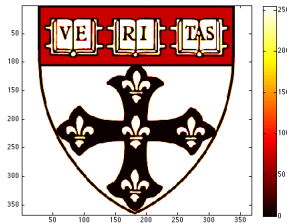


BIO 226, Spring 2015: Lab 2

Using the MIXED procedure in SAS

Instructor: Brent Coull
TA: Fei Li

February 19, 2015



Background

We will use the `MIXED` procedure in SAS to model the blood lead levels in the treated group of the 'Treatment of Lead-Exposed Children' (TLC) dataset.

- Exposure to lead is associated with deficits in cognitive ability.
- Children with high lead levels can be treated with injections which generally require hospitalization.



Background

- In this study, investigators are interested in assessing a new agent, *succimer*, which can be given orally.
- Blood lead levels were measured at baseline, and after 1, 4 and 6 weeks of follow-up.

In the analysis that follows, we first profile the blood lead levels over time in the treated (succimer) group. And then explore the placebo group before examining the full dataset.

Read in Dataset

Use the following code to read-in the dataset and print the first few observations:

```
DATA lead;  
  INFILE 'D:\BI0226\tlc.txt';  
  INPUT id trmt $ y1 y2 y3 y4;  
  RUN;
```

```
PROC PRINT DATA=lead (obs=5);  
  RUN;
```

| Obs | id | trmt | y1 | y2 | y3 | y4 |
|-----|----|------|------|------|------|------|
| 1 | 1 | P | 30.8 | 26.9 | 25.8 | 23.8 |
| 2 | 2 | A | 26.5 | 14.8 | 19.5 | 21.0 |
| 3 | 3 | A | 25.8 | 23.0 | 19.1 | 23.2 |
| 4 | 4 | P | 24.7 | 24.5 | 22.0 | 22.5 |
| 5 | 5 | A | 20.4 | 2.8 | 3.2 | 9.4 |

Note that P corresponds to the placebo and A corresponds to 'succimer' treatment.

Read in Dataset - continued

Since our focus is on the 'succimer' group, we will need to subset these observations from the rest of the data. Use the following code to do the subsetting and then print the first few observations:

```
DATA lead2;  
  SET lead;  
  IF trmt = 'A';  
RUN;  
  
PROC PRINT DATA=lead2 (obs=5);  
  RUN;
```

| Obs | id | trmt | y1 | y2 | y3 | y4 |
|-----|----|------|------|------|------|------|
| 1 | 2 | A | 26.5 | 14.8 | 19.5 | 21.0 |
| 2 | 3 | A | 25.8 | 23.0 | 19.1 | 23.2 |
| 3 | 5 | A | 20.4 | 2.8 | 3.2 | 9.4 |
| 4 | 6 | A | 20.4 | 5.4 | 4.5 | 11.9 |
| 5 | 12 | A | 24.8 | 23.1 | 24.6 | 30.9 |

- To obtain the subsetted data, we use another DATA step.
- Within any DATA step, the SET statement allows you to access an existing SAS dataset.
- Within any DATA step, the IF statement allows us to keep the observations matching the given criteria.
- Within any DATA step, the DATA statement specifies the name of the new dataset. In the code on the previous slide, we named the new dataset (only subjects with 'succimer' treatment) lead2 rather than overwrite the original dataset. (In case we need to access the original information again.)

Initial Thoughts on Data

Here the data are in “multivariate” (wide) form and must be transformed to “univariate” (long) form in order to use PROC MIXED. The “multivariate” form allows us to examine the “profile” of blood lead levels over time. To do this, we calculate the empirical mean lead levels at each of the four time points to get a rough idea of the trend over time.

