

# BIOS6643 Longitudinal

## L5 Contrast

EJC

Department of Biostatistics & Informatics

## 1. Overview

- ▶ Foundational material (see other document)
- ▶ Statistical models, different formulations
  - ▶ Means versus effects models
  - ▶ Full rank versus less-than-full-rank models
- ▶ g-inverses and estimability
- ▶ Key theorems and identities
- ▶ Writing custom tests for fixed effects in the LMM
- ▶ Class versus continuous predictors (e.g., time).

Consider an experiment where subjects in 2 groups are observed over 3 times. Write the associated effects and means models.

If different subjects were observed at each  $Group \times Time$  combination, we would have factorial data that would normally be analyzed with 2-way ANOVA. The statistical models would look similar, but in that case I usually write the model for  $Y_{ijk}$ , where  $i$  indexes group,  $j$  indexes time, and  $k$  replicate (right side adjusted accordingly).

## Full-rank versus less-than-full-rank models:

In 6612 you were introduced to full-rank and less-than-full-rank (LTFR) models. The LTFR model includes a column for each level of each categorical variable.

**Full rank approach:** Say we have a continuous variable such as time, and 3 treatment *Groups* (*A*, *B*, *Control*); *Group*  $\times$  *Time* will also be included in the model; *X* has 6 columns. Let  $x_1 = \text{Time}$ ,  $x_2 = 1$  for treatment *A*, 0 otherwise;  $x_3 = 1$  for treatment *B*, 0 otherwise (*Control* will be the reference group – i.e., we have imposed a set-to-0 restriction for the *Control* group). Let  $i$  denote the unique data-wide index for subject, and  $j$  the index for *Time*. Since times of measurement may not be the same for subjects, I keep the  $i$  index on the time variable as well as  $j$ . The model can then be expressed as:

$$Y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{1ij} \cdot x_{2i} + \beta_5 x_{1ij} \cdot x_{3i} + \varepsilon_{ij}$$

Note that we can use like notation for all predictors, but we will have to ‘manually construct’ the dummy variables for treatment group. We have a model with full rank *X*, so don’t need to worry about generalized inverses.

**Less-than-full-rank approach:** There are 8 columns in  $X$ ; let's define the treatment effects as  $\kappa_h$  for *Group* 1(*A*), 2(*B*) and 3(*Control*), and let  $\gamma_h$  denote the *Group*  $\times$  *Time* interaction for  $h = 1, 2, 3$ . Since there is only one interaction term, we don't need to add another index on the effect other than one for *Group*. The model is:

$$Y_{hij} = \beta_0 + \beta_1 x_{1ij} + \kappa_h + \gamma_h x_{1ij} + \epsilon_{ij}$$

for *Group*  $h$ , *Subject*  $i$  and *Time*  $j$ . The statistical model has mixed notation, and the associated matrix has less-than-full rank (i.e., dependency in columns).

In computing estimates, the issue with the LTFR models is that if we include a column in  $\mathbf{X}$  for each of  $c$  levels of a categorical variable,  $\mathbf{X}$  will not have full rank, nor will  $\mathbf{X}^\top \mathbf{X}$ , and hence a regular inverse cannot be used when calculating the estimator for  $\hat{\beta}$ .

Two basic approaches can be used to handle this issue.

- ▶ Force a constraint on the model up front (set-to-0 or sum-to-0), so that  $\mathbf{X}$  has full rank and regular inverses can be used. I.e. create a full-rank model for the situation.
- ▶ Introduce a new type of inverse that will handle the singularity issue.
- ▶ In some cases these approaches are essentially the same. (E.g., SAS's g-inverse approach is essentially like setting the highest levels of factors to 0.)

**For more information, see the course notes, or refer to your BIOS6612 material.**

- ▶ For LTFR models,  $\hat{\beta}$  is not unique and depends on the g-inverse used. Theory of estimability tells us which linear combination of parameters are estimable in the LTFR model.
- ▶ Some might find using LTFR models a little confusing, however doing so eliminates the need to manually create indicator variables and helps streamline modeling and programming code to get estimates related to all levels of categorical variables.
- ▶ When you use the CLASS statement in PROC MIXED or PROC GLM, you are fitting a LTFR model.
- ▶ In R, using 'factor' in the lme function is like using CLASS in PROC MIXED, except this approach is equivalent to setting the 1st level of a factor as the reference, rather than the last.
- ▶ To determine whether  $L\beta$  is estimable, see if  $L = LH$  for the  $L$  in question. If the equation holds, then  $L\hat{\beta}$  has a unique distribution (despite which g-inverse is used for  $X^T X$ ) and  $L\hat{\beta}$  is estimable.
- ▶ When including the 'e' option in the MODEL statement, SAS produces the form of estimable functions for generic  $L$ .



