Example 3a: From Between-Person to Within-Person Models for Longitudinal Data
(complete syntax, data, and output available for SAS, STATA, and R electronically)

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. For an example results section, please see the end of chapter 3.

SAS Syntax for Data Import and Manipulation:

```
* Define global variable for file location to be replaced in code below;
%LET filesave = C:\Dropbox\22_PSQF6271\PSQF6271_Example3a;
* 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.Example3a; SET filesave.SAS_Chapter3a;
occasion=1; outcome=outcome1; OUTPUT;
occasion=2; outcome=outcome2; OUTPUT;
DROP outcome1 outcome2;
LABEL occasion = "occasion: 1=pre-test, 2=post-test"
      outcome = "outcome: Learning Outcome"; RUN;

* Center predictors for analysis;
DATA work.Example3a; SET work.Example3a;
time = occasion - 1;  treat = group - 1;
LABEL time  = "time: 0=pre-test, 1=post-test"
      treat = "treat: 0=control, 1=treatment"; RUN;

* Sort by person and occasion;
PROC SORT DATA=work.Example3a; BY PersonID occasion; RUN;
```

STATA Syntax for Data Import and Manipulation:

```
// Define global variable for file location to be replaced in code below
   global filesave " C:\Dropbox\22_PSQF6271\PSQF6271_Example3a "

// Import and stack chapter 3a two-occasion multivariate data
// List time-varying variables first, i(level2ID) j(newtimeID)
   use "$filesave\STATA_Chapter3a.dta", clear
   reshape long outcome, i(personid) j(occasion)
   label variable occasion "occasion: 1=pre-test, 2=post-test"
   label variable outcome "outcome: Learning Outcome"

// Center predictors for analysis
   gen time  = time - 1
   gen treat = group - 1
   label variable time  "time: 0=pre-test, 1=post-test"
   label variable treat "treat: 0=control, 1=treatment"

// Sort by person and occasion
   sort personid occasion
```

R Syntax for Data Import and Manipulation (see original file online for loading libraries):

```
# Define variables for working directory and data name
filesave = "C:\\Dropbox/22_PSQF6271/PSQF6271_Example3a/"
filename = "SAS_Chapter3a.sas7bdat"
setwd(dir=filesave)

# Import chapter 3 two-occasion multivariate data with labels
Example3a_wide = read_sas(data_file=paste0(filesave,filename))
# Convert to data frame as data frame without labels to use for analysis
Example3a_wide = as.data.frame(Example3a_wide)
```

```
# Stack into long format (one row per occasion per person)
Example3a = reshape(Example3a_wide, direction="long",
                    v.names="outcome", idvar="PersonID", timevar="occasion",
                    varying=c("outcome1","outcome2"), times=c(1,2))
# Center predictors for analysis
Example3a$time  = Example3a$occasion-1
Example3a$treat = Example3a$group-1
# Labels as comments only
# occasion: 1=pre-test, 2=post-test
# outcome: Learning Outcome
# time: 0=pre-test, 1=post-test
# treat: 0=control, 1=treatment

# Sort by person and occasion
Example3a = sort_asc(data=Example3a, PersonID,occasion)
```

SAS, STATA, and R Syntax for Descriptive Statistics:

```
* CLASS= means per group and time, WAYS= means overall=0, per category=1, per cell=2;
TITLE1 "Chapter 3a Example: Means by group and occasion for learning outcome";
PROC MEANS NDEC=2 MEAN STDERR MIN MAX DATA=work.Example3a;
     CLASS group occasion; WAYS 0 1 2; VAR outcome;
RUN; TITLE1;
```

```
display "Chapter 3a Example: Means by group and occasion for learning outcome"
tabulate group occasion, summarize(outcome)
```

```
print("Chapter 3a Example: Means by group and occasion for learning outcome")
describeBy(x=Example3a$outcome, digits=2,
           group=list(group=Example3a$group, occasion=Example3a$occasion))
```

SAS Output (that created the means in slide 15 in Lecture 3):