Means Over Time

```
PROC MEANS DATA = lead2;  
  VAR y1-y4;  
  RUN;
```

| Variable | N | Mean | Std Dev | Minimum | Maximum |
|----------|----|------------|-----------|------------|------------|
| y1 | 50 | 26.5400000 | 5.0209358 | 19.7000000 | 41.1000000 |
| y2 | 50 | 13.5220000 | 7.6724870 | 2.8000000 | 39.0000000 |
| y3 | 50 | 15.5140000 | 7.8522065 | 3.0000000 | 40.4000000 |
| y4 | 50 | 20.7620000 | 9.2463316 | 4.1000000 | 63.9000000 |

Note that we could also calculate mean lead levels at each timepoint if the data is in “univariate” form using the BY command.

- Are there any apparent differences in the mean blood levels over time?
- Does the variability appear to be constant over time?

Variance/Covariance Structure

To obtain a covariance or correlation matrix for within subject outcomes, use PROC CORR:

```
PROC CORR cov noprob nosimple data=lead2;  
  VAR y1-y4;  
RUN;
```

| ----- Covariance Matrix, DF = 49 ----- | | | | |
|--|----------|----------|----------|----------|
| | y1 | y2 | y3 | y4 |
| y1 | 25.20972 | 15.46542 | 15.13800 | 22.98542 |
| y2 | 15.46547 | 58.86705 | 44.02907 | 35.96595 |
| y3 | 15.13800 | 44.02907 | 61.65714 | 33.02197 |
| y4 | 22.98547 | 35.96595 | 33.02197 | 85.49464 |

NOPROB - Suppresses p -values

NOSIMPLE - Suppresses descriptive statistics

“Multivariate” to “Univariate” form

We entered the data in “multivariate” form (one line per subject), but in order to use PROC MIXED we need the “univariate” form (one line per observation).

The “multivariate” form for the first two subjects looks like:

| id | y1 | y2 | y3 | y4 |
|----|------|------|------|------|
| 1 | 26.5 | 14.8 | 19.5 | 21.0 |
| 2 | 25.8 | 23.0 | 19.1 | 23.2 |

The “univariate” form for the first two subjects becomes:

| id | y |
|----|------|
| 1 | 26.5 |
| 1 | 14.8 |
| 1 | 19.5 |
| 1 | 21.0 |
| 2 | 25.8 |
| 2 | 23.0 |
| 2 | 19.1 |
| 2 | 23.2 |

“Multivariate” to “Univariate” form - continued

Note that there is no longer an obvious way to distinguish among the four values of blood lead levels for each subject. To keep track of the time, we generate two new variables:

- 1 **t**, that indicates the occasion of measurement.
- 2 **time**, that indicates the number of weeks elapsed since the baseline.

To obtain the “univariate” form, we use another DATA step. Within this new DATA step, we will include statements to create **y**, **t**, and **time**.

“Multivariate” to “Univariate” form - continued

To transform from “multivariate” (one line per subject) to “univariate” form (one line per observation) we use:

```
-----  
Obs    id    trmt    y1    y2    y3    y4  
  1     2     A    26.5   14.8   19.5   21.0  
  2     3     A    25.8   23.0   19.1   23.2  
-----
```

```
DATA lead3;  
  SET lead2;  
  y=y1; time=0; t=1; OUTPUT;  
  y=y2; time=1; t=2; OUTPUT;  
  y=y3; time=4; t=3; OUTPUT;  
  y=y4; time=6; t=4; OUTPUT;  
  DROP y1-y4;  
  RUN;
```

```
PROC PRINT DATA=lead3 (obs=5);  
  RUN;
```

```
-----  
Obs    id    trmt    y    time    t  
  1     2     A    26.5     0     1  
  2     2     A    14.8     1     2  
  3     2     A    19.5     4     3  
  4     2     A    21.0     6     4  
  5     3     A    25.8     0     1  
-----
```

“Multivariate” Dataset

Role of variable `t` in this dataset:

- Essential for the `REPEATED` option in `PROC MIXED` when time is specified as a continuous variable
- Keeps track of the order of repeated measures
- If values are missing, `t` establishes placement in covariance matrix

| Obs | id | trmt | y | time | t |
|-----|----|------|------|------|---|
| 1 | 2 | A | 26.5 | 0 | 1 |
| 2 | 2 | A | 14.8 | 1 | 2 |
| 4 | 2 | A | 21.0 | 6 | 4 |
| 5 | 3 | A | 25.8 | 0 | 1 |

Role of variable `time` in this dataset:

- Keeps record of actual time of measurement
- Used as an explanatory variable in the regression model (can be categorical or continuous)

Visualization of Data

To obtain trajectory plots for each individual, use PROC GPLOT:

```
PROC GPLOT DATA = lead3;  
  SYMBOL1  interpol=join value=triangle;  
  SYMBOL2  interpol=join value=triangle;  
  SYMBOL3  interpol=join value=triangle;  
  SYMBOL4  interpol=join value=triangle;  
  SYMBOL5  interpol=join value=triangle;  
  PLOT y*time=id;  
RUN;
```

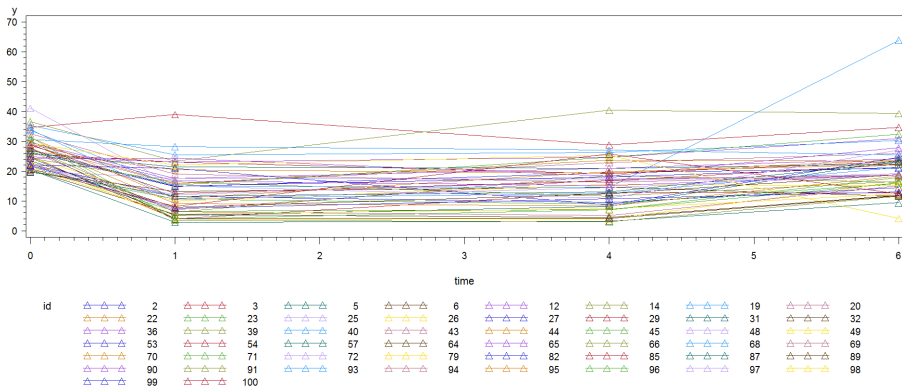
INTERPOL=JOIN connects data points with straight lines.

VALUE a plot symbol for the data points

SYMBOLn will rotate through 5 colors

the " =id" in the PLOT statement tells SAS to rotate through the different colors based on id

Visualization of Data



Structure of PROC MIXED

Now the data are in the correct format for PROC MIXED. The following code can be used to model lead level as a function of time. This is also called the "saturated" model because the number of parameters in the model equals the the number of unique means contained in the data (one for each time point).

```
PROC MIXED DATA=lead3;  
  CLASS id t time;  
  MODEL y = time / SOLUTION CHISQ;  
  REPEATED t / TYPE=UN SUBJECT=id R;  
  RUN;
```


Additional Notes on PROC MIXED

```
PROC MIXED DATA=lead3;  
  CLASS id t time;  
  MODEL y = time / SOLUTION CHISQ;  
  REPEATED t / TYPE=UN SUBJECT=id R;  
RUN;
```

- Parameter estimation - SAS uses the REML method by default.

- CLASS statement:

- 1 *Must* appear before MODEL statement.
- 2 Variables specified here will be treated as categorical.
- 3 Generates appropriate design variables (e.g. dummy variables) when they appear in the model. (similar to the i.variable in STATA).
- 4 To specify subjects within the REPEATED statement, id must be included in the CLASS statement.
- 5 To be used in the REPEATED statement, t must be in the CLASS statement.
- 6 To be included in the model as a categorical variable, time must be in the CLASS statement, otherwise time will be treated as continuous.

Additional Notes on PROC MIXED - continued

```
PROC MIXED DATA=lead3;  
  CLASS id t time;  
  MODEL y = time / SOLUTION CHISQ;  
  REPEATED t / TYPE=UN SUBJECT=id R;  
RUN;
```

- MODEL statement

- 1 Specifies form of the model
- 2 SOLUTION displays parameter estimates, standard errors...

