We return to the Treatment of Lead-Exposed Children (TLC) Trial example. Most of this code can be found in Section 5.9 of your text.

We first need to set the data up such that each observation from each subject has its own row in the dataset. This setup corresponds to the matrix notation you learned in lecture. (Note that the data has already been sorted by trial arm.)

/\*Setting up the data in an analyzable format\*/

**data** tlc; set long.tlc;

y=y0; time=**0**; output; /\*Observation for baseline\*/

y=y1; time=**1**; output; /\*Observation for Week 1\*/

y=y4; time=**4**; output; /\*Observation for Week 4\*/

y=y6; time=**6**; output; /\*Observation for Week 6\*/

drop y0 y1 y4 y6;

**run**;

**proc** **print** data=tlc; **run**;

| **Obs** | **id** | **trt** | **y** | **time** |
| --- | --- | --- | --- | --- |
| **1** | 2 | A | 26.5 | 0 |
| **2** | 2 | A | 14.8 | 1 |
| **3** | 2 | A | 19.5 | 4 |
| **4** | 2 | A | 21.0 | 6 |
| **5** | 3 | A | 25.8 | 0 |
| **6** | 3 | A | 23.0 | 1 |
| **7** | 3 | A | 19.1 | 4 |
| **8** | 3 | A | 23.2 | 6 |

Proc mixed is the common SAS procedure for analyzing correlated continuous outcome data, which we assume have been generated from a multivariate normal distribution. Proc glimmix can also do this, and you will learn how to use glimmix in future lectures.

The following code and output will be from fitting a saturated model, as is done in Table 5.11 of your text. This means that I am fitting a separate mean (eight of them) for each time point and trial arm. Therefore, there are a total of eight regression parameters estimated. REML is used to estimate the covariance parameters. Furthermore, for now, I am going to assume that each subject has the same unstructured working covariance matrix.

/\*Needed for getting correlation and covariance matrices in the correct order\*/

**data** tlc2; set tlc; **run**;

/\*Need to sort by descending time, and use "order=data" in the proc mixed statement, in order for proc mixed to treat baseline (time=0) as the reference group instead of time=6\*/

**proc** **sort** data=tlc;

by id descending time;

**run**;

/\*Need succimer patients at the top of the dataset\*/

**proc** **sort** data=tlc;

by trt;

**run**;

/\*Using proc mixed to fit a saturated model with common unstructured covariance matrix.

By default, REML is used.\*/

**proc** **mixed** data=tlc order=data;

class time trt id; /\*Telling SAS that time, trt, and id should be treated as categorical and not continuous.\*/

model y = trt time trt\*time / s; /\*trt\*time tells SAS we want the interaction of treatment and time in the model. s, or solution, is needed if we want SAS to output parameter estimates, etc.\*/

repeated time / type=un subject=id r rcorr;

**run**;

/\*The repeated statement is used to tell SAS that we are dealing with correlated data. We tell SAS what variable denotes the timing (or spacing) of observations, what type of covariance structure we wish to use, what outcomes are correlated, and we can request some things to be output.

Here, time is the repeated effect. It identifies the measurment occasions. Read your book on this. Always make sure you know what you are doing.

type specifies the covariance structure you are using

subject=variable - variable in our dataset is id; it tells SAS that observations with the same unique id value are correlated; make sure this variable is in the class statement!

r - tells SAS to print the estimated covariance matrix. Make sure this looks correct.

rcorr - tells SAS to print the estimated correlation matrix. Make sure this looks correct.\*/