Consider an  $n \times 1$  random vector  $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$  where  $\Sigma$  has full rank. For an  $m \times n$  matrix  $\mathbf{A}$ , the linear form:

$$\mathbf{AY} \sim \mathcal{N}(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\Sigma\mathbf{A}^\top)$$

if  $r(\mathbf{A}) = m$ , where  $m \leq n$ . Note that this is an  $m$ -variate distribution since  $\mathbf{AY}$  is an  $m \times 1$  vector.

Example: distribution of  $\hat{\boldsymbol{\beta}}$ .

For an  $n \times n$  matrix  $\mathbf{A}$  and  $\mathbf{Y}$  as before,

$$\mathbf{Y}^T \mathbf{A} \mathbf{Y} \sim \chi^2_{\nu}(\lambda)$$

where  $\lambda = \frac{1}{2} \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu}$  is the noncentrality parameter and  $\nu = r(\mathbf{A})$  are degrees of freedom if and only if any of the following conditions hold:

1.  $\mathbf{A}\boldsymbol{\Sigma}$  is idempotent [i.e.,  $(\mathbf{A}\boldsymbol{\Sigma})^2 = \mathbf{A}\boldsymbol{\Sigma}$ ],
2.  $\boldsymbol{\Sigma}\mathbf{A}$  is idempotent [i.e.,  $(\boldsymbol{\Sigma}\mathbf{A})^2 = \boldsymbol{\Sigma}\mathbf{A}$ ],
3.  $\boldsymbol{\Sigma}$  is a generalized inverse of  $\mathbf{A}$ .

The distribution shown above is a non-central  $\chi^2$  distribution. When  $\boldsymbol{\mu} = 0$ , we have  $\lambda = 0$ , which is the central  $\chi^2$  distribution that we're familiar with.

Example: Distribution of  $\hat{\sigma}^2$

Researchers will (or should) have some idea about specific hypotheses they would like to test or estimates then would like to obtain. The inferential methods we've already learned associated with GLM's and LMM's will help us to create custom tests and estimates. Here, we focus on the LMM although the methods are similar in the GLM case. Below is a quick review of the methods, followed by some examples.

- ▶ Likelihood ratio tests (LRT),  $F$ -tests for multiple equations
  - ▶ LRTs or  $F$ -tests can be performed for  $H_0 : \mathbf{c}\beta = \mathbf{h}$  where  $\mathbf{c}$  is a matrix and  $\mathbf{h}$  is a vector of numbers, but usually  $\mathbf{0}$ .
    - ▶ In SAS PROC MIXED, obtain  $F$ -tests using the CONTRAST statement.
    - ▶ In R, equivalent LRT tests can be obtained by using the **anova.lme()** function that follows full and reduced lme fits.
    - ▶ For class variables with more than 2 levels, associated main effect or interaction tests require  $F$ -tests.
  - ▶ Tests are more versatile than  $t$ -tests since the latter can only handle 1 equation at a time.
  
- ▶ Estimate of  $\mathbf{L}\beta$  and associated Wald tests
  - ▶  $\mathbf{L}$  is a row vector (e.g., one row in  $\mathbf{c}$ ).
  - ▶ The estimate of  $\mathbf{L}\beta$  is  $\mathbf{L}\hat{\beta}$ .
  - ▶ The test of  $H_0 : \mathbf{L}\beta = \mathbf{0}$  is carried out using the  $t$ -statistic  $\mathbf{L}\hat{\beta}/se(\mathbf{L}\hat{\beta})$ .
  - ▶ The estimate and test are obtained using the ESTIMATE statement in PROC MIXED. In R, you can obtain these by using the **multcomp::glht()** function in the **multcomp** package.
  
- ▶ The above inferential methods can be carried out on either means or effects models. See the text for examples of main-effect and interaction tests written for means models. Here, we focus on writing coefficients for the effects models and less on computation. More computational details are in the text.