- REPEATED statement

- 1 Specify variable that keeps track of the order of repeated measures
- 2 TYPE option specifies assumed form of covariance matrix
- 3 SUBJECT specifies variable that denotes independent units of observation
- 4 R option displays estimated covariance matrix for individual subject

- SAS Online Documentation

- 1 <http://support.sas.com/onlinedoc/913/docMainpage.jsp>
- 2 SAS/STAT and Base SAS are commonly used sections
- 3 Google “proc mixed sas” - usually the first in the list

Output from PROC MIXED

The Mixed Procedure

Model Information

| | |
|---------------------------|----------------|
| Data Set | WORK.LEAD3 |
| Dependent Variable | y |
| Covariance Structure | Unstructured |
| Subject Effect | id |
| Estimation Method | REML |
| Residual Variance Method | None |
| Fixed Effects SE Method | Model-Based |
| Degrees of Freedom Method | Between-Within |

Class Level Information

| Class | Levels | Values |
|-------|--------|--|
| id | 50 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 |
| t | 4 | 1 2 3 4 |
| time | 4 | 0 1 4 6 |

| Dimensions | |
|-----------------------|----|
| Covariance Parameters | 10 |
| Columns in X | 5 |
| Columns in Z | 0 |
| Subjects | 50 |
| Max Obs Per Subject | 4 |

| Number of Observations | |
|---------------------------------|-----|
| Number of Observations Read | 200 |
| Number of Observations Used | 200 |
| Number of Observations Not Used | 0 |

| Estimated R Matrix for id 1 | | | | |
|-----------------------------|---------|---------|---------|---------|
| Row | Col1 | Col2 | Col3 | Col4 |
| 1 | 25.2098 | 15.4654 | 15.1380 | 22.9854 |
| 2 | 15.4654 | 58.8671 | 44.0291 | 35.9660 |
| 3 | 15.1380 | 44.0291 | 61.6571 | 33.0220 |
| 4 | 22.9854 | 35.9660 | 33.0220 | 85.4946 |

| Covariance Parameter Estimates | | |
|--------------------------------|---------|----------|
| Cov Parm | Subject | Estimate |
| UN(1,1) | id | 25.2098 |
| UN(2,1) | id | 15.4654 |
| UN(2,2) | id | 58.8671 |
| UN(3,1) | id | 15.1380 |
| UN(3,2) | id | 44.0291 |
| UN(3,3) | id | 61.6571 |
| UN(4,1) | id | 22.9854 |
| UN(4,2) | id | 35.9660 |
| UN(4,3) | id | 33.0220 |
| UN(4,4) | id | 85.4946 |

| Fit Statistics | |
|--------------------------|--------|
| -2 Res Log Likelihood | 1280.3 |
| AIC (smaller is better) | 1300.3 |
| AICC (smaller is better) | 1301.5 |
| BIC (smaller is better) | 1319.5 |

Null Model Likelihood Ratio Test

| DF | Chi-Square | Pr > ChiSq |
|----|------------|------------|
| 9 | 86.73 | <.0001 |

Solution for Fixed Effects

| | | Standard | | | | |
|-----------|------|----------|--------|----|---------|---------|
| Effect | time | Estimate | Error | DF | t Value | Pr > t |
| Intercept | | 20.7620 | 1.3076 | 49 | 15.88 | <.0001 |
| time | 0 | 5.7780 | 1.1378 | 49 | 5.08 | <.0001 |
| time | 1 | -7.2400 | 1.2036 | 49 | -6.02 | <.0001 |
| time | 4 | -5.2480 | 1.2736 | 49 | -4.12 | 0.0001 |
| time | 6 | 0 | . | . | . | . |

