3/30/2010 36-402/608 ADA-II H. Seltman
Breakout #18 Comments

These data come from “Estimating actor, partner, and interaction effects for dyadic data using PROC MIXED and HLM: A user-friendly guide”, by Campbell and Kashy, *Personal Relationships*, 9 (2002), p 327.

Here are the first few lines of data:

```
id wdraw asecure agen psecure pgen cond
001 3 7 1 6 -1 -1
001 4 6 -1 7 1 -1
002 6 5 1 4 -1 -1
002 4 4 -1 5 1 -1
```

Heterosexual couples were studied in a lab while discussing either a major or minor problem (randomly assigned). “id” is the dyad identification code, wdraw is the outcome (observer rating of emotional withdrawal; “asecure” and “psecure” are measures of “attachment security” for “actors” and “partners” respectively; “agen” and “pgen” are “gender” for “actors” and “partners” respectively with code 1=male, -1=female; and “cond” is randomly assigned treatment condition with code 1=major problem and 2=minor problem.

Question 1: Why would any model use agen or pgen, but not both? agen and pgen are perfectly correlated, i.e., a regression of one on the other has zero error. This leads to a singular $\mathbf{X}'\mathbf{X}$ matrix, and most programs drop one or the other in that case. Also the model is unidentifiable in the sense that you cannot assign the effect of a change in one (which must result in a change in both).

Here is the SAS code for the first analysis:

```
OPTIONS LINESIZE=70;
DATA IN;
    INFILE "Withdrawal.dat" firstobs=2 termstr=CRLF;
    INPUT id wdraw asecure0 agen psecure0 pgen cond;
RUN;

/* Center "secure" */
PROC SQL;
    CREATE TABLE WD AS
        SELECT id, wdraw, asecure0-mean(asecure0) as asecure,
               agen, psecure0-mean(psecure0) as psecure, cond
               FROM IN;
```

```
QUIT;

PROC MIXED;
    CLASS ID;
    MODEL wdraw = asecure psecure agen cond/ SOLUTION DDFM=SATTERTH;
    RANDOM INT / SUBJECT=id TYPE=UN G  V VCORR;
    TITLE "PROC MIXED example: as random intercept";
RUN;
```

Log file: NOTE: Convergence criteria met.

```
                  Model Information
        Data Set                     WORK.WD
        Dependent Variable           wdraw
        Covariance Structure         Unstructured
        Subject Effect               id
        Estimation Method            REML
        Degrees of Freedom Method    Satterthwaite

               Class Level Information
          Class    Levels    Values
          id           16    1 2 3 4 5 6 7 8 9 10 11 12 13
                             14 15 16

                        Dimensions
                 Covariance Parameters            2
                 Columns in X                     5
                 Columns in Z Per Subject         1
                 Subjects                        16
                 Max Obs Per Subject              2
         Number of Observations Used             32

                Convergence criteria met.

                  Estimated G Matrix

              Row    Effect       id        Col1
                1    Intercept     1      0.8213

                Estimated V Matrix for id 1
                Row        Col1        Col2
                  1      1.2206      0.8213
```

2	0.8213	1.2206

Estimated V Correlation Matrix for id 1

Row	Col1	Col2
1	1.0000	0.6729
2	0.6729	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)	id	0.8213
Residual		0.3993

Fit Statistics

BIC (smaller is better)	96.9

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	5.5625	0.2526	13	22.02	<.0001
asecure	0.1152	0.1502	18.6	0.77	0.4525
psecure	-0.6254	0.1502	18.6	-4.16	0.0005
agen	0.1181	0.1123	14	1.05	0.3111
cond	0.9433	0.3647	13	2.59	0.0226

Note that the Satterthwaite df method (or the similar KR method) are better than the default (in the sense of better achieving appropriate type-1 error rates and improving power.)

Question 2: Which output corresponds to G, V, and VCORR? What calculation from which values in the output gives the correlation value of 0.6729? What does this number mean?

G gives " Estimated G Matrix", V gives "Estimated V Matrix for id 1", and VCORR gives "Estimated V Correlation for id 1". The covariance of any two y' from the same dyad is 0.8213, and the variance of and each of these two y's is 0.3993, so the correlation is

$$\frac{\text{cov}}{\text{sd} \times \text{sd}} = \frac{0.8213}{\sqrt{0.3993}\sqrt{0.3993}} = 0.6729$$

.

Question 3: We must test if the intra-dyad correlation is significant? How?

Compare this model to the model with no RANDOM statement and show that including the random intercept lowers the BIC.

Question 4: What do each of the Estimates tell us, including the intercept? The intercept is the average withdrawal score for the two members of a dyad with average secure status for both members and averaged over the two conditions. Its p-value has no practical meaning (it tests whether that value is different from zero).

The asecure p-value tells us that actor security does not affect withdrawal (or there is not enough power). The CI of roughly 0.12 +/- 0.30 can be used by a subject matter expert to assess the potential power: if a change in withdrawal of 0.42 is not of an "important" size, then we can say that agent security does not have an important effect.

The p-value of 0.0005 (t=-4.16, df=18.6) for p-secure indicates that partner security has a statistically significant effect on withdrawal. All other things being equal (and assuming that this is the final model, e.g., the need for interactions or transformations have been ruled out and the residual checks are OK), then each one unit rise in partner security lowers an agent's withdrawal level by roughly 0.625 +/- 2(0.15) as a 95% CI.

Gender is not significant, so we conclude (again assuming this is the final model) that their is no effect gender on withdrawal, and the above partner security effect applies to both males and females equally.