N Obs	Mean	Std Error	Minimum	Maximum
100	53.34	0.64	37.53	68.62

occasion: 1=pre-test, 2=post-test	N Obs	Mean	Std Error	Minimum	Maximum
1	50	49.92	0.73	37.53	62.13
2	50	56.76	0.79	44.56	68.62

group: 1=control, 2=treatment	N Obs	Mean	Std Error	Minimum	Maximum
1	50	51.99	0.89	37.53	67.11
2	50	54.69	0.87	40.53	68.62

group: (1=control, 2=treatment	occasion: 1=pre-test, 2=post-test	N Obs	Mean	Std Error	Minimum	Maximum
1	1	25	49.08	1.14	37.53	59.55
	2	25	54.90	1.13	44.56	67.11
2	1	25	50.76	0.91	40.53	62.13
	2	25	58.62	0.99	47.43	68.62

SAS, STATA, and R Syntax for 3.1: Between-Person Empty Model $y_{ti} = \beta_0 + e_{ti}$

```
TITLE "Eq 3.1: Empty Between-Person model via SAS MIXED";
PROC MIXED DATA=work.Example3a NOCLPRINT COVTEST NAMELEN=100 IC METHOD=REML;
     CLASS PersonID occasion;
     MODEL outcome = / SOLUTION CL ALPHA=.05 DDFM=BW;
     REPEATED occasion / R RCORR TYPE=VC SUBJECT=PersonID;
RUN; TITLE;
```

The pattern of variances and covariances over time (= ID variable) is controlled by SAS REPEATED, STATA RESIDUALS, and R CORRELATION.

```
display "Eq 3a.1: Empty Between-Person Model via STATA MIXED"
mixed outcome , ///
         || personid: , noconstant variance reml ///
         residuals(independent,t(occasion)) dfmethod(residual) dftable(pvalue)
      display "-2LL = " e(ll)*-2    // Print -2LL for model
      estat ic, n(50)               // Information criteria using level-2 N
```

STATA MIXED includes a random intercept by default,

Use `dftable(default)` to get CIs without DF instead

```
print("Eq 3.1: Empty Between-Person model via R GLS (no correlation, =LM)")
empty_BP = gls(data=Example3a, method="REML", model=outcome~1,
               correlation=NULL) # VC R matrix (no correlation)
print("Show results with -2LL, total original variance, and 95% CI")
summary(empty_BP); -2*logLik(empty_BP); summary(empty_BP)$sigma^2; confit(empty_BP, level=.95)
```

SAS Output:

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.

Iteration History

Iteration	Evaluations	-2 Res Log Like	
0	1	651.58043111	
1	1	**651.58043111**	0.00000000

Use this table to get −2LL with enough precision for your HW.

Estimated R Matrix for PersonID 1

Row	Col1	Col2
1	40.3353	
2		40.3353

This **R** matrix is the total ("marginal") model-predicted covariance matrix over time. Here, it predicts equal variance and no covariance over time. The **RCORR** matrix is the correlation version instead.

Estimated R Correlation Matrix for PersonID 1

Row	Col1	Col2
1	1.0000	
2		1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
occasion	PersonID	40.3353	5.7330	7.04	<.0001

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

Information Criteria

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

In REML using SAS, model df = # for calculating AIC and BIC only includes parameters in the model for the variance (1 here).

Solution for Fixed Effects

Effect	Estimate	Standard 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 (CI not shown to save space).

