.sas. In particular,
     (a) Run the command
     `%include 'C:\ < put path to macro > \ hlmrsq.sas';`
     OR open up the macro in a program window and run it.
     (b) Add the option `namelen=200` to the PROC MIXED statement.
     (c) Add the option `g` to the RANDOM statement.
     (d) On the line after the RANDOM statement, add the command
     `ods output CovParms=cov G=gmat ModelInfo=mod SolutionF=solf;`

(e) After running PROC MIXED, run the macro using
`%hlmrsq(CovParms=cov,GMatrix=gmat,ModelInfo=mod,SolutionF=solf);`

Note: This macro is a bit finicky. Although it should work, do not use the macro on a random intercept model (they're easy to compute without a macro). Only use it on random slope models.

2. Copy your code for fitting model (s) from computer lab 2 and re-run using the "EMPIRICAL" option as an options to the PROC MIXED statement (and remove the DDFM= SATTERTH option from the model statement—if you don't remove it SAS will just ignore it and issue a warning message). Use ddfm=betwithin. This will produce robust estimates of the standard errors.

Recall that model (s) was

Hierarchical model :

Level 1 :

$$\begin{aligned}\texttt{(science)}_{ij} &= \beta_{0j} + \beta_{1j}\texttt{(grpCmath)}_{ij} + \beta_{2j}\texttt{(gender)}_{ij} + \beta_{3j}\texttt{(grade)}_{ij} \\ &\quad +\beta_{4j}\texttt{(hours\_TV)}_{ij} + \beta_{5j}\texttt{(hours\_computer\_games)}_{ij} + R_{ij}\end{aligned}$$

where $R_{ij} \sim \mathcal{N}(0, \sigma^2)$ and independent.

Level 2 :

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \gamma_{01}\texttt{(grpMmath)}_j + \gamma_{02}\texttt{(isolated)}_j + \gamma_{03}\texttt{(rural)}_j \\ &\quad +\gamma_{04}\texttt{(suburb)}_j + U_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}\texttt{(grpMmath)}_j + \gamma_{12}\texttt{(isolated)}_j + \gamma_{13}\texttt{(rural)}_j \\ &\quad +\gamma_{14}\texttt{(suburb)}_j + U_{1j} \\ \beta_{2j} &= \gamma_{20} \\ \beta_{3j} &= \gamma_{30} \\ \beta_{4j} &= \gamma_{40} \\ \beta_{5j} &= \gamma_{5j}\end{aligned}$$

where

$$\begin{pmatrix} U_{0j} \\ U_{1j} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{pmatrix}\right) \qquad i.i.d.$$

and independent of R_{ij}.

Linear mixed model :

$$\begin{aligned}\texttt{(science)}_{ij} &= \gamma_{00} + \gamma_{10}\texttt{(grpCmath)}_{ij} + \gamma_{20}\texttt{(gender)}_{ij} + \gamma_{30}\texttt{(grade)}_{ij} \\ &\quad +\gamma_{40}\texttt{(hours\_TV)}_{ij} + \gamma_{5j}\texttt{(hours\_computer\_games)}_{ij} \\ &\quad +\gamma_{01}\texttt{(grpMmath)}_j + \gamma_{02}\texttt{(isolated)}_j + \gamma_{03}\texttt{(rural)}_j + \gamma_{04}\texttt{(suburb)}_j\end{aligned}$$

$$+\gamma_{11}(\texttt{grpMmath})_j(\texttt{grpCmath})_{ij} + \gamma_{12}(\texttt{isolated})_j(\texttt{grpCmath})_{ij}$$
$$+\gamma_{13}(\texttt{rural})_j(\texttt{grpCmath})_{ij} + \gamma_{14}(\texttt{suburb})_j(\texttt{grpCmath})_{ij}$$
$$+U_{0j} + U_{1j}(\texttt{grpCmath})_{ij} + R_{ij}$$

Furthermore, recall that...

