From Between-Person to Within-Person Models for Longitudinal Data

The models for this example come from Hoffman (2015) chapter 3 example 3a. We will be examining the extent to which a learning achievement outcome) can be predicted from group (control as the reference vs. treatment) and time (pre-test as the reference vs. post-test) in a sample of 50 children.

SAS Syntax and Output for Data Manipulation:

```
* Defining global variable for file location to be replaced in code below;
%LET filesave= C:\Dropbox\17_CLDP944\CLP944_Example03a;
* Location for SAS files for these models (uses macro variable filesave);
LIBNAME filesave "&filesave.";
* Import and stack chapter 3 two-occasion multivariate data;
* Create new variable on left from old variable on right, OUTPUT writes data;
DATA work.Chapter3a; SET filesave.SAS_Chapter3a;
time=1; outcome=outcome1; OUTPUT;
time=2; outcome=outcome2; OUTPUT;
DROP outcome1 outcome2;
LABEL  time = "time: Occasion (1=pre-test, 2=post-test)"
       outcome = "outcome: Learning Outcome"; RUN;
* Center predictors for analysis;
DATA work.Chapter3a; SET work.Chapter3a;
time1 = time - 1;  treat = group - 1;
LABEL  time1 = "time1: Time (0=pre-test, 1=post-test)"
       treat = "treat: Treatment Group (0=control, 1=treatment)"; RUN;

* CLASS= means per group and time, WAYS= means overall=0, per category=1, per cell=2;
TITLE1 "Chapter 3a Example: Means by group and time for learning outcome";
PROC MEANS MEAN STDERR MIN MAX DATA=work.Chapter3a;
    CLASS group time; WAYS 0 1 2; VAR outcome; RUN; TITLE1;
```

Cell means by group and time for y outcome

Treatment Group (1=control, 2=treatment)	Time (1=pre-test 2=post-test)	N Obs	Mean	Std Error	Minimum	Maximum
1	1	25	49.0767977	1.1370576	37.5335041	59.5504810
	2	25	54.8991630	1.1256529	44.5615778	67.1060321
2	1	25	50.7587396	0.9070808	40.5321932	62.1309134
	2	25	58.6236314	0.9864754	47.4303443	68.6163028

Marginal means by group for y outcome

Treatment Group (1=control, 2=treatment)	N Obs	Mean	Std Error	Minimum	Maximum
1	50	51.9879804	0.8943692	37.5335041	67.1060321
2	50	54.6911855	0.8691455	40.5321932	68.6163028

Marginal means by time for y outcome

Time (1=pre-test 2=post-test)	N Obs	Mean	Std Error	Minimum	Maximum
1	50	49.9177687	0.7297690	37.5335041	62.1309134
2	50	56.7613972	0.7870204	44.5615778	68.6163028

Grand mean for y outcome

Mean	Std Error	Minimum	Maximum
53.3395829	0.6351006	37.5335041	68.6163028

3.1: Between-Person Empty Model $y_{ti} = \beta_0 + e_{ti}$

```
TITLE "Eq 3.1: Empty Between-Person model via MIXED";
PROC MIXED DATA=work.Chapter3a NOCLPRINT COVTEST METHOD=REML;
      CLASS PersonID time;
      MODEL outcome = / SOLUTION DDFM=BW;
      REPEATED time / R RCORR TYPE=VC SUBJECT=PersonID; RUN; TITLE;
```

```
              Dimensions
Covariance Parameters            1
Columns in X                     1
Columns in Z                     0
Subjects                        50
Max Obs Per Subject              2
```

> This table tells you how many parameters are in your model for the means ("columns in x", the fixed effects, or 1 fixed intercept here) and in your model for the variance ("covariance parameters", or 1 residual variance here). It also tells you how many observations were read per subject, as defined by SUBJECT= on the REPEATED line.

```
           Number of Observations
Number of Observations Read          100
Number of Observations Used          100
Number of Observations Not Used        0
```

```
      Estimated R Matrix
        for PersonID 1
  Row       Col1         Col2
   1      40.3353
   2                   40.3353
```

```
   Estimated R Correlation
    Matrix for PersonID 1
  Row       Col1         Col2
   1       1.0000
   2                    1.0000
```

```
              Covariance Parameter Estimates
Cov                            Standard      Z
Parm      Subject    Estimate    Error     Value      Pr > Z
time     PersonID    40.3353    5.7330     7.04      <.0001
```

