Statistical Computing Seminar
Introduction to Multilevel Modeling Using SAS

This seminar is based on the paper Using SAS Proc Mixed to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models
by Judith Singer and can be downloaded from Professor Singer's web site at
http://gseweb.harvard.edu/~faculty/singer/Papers/sasprocmixed.pdf .

SAS data files, hsb12.sas7bdat and willett.sas7bdat and the SAS program code is here.

Outline

"The purpose of this paper is to show educational and behavioral statisticians and researchers how they can use **proc mixed** to fit many common types of multilevel models."

There are two types of models that this paper has focused on: (a) *school effects* models and (b) *individual growth* models.

- A school effect model using data file hsb12.sas7bdat
    - modeling organizational research;
    - students nested within classes, children nested within families, patients nested within hospitals;
        - Model 1: Unconditional Means Model
        - Model 2: Including Effects of School Level (level 2) Predictors
        - Model 3**:** Including Effects of Student-Level Predictors
        - Model 4: Including Both Level-1 and Level-2 Predictors
- Growth model using data file willett.sas7bdat
    - modeling individual change
    - multiple observations on each individual as nested within the person;
        - Model 1 :Unconditional Linear Growth Model
        - Model 2: A Linear Growth Model with a Person-Level Covariance
        - Model 3: Exploring the Structure of Variance Covariance Matrix Within Persons

---

School Effect Model

A segment of the data file:

```
SCHOOL     MATHACH       SES      MEANSES     SECTOR
 1296        6.588    -0.178      -0.420         0
 1296       11.026     0.392      -0.420         0
 1296        7.095    -0.358      -0.420         0
 1296       12.721    -0.628      -0.420         0
 1296        5.520    -0.038      -0.420         0
 1296        7.353     0.972      -0.420         0
```

```
 1296       7.095     0.252     -0.420       0
 1296       9.999     0.332     -0.420       0
 1296      10.715    -0.308     -0.420       0
 1308      13.233     0.422      0.534       1
 1308      13.952     0.562      0.534       1
 1308      13.757    -0.058      0.534       1
 1308      13.970     0.952      0.534       1
 1308      23.434     0.622      0.534       1
 1308       9.162     0.832      0.534       1
 1308      23.818     1.512      0.534       1
 1308      15.998     0.622      0.534       1
 1308      16.039     0.332      0.534       1
 1308      24.993     0.442      0.534       1
 1308      15.657     0.582      0.534       1
 1308      16.258     1.102      0.534       1
```

The data file is a subsample from the 1982 High School and Beyond Survey and is used extensively in *Hierarchical Linear Models* by Raudenbush and Bryk. The data file consists of 7185 students nested in 160 schools. The outcome variable of interest is student-level math achievement score (**MATHACH**). Variable **SES** is social-economic-status of a student and therefore is a student-level variable. Variable **MEANSES** is the group mean of **SES** and therefore is a school-level variable. Both **SES** and **MEANSES** are centered at the grand mean (they both have means of 0). Variable **SECTOR** is an indicator variable indicating if a school is public or catholic and is therefore a school-level variable. There are 90 public schools (SECTOR=0) and 70 catholic schools (SECTOR=1) in the sample.

Model 1: Unconditional Means Model

This model is referred as a one-way ANOVA with random effects and is the simplest possible random effect linear model and is discussed in detail by Raudenbush and Bryk. The motivation for this model is the question on how much schools vary in their mean mathematics achievement. In terms of regression equations, we have the following, where $r_{ij} \sim N(0, \sigma^2)$ and $u_{0j} \sim N(0, \tau^2)$,

$MATHACH_{ij} = \beta_{0j} + r_{ij}$
$\beta_{0j} = \gamma_{00} + u_{0j}$

Combining the two equations into one by substituting the level-2 equation to level-1 equation, we have

$MATHACH_{ij} = \gamma_{00} + u_{0j} + r_{ij}$

```
proc mixed data = in.hsb12 covtest noclprint;
   class school;
   model mathach = / solution;
   random intercept / subject = school;
run;
                  Covariance Parameter Estimates
                                     Standard         Z
Cov Parm      Subject    Estimate       Error     Value        Pr Z
Intercept     SCHOOL       8.6097      1.0778      7.99      <.0001
Residual                  39.1487      0.6607     59.26      <.0001
           Fit Statistics
-2 Res Log Likelihood        47116.8
AIC (smaller is better)      47120.8
```

```
AICC (smaller is better)      47120.8
BIC (smaller is better)       47126.9
                   Solution for Fixed Effects
                         Standard
Effect       Estimate       Error      DF    t Value    Pr > |t|
Intercept     12.6370      0.2443     159      51.72      <.0001
```