SAS, STATA, and R Syntax for 3.2: Within-Person Empty Model $y_{ti} = \beta_0 + U_{0i} + e_{ti}$

```
TITLE "Eq 3.2: Empty Within-Person model via SAS MIXED";
PROC MIXED DATA=work.Example3a NOCLPRINT COVTEST NAMELEN=100 IC METHOD=REML;
     CLASS PersonID occasion;
     MODEL outcome = / SOLUTION CL ALPHA=.05 DDFM=BW;
     REPEATED occasion / R RCORR TYPE=CS SUBJECT=PersonID;
RUN; TITLE;
```

```
display "Eq 3a.2: Empty Within-Person Model via STATA MIXED"
mixed outcome , ///
         || personid: , noconstant variance reml ///
         residuals(exchangeable,t(occasion)) dfmethod(repeated) dftable(pvalue)
      display "-2LL = " e(ll)*-2       // Print -2LL for model
      estat ic, n(50)                  // Information criteria using level-2 N
      estat wcorrelation, covariance   // RCOV matrix
      estat wcorrelation               // RCORR matrix
```

```
print("Eq 3.2: Empty Within-Person model via R GLS (CS correlation)")
empty_WP = gls(data=Example3a, method="REML", model=outcome~1,
               correlation=corCompSymm(form=~1|PersonID))
print("Show results with -2LL, total original variance, and 95% CI")
summary(empty_WP); -2*logLik(empty_WP); summary(empty_WP)$sigma^2; confit(empty_BP, level=.95)

print("Show RCOV and RCORR matrices")
getVarCov(empty_WP); corMatrix(empty_WP$modelStruct$corStruct)[[5]]
print("Show likelihood ratio test comparing empty means variance model fit")
anova(empty_WP, empty_BP)
```

This LRT (in order of "more", "fewer") is done for you in SAS and STATA.

SAS Output:

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.

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	651.58043111	
1	1	**646.80947282**	0.00000000

Use this table to get −2LL with enough precision for your HW.

Estimated R Matrix for PersonID 1

Row	Col1	Col2
1	40.4590	12.2526
2	12.2526	40.4590

R matrix for total variance and covariance over time now has a "compound symmetry" structure (equal variance, "CS" covariance)

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

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

Cov Parm	Subject	Estimate	Standard Error	Z 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

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	4.77	0.0289

This is the LRT model comparison of BP ("fewer") – WP ("more") for the empty means model: $-2\Delta LL(df = 2–1 = 1) = 657.58 - 646.81 = 4.77$

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

Effect	Estimate	Standard 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?

SAS, STATA, and R Syntax and SAS Output for 3.7 (top): Between-Person Conditional Model

$$y_{ti} = \beta_0 + \beta_1(\text{Time}_{ti}) + \beta_2(\text{Treat}_i) + \beta_3(\text{Time}_{ti})(\text{Treat}_i) + e_{ti}$$

Top: ALL fixed effects contribute to predicted outcomes.

Bottom: ONLY fixed effects containing the slope of interest contribute to that predicted slope.

Control Group Mean at Pre-Test: $\hat{y}_{ti} = \beta_0 + \beta_1(0) + \beta_2(0) + \beta_3(0)(0)$

Control Group Mean at Post-Test: $\hat{y}_{ti} = \beta_0 + \beta_1(1) + \beta_2(0) + \beta_3(0)(0)$

Treatment Group Mean at Pre-Test: $\hat{y}_{ti} = \beta_0 + \beta_1(0) + \beta_2(1) + \beta_3(0)(0)$

Treatment Group Mean at Post-Test: $\hat{y}_{ti} = \beta_0 + \beta_1(1) + \beta_2(1) + \beta_3(1)(1)$

Time Slope for Control Group: $\beta_1(\text{Time}_{ti}) + \beta_3(\text{Time}_{ti})(\text{Treat}_i) \rightarrow [\beta_1 + \beta_3(0)](\text{Time}_{ti})$

Time Slope for Treatment Group: $\beta_1(\text{Time}_{ti}) + \beta_3(\text{Time}_{ti})(\text{Treat}_i) \rightarrow [\beta_1 + \beta_3(1)](\text{Time}_{ti})$

Group Slope at Pre-Test: $\beta_2(\text{Treat}_i) + \beta_3(\text{Time}_{ti})(\text{Treat}_i) \rightarrow [\beta_2 + \beta_3(0)](\text{Treat}_i)$