Consider the model  $Y_{hij} = \mu + \alpha_h + \tau_j + \gamma_{hj} + \varepsilon_{hij}$ , where  $h$  denotes *Group*,  $i$  denotes *Subject*, and  $j$  denotes *Time*. We can impose a specific covariance structure for the errors (e.g., AR(1)) to account for repeated measures. If there are 2 *Groups* and 3 *Times*, then there are 12 parameters:  $\beta^T = (\mu \alpha_1 \alpha_2 \tau_1 \tau_2 \tau_3 \gamma_{11} \gamma_{12} \gamma_{13} \gamma_{21} \gamma_{22} \gamma_{23})$

For estimates and tests involving  $L\beta$ ,  $L$  will be a  $1 \times 12$  vector. When we write ESTIMATE or CONTRAST statements in SAS, the coefficients in  $L$  are broken down by factors. For example, say we want to estimate the mean for *Group* 1 at the 3rd *Time*. For this,  $L = (1 \ 1 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0)$ .

In SAS we would write:

```
ESTIMATE 'C at 72 hrs' intercept 1 group 1 0 time 0 0 1  
group*time 0 0 1 0 0 0;
```

If certain factors do not come into play, then we do not need to include them in the ESTIMATE or CONTRAST statement. For example, say we want to compare means for between groups at the first time point. Then  $L = (0 \ 1 \ -1 \ 0 \ 0 \ 1 \ 0 \ 0 \ -1 \ 0 \ 0)$ . In SAS, the estimate statement is:

```
ESTIMATE 'group diffs at 24 hrs' intercept 0 group 1 -1 time 0 0 0
group*time 1 0 0 -1 0 0;
```

But since we have 0's for two of the factors, we can just write:

```
ESTIMATE 'group diffs at 24 hrs' group 1 -1 group*time 1 0 0 -1 0 0;
```

Note: in order to figure out how to write  $L$  above, you might find it easier to consider each group first, and then take the difference:

$(Int | Group | Time | Group \times Time)$

Control group at 24 hours:  $L = (1 | 1 0 | 1 0 0 | 1 0 0 0 0 0)$

Myostatin group at 24 hours:  $L = (1 | 0 1 | 1 0 0 | 0 0 0 1 0 0)$

Difference:  $L = (0 | 1 - 1 | 0 0 0 | 1 0 0 - 1 0 0)$

For practice: write out the null hypothesis for the main effect tests and interaction test based on the 2-way model.

Ramus data: The size of the ramus bone in the jaw was measured for 20 boys at 4 ages (per boy). Write the model that includes a class variable for age and a random intercept for boy.



## Write polynomial contrasts for age.

Some questions:

1. With 4 ages, we can go up to which degree of polynomial?
2. What are the coefficients for the polynomial contrasts?
3. What is an alternative way to write the model that will yield the same  $-2\ln\mathcal{L}$  value (considering ML estimation)? Why is this?
4. Compare the  $F$ -test that use polynomial contrasts as rows for  $c$  versus the  $F$ -test that compares *Time1* versus each successive time point. Why are they equivalent? (Full rank reparameterization theorem.) (It is also the same as the overall test for age.)



```
proc mixed data=long.ramus_uni method=ml;  
  class boy age;  
  model height= age / soluion;  
  random intercept / subject=boy;  
  contrast 't2 versus t1' age -1 1 0 0;  
  contrast 't3 versus t1' age -1 0 1 0;  
  contrast 't4 versus t1' age -1 0 0 1;  
  contrast 'all 3' age -1 1 0 0,  
                  age -1 0 1 0,  
                  age -1 0 0 1;  
  estimate 'linear' age -3 -1 1 3;  
  estimate 'quadratic' age 1 -1 -1 1;  
  estimate 'cubic' age -1 3 -3 1;  
  contrast 'linear' age -3 -1 1 3;  
  contrast 'quadratic' age 1 -1 -1 1;  
  contrast 'cubic' age -1 3 -3 1;  
  contrast 'deviations' age 1 -1 -1 1,  
                        age -1 3 -3 1;run;
```

## Covariance Parameter Estimates

Cov Parm	Subject	Estimate
Intercept	boy	5.7851
Residual		0.6659

## Fit Statistics

-2 Log Likelihood	266.0
AIC (smaller is better)	278.0

## Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
age	3	57	43.60	<.0001

## Estimates

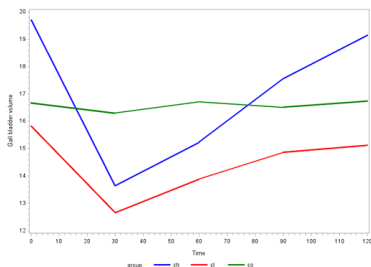
Label	Estimate	SE	DF	t Value	Pr >  t
linear	9.3300	0.8160	57	11.43	<.0001
quadratic	-0.09000	0.3649	57	-0.25	0.8061
cubic	-0.04000	0.8160	57	-0.05	0.9611