Comments:

1. In proc mixed, the statement **MODEL** includes intercept as default. Therefore, we can further request that intercept be random in the **random** statement.
2. There are different estimation methods that proc mixed can use. The default is residual (restricted) maximum likelihood and is the method that we use here. This is also the default for HLM program.
3. The option **solution** in the model statement gives the parameter estimates for the fixed effect.
4. The option **covtest** requests for the standard error for the covariance-variance parameter estimates and the corresponding z-test.
5. The option **noclprint** requests that SAS not print the class information.
6. The estimated between variance, τ^2 corresponds to the term INTERCEPT in the output of Covariance Parameter Estimates and the estimated within variance, σ^2, corresponds to the term RESIDUAL in the same output section.
7. Based on the covariance estimates, we can compute the intraclass correlation: 8.6097/(8.6097+39.1487) = .18027614. This tells us the portion of the total variance that occurs between schools.
8. To measure the magnitude of the variation among schools in their mean achievement levels, we can calculate the *plausible values range* for these means, based on the between variance we obtained from the model: $12.637 \pm 1.96*(8.61)^{1/2} = (6.89, 18.39)$.

Model 2: Including Effects of School Level (level 2) Predictors -- predicting **mathach** from **meanses**

This model is referred as regression with Means-as-Outcomes by Raudenbush and Bryk. The motivation of this model is the question on if the schools with high **MEANSES** also have high math achievement. In other words, we want to understand why there is a school difference on mathematics achievement. In terms of regression equations, we have the following.

$MATHACH_{ij} = \beta_{0j} + r_{ij}$
$\beta_{0j} = \gamma_{00} + \gamma_{01}(MEANSES) + u_{0j}$

Combining the two equations into one by substituting the level-2 equation to level-1 equation, we have

$MATHACH_{ij} = \gamma_{00} + \gamma_{01}(MEANSES) + u_{0j} + r_{ij}$

```
proc mixed data = in.hsb12 covtest noclprint;
   class school;
   model mathach = meanses / solution ddfm = bw;
   random intercept / subject = school;
run;
```

```
Covariance Parameter Estimates
                                   Standard         Z
Cov Parm      Subject    Estimate     Error     Value        Pr Z
Intercept     SCHOOL       2.6357    0.4036      6.53      <.0001
Residual                  39.1578    0.6608     59.26      <.0001
           Fit Statistics
-2 Res Log Likelihood         46961.3
AIC (smaller is better)       46965.3
AICC (smaller is better)      46965.3
BIC (smaller is better)       46971.4
                   Solution for Fixed Effects
                         Standard
Effect       Estimate       Error      DF    t Value    Pr > |t|
Intercept     12.6495      0.1492     158      84.77      <.0001
MEANSES        5.8635      0.3613     158      16.23      <.0001
        Type 3 Tests of Fixed Effects
              Num     Den
Effect         DF      DF    F Value    Pr > F
MEANSES         1     158     263.37    <.0001
```

Comments:

1. The coefficient for the constant is the predicted math achievement when all predictors are 0, so when the average school SES is 0, the students math achievement is predicted to be 12.65.
2. The variance component representing variation between schools decreases greatly (from 8.6097 to 2.6357). This means that the level-2 variable **meanses** explains a large portion of the school-to-school variation in mean math achievement. More precisely, the proportion of variance explained by **meanses** is (8.6097 - 2.6357)/8.6097 = .694, that is about 69% of the explainable variation in school mean math achievement scores is explained by **meanses**.
3. A range of plausible values for school means, given that all schools have MEANSES of zero, is $12.65 \pm 1.96 *(2.64)^{1/2} = (9.47, 15.83)$.
4. We can also calculate the conditional intraclass correlation conditional on the values of MEANSES. 2.64/(2.64 + 39.16) = .06 measures the degree of dependence among observations within schools that are of the same MEANSES.
5. Do school achievement means still vary significantly once MEANSES is controlled? From the output of Covariance Parameter Estimates, we see that the test that between variance is zero is highly significant. Therefore, we conclude that after controlling for MEANSES, significant variation among school mean math achievement still remains to be explained.
6. Notice though, the standard error used to perform the above hypothesis test is based on large-sample theory of the maximum likelihood estimates and in many cases the normality approximation will be extremely poor. We will only use these results as guidance for further analysis, rather than definitive results. In SAS version 8 and later, SAS uses one-tailed z-test on variance and two-tailed z-test on covariance, trying to avoid misleading results by previously used two-tailed test for both.
7. The option **ddfm = bw** (between and within method) used in the model statement is to request SAS to use between and within method for computing the denominator degrees of freedom for the tests of fixed effects, instead of the default, containment method. This option is especially useful when there are large number of random effects in the model and the design is severely unbalanced. The default, on the other hand, matches the tests performed for balanced split-plot designs and should be adequate for moderately unbalanced designs.

Model 3: Including Effects of Student-Level Predictors--predicting **mathach** from centered student-level ses, **cses**

This model is referred as a random-coefficient model by Raudenbush and Bryk. Pretend that we run regression of **mathach** on centered **ses** on each school, that is we are going to run 160 regressions.

1. What would be the average of the 160 regression equations (both intercept and slope)?
2. How much do the regression equations vary from school to school?
3. What is the correlation between the intercepts and slopes?

These are some of the questions that motivates the following model.

$MATHACH_{ij} = \beta_{0j} + \beta_{1j}$ (SES - MEANSES) + r_{ij}
$\beta_{0j} = \gamma_{00} + u_{0j}$
$\beta_{1j} = \gamma_{10} + u_{1j}$

Combining the two equations into one by substituting the level-2 equation to level-1 equation, we have

$MATHACH_{ij} = \gamma_{00} + \gamma_{10}$(SES - MEANSES) + $u_{0j} + u_{1j}$(SES - MEANSES) + r_{ij}

```
data hsbc;
  set in.hsb12;
    cses = ses - meanses;
run;
proc mixed data = hsbc noclprint covtest noitprint;
  class school;
  model mathach = cses / solution ddfm = bw notest;
  random intercept cses / subject = school type = un gcorr;
run;
```

```
           Estimated G Correlation Matrix
 Row    Effect       SCHOOL        Col1        Col2
   1    Intercept    1224        1.0000     0.02068
   2    cses         1224       0.02068      1.0000
                  Covariance Parameter Estimates
                                    Standard         Z
Cov Parm     Subject    Estimate       Error     Value        Pr Z
UN(1,1)      SCHOOL       8.6769      1.0786      8.04      <.0001
UN(2,1)      SCHOOL      0.05075      0.4062      0.12      0.9006
UN(2,2)      SCHOOL       0.6940      0.2808      2.47      0.0067
Residual                 36.7006      0.6258     58.65      <.0001

           Fit Statistics
-2 Res Log Likelihood         46714.2
AIC (smaller is better)       46722.2
AICC (smaller is better)      46722.2
BIC (smaller is better)       46734.5
  Null Model Likelihood Ratio Test
    DF    Chi-Square      Pr > ChiSq
     3       1065.70          <.0001
                   Solution for Fixed Effects
                         Standard
Effect       Estimate       Error       DF    t Value    Pr > |t|
Intercept     12.6493      0.2445      159      51.75      <.0001
cses           2.1932      0.1283     7024      17.10      <.0001
```

Comments:

1. Specifying level-1 predictor **cses** as random effect, we formulate that effect of **cses** can vary across schools.
2. The option **type = un** in the random statement allows us to estimate the three parameters (the variance of **intercept** and the variance of slopes for **cses** and the covariance between them) from the data.
3. Option **gcorr** displays the correlation matrix corresponding to the estimated variance-covariance matrix, called G matrix.
4. The covariance estimate is 0.05075 with standard error 0.4062. That yields a p-vlaue of 0.9006. This is saying that there is no evidence that the effect of **cses** depending upon the average math achievement in the school.
5. In the output of Covariance Parameter Estimates, the parameter corresponding to UN(2,2) is the variability in slopes of **cses**. The estimate is 0.6940 with standard error 0.2808. That yields a p-value of 0.0067 for 1-tailed test. The test being significant tells us that we can not accept the hypothesis that there is no difference in slopes among schools.
6. The 95% plausible value range for the school means is $12.65 \pm 1.96 *(8.68)^{1/2} = (6.87, 18.41)$.
7. The 95% plausible value range for the SES-achievement slope is $2.19 \pm 1.96 *(.69)^{1/2} = (.56, 3.82)$.
8. Notice that the residual variance is now 36.70, comparing with the residual variance of 39.15 in the one-way ANOVA with random effects model. We can compute the proportion variance explained at level 1 by (39.15 - 36.70) / 39.15 = .063. This means using student-level SES as a predictor of math achievement reduced the within-school variance by 6.3%.

Model 4: Including Both Level-1 and Level-2 Predictors --predicting **mathach** from **meanses, sector, cses** and the cross level interaction of **meanses** and **sector** with **cses**

This model is referred as an intercepts and slopes-as-outcomes model by Raudenbush and Bryk. We have examined the variability of the regression equations across schools. Now we will build an explanatory model to account for the variability. That is we want to model the following:

$$\mathrm{MATHACH}_{ij} = \beta_{0j} + \beta_{1j}\,(\mathrm{SES} - \mathrm{MEANSES}) + r_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\mathrm{MEANSES}) + \gamma_{02}(\mathrm{SECTOR}) + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\mathrm{MEANSES}) + \gamma_{12}(\mathrm{SECTOR}) + u_{1j}$$

Combining the two equations into one by substituting the level-2 equation to level-1 equation, we have

$$\begin{aligned}\mathrm{MATHACH}_{ij} = {} & \gamma_{00} + \gamma_{01}(\mathrm{MEANSES}) + \gamma_{02}(\mathrm{SECTOR}) + \gamma_{10}\,(\mathrm{SES} - \mathrm{MEANSES}) + \\ & \gamma_{11}(\mathrm{MEANSES})*\,(\mathrm{SES} - \mathrm{MEANSES}) + \gamma_{12}(\mathrm{SECTOR})*\,(\mathrm{SES} - \mathrm{MEANSES}) + \\ & u_{0j} + u_{1j}(\mathrm{SES}-\mathrm{MEANSES}) + r_{ij}\end{aligned}$$

The questions that we are interested in are:

1. Do MEANSES and SECTOR significantly predict the intercept?
2. Do MEANSES and SECTOR significantly predict the within-school slopes?
3. How much variation in the intercepts and the slopes is explained by MEANSES and SECTOR?

```
proc mixed data = hsbc noclprint covtest noitprint;
  class school;
  model mathach = meanses sector cses meanses*cses sector*cses
                / solution ddfm = bw notest;
  random intercept cses / subject = school type = un;
run;
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	SCHOOL	2.3817	0.3717	6.41	<.0001
UN(2,1)	SCHOOL	0.1926	0.2045	0.94	0.3464
UN(2,2)	SCHOOL	0.1014	0.2138	0.47	0.3177
Residual		36.7212	0.6261	58.65	<.0001

Fit Statistics

-2 Res Log Likelihood	46503.7
AIC (smaller is better)	46511.7
AICC (smaller is better)	46511.7
BIC (smaller is better)	46524.0

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
3	220.57	<.0001

Solution for Fixed Effects

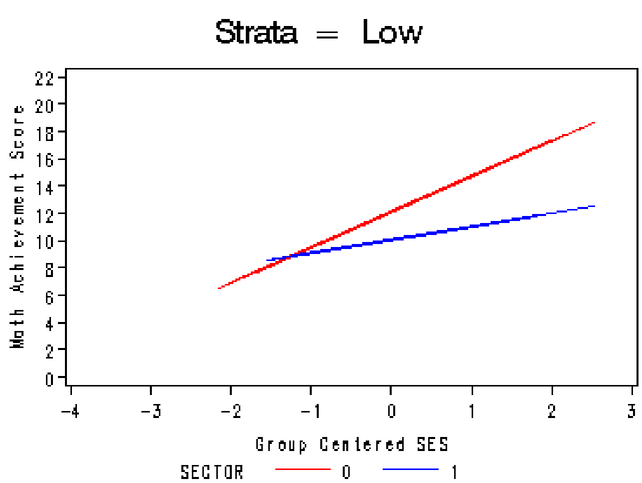
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	12.1136	0.1988	157	60.93	<.0001
MEANSES	5.3391	0.3693	157	14.46	<.0001
SECTOR	1.2167	0.3064	157	3.97	0.0001
cses	2.9388	0.1551	7022	18.95	<.0001
MEANSES*cses	1.0389	0.2989	7022	3.48	0.0005
SECTOR*cses	-1.6426	0.2398	7022	-6.85	<.0001

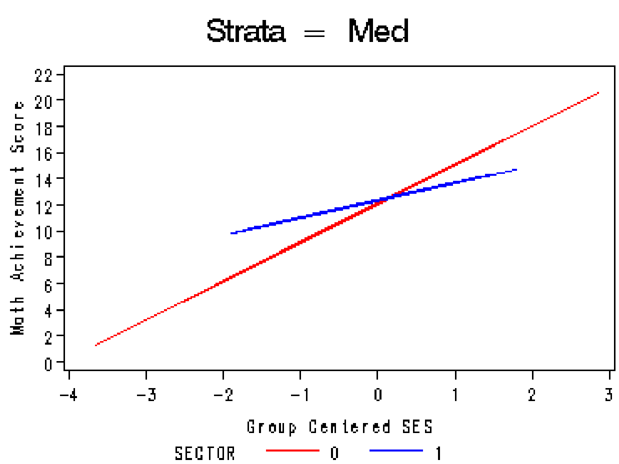
Comments:

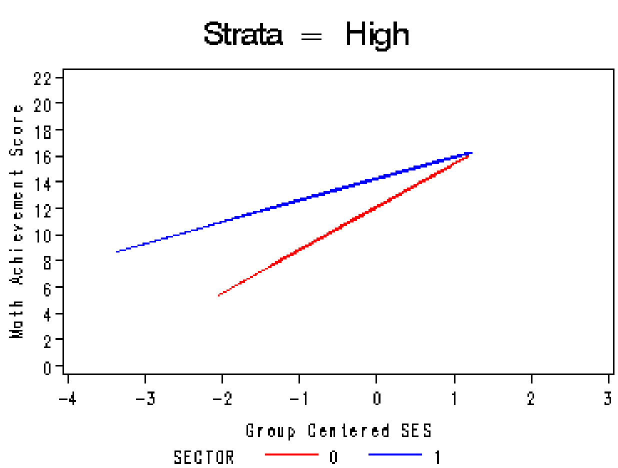
1. First take a look at the output of Solutions for Fixed Effects. The first three parameters are about the intercept, or more precisely about the mean math achievement across schools. We see that MEANSES is positively related to math achievement and catholic schools have significantly higher mean math achievement than public schools, controlling for other effects.
2. The last three parameters in the output are about the slopes. Schools of high MEANSES tend to have larger slopes and catholic schools have significantly weaker slopes, on the average, than public schools.
3. Variable **sector** and its interaction with **cses** are significant in the model, indicating that the intercepts and the slopes for **cses** are different for Catholic and public schools. This can also be shown by plotting the predicted math achievement scores constraining the meanses to low, medium and high. We use 25th/50th/75th percentiles to define the strata of low, medium and high.

```
4.   proc univariate data = hsbc;
5.     var meanses;
6.   run;
7.   /*
8.   90%              0.523
9.   75% Q3           0.333
10.  50% Median       0.038
11.  25% Q1          -0.317
12.  10%             -0.579
13.  5%              -0.696
14.  1%              -1.043
15.  0% Min          -1.188
```

```
*/
data toplot;
  set hsbc;
  if meanses <= -0.317 then do;
         ms = -0.317;
         strata = "Low";   end;
  else if meanses >= 0.333 then do;
         ms = 0.333;
         strata = "Hig";   end;
  else do; ms = 0.038; strata = "Med" ; end;
  predicted = 12.1136 + 5.3391*ms + 1.2167*sector + 2.9388*cses +
              1.0389*ms*cses - 1.6426*sector*cses;
run;
proc sort data = toplot;
   by strata;
run;
goptions reset = all;
symbol1 v = none i = join c = red ;
symbol2 v = none i = join c = blue  ;
axis1 order = (-4 to 3 by 1) minor = none label=("Group Centered SES");
axis2 order = (0 to 22 by 2) minor = none label=(a = 90 "Math Achievement
Score");
proc gplot data = toplot;
   by strata;
   plot predicted*cses = sector / vaxis = axis2 haxis = axis1;
run;
quit;
```







16. Possibly there would be two-way interaction between **meanses** and **sector** and a three way interaction between **meanses**, **cses** and **sector**. We can test it by adding the interaction into the model. For example,