The "major issue" condition is estimated to raise the withdrawal level by 2(0.943) compared to the "minor issue" condition (p=0.0226, df=13). Or you could say that the intercept for minor issues is estimated at 5.562-0.943 and the intercept from major issues is estimated at 5.562+0.943.

Here is an alternative analysis:

```
PROC MIXED;
    CLASS id;
    MODEL wdraw = asecure psecure agen cond/ SOLUTION DDFM=SATTERTH;
    REPEATED / TYPE=CS SUBJECT=id R RCORR;
    TITLE "PROC MIXED example: model includes only main effects";
RUN;
```

```
                          Model Information
        Covariance Structure         Compound Symmetry
        Subject Effect               id
        Estimation Method            REML
        Degrees of Freedom Method    Satterthwaite

                             Dimensions
                 Covariance Parameters            2
```

```
Columns in X                        5
Columns in Z                        0

     Convergence criteria met.

     Estimated R Matrix for id 1
     Row        Col1        Col2
       1      1.2206      0.8213
       2      0.8213      1.2206

       Estimated R Correlation
           Matrix for id 1
     Row        Col1        Col2
       1      1.0000      0.6729
       2      0.6729      1.0000

   Covariance Parameter Estimates
  Cov Parm     Subject    Estimate
  CS           id           0.8213
  Residual                  0.3993

            Fit Statistics
 BIC (smaller is better)         96.9


                   Solution for Fixed Effects
                             Standard
Effect        Estimate          Error        DF     t Value     Pr > |t|
Intercept       5.5625         0.2526        13       22.02       <.0001
asecure         0.1152         0.1502      18.6        0.77       0.4525
psecure        -0.6254         0.1502      18.6       -4.16       0.0005
agen            0.1181         0.1123        14        1.05       0.3111
cond            0.9433         0.3647        13        2.59       0.0226
```

Question 5: What is the same and what is different from the first analysis? Using today's handout, explain what is going on.

The only difference is use of the G vs. R matrix and whether the covariance of y's for the same subject is labeled UN(1,1) or CS. The random intercept model sets Z (and G) and "induces" the equicorrelation of Y through the equation var(y)=ZGZ'+R, while the use of the Compound-Symmetry type for the REPEATED statement directly sets R to have a block diagonal, equicorrelation structure. Because both models have the same mean and variance, they ARE the same model.

Here is another analysis:

```
DATA WD;
    SET WD;
    aconsec = cond*asecure;
    pconsec = cond*psecure;
run;

PROC MIXED;
    CLASS id;
    MODEL wdraw = asecure psecure agen cond aconsec pconsec /
        SOLUTION DDFM=SATTERTH;
    REPEATED / TYPE=CS SUBJECT=id RCORR;
    TITLE "Mystery model";
RUN;
```

Question 6: What does this code model?

This models interactions between treatment condition and security of both the agent and the partner.

```
                 Columns in X                      7
                      Convergence criteria met.

                        Estimated R Correlation
                      Row          Col1         Col2
                        1        1.0000       0.5543
                        2        0.5543       1.0000

                BIC (smaller is better)          95.9
                      Solution for Fixed Effects
                            Standard
   Effect        Estimate        Error        DF     t Value     Pr > |t|
   Intercept       6.1369       0.3158        12       19.43       <.0001
   asecure        0.03733       0.1372      18.9        0.27       0.7885
   psecure        -0.7148       0.1372      18.9       -5.21       <.0001
   agen            0.1101       0.1161        13        0.95       0.3605
   cond            0.7812       0.3158        12        2.47       0.0293
   aconsec         0.2597       0.1372      18.9        1.89       0.0738
   pconsec         0.3333       0.1372      18.9        2.43       0.0253
```

Question 7: How do you interpret the results?

BIC prefers the model with the interactions (we should do more careful, detailed model selection).

The partner security : condition interaction is statistically signficant while the corresponding agent interaction is borderline. Interpretation of interaction is difficult, other than saying that the effects of partner security and condition are not additive.

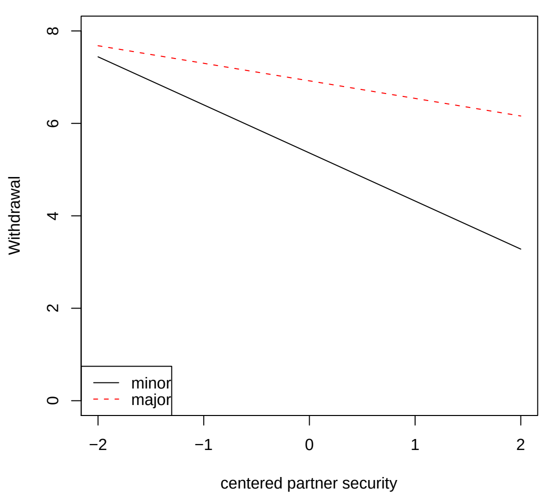
We need to, e.g., make a means plot and/or table to communicate the conclusions. The model for average agent security and averaged over both genders is:

E(Withdrawal) = 6.14 -0.71PS +0.78C +0.33PS:C

so for representative PS values of -2, 0, and 2 we get:

psecure	Cond	mean(withdrawal)
-2	minor	7.44
0	minor	5.36
2	minor	3.28
-2	major	7.68
0	major	6.92
2	major	6.16

With the -1 vs +1 coding for condition, we see the 0.33 estimate in the fact that a rise of 1 in psecure is associated with a change in withdrawal that is 2*0.33=0.66 bigger. For a rise of 2 in psecure this is 1.32 bigger as seen here: (6.16-3.28)-(6.92-5.36)=1.32.



We conclude that the effect of partner security is stronger for minor than major issues, while for all parter security levels between -2 and +2, the mean withdrawal is greater for major issues.