The following is the output you will obtain:

|  |
| --- |
| The SAS System |

The Mixed Procedure

| **Model Information** | |
| --- | --- |
| **Data Set** | WORK.TLC |
| **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** |
| **time** | 4 | 0 1 4 6 |
| **trt** | 2 | A P |
| **id** | 100 | 2 3 5 6 12 14 19 20 22 23 25 26 27 29 31 32 36 39 40 43 44 45 48 49 53 54 57 64 65 66 68 69 70 71 72 79 82 85 87 89 90 91 93 94 95 96 97 98 99 100 1 4 7 8 9 10 11 13 15 16 17 18 21 24 28 30 33 34 35 37 38 41 42 46 47 50 51 52 55 56 58 59 60 61 62 63 67 73 74 75 76 77 78 80 81 83 84 86 88 92 |

We fit an unstructured covariance matrix, and there were 4 time points. Therefore, there were six covariance parameters to estimate, and four variance parameters to estimate.

| **Dimensions** | |
| --- | --- |
| **Covariance Parameters** | 10 |
| **Columns in X** | 15 |
| **Columns in Z** | 0 |
| **Subjects** | 100 |
| **Max Obs Per Subject** | 4 |

| **Number of Observations** | |
| --- | --- |
| **Number of Observations Read** | 400 |
| **Number of Observations Used** | 400 |
| **Number of Observations Not Used** | 0 |

| **Iteration History** | | | |
| --- | --- | --- | --- |
| **Iteration** | **Evaluations** | **-2 Res Log Like** | **Criterion** |
| **0** | 1 | 2626.25517748 |  |
| **1** | 1 | 2416.07594087 | 0.00000000 |

|  |
| --- |
| Convergence criteria met. |

Estimated unstructured covariance matrix. We have modeled it to be the same for everyone.

Note that because we have sorted time in descending order, the covariance and correlation matrices will be reversed. For instance, instead of row 1 column 1 corresponding to baseline (week 0), they correspond to the last time point (week 6). If we had not sorted the data (which we needed to do to force proc mixed to make baseline the reference group), then the correlation and covariance matrices would be correct and correspond to increasing time. For instance, the below matrices are from running the same code but using the dataset tlc2.

| **Estimated R Matrix for id 2** | | | | |
| --- | --- | --- | --- | --- |
| **Row** | **Col1** | **Col2** | **Col3** | **Col4** |
| **1** | 25.2257 | 19.1074 | 19.6995 | 22.2016 |
| **2** | 19.1074 | 44.3458 | 35.5351 | 29.6750 |
| **3** | 19.6995 | 35.5351 | 47.3778 | 30.6205 |
| **4** | 22.2016 | 29.6750 | 30.6205 | 58.6510 |

Estimated unstructured correlation matrix. We have modeled it to be the same for everyone.

| **Estimated R Correlation Matrix for id 2** | | | | |
| --- | --- | --- | --- | --- |
| **Row** | **Col1** | **Col2** | **Col3** | **Col4** |
| **1** | 1.0000 | 0.5713 | 0.5698 | 0.5772 |
| **2** | 0.5713 | 1.0000 | 0.7753 | 0.5819 |
| **3** | 0.5698 | 0.7753 | 1.0000 | 0.5809 |
| **4** | 0.5772 | 0.5819 | 0.5809 | 1.0000 |

| **Covariance Parameter Estimates** | | |
| --- | --- | --- |
| **Cov Parm** | **Subject** | **Estimate** |
| **UN(1,1)** | **id** | 25.2257 |
| **UN(2,1)** | **id** | 19.1074 |
| **UN(2,2)** | **id** | 44.3458 |
| **UN(3,1)** | **id** | 19.6995 |
| **UN(3,2)** | **id** | 35.5351 |
| **UN(3,3)** | **id** | 47.3778 |
| **UN(4,1)** | **id** | 22.2016 |
| **UN(4,2)** | **id** | 29.6750 |
| **UN(4,3)** | **id** | 30.6205 |
| **UN(4,4)** | **id** | 58.6510 |

| **Fit Statistics** | |
| --- | --- |
| **-2 Res Log Likelihood** | 2416.1 |
| **AIC (smaller is better)** | 2436.1 |
| **AICC (smaller is better)** | 2436.7 |
| **BIC (smaller is better)** | 2462.1 |