## Contrasts

Label	Num DF	Den DF	F Value	Pr > F
t2 versus t1	1	57	14.13	0.0004
t3 versus t1	1	57	55.07	<.0001
t4 versus t1	1	57	117.32	<.0001
all 3	3	57	43.60	<.0001
deviations	2	57	0.03	0.9689

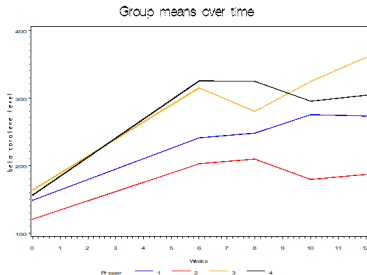
# Tests for interaction in the Dog data

In Sterczer, Voros, and Karsai (1996), the effect of cholagogues on changes in gallbladder volume (GBV) in dogs was studied by two-dimensional ultrasonography. Three different kinds of cholagogues and tap water (administered orally) as control were used in the experiment. Six healthy dogs were treated with each substance. GBV was determined immediately before the administration of the test substance and at 10-minute intervals for 120 minutes thereafter. CH=cholagogue 1 (blue), CL=cholagogue 2 (red); CO=control (green).



1. Write a model that includes group and time as class variables, plus group  $\times$  time interaction.
2. Write a test to compare is if trends over time differ between treatment (cholagogue) groups.
3. Write a test to compare changes from beginning to end among the 3 groups.
4. Write a test to compare the quadratic trend of the average of treatment groups compared with the control.

A clinical trial was planned comparing the incidence of cancer in a group taking beta-carotene in capsule form compared with a group taking beta-carotene placebo capsules. One issue in planning such a study is which preparation to use for the beta-carotene capsules. Four preparations were considered: (1) Solatene (30mg capsules), (2) Roche (60mg capsules), (3) BASF (30mg capsules), (4) BASF (60mg capsules). To test efficacy of the four agents in raising plasma-carotene levels, a small bioavailability study was conducted. After two consecutive-day fasting blood samples, 23 volunteers were randomized to one of the four preparations mentioned above, taking 1 pill every other day for 12 weeks. The primary endpoint was level of plasma carotene attained after moderately prolonged steady ingestion. For this purpose, blood samples were drawn at 6, 8, 10 and 12 weeks. In the plot below, the 2nd baseline measure was used at Time 0. So in total, there are measurements at 0, 6, 8, 10 and 12 weeks.



1. Look at means by time first. What are some potential models?
2. Write out two models, one where time is a class variable, and one where it is continuous.
3. Perform a test of interest for the model with time as a class variable.
4. Perform a test of interest for the model with time as a continuous variable.



- ▶ How do the estimates change if Time is treated as a continuous variable relative to Time as class?
- ▶ If modeling Time as continuous, are higher order terms necessary?
- ▶ What should you think about when deciding between Time as class versus Time as continuous?
- ▶ In SAS PROC GLM and PROC MIXED, time can be modeled as a continuous variable by simply leaving the Time variable out of the CLASS statement (or by not using the factor argument in R).
- ▶ For an indicator variable (e.g.,  $0 = \textit{Male}$  and  $1 = \textit{Female}$ ), do you need to define it as a class variable? What is the difference when you do or don't?
- ▶ When modeling time as a class variable, the linearity assumption can be checked informally by inspecting graphs to see if the patterns look linear, or add higher order terms and see if they are significant. Can perform 'lack of fit' test for linearity.

Modeling time as a class versus continuous variable is an important issue that we will discuss throughout the course.

- ▶ Time as a class variable...
  - ▶ offers the most flexibility
  - ▶ imposes no parametric constraints imposed across levels of time
  - ▶ uses more degrees of freedom in the model (e.g., with 4 times there are 3 *d.f.*, 1 *d.f.* if you have a simple linear term for time as continuous).
  - ▶ recommended when there are relatively few times (say, five or less), for which tests for polynomial trends can still be conducted (see course notes for details).
- ▶ Time as continuous is recommended when...
  - ▶ there are many times of observation, possibly unequally spaced
  - ▶ there are different times of measurement for subjects
  - ▶ interpolating estimates and predicted values may be of interest

Equivalence of model using time as class and model using time as continuous with  $(t-1)$ th degree polynomials (if time has  $t$  levels)

- ▶ Same log likelihood
- ▶ Same  $F$ -test for Time