Group Slope at Post-Test: $\beta_2(\text{Treat}_i) + \beta_3(\text{Time}_{ti})(\text{Treat}_i) \rightarrow [\beta_2 + \beta_3(1)](\text{Treat}_i)$

```
TITLE1 "Eq 3.7 (top): Between-Person Conditional (Predictor) Model via SAS MIXED";
TITLE2 "Not using CLASS statement, manually dummy coding group and time";
PROC MIXED DATA=work.Example3a NOCLPRINT COVTEST NAMELEN=100 IC METHOD=REML;
     CLASS PersonID occasion;
     MODEL outcome = time treat time*treat / SOLUTION CL ALPHA=.05 DDFM=BW;
     REPEATED occasion / R RCORR TYPE=VC SUBJECT=PersonID; * Est: Add / CL ALPHA=.05 for 95% CI;
     ESTIMATE "Control Group Mean at Pre-Test"      intercept 1 time 0 treat 0 time*treat 0;
     ESTIMATE "Control Group Mean at Post-Test"     intercept 1 time 1 treat 0 time*treat 0;
     ESTIMATE "Treatment Group Mean at Pre-Test"    intercept 1 time 0 treat 1 time*treat 0;
     ESTIMATE "Treatment Group Mean at Post-Test"   intercept 1 time 1 treat 1 time*treat 1;
     ESTIMATE "Time Slope for Control Group"        time  1 time*treat 0;
     ESTIMATE "Time Slope for Treatment Group"      time  1 time*treat 1;
     ESTIMATE "Group Slope at Pre-Test"             treat 1 time*treat 0;
     ESTIMATE "Group Slope at Post-Test"            treat 1 time*treat 1;
     ESTIMATE "Time*Group Interaction Slope"        time*treat 1;
     CONTRAST "DF=3 Joint F-test for Model R2"      time 1, treat 1, time*treat 1 / CHISQ;
RUN; TITLE1; TITLE2;

display "Eq 3.7 (top): Between-Person Conditional (Predictor) Model via STATA MIXED"
display "Not using i., manually dummy coding group and time"
mixed outcome c.time c.treat c.time#c.treat,      ///
         || personid: , noconstant variance reml  ///
         residuals(independent,t(occasion)) dfmethod(residual) dftable(pvalue)
      display "-2LL = " e(ll)*-2      // Print -2LL for model
```

```
estat ic, n(50)                  // Information criteria using level-2 N
// Control Group Mean at Pre-Test
lincom _cons*1 + c.time*0 + c.treat*0 + c.time#c.treat*0, small
// Control Group Mean at Post-Test
lincom _cons*1 + c.time*1 + c.treat*0 + c.time#c.treat*0, small
// Treatment Group Mean at Pre-Test
lincom _cons*1 + c.time*0 + c.treat*1 + c.time#c.treat*0, small
// Treatment Group Mean at Post-Test
lincom _cons*1 + c.time*1 + c.treat*1 + c.time#c.treat*1, small
// Time Slope for Control Group
lincom c.time*1 + c.time#c.treat*0, small
// Time Slope for Treatment Group
lincom c.time*1 + c.time#c.treat*1, small
// Group Slope at Pre-Test
lincom c.treat*1 + c.time#c.treat*0, small
// Group Slope at Post-Test
lincom c.treat*1 + c.time#c.treat*1, small
// Time*Group Interaction Slope
lincom c.time#c.treat*1, small
// DF=3 Joint F-test for Model R2
test (c.time=0)(c.treat=0)(c.time#c.treat=0), small
```

```
print("Eq 3.7 (top): Between-Person Conditional (Predictor) Model via R GLS (no corr, =LM)")
print("Not using factor variables, manually dummy coding group and time")
cond_BP = gls(data=Example3a, method="REML", model=outcome~1+time+treat+time:treat,
              correlation=NULL)
print("Show results with -2LL, total leftover variance, and 95% CI")
summary(cond_BP); -2*logLik(cond_BP); summary(cond_BP)$sigma^2; confit(cond_BP, level=.95)

print("Get linear combination estimates, add correct denominator DF and 95% CIs")
estimates_BP = glht(model=cond_BP, df=48, linfct=rbind(
       "Control Group Mean at Pre-Test"     =c(1,0,0,0), # Values= multipliers of ordered fixed effects
       "Control Group Mean at Post-Test"    =c(1,1,0,0),
       "Treatment Group Mean at Pre-Test"   =c(1,0,1,0),
       "Treatment Group Mean at Post-Test"  =c(1,1,1,1),
       "Time Slope for Control Group"       =c(0,1,0,0),
       "Time Slope for Treatment Group"     =c(0,1,0,1),
       "Group Slope at Pre-Test"            =c(0,0,1,0),
       "Group Slope at Post-Test"           =c(0,0,1,1),
       "Time*Group Interaction Slope"       =c(0,0,0,1)))
Summary(estimates_BP, test=adjusted("none"))
confint(estimates_BP, level=.95, calpha=univariate_calpha())

print("Get and show joint test of model R2 using correct denominator DF")
modelF_BP=glht(model=cond_BP, df=48, linfct=c("time=0","treat=0","time:treat=0"))
BP=summary(modelF_BP, test=Ftest()); BP
print("Get and show hidden results for F, dfnum, dfden, and p-value")
BP$test$fstat; BP$test$df; BP$df
pf(BP$test$fstat,df1=BP$test$df,df2=BP$df,lower.tail=FALSE)
```