| **Null Model Likelihood Ratio Test** | | |
| --- | --- | --- |
| **DF** | **Chi-Square** | **Pr > ChiSq** |
| 9 | 210.18 | <.0001 |

| **Solution for Fixed Effects** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Effect** | **trt** | **time** | **Estimate** | **Standard Error** | **DF** | **t Value** | **Pr > |t|** |
| **Intercept** |  |  | 26.2720 | 0.7103 | 98 | 36.99 | <.0001 |
| **trt** | **A** |  | 0.2680 | 1.0045 | 98 | 0.27 | 0.7902 |
| **trt** | **P** |  | 0 | . | . | . | . |
| **time** |  | **6** | -2.6260 | 0.8885 | 98 | -2.96 | 0.0039 |
| **time** |  | **4** | -2.2020 | 0.8149 | 98 | -2.70 | 0.0081 |
| **time** |  | **1** | -1.6120 | 0.7919 | 98 | -2.04 | 0.0445 |
| **time** |  | **0** | 0 | . | . | . | . |
| **time\*trt** | **A** | **6** | -3.1520 | 1.2566 | 98 | -2.51 | 0.0138 |
| **time\*trt** | **P** | **6** | 0 | . | . | . | . |
| **time\*trt** | **A** | **4** | -8.8240 | 1.1525 | 98 | -7.66 | <.0001 |
| **time\*trt** | **P** | **4** | 0 | . | . | . | . |
| **time\*trt** | **A** | **1** | -11.4060 | 1.1199 | 98 | -10.18 | <.0001 |
| **time\*trt** | **P** | **1** | 0 | . | . | . | . |
| **time\*trt** | **A** | **0** | 0 | . | . | . | . |
| **time\*trt** | **P** | **0** | 0 | . | . | . | . |

| **Type 3 Tests of Fixed Effects** | | | | |
| --- | --- | --- | --- | --- |
| **Effect** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **trt** | 1 | 98 | 25.43 | <.0001 |
| **time** | 3 | 98 | 61.49 | <.0001 |
| **time\*trt** | 3 | 98 | 35.93 | <.0001 |

Now I will fit the model using maximum likelihood, and present only results I want to compare with the above REML results (to obtain correlation and covariance matrices, use the dataset tlc2; to obtain parameter estimates with baseline as the reference, use the dataset tlc).

/\*Using maximum likelihood\*/

**proc** **mixed** data=tlc2 method=ml order=data;

class time trt id;

model y = trt time trt\*time / s;

repeated time / type=un subject=id r rcorr;

**run**;

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

| **Estimated R Matrix for id 2** | | | | |
| --- | --- | --- | --- | --- |
| **Row** | **Col1** | **Col2** | **Col3** | **Col4** |
| **1** | 24.7212 | 18.7253 | 19.3055 | 21.7576 |
| **2** | 18.7253 | 43.4589 | 34.8244 | 29.0815 |
| **3** | 19.3055 | 34.8244 | 46.4303 | 30.0081 |
| **4** | 21.7576 | 29.0815 | 30.0081 | 57.4780 |

| **Estimated R Correlation Matrix for id 2** | | | | |
| --- | --- | --- | --- | --- |
| **Row** | **Col1** | **Col2** | **Col3** | **Col4** |
| **1** | 1.0000 | 0.5713 | 0.5698 | 0.5772 |
| **2** | 0.5713 | 1.0000 | 0.7753 | 0.5819 |
| **3** | 0.5698 | 0.7753 | 1.0000 | 0.5809 |
| **4** | 0.5772 | 0.5819 | 0.5809 | 1.0000 |

| **Covariance Parameter Estimates** | | |
| --- | --- | --- |
| **Cov Parm** | **Subject** | **Estimate** |
| **UN(1,1)** | **id** | 24.7212 |
| **UN(2,1)** | **id** | 18.7253 |
| **UN(2,2)** | **id** | 43.4589 |
| **UN(3,1)** | **id** | 19.3055 |
| **UN(3,2)** | **id** | 34.8244 |
| **UN(3,3)** | **id** | 46.4303 |
| **UN(4,1)** | **id** | 21.7576 |
| **UN(4,2)** | **id** | 29.0815 |
| **UN(4,3)** | **id** | 30.0081 |
| **UN(4,4)** | **id** | 57.4780 |

Now I want to compare the covariance estimates from REML and ML. I have copied them below again.

