Topics for this lecture:

- *Nesting and crossing*

Associated reading: Course notes, '7 LMM nesting and crossing' chapter, Sections 3.2.3, 4; Hedeker, Ch. 13 (for hierarchical models)

3.2.3 Case study: Mouse and tumor data

- Consider an experiment performed at the university involving trials on mice (Dr. Kian Behbakht, PI).
    - Each mouse in the experiment was assigned to receive a treatment (A, B, Control), and then two tumors were planted within each mouse.
    - Tumor volume measurements (unspecified units) were then taken five times on each tumor.
    - For these data,
        - the mouse is the level-3 data;
        - tumors within mice are level-2 data,
        - the repeated measures are level-1 data.

- Treatment A and B tend to actually maintain the tumor size, while those in the Control tend to have tumors that shrink over time.

- However whether these differences are significant remains to be seen, and is the purpose for fitting a linear mixed model.

- Since the design follows a 3-level nested pattern, we can apply a model that uses nested random effects plus an error covariance structure for repeated measures over time (Approach 1).

- If tumors were placed in mice systematically such that ‘Tumor 1’ and ‘Tumor 2’ had consistent meanings across mice (e.g., ‘Tumor 1’ was always near the brain and ‘Tumor 2’ always in the abdomen), then we could consider tumor and time as crossed factors, and consider modeling the data using a Kronecker Product structure.

- Even if tumors and time were not crossed (which is what I believe to be the case), we could consider the Kronecker Product structure as an approximate covariance structure for $\mathbf{Y}_i$ despite the 3-level design (Approach 2).

- In my mind, this modeling approach results in a structure that does make intuitive sense, particularly because it allows for a decay in correlation between tumor measurements within a mouse, as time between measurements is increased. Approach 1 does not allow for this.

- Final analyses were performed on log transformed tumor volumes for two reasons: (i) log volumes were more normally distributed, and (ii) results were not as sensitive to model specifications, e.g., ‘Approach 1’ (A1) versus ‘Approach 2’ (A2).

- The structures for these modeling approaches are shown below, followed by actual fits with the data. To simplify notation below, I considered 3 repeated measures within mice rather than 5. For actual fits, what complicates matters is that only Treatment B had all 3 mice with 2 tumors measured; the other groups only had 1 of 3 mice with 2 tumors measured (the remaining mice just had 1).

- Here is the SAS code to carry out model fits for the approaches, followed by more detail for each one.

  Approach 1 (nested):

```
PROC MIXED data=mouse;
CLASS mouseno tumor group time;
MODEL vol = time group time*group;
RANDOM mouse tumor(mouseno);
REPEATED / subject=tumor(mouseno) type=ar(1);
RUN;
```

  Approach 2 (Kronecker):

```
PROC MIXED data=mouse;
CLASS mouseno tumor group time;
MODEL vol = time group time*group;
REPEATED tumor time / subject=mouseno type=un@ar(1);
RUN;
```

Approach 1:

- Treat tumors as nested within mice, and repeated measures as nested within tumors (3-level data).

- The model can be expressed as $Y_{hij} = \mathbf{x}_{hij}\boldsymbol{\beta} + b_h + b_{i(h)} + \varepsilon_{hij}$, where $b_h \sim N(0, \sigma_M^2)$, $b_{i(h)} \sim N(0, \sigma_T^2)$ and ε_{hij} follows an AR(1) process (within tumor), other.

- For practice, write out the design matrix for the random effects for largest unit, mouse: $\mathbf{Z}_h$, and for the full data: $\mathbf{Z}$.

- The resulting $\mathbf{V}_h$ matrix would be:

$$\begin{pmatrix}
\sigma_M^2+\sigma_T^2+\sigma_\varepsilon^2 & \sigma_M^2+\sigma_T^2+\sigma_\varepsilon^2\phi & \sigma_M^2+\sigma_T^2+\sigma_\varepsilon^2\phi^2 & \sigma_M^2 & \sigma_M^2 & \sigma_M^2 \\
\sigma_M^2+\sigma_T^2+\sigma_\varepsilon^2\phi & \sigma_M^2+\sigma_T^2+\sigma_\varepsilon^2 & \sigma_M^2+\sigma_T^2+\sigma_\varepsilon^2\phi & \sigma_M^2 & \sigma_M^2 & \sigma_M^2 \\
\sigma_M^2+\sigma_T^2+\sigma_\varepsilon^2\phi^2 & \sigma_M^2+\sigma_T^2+\sigma_\varepsilon^2\phi & \sigma_M^2+\sigma_T^2+\sigma_\varepsilon^2 & \sigma_M^2 & \sigma_M^2 & \sigma_M^2 \\
\sigma_M^2 & \sigma_M^2 & \sigma_M^2 & \sigma_M^2+\sigma_T^2+\sigma_\varepsilon^2 & \sigma_M^2+\sigma_T^2+\sigma_\varepsilon^2\phi & \sigma_M^2+\sigma_T^2+\sigma_\varepsilon^2\phi^2 \\
\sigma_M^2 & \sigma_M^2 & \sigma_M^2 & \sigma_M^2+\sigma_T^2+\sigma_\varepsilon^2\phi & \sigma_M^2+\sigma_T^2+\sigma_\varepsilon^2 & \sigma_M^2+\sigma_T^2+\sigma_\varepsilon^2\phi \\
\sigma_M^2 & \sigma_M^2 & \sigma_M^2 & \sigma_M^2+\sigma_T^2+\sigma_\varepsilon^2\phi^2 & \sigma_M^2+\sigma_T^2+\sigma_\varepsilon^2\phi & \sigma_M^2+\sigma_T^2+\sigma_\varepsilon^2
\end{pmatrix}$$

- One of the problems that I have with this structure is that the covariance for measurements between tumors at different times stays the same as the times are more spread out. This was not the case for Approach 1 – we allow for a decay in the measurements.

- Model fit: AIC=188.5. This model estimates that there is no covariance between tumors within mice.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
mouseno		0.01654
TUMOR(mouseno)		0.1325
AR(1)	TUMOR(mouseno)	0.4670
Residual		1.1897

Estimated V Correlation Matrix for one mouseno 6

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10
1	1	0.5263	0.3051	0.2019	0.1536	0.01235	0.01235	0.01235	0.01235	0.01235
2	0.5263	1	0.5263	0.3051	0.2019	0.01235	0.01235	0.01235	0.01235	0.01235
3	0.3051	0.5263	1	0.5263	0.3051	0.01235	0.01235	0.01235	0.01235	0.01235
4	0.2019	0.3051	0.5263	1	0.5263	0.01235	0.01235	0.01235	0.01235	0.01235
5	0.1536	0.2019	0.3051	0.5263	1	0.01235	0.01235	0.01235	0.01235	0.01235
6	0.01235	0.01235	0.01235	0.01235	0.01235	1	0.5263	0.3051	0.2019	0.1536
7	0.01235	0.01235	0.01235	0.01235	0.01235	0.5263	1	0.5263	0.3051	0.2019
8	0.01235	0.01235	0.01235	0.01235	0.01235	0.3051	0.5263	1	0.5263	0.3051
9	0.01235	0.01235	0.01235	0.01235	0.01235	0.2019	0.3051	0.5263	1	0.5263
10	0.01235	0.01235	0.01235	0.01235	0.01235	0.1536	0.2019	0.3051	0.5263	1

Approach 2:

- Kronecker Product structure, using the UN structure for tumors within mice, and repeated measures over time using the AR(1) structure.

- The model is $Y_{hij} = \mathbf{x}_{hij}\boldsymbol{\beta} + \varepsilon_{hij}$, where *h* indexes mouse, *i* indexes tumor and *j* indexes time, $\mathbf{x}_{hij}$ is a row vector containing the predictors. We will consider covariance structures relative to 'mouse' (with index *h*), since that is the largest experimental unit. Thus, we need to derive $\mathbf{R}_h$, the covariance matrix for vector $\boldsymbol{\varepsilon}_h$.