> This is the estimate of the residual variance σ_e^2. It is labeled "time" because that is how the R matrix is structured via the REPEATED line.

```
           Fit Statistics
-2 Res Log Likelihood            651.6
AIC (smaller is better)          653.6
AICC (smaller is better)         653.6
BIC (smaller is better)          655.5
```

```
  Null Model Likelihood Ratio Test
   DF     Chi-Square      Pr > ChiSq
    0        0.00           1.0000
```

> This "null model" LRT examines the need for any random effects variances and covariances. Because we don't have any (yet), df = 0.

```
                    Information Criteria
Neg2LogLike     Parms      AIC      AICC      HQIC       BIC      CAIC
    651.6         1       653.6     653.6     654.3     655.5     656.5
```

> In REML, model df = # for calculating AIC and BIC only includes parameters in the model for the variance.

```
           Solution for Fixed Effects
                    Standard
Effect      Estimate    Error     DF    t Value    Pr > |t|
Intercept    53.3396    0.6351    49     83.99      <.0001
```

> This is the estimate of the fixed intercept β_0.

3.2: Within-Person Empty Model $y_{ti} = \beta_0 + U_{0i} + e_{ti}$

```
TITLE "Eq 3.2: Empty Within-Person model via MIXED";
PROC MIXED DATA=work.Chapter3a NOCLPRINT COVTEST METHOD=REML;
      CLASS PersonID time;
      MODEL outcome = / SOLUTION DDFM=BW;
      REPEATED time / R RCORR TYPE=CS SUBJECT=PersonID; RUN; TITLE;
```

```
              Dimensions
Covariance Parameters          2
Columns in X                   1
Columns in Z                   0
Subjects                      50
Max Obs Per Subject            2
```

We still have 1 fixed effect, the fixed intercept, but now the model for the variance includes random intercept variance and residual variance.

```
          Number of Observations
Number of Observations Read          100
Number of Observations Used          100
Number of Observations Not Used        0
```

```
      Estimated R Matrix
        for PersonID 1
 Row        Col1        Col2
  1      40.4590     12.2526
  2      12.2526     40.4590
```

$$\begin{bmatrix} \sigma_e^2 + \tau_{u_0}^2 & \tau_{u_0}^2 \\ \tau_{u_0}^2 & \sigma_e^2 + \tau_{u_0}^2 \end{bmatrix}$$

```
   Estimated R Correlation
    Matrix for PersonID 1
 Row        Col1        Col2
  1       1.0000      0.3028
  2       0.3028      1.0000
```

$$\begin{bmatrix} 1 & ICC \\ ICC & 1 \end{bmatrix}$$

```
              Covariance Parameter Estimates
                              Standard        Z
Cov Parm    Subject    Estimate    Error    Value      Pr Z
CS          PersonID    12.2526    6.0256    2.03     0.0420
Residual                28.2064    5.6413    5.00     <.0001
```

CS = Random Intercept Variance $\tau_{U_0}^2$
Residual = Residual Variance σ_e^2

```
          Fit Statistics
-2 Res Log Likelihood          646.8
AIC (smaller is better)        650.8
AICC (smaller is better)       650.9
BIC (smaller is better)        654.6
```

```
  Null Model Likelihood Ratio Test
   DF     Chi-Square     Pr > ChiSq
    1        4.77          0.0289
```

Now we have a random intercept variance, so df=1. This is the model comparison of BP vs. WP. Who wins?

```
                    Information Criteria
Neg2LogLike    Parms     AIC      AICC      HQIC      BIC      CAIC
    646.8        2      650.8     650.9     652.3    654.6    656.6
```

Now the model for the variance df=2.

```
              Solution for Fixed Effects
                    Standard
Effect      Estimate    Error    DF    t Value    Pr > |t|
Intercept   53.3396    0.7260    49     73.47      <.0001
```

This is still the estimate of the fixed intercept β_0, but note the SE differs.

Which is the better empty model, and how do you know?

What is the ICC for these data and what does it mean?