For REML:

| **Estimated R Matrix for id 2** | | | | |
| --- | --- | --- | --- | --- |
| **Row** | **Col1** | **Col2** | **Col3** | **Col4** |
| **1** | 25.2257 | 19.1074 | 19.6995 | 22.2016 |
| **2** | 19.1074 | 44.3458 | 35.5351 | 29.6750 |
| **3** | 19.6995 | 35.5351 | 47.3778 | 30.6205 |
| **4** | 22.2016 | 29.6750 | 30.6205 | 58.6510 |

For ML:

| **Estimated R Matrix for id 2** | | | | |
| --- | --- | --- | --- | --- |
| **Row** | **Col1** | **Col2** | **Col3** | **Col4** |
| **1** | 24.7212 | 18.7253 | 19.3055 | 21.7576 |
| **2** | 18.7253 | 43.4589 | 34.8244 | 29.0815 |
| **3** | 19.3055 | 34.8244 | 46.4303 | 30.0081 |
| **4** | 21.7576 | 29.0815 | 30.0081 | 57.4780 |

Is there a notable difference between estimates? Why or why not?

| **Solution for Fixed Effects** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Effect** | **trt** | **time** | **Estimate** | **Standard Error** | **DF** | **t Value** | **Pr > |t|** |
| **Intercept** |  |  | 26.2720 | 0.7032 | 98 | 37.36 | <.0001 |
| **trt** | **A** |  | 0.2680 | 0.9944 | 98 | 0.27 | 0.7881 |
| **trt** | **P** |  | 0 | . | . | . | . |
| **time** |  | **6** | -2.6260 | 0.8796 | 98 | -2.99 | 0.0036 |
| **time** |  | **4** | -2.2020 | 0.8067 | 98 | -2.73 | 0.0075 |
| **time** |  | **1** | -1.6120 | 0.7840 | 98 | -2.06 | 0.0424 |
| **time** |  | **0** | 0 | . | . | . | . |
| **time\*trt** | **A** | **6** | -3.1520 | 1.2439 | 98 | -2.53 | 0.0129 |
| **time\*trt** | **P** | **6** | 0 | . | . | . | . |
| **time\*trt** | **A** | **4** | -8.8240 | 1.1409 | 98 | -7.73 | <.0001 |
| **time\*trt** | **P** | **4** | 0 | . | . | . | . |
| **time\*trt** | **A** | **1** | -11.4060 | 1.1087 | 98 | -10.29 | <.0001 |
| **time\*trt** | **P** | **1** | 0 | . | . | . | . |
| **time\*trt** | **A** | **0** | 0 | . | . | . | . |
| **time\*trt** | **P** | **0** | 0 | . | . | . | . |

Note that the parameter estimates have not changed. This is because we have fitted a saturated model. The estimated means are just equal to the sample means. These model-based standard error (SE) estimates are slightly smaller. This is because they use the negatively biased ML estimates for the covariance parameters, and therefore the SE estimates are negatively biased.

From a previous lecture, we saw that the covariance structure may be different for the two trial arms. If this is the case, then the previous model-based SE estimates from the models using a common unstructured covariance matrix can be biased. However, due to fitting a saturated model, the parameter estimates will not change.