```
17.  proc mixed data = hsbc noclprint covtest noitprint;
18.    class school;
19.    model mathach = meanses sector cses meanses*sector
20.                    meanses*cses sector*cses meanses*sector*cses
21.                    / solution ddfm = bw notest;
22.    random intercept cses / subject = school type = un;
   run;
```

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	12.1842	0.2030	156	60.01	<.0001
MEANSES	5.8732	0.5065	156	11.60	<.0001
SECTOR	1.2430	0.3052	156	4.07	<.0001
cses	2.9513	0.1616	7021	18.26	<.0001
MEANSES*SECTOR	**-1.1276**	**0.7355**	**156**	**-1.53**	**0.1273**
MEANSES*SECTOR*cses	**-0.1888**	**0.5997**	**7021**	**-0.31**	**0.7528**
MEANSES*cses	1.1289	0.4232	7021	2.67	0.0077
SECTOR*cses	-1.6407	0.2406	7021	-6.82	<.0001

23. Since the variance component for slopes is very small and its corresponding p-value is 0.3177. We cannot reject the hypothesis that the slopes do not differ across schools. Similarly, we can not reject the hypothesis that the covariance between intercepts and slopes is zero. Therefore, a simpler model can be used:

```
24.  proc mixed data = hsbc noclprint covtest noitprint;
25.    class school;
26.    model mathach = meanses sector cses meanses*cses sector*cses / solution
   ddfm = bw notest;
27.    random intercept / subject = school;
   run;
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
Intercept	SCHOOL	2.3752	0.3709	6.40	<.0001
Residual		36.7661	0.6207	59.24	<.0001

Fit Statistics

-2 Res Log Likelihood	46504.8
AIC (smaller is better)	46508.8
AICC (smaller is better)	46508.8
BIC (smaller is better)	46514.9

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	12.1138	0.1986	157	60.98	<.0001
MEANSES	5.3429	0.3690	157	14.48	<.0001
SECTOR	1.2146	0.3061	157	3.97	0.0001
cses	2.9358	0.1507	7022	19.48	<.0001
MEANSES*cses	1.0441	0.2910	7022	3.59	0.0003
SECTOR*cses	-1.6421	0.2331	7022	-7.04	<.0001

To compare the original model with this simplified one, we can compare their -2LL's, since the fixed portion of these two models are the same.

Model	Number of parameters	-2 LL
restricted	2	46504.8
Unrestricted	4	46503.7

Approximately, the difference in -2LL's is a χ^2 distribution with two degrees of freedom (corresponding to the difference in the number of parameters). The p-value is .577. This justifies the use of the simpler model. The SAS program is shown below.

```
data pvalue;
  df = 2; chisq = 46504.8 - 46503.7;
  pvalue = 1 - probchi(chisq, df);
run;
proc print data = pvalue noobs;
run;
df    chisq     pvalue
 2     1.1     0.57695
```

Linear Growth Model

A segment of the data file:

```
id    time    cons    covar     y
 1     0       1       137     205
 1     1       1       137     217
 1     2       1       137     268
 1     3       1       137     302
 2     0       1       123     219
 2     1       1       123     243
 2     2       1       123     279
 2     3       1       123     302
 3     0       1       129     142
 3     1       1       129     212
 3     2       1       129     250
 3     3       1       129     289
 4     0       1       125     206
 4     1       1       125     230
 4     2       1       125     248
 4     3       1       125     273
 5     0       1        81     190
 5     1       1        81     220
 5     2       1        81     229
 5     3       1        81     220
```

Model 1: Unconditional Linear Growth Model -- page 340

```
proc mixed data = willett noclprint covtest;
  class id;
  model y = time /solution ddfm = bw notest;
  random intercept time / subject = id type = un;
run;
                  Covariance Parameter Estimates
```