SAS Output:

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).

Estimated R Matrix for PersonID 1

Row	Col1	Col2
1	27.2245	
2		27.2245

This **R** matrix is the total ("marginal") model-predicted covariance matrix over time. Here, it predicts equal **leftover** variance and no covariance over time.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
occasion	PersonID	27.2245	3.9295	6.93	<.0001

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

Information Criteria

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

–2LL = 602.507 from Iteration History

BP Solution for Fixed Slopes

Slope	Estimate	Standard Error	DF	t Value	Pr > |t|	
Intercept	**49.0768**	**1.0435**	**48**	**47.03**	**<.0001**	**beta0**
time	**5.8224**	**1.4758**	**48**	**3.95**	**0.0003**	**beta1**
treat	**1.6819**	**1.4758**	**48**	**1.14**	**0.2601**	**beta2**
time*treat	**2.0425**	**2.0871**	**48**	**0.98**	**0.3327**	**beta3**

BP Model Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	
Control Group Mean at Pre-Test	49.0768	1.0435	48	47.03	<.0001	
Control Group Mean at Post-Test	54.8992	1.0435	48	52.61	<.0001	
Treatment Group Mean at Pre-Test	50.7587	1.0435	48	48.64	<.0001	
Treatment Group Mean at Post-Test	58.6236	1.0435	48	56.18	<.0001	
Time Slope for Control Group	5.8224	1.4758	48	3.95	0.0003	beta1
Time Slope for Treatment Group	7.8649	1.4758	48	5.33	<.0001	beta1 + beta3*1
Group Slope at Pre-Test	1.6819	1.4758	48	1.14	0.2601	beta2
Group Slope at Post-Test	3.7245	1.4758	48	2.52	0.0150	beta2 + beta3*1
Time*Group Interaction Slope	2.0425	2.0871	48	0.98	0.3327	beta3

BP Model Contrasts

Label	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
DF=3 Joint F-test for Model R2	3	48	50.68	**16.89**	<.0001	<.0001