- Structure for 2 tumors: $\mathbf{R}_{h1} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$ 3 times: $\mathbf{R}_{h2} = \sigma_\varepsilon^2 \begin{pmatrix} 1 & \phi & \phi^2 \\ \phi & 1 & \phi \\ \phi^2 & \phi & 1 \end{pmatrix}$

- The combined (Kronecker product structure):

$$\mathbf{R}_h = \mathrm{R}_{h1} \otimes \mathrm{R}_{h2} \begin{pmatrix} \sigma_1^2 & \sigma_1^2\phi & \sigma_1^2\phi^2 & \sigma_{12} & \sigma_{12}\phi & \sigma_{12}\phi^2 \\ \sigma_1^2\phi & \sigma_1^2 & \sigma_1^2\phi & \sigma_{12}\phi & \sigma_{12} & \sigma_{12}\phi \\ \sigma_1^2\phi^2 & \sigma_1^2\phi & \sigma_1^2 & \sigma_{12}\phi^2 & \sigma_{12}\phi & \sigma_{12} \\ \sigma_{12} & \sigma_{12}\phi & \sigma_{12}\phi^2 & \sigma_2^2 & \sigma_2^2\phi & \sigma_2^2\phi^2 \\ \sigma_{12}\phi & \sigma_{12} & \sigma_{12}\phi & \sigma_2^2\phi & \sigma_2^2 & \sigma_2^2\phi \\ \sigma_{12}\phi^2 & \sigma_{12}\phi & \sigma_{12} & \sigma_2^2\phi^2 & \sigma_2^2\phi & \sigma_2^2 \end{pmatrix} = \mathbf{V}_h$$

- The σ_ε^2 on the AR(1) structure is not included because it becomes redundant once we take the direct product, i.e., it is absorbed into parameters in the other matrix.

- Model fit: AIC=188.1. Here, we get a covariance between tumors within mice that decays as the time between measurements is increased.

Cov Parm	Subject	Estimate
TUMOR UN(1,1)	mouseno	1.1732
UN(2,1)	mouseno	0.03573
UN(2,2)	mouseno	1.5740
Day AR(1)	mouseno	0.5111

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10
1	1	0.5111	0.2612	0.1335	0.0682	0.0263	0.0134	0.0069	0.0035	0.0018
2	0.5111	1	0.5111	0.2612	0.1335	0.0134	0.0263	0.0134	0.0069	0.0035
3	0.2612	0.5111	1	0.5111	0.2612	0.0069	0.0134	0.0263	0.0134	0.0069
4	0.1335	0.2612	0.5111	1	0.5111	0.0035	0.0069	0.0134	0.0263	0.0134
5	0.0682	0.1335	0.2612	0.5111	1	0.0018	0.0035	0.0069	0.0134	0.0263
6	0.0263	0.0134	0.0069	0.0035	0.0018	1	0.5111	0.2612	0.1335	0.0682
7	0.0134	0.0263	0.0134	0.0069	0.0035	0.5111	1	0.5111	0.2612	0.1335
8	0.0069	0.0134	0.0263	0.0134	0.0069	0.2612	0.5111	1	0.5111	0.2612
9	0.0035	0.0069	0.0134	0.0263	0.0134	0.1335	0.2612	0.5111	1	0.5111
10	0.0018	0.0035	0.0069	0.0134	0.0263	0.0682	0.1335	0.2612	0.5111	1

- Although the model fits have similar AIC values, Approach 2 yields a slightly better model fit.

- Below is full SAS code and some of the tests of interest for Approach 2. The PI for the project had specifically asked for a comparison of groups at the last time (time 18).

```
/* Simplified labels:  Control-etoposide = Control (Group ‘2’);
                       FADD-etoposide = Trt A (Group ‘4’);
                       FADD V108E-etoposide = Trt B (Group ‘6’) */
proc sort data=mouse.BJABstudy3rd; by group tumorno day;
proc mixed data=mouse.BJABstudy3rd order=internal;
class group tumor mouseno day;
model log_tumor_volume=group day group*day / noint solution ddfm=kr;
repeated tumor day / subject=mouseno type=un@ar(1);
where group in (2,4,6) and day~=0;
lsmeans group*day/slice=day;
estimate 'Control vs Trt A, day 18' group 1 -1 0
         group*day 0 0 0 0 1 0 0 0 0 -1 0 0 0 0 0;
estimate 'Control vs Trt B, day 18' group 1 0 -1
         group*day 0 0 0 0 1 0 0 0 0 0 0 0 0 0 -1;
estimate 'Cont. vs average of Trt A and B, day 18' group 1 -0.5 -0.5
         group*day 0 0 0 0 1 0 0 0 0 -0.5 0 0 0 0 -0.5;
contrast 'linear' day -2 -1 0 1 2;
contrast 'quadratic' day 2 -1 -2 -1 2;
contrast 'cubic' day -1 2 0 -2 1;
contrast 'quartic' day 1 -4 6 -4 1;
```

```
contrast 'lxl' group*day -2 -1 0 1 2 2 1 0 -1 -2 0 0 0 0 0,
                         group*day -2 -1 0 1 2 0 0 0 0 0 2 1 0 -1 -2;
contrast 'qxq' group*day 2 -1 -2 -1 2 -2 1 2 1 -2 0 0 0 0 0,
                         group*day 2 -1 -2 -1 2 0 0 0 0 0 -2 1 2 1 -2;
contrast 'cxc' group*day -1 2 0 -2 1 1 -2 0 2 -1 0 0 0 0 0,
                         group*day -1 2 0 -2 1 0 0 0 0 0 1 -2 0 2 -1;
contrast '4x4' group*day 1 -4 6 -4 1 -1 4 -6 4 -1 0 0 0 0 0,
                         group*day 1 -4 6 -4 1 0 0 0 0 0 -1 4 -6 4 -1;
ods output lsmeans=estimates; run;

*Graph of estimates from the mixed model;
symbol1 i=join; axis1 order=(0 to 110 by 10);
proc gplot data=estimates; plot estimate*day=group / vaxis=axis1;
run;
```