3.7 (top): Between-Person Conditional Model
$$y_{ti} = \beta_0 + \beta_1(Time_{ti}) + \beta_2(Group_i) + \beta_3(Time_{ti})(Group_i) + e_{ti}$$

```
TITLE1 "Eq 3.7 (top): Between-Person Conditional (Predictor) Model via MIXED";
TITLE2 "Not using CLASS statement, manually dummy coding group and time";
PROC MIXED DATA=work.Chapter3a NOCLPRINT COVTEST METHOD=REML;
     CLASS PersonID time;
     MODEL outcome = time1 treat time1*treat / SOLUTION DDFM=BW;
     REPEATED time / R RCORR TYPE=VC SUBJECT=PersonID;
ESTIMATE "Mean: Control Group at Pre-Test"     intercept 1 time1 0 treat 0 time1*treat 0;
ESTIMATE "Mean: Control Group at Post-Test"    intercept 1 time1 1 treat 0 time1*treat 0;
ESTIMATE "Mean: Treatment Group at Pre-Test"   intercept 1 time1 0 treat 1 time1*treat 0;
ESTIMATE "Mean: Treatment Group at Post-Test"  intercept 1 time1 1 treat 1 time1*treat 1;
ESTIMATE "Time Effect for Control Group"       time1 1 time1*treat 0;
ESTIMATE "Time Effect for Treatment Group"     time1 1 time1*treat 1;
ESTIMATE "Group Effect at Pre-Test"            treat 1 time1*treat 0;
ESTIMATE "Group Effect at Post-Test"           treat 1 time1*treat 1;
RUN; TITLE1; TITLE2;
```

```
             Dimensions
Covariance Parameters          1
Columns in X                   4
Columns in Z                   0
Subjects                      50
Max Obs Per Subject            2
```

Now we have 4 parameters in the model for the means and 1 parameter in the model for the variance (σ_e^2).

```
           Number of Observations
Number of Observations Read          100
Number of Observations Used          100
Number of Observations Not Used        0
```

```
      Estimated R Matrix
       for PersonID 1
  Row       Col1        Col2
    1    27.2245
    2                27.2245
```

```
   Estimated R Correlation
    Matrix for PersonID 1
  Row       Col1        Col2
    1     1.0000
    2                 1.0000
```

```
              Covariance Parameter Estimates
Cov                              Standard        Z
Parm     Subject     Estimate      Error      Value      Pr > Z
time     PersonID    27.2245      3.9295      6.93       <.0001
```

This is the estimate of the residual variance σ_e^2. It is labeled "time" because that is how the R matrix is structured via the REPEATED line.

```
           Fit Statistics
-2 Res Log Likelihood          602.5
AIC (smaller is better)        604.5
AICC (smaller is better)       604.5
BIC (smaller is better)        606.4
```

```
  Null Model Likelihood Ratio Test
   DF    Chi-Square      Pr > ChiSq
    0         0.00          1.0000
```

This "null model" LRT examines the need for any random effects variances and covariances. Because we don't have any (yet), df = 0.

```
                    Information Criteria
Neg2LogLike     Parms      AIC      AICC      HQIC       BIC       CAIC
     602.5         1      604.5     604.5     605.2     606.4     607.4
```

```
                    BP Solution for Fixed Effects
                         Standard
Effect          Estimate     Error     DF    t Value    Pr > |t|
Intercept       49.0768     1.0435     48     47.03       <.0001   beta0
time1            5.8224     1.4758     48      3.95       0.0003   beta1
treat            1.6819     1.4758     48      1.14       0.2601   beta2
time1*treat      2.0425     2.0871     48      0.98       0.3327   beta3


            Type 3 Tests of Fixed Effects
                  Num     Den
Effect            DF      DF    F Value    Pr > F
time1              1      48     15.56     0.0003
treat              1      48      1.30     0.2601
time1*treat        1      48      0.96     0.3327


                                    Estimates
                                        Standard
Label                            Estimate     Error     DF    t Value    Pr > |t|
Mean: Control Group at Pre-Test     49.0768    1.0435    48     47.03      <.0001
Mean: Control Group at Post-Test    54.8992    1.0435    48     52.61      <.0001
Mean: Treatment Group at Pre-Test   50.7587    1.0435    48     48.64      <.0001
Mean: Treatment Group at Post-Test  58.6236    1.0435    48     56.18      <.0001
Time Effect for Control Group        5.8224    1.4758    48      3.95      0.0003 beta1
Time Effect for Treatment Group      7.8649    1.4758    48      5.33      <.0001 beta1 + beta3*1
Group Effect at Pre-Test             1.6819    1.4758    48      1.14      0.2601 beta2
Group Effect at Post-Test            3.7245    1.4758    48      2.52      0.0150 beta2 + beta3*1
```

These results assume independent observations… what happens if that's not the case?