TIMSS	SAS data	Values	Description
ACBGCOMM	TYPE_COMMUNITY	1 =	a geographically isolated area
		2 =	village or rural (farm) area
		3 =	one on the outskirts of a town/city
		4 =	one close to the center of a town/city

3. Consider model (s) from your last computer/homework assignment (see above). Fit a model that you can use to check whether you need a random slope for $(\texttt{grpCmath})_{ij}$.

4. You will need to report a test of the random slope for `grpCmath`, to do this you need to fit model with and without a random slope and then...

```
data pvalue;
  LR = <put -2lnlike for null> - <put -2lnlike for full model>;
  df = <fill in standard value>;
  dfless = df-1;
  pdf = 1 - cdf('Chisquare',LR,df);
  pdfless  = 1 - cdf('chisquare',LR,dfless);
  pvalue = .5*(pdf + pdfless);
run;
```

You can use EXPLORER to look at results, or if you want to record the results in your output,

```
proc print data=pvalue;
run;
```

5. Examine the parameter estimates for type of community. Using a contrast statement, test whether the difference between the parameters for urban and suburban are statistically different.

The CONTRAST statement that will do this is

```
CONTRAST  'title for the contrast'  type_community 0 0 1 -1;
```

Some other contrasts you may want to try is

```
CONTRAST 'isolated vs sub/urban' type_community  1 0 -.5 -.5;
CONTRAST 'rural vs sub/urban'    type_community  0 1 -.5 -.5;
```

Note: If you want to enter more than one row of $\mathbf{L}$ where the null hypothesis is $H_o : \mathbf{L\Gamma} = \mathbf{0}$, then, for example,

```
CONTRAST 'example of df=2 contrast'
    type_community  1 -1  0 0,
    type_community  1  0 -1 0;
```

Suppose that the two parameters are not statistically different, for example, sub-urban and urban. The variable TYPE_COMMUNITY can be recoded and the model re-fit to the data used the recoded variable. In this case, to re-code, the following commands will to the trick:

```
 DATA lab3;
    SET clab2;                *<- put name of current data set here;
    community=type_community;
* Recode so that urba-n & suburban are the same;
    IF community=4 THEN community=3;
 run;
```

Decide best way to recode (if at all) type of community and use this in all subsequent models. Note you should also check the interaction with type of community.

Further Re-call that

TIMSS	SAS data	Values	Description
ASBGDAY1	HOURS_TV		Time spent watching TV or videos
		1 =	no time
		2 =	$<$ 1 hr.
		3 =	1-2 hrs.
		4 =	3-4 hrs.
		5 =	$>$4 hrs.
ASBGDAY2	HOURS_COMPUTER_GAMES		Time spent playing computer games
		1 =	no time
		2 =	less than 1 hour"
		3 =	1–2 hours
		4 =	3–4 hours
		5 =	more than 4 hours

6. We have been considering HOURS_TV as a numerical variable.
   (a) Treat it as a CLASS variable (i.e., a nominal, categorical variable).
   (b) Examine the parameter estimates for HOURS_TV.
   (c) Test any contrasts that are suggested by the parameter estimates and that make sense).
   (d) If warranted, re-code HOURS_TV and re-fit the model (substantively). (If you re-code, you should do a likelihood ratio test of whether re-coding had an effect or not). If you decide to re-code, then use this in subsequent models.
7. We have also been considering HOURS_COMPUTER_GAMES as a numerical a variable.
   (a) Treat it as a CLASS variable (i.e., a nominal, categorical variable).
   (b) Examine the parameter estimates for HOURS_COMPUTER_GAMES.
   (c) Test any contrasts that are suggested by the parameter estimates and that make sense (substantively).
   (d) If warranted, re-code HOURS_COMPUTER_GAMES and re-fit the model. (If you re-code, you should do a likelihood ratio test of whether re-coding had an effect or not).
8. Fit any other models that you think may be useful or needed (see homework).