Abbreviated output:

Dimensions

Covariance Parameters	4
Columns in X	23
Columns in Z	0
Subjects	9
Max Obs Per Subject	10
Number of Observations Used	70

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
GROUP	2	7	5.24	0.0407
Day	4	24	6.87	0.0008
GROUP*Day	8	24	1.98	0.0933

Tests of Effect Slices

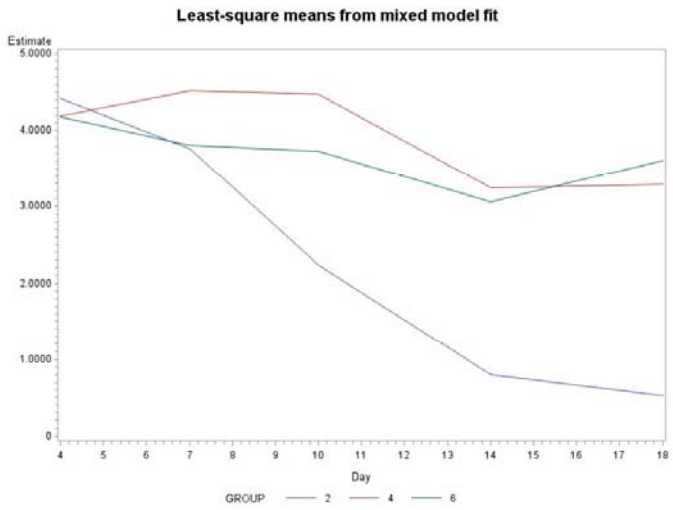
Effect	Day	Num DF	Den DF	F Value	Pr>F
GROUP*Day	4	2	24	0.06	0.9399
GROUP*Day	7	2	24	0.60	0.5570
GROUP*Day	10	2	24	4.13	0.0287
GROUP*Day	14	2	24	6.13	0.0071
GROUP*Day	18	2	24	9.75	0.0008

Contrasts

Label	Num DF	Den DF	F Value	Pr>F
linear	1	24	23.94	<.0001
quadratic	1	24	0.01	0.9067
cubic	1	24	5.17	0.0321
quartic	1	24	1.67	0.2080
lxl	2	24	6.62	0.0051
qxq	2	24	0.93	0.4092
cxc	2	24	0.26	0.7756
4x4	2	24	0.35	0.7062

Estimates

Label	Estimate	SE	DF	t Value	Pr>|t|
Control vs Trt A, day 18	-2.7595	0.7962	24	-3.47	0.0020
Control vs Trt B, day 18	-3.0841	0.7395	24	-4.17	0.0003
Control vs average of Trt A and B, day 18	-2.9218	0.6736	24	-4.34	0.0002



- There is a general cubic pattern (flattish, drop, flattish), but since this pattern exists more or less for each group (after accounting for linear trends), we do not see a cubic-by-cubic interaction.

- There is a clear linear trend as well as linear-by-linear interaction.

- Finally, control differs significantly from each of the other groups, particularly at later times. Mean log volume for control was significantly different than Treatments A and B at 18 days.