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	id	1198.78	318.38	3.77	<.0001
UN(2,1)	id	-179.26	88.9634	-2.01	0.0439
UN(2,2)	id	132.40	40.2107	3.29	0.0005
Residual		159.48	26.9566	5.92	<.0001

Fit Statistics

-2 Res Log Likelihood	1266.8
AIC (smaller is better)	1274.8
AICC (smaller is better)	1275.1
BIC (smaller is better)	1281.0

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
3	120.90	<.0001

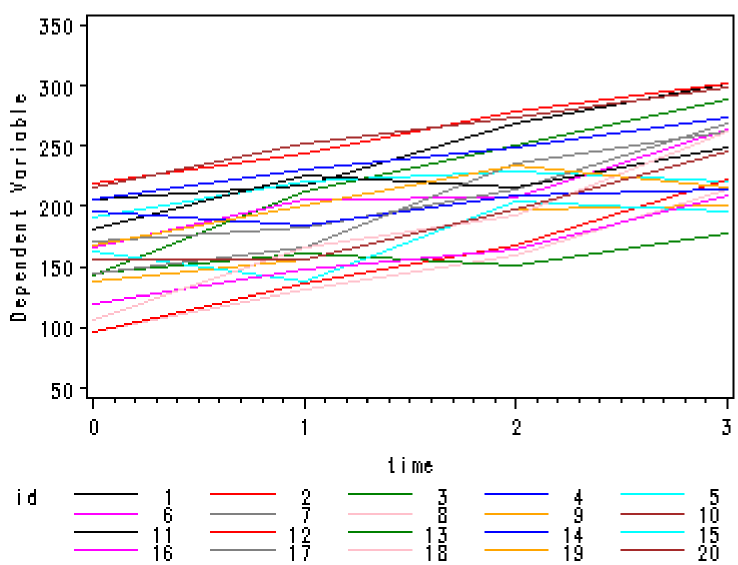
Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	164.37	6.1188	34	26.86	<.0001
time	26.9600	2.1666	104	12.44	<.0001

Comments:

1. Notice that variable **time** is coded 0, 1, 2 and 3. Therefore, the intercept is the estimate of the initial value and the slope is the estimate of the rate of change across occasions.
2. We may want to visually see the relationship between the dependent variable and time by subject. This gives us a good sense if the linear relationship holds across all the subjects and if the slopes vary across all the subjects.

```
3.  proc gplot data = willett;
4.    plot y*time = id;
5.    where id <=20;
6.  run;
    quit;
```



Model 2: A Linear Growth Model with a Person-Level Covariance -- predicting y by **time** and centered **covar** -- page 344

```
data willett;
  set in.willett;
  wave = time;
  ccovar = covar -  113.4571429;
run;
proc mixed data = willett noclprint covtest;
  class id;
  model y = time ccovar time*ccovar /solution ddfm = bw notest;
  random intercept time / subject = id type = un gcorr;
run;
```

```
         Estimated G Correlation Matrix
 Row    Effect       id        Col1        Col2
   1    Intercept     1      1.0000     -0.4895
   2    time          1     -0.4895      1.0000
                   Covariance Parameter Estimates
                                     Standard         Z
Cov Parm     Subject    Estimate        Error     Value        Pr Z
UN(1,1)      id          1236.41       332.40      3.72      <.0001
UN(2,1)      id          -178.23      85.4298     -2.09      0.0370
UN(2,2)      id           107.25      34.6767      3.09      0.0010
Residual                  159.48      26.9566      5.92      <.0001
           Fit Statistics
-2 Res Log Likelihood          1260.3
AIC (smaller is better)        1268.3
AICC (smaller is better)       1268.6
BIC (smaller is better)        1274.5
  Null Model Likelihood Ratio Test
    DF    Chi-Square      Pr > ChiSq
     3        120.72          <.0001
                     Solution for Fixed Effects
                            Standard
Effect         Estimate        Error      DF    t Value    Pr > |t|
Intercept        164.37       6.2061      33      26.49      <.0001
time            26.9600       1.9939     103      13.52      <.0001
ccovar          -0.1136       0.5040      33      -0.23      0.8231
time*ccovar      0.4329       0.1619     103       2.67      0.0087
```

Comments:

1. Variable **wave** created in the data step will be used in our next model.
2. Estimated correlation matrix among the random effect is requested by using the option **gcorr**.
3. Comparing with the model of unconditional growth, this model

---

Model 3: Exploring the Structure of Variance Covariance Matrix Within Persons

A. Compound Symmetry

```
proc mixed data = willett covtest noitprint;
  class id wave;
  model y = time / s notest;
  repeated wave /type = cs subject = id r;
run;
```

```
             Estimated R Matrix for id 1
 Row        Col1        Col2        Col3        Col4
   1     1280.71      904.81      904.81      904.81
```

```
   2      904.81     1280.71      904.81      904.81
   3      904.81      904.81     1280.71      904.81
   4      904.81      904.81      904.81     1280.71
                  Covariance Parameter Estimates
                                    Standard        Z
Cov Parm     Subject    Estimate       Error     Value        Pr Z
CS           id           904.81      242.59      3.73      0.0002
Residual                  375.90     52.1281      7.21      <.0001
           Fit Statistics
-2 Res Log Likelihood          1300.3
AIC (smaller is better)        1304.3
AICC (smaller is better)       1304.4
BIC (smaller is better)        1307.5
  Null Model Likelihood Ratio Test
    DF    Chi-Square      Pr > ChiSq
     1         87.39          <.0001
                   Solution for Fixed Effects
                         Standard
Effect       Estimate       Error      DF    t Value    Pr > |t|
Intercept      164.37      5.7766      34      28.45      <.0001
time          26.9600      1.4656     104      18.40      <.0001
```

B.Unstructured

```
proc mixed data = willett covtest noitprint;
  class id wave;
  model y = time / s notest;
  repeated wave /type = un subject = id r;
run;
            Estimated R Matrix for id 1
 Row        Col1        Col2        Col3        Col4
   1     1307.96      977.17      921.87      563.54
   2      977.17     1120.32     1018.97      855.53
   3      921.87     1018.97     1289.47     1081.77
   4      563.54      855.53     1081.77     1415.03
                 Covariance Parameter Estimates
                                   Standard        Z
Cov Parm    Subject    Estimate       Error     Value        Pr Z
UN(1,1)     id          1307.96      316.95      4.13      <.0001
UN(2,1)     id           977.17      266.55      3.67      0.0002
UN(2,2)     id          1120.32      270.69      4.14      <.0001
UN(3,1)     id           921.87      272.81      3.38      0.0007
UN(3,2)     id          1018.97      269.55      3.78      0.0002
UN(3,3)     id          1289.47      312.07      4.13      <.0001
UN(4,1)     id           563.54      252.45      2.23      0.0256
UN(4,2)     id           855.53      260.70      3.28      0.0010
UN(4,3)     id          1081.77      296.64      3.65      0.0003
UN(4,4)     id          1415.03      343.17      4.12      <.0001
           Fit Statistics
-2 Res Log Likelihood          1263.4
AIC (smaller is better)        1283.4
AICC (smaller is better)       1285.2
BIC (smaller is better)        1299.0
  Null Model Likelihood Ratio Test
    DF    Chi-Square      Pr > ChiSq
     9        124.30          <.0001
                   Solution for Fixed Effects
                         Standard
Effect       Estimate       Error      DF    t Value    Pr > |t|
Intercept      165.83      5.8668      34      28.27      <.0001
```

```
time           26.5846      2.1215      34      12.53      <.0001
```

C. AR(1)

```
proc mixed data = willett covtest noitprint;
  class id wave;
  model y = time / s notest;
  repeated wave /type = ar(1) subject = id r;
run;
```

Estimated R Matrix for id 1

Row	Col1	Col2	Col3	Col4
1	1323.77	1092.07	900.93	743.24
2	1092.07	1323.77	1092.07	900.93
3	900.93	1092.07	1323.77	1092.07
4	743.24	900.93	1092.07	1323.77

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
AR(1)	id	0.8250	0.03937	20.96	<.0001
Residual		1323.77	258.56	5.12	<.0001

Fit Statistics

-2 Res Log Likelihood	1273.5
AIC (smaller is better)	1277.5
AICC (smaller is better)	1277.6
BIC (smaller is better)	1280.6

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	114.26	<.0001

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	164.34	6.1371	34	26.78	<.0001
time	27.1979	1.9198	104	14.17	<.0001