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

SAS, STATA, and R Syntax for 3.7 (bottom): Within-Person Conditional Model

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

```
TITLE1 "Eq 3.7 (bottom): Within-Person Conditional (Predictor) Model via SAS MIXED";
TITLE2 "Not using CLASS statement, manually dummy coding group and time";
PROC MIXED DATA=work.Example3a NOCLPRINT COVTEST NAMELEN=100 IC METHOD=REML;
     CLASS PersonID occasion;
     MODEL outcome = time treat time*treat / SOLUTION CL ALPHA=.05 DDFM=BW;
     REPEATED occasion / R RCORR TYPE=CS SUBJECT=PersonID; * Est: Add / CL ALPHA=.05 for 95% CI;
     ESTIMATE "Control Group Mean at Pre-Test"      intercept 1 time 0 treat 0 time*treat 0;
     ESTIMATE "Control Group Mean at Post-Test"     intercept 1 time 1 treat 0 time*treat 0;
     ESTIMATE "Treatment Group Mean at Pre-Test"    intercept 1 time 0 treat 1 time*treat 0;
     ESTIMATE "Treatment Group Mean at Post-Test"   intercept 1 time 1 treat 1 time*treat 1;
     ESTIMATE "Time Slope for Control Group"        time  1 time*treat 0;
     ESTIMATE "Time Slope for Treatment Group"      time  1 time*treat 1;
     ESTIMATE "Group Slope at Pre-Test"             treat 1 time*treat 0;
     ESTIMATE "Group Slope at Post-Test"            treat 1 time*treat 1;
     ESTIMATE "Time*Group Interaction Slope"        time*treat 1;
     CONTRAST "DF=3 Joint F-test for Model R2"      time 1, treat 1, time*treat 1 / CHISQ;
RUN; TITLE1; TITLE2;
```

```
display "Eq 3.7 (bottom): Within-Person Conditional (Predictor) Model via STATA MIXED"
display "Not using i., manually dummy coding group and time"
mixed outcome c.time c.treat c.time#c.treat,      ///
         || personid: , noconstant variance reml  ///
         residuals(exchangeable,t(occasion)) dfmethod(repeated) dftable(pvalue)
      display "-2LL = " e(ll)*-2       // Print -2LL for model
      estat ic, n(50)                  // Information criteria using level-2 N
```

```
estat wcorrelation, covariance  // RCOV matrix
estat wcorrelation              // RCORR matrix
// Control Group Mean at Pre-Test
lincom _cons*1 + c.time*0 + c.treat*0 + c.time#c.treat*0, small
// Control Group Mean at Post-Test
lincom _cons*1 + c.time*1 + c.treat*0 + c.time#c.treat*0, small
// Treatment Group Mean at Pre-Test
lincom _cons*1 + c.time*0 + c.treat*1 + c.time#c.treat*0, small
// Treatment Group Mean at Post-Test
lincom _cons*1 + c.time*1 + c.treat*1 + c.time#c.treat*1, small
// Time Slope for Control Group
lincom c.time*1 + c.time#c.treat*0, small
// Time Slope for Treatment Group
lincom c.time*1 + c.time#c.treat*1, small
// Group Slope at Pre-Test
lincom c.treat*1 + c.time#c.treat*0, small
// Group Slope at Post-Test
lincom c.treat*1 + c.time#c.treat*1, small
// Time*Group Interaction Slope
lincom c.treat*1 + c.time#c.treat*1, small
// DF=3 Joint F-test for Model R2
test (c.time=0)(c.treat=0)(c.time#c.treat=0), small
```

```
print("Eq 3.7 (bottom): Within-Person Conditional (Predictor) Model via gls (CS correlation")
print("Not using factor variables, manually dummy coding group and time")
cond_WP = gls(data=Example3a, method="REML", model=outcome~1+time+treat+time:treat,
              correlation=corCompSymm(form=~1|PersonID) )
print("Show results with -2LL, total leftover variance, and 95% CI")
summary(cond_WP); -2*logLik(cond_WP); summary(cond_WP)$sigma^2; confit(cond_WP, level=.95)

print("Show RCOV and RCORR matrices")
getVarCov(cond_WP); corMatrix(cond_WP$modelStruct$corStruct)[[5]]

print("Get linear combination estimates, add correct denominator DF and 95% CIs")
estimates_WP = glht(model=cond_BP, df=48, linfct=rbind(
       "Control Group Mean at Pre-Test"     =c(1,0,0,0), # Values= multipliers of ordered fixed effects
       "Control Group Mean at Post-Test"    =c(1,1,0,0),
       "Treatment Group Mean at Pre-Test"   =c(1,0,1,0),
       "Treatment Group Mean at Post-Test"  =c(1,1,1,1),
       "Time Slope for Control Group"       =c(0,1,0,0),
       "Time Slope for Treatment Group"     =c(0,1,0,1),
       "Group Slope at Pre-Test"            =c(0,0,1,0),
       "Group Slope at Post-Test"           =c(0,0,1,1),
       "Time*Group Interaction Slope"       =c(0,0,0,1)))
Summary(estimates_WP, test=adjusted("none"))
confint(estimates_WP, level=.95, calpha=univariate_calpha())

print("Get and show joint test of model R2 using correct denominator DF")
modelF_WP=glht(model=cond_WP, df=48, linfct=c("time=0","treat=0","time:treat=0"))
WP=summary(modelF_WP, test=Ftest()); WP # Print results
print("Get and show hidden results for F, dfnum, dfden, and p-value")
WP$test$fstat; WP$test$df; WP$df
pf(WP$test$fstat,df1=WP$test$df,df2=WP$df,lower.tail=FALSE)
print("Show likelihood ratio test comparing conditional means variance model fit")
anova(cond_WP, cond_BP)
```

This LRT (in order of "more", "fewer") is done for you in SAS and STATA.