- *Discussion*: If the design is really a nested one and tumors are not crossed but the Kronecker Product structure is still used as an approximate model, it may make more sense to use CS $\otimes$ AR(1). (If tumors '1' and '2' are arbitrarily assigned across mice, we probably wouldn't want to use separate variances for them, such as is done with the UN structure.) However, right now SAS does not have a canned structure for that. Thus, we are probably spending an extra DF for no real reason. Nevertheless, we only have 4 covariance parameters, and it's the same number used for Approach 1. Hopefully later versions will have that capability.

- In general I would recommend that you stick with the model that is consistent with the design UNLESS there is good reason to change. So, if you have a nested design, stick with the nested factors, if you have a crossed design, use crossed factors. In this case, we used a Kronecker Product structure for the nested data in Approach 2. My reason for doing this was because the resulting covariance matrix made more sense to me, intuitively, and was supported by the AIC (albeit a small difference). It would be nice to be able to have a common variance for tumors within mice since there is no real reason to believe that they should be different. Even so, we got a slightly better fit using the Kronecker Product structure. But if you are a design 'purest', you could stick with the nested model – the AIC's were not really that different.

- In some cases a researcher may want to have each subject try multiple treatments in an experiment, rather than just one. In the simplest case, there are 2 treatments, which can be assigned to each subject in a *2 period, 2 treatment (2x2) crossover design*.

- For the 2×2 design, subjects are usually randomly assigned an order of treatments, *AB* or *BA*, in equal amount. This helps to eliminate confounders associated with time.

- If there are 3 treatments, then one may set up a 3 period, 3 treatment crossover design.

- Crossover designs are often used in clinical trials when the cost of tracking subjects longitudinally for extended time does not impose major difficulties.

- Carry-over effects: One limitation of crossover designs is that receiving one treatment first may have an influence on subjects' responses in the subsequent period in which they receive the other treatment.

  - If this *carry-over effect* differs between treatment sequences, then estimates of effect of interest may be biased.

  - The difficulty with the 2×2 design is that carry-over effect estimates are *aliased* with other effects (i.e., they are completely confounded with each other). Specifically, the *sequence*, *carry-over* and *period*treatment* effects are aliased.

  - If *sequence* and *period*treatment* effects are assumed to not exist, then we can test for carry-over effects by including the sequence term in the model. But the validity of the test relies on that assumption…

- In more complex models, it is easier to estimate carryover effects by examining interactions. Including a term in the model for treatment used in the previous period may help in estimating (differential) carryover effects.

- For any crossover design, including a washout period of suitable length between treatment periods may help to eliminate carryover effects that a treatment might have. Most researchers do include some washout period in their crossover experiment, however one of the issues that arises is planning in advance how long this should be since it is often uncertain how long it will take to 'wash out' the treatment.

- If some carryover effects are expected for a given study or experiment, then the researcher may also consider using alternative designs. Here, we focus on crossover experiments with repeated measures within periods.

- For more examples and details about modeling data from crossover designs, see Littell et al, *SAS System for Mixed Models*, and Jones and Kenward, *Design and Analysis of Cross-Over Trials* (in particular, see Chapter 5).

- Consider a crossover experiment that was performed and reported in Connolly et al. (2006), entitled *Efficacy of a tart cherry juice blend in preventing the symptoms of muscle damage* (British Journal of Sports Medicine, 40: 679-683).
  - In the experiment, subjects were randomized to receive cherry juice twice a day or placebo drink for 8 consecutive days.
  - At day 4, subjects performed 'eccentric elbow flexion contractions'.
  - Measures of strength and pain after the challenge (relative to baseline) were then taken on subjects on the last 4 days of the period, after the challenge.

- Subjects then repeated the experiment with the treatment they did not have in Period 1, using the opposite arm. Mean strength was greater and pain was less when subjects had the cherry drink, relative to placebo. Strength loss relative to BL was 22% for placebo but only 4% for cherry juice.
- This is considered a 2-period, 2-treatment crossover design, with repeated measures.

- In the spirit of this experiment, consider a hypothetical data set involving muscle soreness measurements on 4 successive days after an exercise challenge. This soreness score ranges from 0 to 10 but is typically in a range of 1 to 4. These scores are adjusted for baseline soreness before the experiment (e.g., if a subject has a soreness score of 2 coming into the study and a score of 6 one day after the challenge, then their soreness score on that day would be 4). This was designed like the reported experiment (2×2 crossover, 4 repeated measures within each period).

- Below is a description of the predictors in the model and what they can be used to test:

Period: 1 or 2; test accounts for differences between first and second time periods.