Type 3 Tests of Fixed Effects

| | | Num | Den | | | |
|--------|----|-----|------------|---------|------------|--------|
| Effect | DF | DF | Chi-Square | F Value | Pr > ChiSq | Pr > F |
| time | 3 | 49 | 163.72 | 54.57 | <.0001 | <.0001 |

Changing the reference group in an analysis

What if you want to change the model slightly so that the lowest level of time is the reference group (instead of the highest level)? Use the following code:

```
DATA lead4;  
  SET lead3;  
  IF time=0 THEN timecat=11;  
  IF time=1 THEN timecat=1;  
  IF time=4 THEN timecat=4;  
  IF time=6 THEN timecat=6;  
RUN;  
  
PROC MIXED DATA=lead4;  
  CLASS id t timecat;  
  MODEL y = timecat / SOLUTION CHISQ;  
  REPEATED t / TYPE=un SUBJECT=id R;  
RUN;
```

Solution for Fixed Effects

| Effect | timecat | Estimate | Standard Error | DF | t Value | Pr > t |
|-----------|---------|----------|-------------------|----|---------|---------|
| Intercept | | 26.5400 | 0.7101 | 49 | 37.38 | <.0001 |
| timecat | 1 | -13.0180 | 1.0310 | 49 | -12.63 | <.0001 |
| timecat | 4 | -11.0260 | 1.0639 | 49 | -10.36 | <.0001 |
| timecat | 6 | -5.7780 | 1.1378 | 49 | -5.08 | <.0001 |
| timecat | 11 | 0 | . | . | . | . |

Type 3 Tests of Fixed Effects

| Effect | Num DF | Den DF | Chi-Square | F Value | Pr > ChiSq | Pr > F |
|---------|-----------|-----------|------------|---------|------------|--------|
| timecat | 3 | 49 | 163.72 | 54.57 | <.0001 | <.0001 |

Changing the reference group in an analysis - continued

Alternatively, you can make use of the ORDER option in the PROC MIXED statement (like we did in Lab 1!).

```
PROC SORT DATA=lead3;  
  BY id DESCENDING time;  
  RUN;  
  
PROC MIXED DATA=lead3 ORDER=DATA;  
  CLASS id time t;  
  MODEL y = time / SOLUTION CHISQ;  
  REPEATED t / TYPE=UN SUBJECT=id R;  
  RUN;
```


Fitting a model using a linear trend in time (continuous)

What if you believe that lead levels decrease **linearly** with time? By removing time from the CLASS statement, we tell SAS that time is no longer a categorical covariate, but should be considered a continuous covariate. To model time linearly, we fit the model:

```
PROC MIXED DATA=lead3 METHOD = ML;
  CLASS id t;
  MODEL y = time / SOLUTION CHISQ;
  REPEATED t / TYPE=UN SUBJECT=id R;
RUN;
```

Select output from the model is shown below:

| ----- | | | | | |
|----------------------------|----------|----------------|--------|---------|---------|
| Fit Statistics | | | | | |
| -2 Log Likelihood | | | 1357.4 | | |
| ----- | | | | | |
| Solution for Fixed Effects | | | | | |
| | Estimate | Standard Error | DF | t Value | Pr > t |
| Intercept | 24.1376 | 0.6731 | 49 | 35.86 | <.0001 |
| time | -0.4823 | 0.1490 | 49 | -3.24 | 0.0022 |
| ----- | | | | | |

Fitting a model using a linear trend in time (categorical)

Let's compare with the model we previously considered when time was treated as a categorical variable, rather than continuous.

```
PROC MIXED DATA=lead3 METHOD = ML;  
  CLASS id t time;  
  MODEL y = time / SOLUTION CHISQ;  
  REPEATED t / TYPE=UN SUBJECT=id R;  
RUN;
```