The following is code (using the tlc2 dataset) to fit a separate unstructured covariance matrix for each trial arm:

/\*Using a different unstructured covariance matrix for each trial arm.

group= tells SAS what groups have their own unique covariance structure

=1 and =51 is telling SAS to print the estimated covariance and correlation matrices for subjects 1 and 51. These subjects are in different trial arms. Note that 1 and 51 do not correspond to the unique numerical subject id, but the order in which these subjects appear in the dataset. Note that the 1 subject is id 1, and the 51st subject is id 2 (the unique numerical subject ids) \*/

**proc** **mixed** data=tlc2 order=data;

class time trt id;

model y = trt time trt\*time / s;

repeated time / type=un subject=id group=trt r=**1** r=**51** rcorr=**1** rcorr=**51**;

**run**;

id 2: The first subject in the tlc2 dataset has a numerical subject id of 2. This subject is in the succimer group.

| **Estimated R Matrix for id 2** | | | | |
| --- | --- | --- | --- | --- |
| **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 |

| **Estimated R Correlation Matrix for id 2** | | | | |
| --- | --- | --- | --- | --- |
| **Row** | **Col1** | **Col2** | **Col3** | **Col4** |
| **1** | 1.0000 | 0.4015 | 0.3840 | 0.4951 |
| **2** | 0.4015 | 1.0000 | 0.7308 | 0.5070 |
| **3** | 0.3840 | 0.7308 | 1.0000 | 0.4548 |
| **4** | 0.4951 | 0.5070 | 0.4548 | 1.0000 |

id 1: The 51st subject in the tlc2 dataset has a numerical subject id of 1. This subject is in the placebo group.

| **Estimated R Matrix for id 1** | | | | |
| --- | --- | --- | --- | --- |
| **Row** | **Col1** | **Col2** | **Col3** | **Col4** |
| **1** | 25.2416 | 22.7495 | 24.2610 | 21.4178 |
| **2** | 22.7495 | 29.8245 | 27.0412 | 23.3841 |
| **3** | 24.2610 | 27.0412 | 33.0985 | 28.2190 |
| **4** | 21.4178 | 23.3841 | 28.2190 | 31.8074 |

| **Estimated R Correlation Matrix for id 1** | | | | |
| --- | --- | --- | --- | --- |
| **Row** | **Col1** | **Col2** | **Col3** | **Col4** |
| **1** | 1.0000 | 0.8291 | 0.8394 | 0.7559 |
| **2** | 0.8291 | 1.0000 | 0.8607 | 0.7592 |
| **3** | 0.8394 | 0.8607 | 1.0000 | 0.8697 |
| **4** | 0.7559 | 0.7592 | 0.8697 | 1.0000 |

| **Dimensions** | | | |
| --- | --- | --- | --- |
| **Covariance Parameters** | | 20 | |
| **Columns in X** | | 15 | |
| **Columns in Z** | | 0 | |
| **Subjects** | | 100 | |
| **Max Obs Per Subject** | | 4 | |
| **Covariance Parameter Estimates** | | | |
| **Cov Parm** | **Subject** | **Group** | **Estimate** |
| **UN(1,1)** | **id** | **trt A** | 25.2098 |
| **UN(2,1)** | **id** | **trt A** | 15.4654 |
| **UN(2,2)** | **id** | **trt A** | 58.8671 |
| **UN(3,1)** | **id** | **trt A** | 15.1380 |
| **UN(3,2)** | **id** | **trt A** | 44.0291 |
| **UN(3,3)** | **id** | **trt A** | 61.6571 |
| **UN(4,1)** | **id** | **trt A** | 22.9854 |
| **UN(4,2)** | **id** | **trt A** | 35.9660 |
| **UN(4,3)** | **id** | **trt A** | 33.0220 |
| **UN(4,4)** | **id** | **trt A** | 85.4946 |
| **UN(1,1)** | **id** | **trt P** | 25.2416 |
| **UN(2,1)** | **id** | **trt P** | 22.7495 |
| **UN(2,2)** | **id** | **trt P** | 29.8245 |
| **UN(3,1)** | **id** | **trt P** | 24.2610 |
| **UN(3,2)** | **id** | **trt P** | 27.0412 |
| **UN(3,3)** | **id** | **trt P** | 33.0985 |
| **UN(4,1)** | **id** | **trt P** | 21.4178 |
| **UN(4,2)** | **id** | **trt P** | 23.3841 |
| **UN(4,3)** | **id** | **trt P** | 28.2190 |
| **UN(4,4)** | **id** | **trt P** | 31.8074 |