3.7 (bottom): Within-Person Conditional Model

$$y_{ti} = \beta_0 + \beta_1(\text{Time}_{ti}) + \beta_2(\text{Group}_i) + \beta_3(\text{Time}_{ti})(\text{Group}_i) + U_{0i} + e_{ti}$$

```
TITLE1 "Eq 3.7 (bottom): Within-Person Conditional (Predictor) Model via MIXED";
TITLE2 "Not using CLASS statement, manually dummy coding group and time";
PROC MIXED DATA=work.Chapter3a NOCLPRINT IC COVTEST METHOD=REML;
      CLASS PersonID time;
      MODEL outcome = time1 treat time1*treat / SOLUTION DDFM=BW;
      REPEATED time / R RCORR TYPE=CS SUBJECT=PersonID;
ESTIMATE "Mean: Control Group at Pre-Test"     intercept 1 time1 0 treat 0 time1*treat 0;
ESTIMATE "Mean: Control Group at Post-Test"    intercept 1 time1 1 treat 0 time1*treat 0;
ESTIMATE "Mean: Treatment Group at Pre-Test"   intercept 1 time1 0 treat 1 time1*treat 0;
ESTIMATE "Mean: Treatment Group at Post-Test"  intercept 1 time1 1 treat 1 time1*treat 1;
ESTIMATE "Time Effect for Control Group"       time1 1 time1*treat 0;
ESTIMATE "Time Effect for Treatment Group"     time1 1 time1*treat 1;
ESTIMATE "Group Effect at Pre-Test"            treat 1 time1*treat 0;
ESTIMATE "Group Effect at Post-Test"           treat 1 time1*treat 1;
RUN; TITLE1; TITLE2;


              Dimensions
Covariance Parameters          2
Columns in X                   4
Columns in Z                   0
Subjects                      50
Max Obs Per Subject            2


            Number of Observations
Number of Observations Read          100
Number of Observations Used          100
Number of Observations Not Used        0
```

We still have 4 parameters in the model for the means, but now we have 2 parameters in the model for the variance ($\tau_{U_0}^2$ and σ_e^2).

```
        Estimated R Matrix
          for PersonID 1
Row        Col1        Col2
  1      27.2245     22.7794
  2      22.7794     27.2245
```

$$\begin{bmatrix} \sigma_e^2 + \tau_{u_0}^2 & \tau_{u_0}^2 \\ \tau_{u_0}^2 & \sigma_e^2 + \tau_{u_0}^2 \end{bmatrix}$$

```
        Estimated R Correlation
         Matrix for PersonID 1
Row        Col1        Col2
  1      1.0000      0.8367
  2      0.8367      1.0000
```

$$\begin{bmatrix} 1 & ICC \\ ICC & 1 \end{bmatrix}$$

```
                  Covariance Parameter Estimates
                                  Standard      Z
Cov Parm      Subject    Estimate   Error     Value     Pr Z
CS            PersonID    22.7794   5.1236     4.45    <.0001
Residual                  4.4451    0.9073     4.90    <.0001
```

CS = Random Intercept Variance $\tau_{U_0}^2$
Residual = Residual Variance σ_e^2

```
            Fit Statistics
-2 Res Log Likelihood          544.7
AIC (smaller is better)        548.7
AICC (smaller is better)       548.8
BIC (smaller is better)        552.5
```

```
  Null Model Likelihood Ratio Test
   DF     Chi-Square      Pr > ChiSq
    1       57.81           <.0001
```

Now we have a random intercept variance, so df=1. This is the model comparison of conditional BP vs. WP. Who wins?

```
                        Information Criteria
Neg2LogLike    Parms      AIC      AICC      HQIC      BIC      CAIC
    544.7        2       548.7     548.8     550.2     552.5    554.5
```

```
              WP Solution for Fixed Effects
                        Standard
Effect        Estimate    Error    DF   t Value   Pr > |t|
Intercept      49.0768   1.0435    48    47.03    <.0001 beta0
time1           5.8224   0.5963    48     9.76    <.0001 beta1
treat           1.6819   1.4758    48     1.14    0.2601 beta2
time1*treat     2.0425   0.8433    48     2.42    0.0193 beta3
```