Select output from the model is shown below:

| Fit Statistics | | | | | |
|----------------------------|----------|----------------|--------|---------|---------|
| -2 Log Likelihood | | | 1286.5 | | |
| Solution for Fixed Effects | | | | | |
| Effect | Estimate | Standard Error | DF | t Value | Pr > t |
| Intercept | 20.7620 | 1.2945 | 49 | 16.04 | <.0001 |
| time 0 | 5.7780 | 1.1264 | 49 | 5.13 | <.0001 |
| time 1 | -7.2400 | 1.1915 | 49 | -6.08 | <.0001 |
| time 4 | -5.2480 | 1.2608 | 49 | -4.16 | 0.0001 |
| time 6 | 0 | . | . | . | . |

Nested Models

We have two competing models. One is a full model where the outcome at each time point is estimated separately, and the other reduced model is one where we assume that the outcome changes linearly with continuous time.

$$1 \quad E[lead_{ij}] = \beta_0^c + \beta_1^c I(time_{ij} = 0) + \beta_2^c I(time_{ij} = 1) + \beta_3^c I(time_{ij} = 4)$$

$$2 \quad E[lead_{ij}] = \beta_0 + \beta_1 time_{ij}$$

These two models are nested. Why?

Nested Models

$$1 \quad E[\text{lead}_{ij}] = \beta_0^c + \beta_1^c I(\text{time}_{ij} = 0) + \beta_2^c I(\text{time}_{ij} = 1) + \beta_3^c I(\text{time}_{ij} = 4)$$

$$2 \quad E[\text{lead}_{ij}] = \beta_0 + \beta_1 \text{time}_{ij}$$

$$\text{at time 0 } E[\text{lead}] = \beta_0^c + \beta_1^c = \beta_0$$

$$\text{at time 1 } E[\text{lead}] = \beta_0^c + \beta_2^c = \beta_0 + \beta_1$$

$$\text{at time 4 } E[\text{lead}] = \beta_0^c + \beta_3^c = \beta_0 + 4\beta_1$$

$$\text{at time 6 } E[\text{lead}] = \beta_0^c = \beta_0 + 6\beta_1$$

Then if we choose

$$\beta_0^c = \beta_0 + 6\beta_1$$

$$\beta_1^c = -6\beta_1$$

$$\beta_2^c = -5\beta_1$$

$$\beta_3^c = -2\beta_1$$

we see that model 2 is a special case of model 1. Model 2 is nested in model 1.

Testing linear effect of time

To test whether the relationship between time and lead levels is linear or not, we can perform a likelihood ratio test with 2 degrees of freedom. For nested models, the likelihood ratio test statistic is:

$$\chi_p^2 = 2(\log \hat{L}_{full} - \log \hat{L}_{reduced}) = (-2 \log \hat{L}_{reduced}) - (-2 \log \hat{L}_{full}).$$

Under the null hypothesis that the reduced model is an adequate fit for the data, this test statistic is asymptotically chi-square distributed with p degrees of freedom. Our test has 2 degrees of freedom since the number of parameters in the reduced (linear) model is two less than the number of parameters in the full (saturated) model.

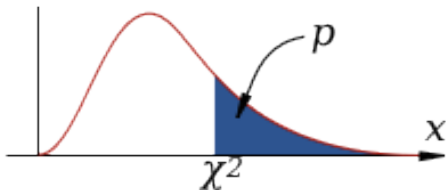
Our test statistics is $\chi_2^2 = 1357.4 - 1286.5 = 70.9$. The corresponding p-value is less than 0.001, and we conclude that the saturated model is a better fit. Since we perform a likelihood ratio test, we must specify METHOD = ML. Why?