Treatment: placebo vs. cherry drink; test is for main effect of treatment (comparing treatment means).

Time: the 4 days that measures were taken following the exercise challenge; test is for main effect of time (comparing means for 4 days following the exercise challenge); time modeled as a class variable.

*Period*time*: Will test for differences between time patterns between the two periods. Can be thought of as the general time variable.

*Treatment*time*: Will test whether changes over time (with a period) differ between the placebo and cherry drink. If treatment*time is significant, then comparisons can be made between treatments for individual days (with multiple comparison adjustments, if desired).

Sequence: Compares *AB* versus *BA* treatments. Since there are only 2 treatments, this sequence effect is aliased with carry-over effects. We can use this term to test for carry-over effects assuming that there are no true *treatment×period* or (other) sequence effects.

Here is the SAS code for the analysis.

```
data cross; format trt $9.;
input id pd trt $ time seq y @@; datalines;
1 1 Control 1 1 1.7 1 1 Control 2 1 2.9 1 1 Control 3 1 3.4
1 1 Control 4 1 2.8 1 2 Treatment 1 1 1.5 1 2 Treatment 2 1 3.0
1 2 Treatment 3 1 3.1 1 2 Treatment 4 1 1.9 2 1 Control 1 1 1.5
. . .
8 1 Treatment 2 2 3.4 8 1 Treatment 3 2 3.0 8 1 Treatment 4 2 1.7
8 2 Control 1 2 2.0 8 2 Control 2 2 4.2 8 2 Control 3 2 3.3
8 2 Control 4 2 2.8
; run;
```

```
proc sort data=cross_mv; by time;
proc gplot data=cross;
  plot y*time=trt / vaxis=axis1 haxis=axis2;
  axis1 label=(h=2 angle=90 'muscle soreness score')
     order=1 to 4 by 1 value=(h=2);
  axis2 label=(h=2 'time (days after treatment)') value=(h=2);
  symbol1  i=std1mtj l=1 c=blue   v=none w=2 mode=include w=2;
  symbol2  i=std1mtj l=1 c=red    v=none w=2 mode=include w=2; run;
proc mixed data=cross order=data;
 class id pd trt time seq;
 model y = pd trt time pd*time trt*time seq / dfm=kr solution;
 random id;  repeated time / type=ar(1) subject=id*pd;
 lsmeans trt*time; run;
```

Abbreviated output:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
id		0.2434
AR(1)	id*pd	0.7168
Residual		0.5308

Fit Statistics

-2 Res Log Likelihood	113.4
AIC (smaller is better)	119.4
AICC (smaller is better)	120.0
BIC (smaller is better)	119.7

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
pd	1	5.7	0.74	0.4247
trt	1	5.7	4.81	0.0733
time	3	36.1	19.96	<.0001
pd*time	3	36.1	0.15	0.9261
trt*time	3	36.1	2.72	0.0584
seq	1	6	0.18	0.6836

Least Squares Means

Effect	trt	time	Estimate	Standard Error	DF	t Value	Pr > |t|
trt*time	Control	1	1.9500	0.3111	16.4	6.27	<.0001
trt*time	Control	2	3.1000	0.3111	16.4	9.97	<.0001
trt*time	Control	3	2.9875	0.3111	16.4	9.60	<.0001
trt*time	Control	4	2.6750	0.3111	16.4	8.60	<.0001
trt*time	Treatment	1	1.7875	0.3111	16.4	5.75	<.0001
trt*time	Treatment	2	2.6375	0.3111	16.4	8.48	<.0001
trt*time	Treatment	3	2.3750	0.3111	16.4	7.63	<.0001
trt*time	Treatment	4	1.3625	0.3111	16.4	4.38	0.0004

Differences of Least Squares Means

Effect	trt	time	_trt	_time	Estimate	Standard Error	DF	t Value	Pr > |t|
trt*time	Control	1	Treatment	1	0.1625	0.3643	10.8	0.45	0.6643
trt*time	Control	2	Treatment	2	0.4625	0.3643	10.8	1.27	0.2308
trt*time	Control	3	Treatment	3	0.6125	0.3643	10.8	1.68	0.1212
trt*time	Control	4	Treatment	4	1.3125	0.3643	10.8	3.60	0.0042

- Interpretations of the results?

Graph of sample means and *SD* error bars.