Which results differ from the BP model, and why?

```
        Type 3 Tests of Fixed Effects
              Num    Den
Effect        DF     DF    F Value    Pr > F
time1          1     48    95.33      <.0001
treat          1     48     1.30      0.2601
time1*treat    1     48     5.87      0.0193
```

```
                              Estimates
                                        Standard
Label                         Estimate    Error    DF   t Value   Pr > |t|
Mean: Control Group at Pre-Test    49.0768   1.0435    48    47.03    <.0001
Mean: Control Group at Post-Test   54.8992   1.0435    48    52.61    <.0001
Mean: Treatment Group at Pre-Test  50.7587   1.0435    48    48.64    <.0001
Mean: Treatment Group at Post-Test 58.6236   1.0435    48    56.18    <.0001
Time Effect for Control Group       5.8224   0.5963    48     9.76    <.0001 beta1
Time Effect for Treatment Group     7.8649   0.5963    48    13.19    <.0001 beta1 + beta3*1
Group Effect at Pre-Test            1.6819   1.4758    48     1.14    0.2601 beta2
Group Effect at Post-Test           3.7245   1.4758    48     2.52    0.0150 beta2 + beta3*1
```

What other terms that could possibly be included are missing? Are they really missing?

What if we had used the CLASS statement instead for our conditional within-person model?

```
TITLE1 "Eq 3.7 (bottom): Within-Person Conditional (Predictor) Model via MIXED";
TITLE2 "NOW using CLASS statement";
PROC MIXED DATA=work.Chapter3a NOCLPRINT IC COVTEST METHOD=REML;
      CLASS PersonID time time1 treat;
      MODEL outcome = time1 treat time1*treat / SOLUTION DDFM=BW;
      REPEATED time / R RCORR TYPE=CS SUBJECT=PersonID;
      LSMEANS time1 treat time1*treat / DIFF=ALL;
RUN; TITLE1; TITLE2;
```

Solution for Fixed Effects

Effect	Time (0=pre-test, 1= post-test)	Treatment Group (0=control, 1=treatment)	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			58.6236	1.0435	48	56.18	<.0001
time1	0		-7.8649	0.5963	48	-13.19	<.0001
time1	1		0	.	.	.	.
treat		0	-3.7245	1.4758	48	-2.52	0.0150
treat		1	0	.	.	.	.
time1*treat	0	0	2.0425	0.8433	48	2.42	0.0193
time1*treat	0	1	0	.	.	.	.
time1*treat	1	0	0	.	.	.	.
time1*treat	1	1	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
time1	1	48	263.41	<.0001
treat	1	48	3.65	0.0619
time1*treat	1	48	5.87	0.0193

Note that the *p*-values from the solution for fixed effects (simple effects) and Type 3 tests of fixed effects (marginal effects) do not match because they mean different things (stay tuned).

Least Squares Means

Effect	Time (0=pre-test, 1= post-test)	Treatment Group (0=control, 1=treatment)	Estimate	Standard Error	DF	t Value	Pr > |t|
time1	0		49.9178	0.7379	48	67.65	<.0001
time1	1		56.7614	0.7379	48	76.92	<.0001
treat		0	51.9880	1.0000	48	51.99	<.0001
treat		1	54.6912	1.0000	48	54.69	<.0001
time1*treat	0	0	49.0768	1.0435	48	47.03	<.0001
time1*treat	0	1	50.7587	1.0435	48	48.64	<.0001
time1*treat	1	0	54.8992	1.0435	48	52.61	<.0001
time1*treat	1	1	58.6236	1.0435	48	56.18	<.0001

Differences of Least Squares Means

Effect	Time (0=pre-test, 1= post-test)	Treatment Group (0=control, 1=treatment)	Time (0=pre-test, 1= post-test)	Treatment Group (0=control, 1=treatment)	Estimate	Standard Error	DF	t Value	Pr > |t|
time1	0		1		-6.8436	0.4217	48	-16.23	<.0001
treat		0		1	-2.7032	1.4143	48	-1.91	0.0619
time1*treat	0	0	0	1	-1.6819	1.4758	48	-1.14	0.2601
time1*treat	0	0	1	0	-5.8224	0.5963	48	-9.76	<.0001
time1*treat	0	0	1	1	-9.5468	1.4758	48	-6.47	<.0001
time1*treat	0	1	1	0	-4.1404	1.4758	48	-2.81	0.0072
time1*treat	0	1	1	1	-7.8649	0.5963	48	-13.19	<.0001
time1*treat	1	0	1	1	-3.7245	1.4758	48	-2.52	0.0150