Obtaining chi-square p-values

To get p-values, use the SDF function in a DATA step. The SDF function calculates upper-tail probabilities for a specified distribution.

```
DATA pvalues;  
  chsq = SDF('chisquare',70.9,2);  
RUN;  
PROC PRINT DATA=pvalues;  
RUN;
```

| Obs | chsq |
|-----|------------|
| 1 | 4.0203E-16 |



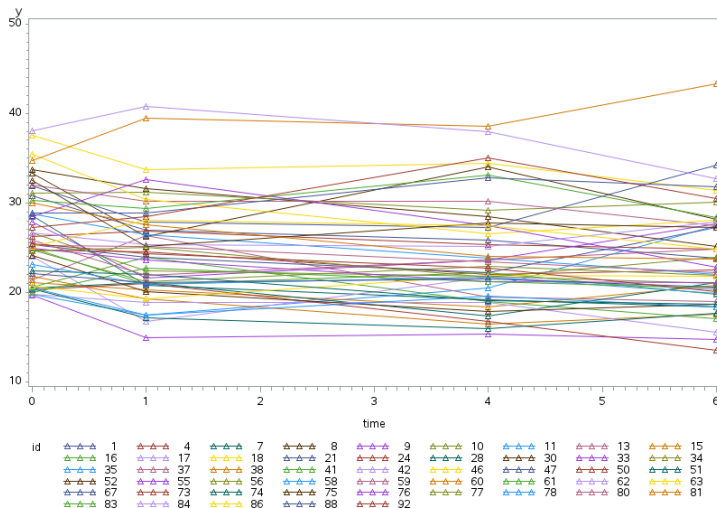
Placebo Group: Data subset

Consider examining only the 'placebo' group now. Let's subset the original data to get only those subjects who received 'placebo.' To conserve our DATA steps, let's also simultaneously get the data in long form.

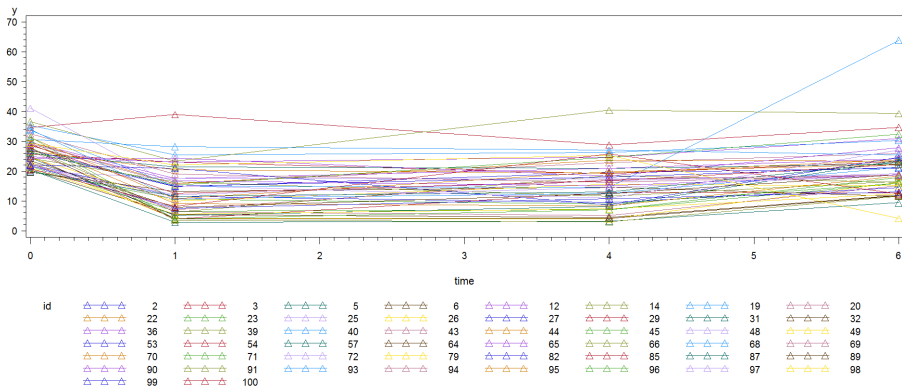
```
DATA lead4;  
  SET lead;  
  IF trmt = 'P';  
  y=y1; time=0; t=1; OUTPUT;  
  y=y2; time=1; t=2; OUTPUT;  
  y=y3; time=4; t=3; OUTPUT;  
  y=y4; time=6; t=4; OUTPUT;  
  DROP y1-y4;  
  RUN;  
  
PROC PRINT DATA=lead4 (obs=5);  
  RUN;
```

```
-----  
Obs    id    trmt      y      time    t  
  1      1      P    30.8        0      1  
  2      1      P    26.9        1      2  
  3      1      P    25.8        4      3  
  4      1      P    23.8        6      4  
  5      4      P    24.7        0      1  
-----
```

Placebo Profiles



Reminder: Succimer Profiles



Fitting a model with linear trend in time for Placebo Group

Now let's examine the effect of time (treated linearly) on mean lead levels.

```
PROC MIXED DATA=lead4;  
  CLASS id t;  
  MODEL y = time / SOLUTION CHISQ;  
  REPEATED t / TYPE=UN SUBJECT=id R;  
  RUN;
```

Select output from the model is shown below:

| ----- | | | | | |
|----------------------------|----------|----------|----|---------|---------|
| Solution for Fixed Effects | | | | | |
| | | Standard | | | |
| Effect | Estimate | Error | DF | t-Value | Pr > t |
| Intercept | 26.0551 | 0.6917 | 49 | 37.67 | <.0001 |
| time | -0.3967 | 0.08369 | 49 | -4.74 | <.0001 |
| ----- | | | | | |