SAS Output:

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

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

Estimated R Matrix for PersonID 1

Row	Col1	Col2
1	27.2245	22.7794
2	22.7794	27.2245

$$\begin{bmatrix} \sigma^2_e + \tau^2_{u_0} & \tau^2_{u_0} \\ \tau^2_{u_0} & \sigma^2_e + \tau^2_{u_0} \end{bmatrix}$$

R matrix for total (marginal) **leftover** variance and covariance over time now has a "compound symmetry" structure (equal variance, "CS" covariance). The RCORR matrix below shows the residual correlation (conditional ICC) over time.

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

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

CS = Random Intercept Variance $\tau^2_{U_0}$
Residual = Residual Variance σ^2_e

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	57.81	<.0001

This is the LRT model comparison of BP ("fewer") – WP ("more") for the conditional means model: $-2\Delta LL(df = 2-1 = 1) = 602.51 - 544.70 = 57.81$

Information Criteria

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

–2LL = 544.699 from Iteration History

BP Solution for Fixed Slopes (REPEATED FROM ABOVE FOR COMPARISON)

Slope	Estimate	Standard Error	DF	t Value	Pr > |t|	
Intercept	**49.0768**	**1.0435**	**48**	**47.03**	**<.0001**	**beta0**
time	**5.8224**	**1.4758**	**48**	**3.95**	**0.0003**	**beta1**
treat	**1.6819**	**1.4758**	**48**	**1.14**	**0.2601**	**beta2**
time*treat	**2.0425**	**2.0871**	**48**	**0.98**	**0.3327**	**beta3**

WP Solution for Fixed Slopes

Which results differ from the BP model, and why?

Slope	Estimate	Standard Error	DF	t Value	Pr > |t|	
Intercept	**49.0768**	**1.0435**	**48**	**47.03**	**<.0001**	**beta0**
time	**5.8224**	**0.5963**	**48**	**9.76**	**<.0001**	**beta1**
treat	**1.6819**	**1.4758**	**48**	**1.14**	**0.2601**	**beta2**
time*treat	**2.0425**	**0.8433**	**48**	**2.42**	**0.0193**	**beta3**

WP Model Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	
Control Group Mean at Pre-Test	49.0768	1.0435	48	47.03	<.0001	
Control Group Mean at Post-Test	54.8992	1.0435	48	52.61	<.0001	
Treatment Group Mean at Pre-Test	50.7587	1.0435	48	48.64	<.0001	
Treatment Group Mean at Post-Test	58.6236	1.0435	48	56.18	<.0001	
Time Slope for Control Group	5.8224	0.5963	48	9.76	<.0001	beta1
Time Slope for Treatment Group	7.8649	0.5963	48	13.19	<.0001	beta1 + beta3*1
Group Slope at Pre-Test	1.6819	1.4758	48	1.14	0.2601	beta2
Group Slope at Post-Test	3.7245	1.4758	48	2.52	0.0150	beta2 + beta3*1
Time*Group Interaction Slope	2.0425	0.8433	48	2.42	0.0193	beta3

WP Model Contrasts

Label	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
DF=3 Join F-test for Model R2	3	48	272.93	**90.98**	<.0001	<.0001

$$y_{ti} = \beta_0 + \beta_1(Time_{ti}) + \beta_2(Treat_i) + \beta_3(Time_{ti})(Treat_i) + U_{0i} + e_{ti}$$

Think about person as a factor in a three-way factorial design of person by time by group…
What other terms that could possibly be included are missing? Are they really missing???