As you can see, the estimated covariance parameters, and thus estimated correlation parameters, are notably different for the two trial arms. Therefore, it would seem that fitting a common unstructured matrix is incorrect, and can result in biased SE estimates.

Note that we saw in an earlier lecture that the sample covariances and correlations for the two trial arms were notably different. These are given below. Are they the same as the matrices estimated by REML?

For placebo:

| **Covariance Matrix, DF = 49** | | | | |
| --- | --- | --- | --- | --- |
|  | **y0** | **y1** | **y4** | **y6** |
| **y0** | 25.24164898 | 22.74946939 | 24.26097959 | 21.41784490 |
| **y1** | 22.74946939 | 29.82448980 | 27.04122449 | 23.38412245 |
| **y4** | 24.26097959 | 27.04122449 | 33.09846939 | 28.21895918 |
| **y6** | 21.41784490 | 23.38412245 | 28.21895918 | 31.80743265 |

| **Pearson Correlation Coefficients, N = 50  Prob > |r| under H0: Rho=0** | | | | |
| --- | --- | --- | --- | --- |
|  | **y0** | **y1** | **y4** | **y6** |
| **y0** | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.82914 | | <.0001 | | |  | | --- | | 0.83935 | | <.0001 | | |  | | --- | | 0.75588 | | <.0001 | |
| **y1** | |  | | --- | | 0.82914 | | <.0001 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.86067 | | <.0001 | | |  | | --- | | 0.75922 | | <.0001 | |
| **y4** | |  | | --- | | 0.83935 | | <.0001 | | |  | | --- | | 0.86067 | | <.0001 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.86971 | | <.0001 | |
| **y6** | |  | | --- | | 0.75588 | | <.0001 | | |  | | --- | | 0.75922 | | <.0001 | | |  | | --- | | 0.86971 | | <.0001 | | |  | | --- | | 1.00000 | |  | |

For succimer:

| **Covariance Matrix, DF = 49** | | | | |
| --- | --- | --- | --- | --- |
|  | **y0** | **y1** | **y4** | **y6** |
| **y0** | 25.20979592 | 15.46542857 | 15.13800000 | 22.98542857 |
| **y1** | 15.46542857 | 58.86705714 | 44.02907347 | 35.96595510 |
| **y4** | 15.13800000 | 44.02907347 | 61.65714694 | 33.02197143 |
| **y6** | 22.98542857 | 35.96595510 | 33.02197143 | 85.49464898 |

| **Pearson Correlation Coefficients, N = 50  Prob > |r| under H0: Rho=0** | | | | |
| --- | --- | --- | --- | --- |
|  | **y0** | **y1** | **y4** | **y6** |
| **y0** | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.40146 | | 0.0039 | | |  | | --- | | 0.38397 | | 0.0059 | | |  | | --- | | 0.49511 | | 0.0003 | |
| **y1** | |  | | --- | | 0.40146 | | 0.0039 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.73082 | | <.0001 | | |  | | --- | | 0.50697 | | 0.0002 | |
| **y4** | |  | | --- | | 0.38397 | | 0.0059 | | |  | | --- | | 0.73082 | | <.0001 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.45482 | | 0.0009 | |
| **y6** | |  | | --- | | 0.49511 | | 0.0003 | | |  | | --- | | 0.50697 | | 0.0002 | | |  | | --- | | 0.45482 | | 0.0009 | | |  | | --- | | 1.00000 | |  | |