Compare to the treatment group:

| ----- | | | | | |
|----------------------------|----------|----------|----|---------|---------|
| Solution for Fixed Effects | | | | | |
| | | Standard | | | |
| Effect | Estimate | Error | DF | t-Value | Pr > t |
| Intercept | 24.1376 | 0.6731 | 49 | 35.86 | <.0001 |
| time | -0.4823 | 0.1490 | 49 | -3.24 | 0.0022 |
| ----- | | | | | |

Full Dataset

Might be more meaningful to fit one model instead of two. We can assess the effect of treatment on mean blood level and see if treatment modifies time. Consider the model

$$E[\text{lead}_{ij}] = \beta_0 + \beta_1 \text{time}_{ij} + \beta_2 \text{treatment}_{ij} + \beta_3 \text{time}_{ij} \times \text{treatment}_{ij}.$$

A note on time: in the treatment group, we rejected the test of linearity for time. This might make us question using time linearly for the full dataset (and for the analysis of the placebo group above). We certainly could and should test for linearity here. However in a moment we will add an interaction in. So for the sake of demonstration, let's keep time linear.

Reformat Full Dataset

We need to get the original data into the long format. Similar to before, we reshape the data using

```
DATA lead5;  
  SET lead;  
  y=y1; time=0; t=1; OUTPUT;  
  y=y2; time=1; t=2; OUTPUT;  
  y=y3; time=4; t=3; OUTPUT;  
  y=y4; time=6; t=4; OUTPUT;  
  DROP y1-y4;  
RUN;  
  
PROC PRINT DATA=lead3 (obs=5);  
  RUN;
```

| Obs | id | trmt | y | time | t |
|-----|----|------|------|------|---|
| 1 | 1 | P | 30.8 | 0 | 1 |
| 2 | 1 | P | 26.9 | 1 | 2 |
| 3 | 1 | P | 25.8 | 4 | 3 |
| 4 | 1 | P | 23.8 | 6 | 4 |
| 5 | 2 | A | 26.5 | 0 | 1 |

Interaction Model

Back to our model,

$$E[lead_{ij}] = \beta_0 + \beta_1 time_{ij} + \beta_2 treatment_{ij} + \beta_3 time_{ij} \times treatment_{ij},$$

Recall from lecture that adding this interaction creates, in essence, a different intercept and slope for each group. Observe:

- Placebo Group Mean Model, $treatment_{ij} = 0$

$$E[lead_{ij}] = \beta_0 + \beta_1 time_{ij}$$

- Succimer Group Mean Model, $treatment_{ij} = 1$

$$E[lead_{ij}] = \beta_0 + \beta_1 time_{ij} + \beta_2 + \beta_3 time_{ij}$$

$$E[lead_{ij}] = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) time_{ij}$$

Fit the Interaction Model

Modifying the code from Slide 34 allows us to run this model. Just include `trmt` in the class statement and `trmt time*trmt` in the model line:

```
PROC MIXED DATA=lead5;  
  CLASS id t trmt;  
  MODEL y = time trmt time*trmt / SOLUTION CHISQ;  
  REPEATED t / TYPE=UN SUBJECT=id R;  
  RUN;
```

Select output from the model is shown below:

```
-----  
Effect          trmt      Estimate      Standard  
Intercept              26.0395      0.6939      98      37.52      <.0001  
time                  -0.3687      0.1223      98      -3.01      0.0033  
trmt                   A      -1.9394      0.9814      98      -1.98      0.0509  
trmt                   P           0           .           .           .           .  
time*trmt              A      -0.1695      0.1730      98      -0.98      0.3296  
time*trmt              P           0           .           .           .           .  
-----
```

The main effect of treatment is no longer significant. What about the interaction?

Questions?