

Mixed Effects

DEFINE MIXED EFFECTS

GIVE OVERVIEW OF CONDITIONAL DISTRIBUTIONS

Fixed Effect or Random Effect?

Whether an effect is classified as fixed or random depends

- on how the factor levels were selected (were they randomly chosen?)
- the objectives of the experiment (are you interested in testing that effect?)

Example: We are running an agricultural experiment on crop yield. Let's consider two related scenarios:

1. We have a group of farmers that paid for statistical consulting for optimizing their farms' yields
2. We sample from a group of farmers and we run an experiment using the selected farms' yields as the response

We would block by farm in both cases

But in 1. the block is of direct interest while in 2. it's used to control for confounding

Some Definitions

If all of the effects in an analysis are fixed, then it is a **fixed** effects model

If all of the effects in an analysis are random, then it is a **random** effects model

In all other cases, it is a **mixed** effects model

Example: A typical mixed effects model would look like:

Two levels of a treatment factor with means $\mu_j, j = 1, 2$

K levels of a blocking factor $B_k, k = 1, 2, \dots, K$

$$Y_{ijk} = \mu + \mu_j + B_k + \varepsilon_{ijk}$$

At this point, we have only specified the form for the model but made no distributional assumptions..

Adding Distributional Assumptions

Example: A typical mixed effects model would look like:

Two levels of a treatment factor with means $\mu_j, j = 1, 2$

K levels of a blocking factor $B_k, k = 1, 2, \dots, K$

$$Y_{ijk} = \mu + \mu_j + B_k + \varepsilon_{ijk}$$

In (almost) all cases, $B_k \sim N(0, \sigma_B^2)$ (but there will be cases in which there is a covariance between the B_k)

This is like saying that all the cumulative effects of being in block B_k can be summarized in a single number that we consider random

Intuitively, we might model it as random to represent our ignorance about all of these unmeasured effects

(this is somewhat analogous to weather predictions)

Adding Distributional Assumptions

Example: A typical mixed effects model would look like:

Two levels of a treatment factor with means $\mu_j, j = 1, 2$

K levels of a blocking factor $B_k, k = 1, 2, \dots, K$

$$Y_{ijk} = \mu + \mu_j + B_k + \varepsilon_{ijk}$$

In (almost) all cases, $B_k \sim N(0, \sigma_B^2)$

Also, typically we assume that the B_k and ε_{ijk} are independent

If the response is sensibly modeled as a normal random variable, then:

$$\varepsilon_{ijk} \sim N(0, \sigma^2)$$

Now inferences can be made with this model

Conditional Distributions

Conditional Distributions

It will be convenient to refer to conditional probability and distributions

We have been talking about them all along, but not in full generality

Example: Modeling the mean of Y as a function of X is a **conditional mean**: $\mu\{Y|X\}$

Likewise, the “conditional distribution of Y given X ” is written $Y|X$

Conditional probability is straightforward intuitively but can be challenging notationally/theoretically

Intuition: Conditional probability can be thought of as “updating” your probabilities with new information that you now “know”

Example: What is the probability that Dr. Gunst is in his office right now?

What is the probability that Dr. Gunst is in his office right now given that he teaches until 9:30 am Tuesday/Thursday?

Conditional Distributions

Suppose we are asking the question: what is the probability that a random person will have a heart attack in the next year: $\text{Prob}(H)$

This would be fairly easy to estimate: Let H be the heart attack event

$$\widehat{\text{Prob}}(H) = \frac{\text{\# of total heart attacks}}{\text{\# of total people}}$$

(Note that $\text{Prob}(H)$ is usually called the **marginal distribution** of H)

However, this would be pretty useless as not everyone has the same probability

Instead, we need to take into account diet, weight, family history,... Call these D

Now, we would like to report $\widehat{\text{Prob}}(H|D)$

This is much harder to estimate...

Conditional Distributions

One thing that we might have a lot of information about is the “opposite”:

Once someone has a heart attack, what are their diet, weight, family history, ...

$$\widehat{\text{Prob}}(D|H) = \frac{\text{\# of people admitted to hospital with } D \text{ and } H}{\text{\# of total people with } H}$$

A common mistake is to informally “invert” this probability and talk about it like $\widehat{\text{Prob}}(H|D)$

Example: Lots of obese people have heart attacks. Hence, there are public service announcements about the risks of obesity.

But, this is fundamentally a statement about $\widehat{\text{Prob}}(D = \text{obese}|H)$, not the hoped-for implication of $\widehat{\text{Prob}}(H|D = \text{obese})$

(Of course, this is layered on top of another interpretation issue)

Conditional Distributions

Bayes' theorem is the language of conditional distributions

Supposing we wish to know $\text{Prob}(H|D)$ then we can use Bayes' theorem:

$$\text{Prob}(H|D) = \frac{\text{Prob}(D|H)\text{Prob}(H)}{\text{Prob}(D)}$$

Now we know how to estimate two of the quantities on the right-hand side

(we will discuss “logistic regression” later which will provide a better estimate of $\text{Prob}(D|H)$)

We need a reasonable estimator of $\text{Prob}(D)$ (the marginal dist. of D):

$$\widehat{\text{Prob}}(D) = \frac{\text{\# of people with condition } D}{\text{\# of total people}}$$

Now, we can produce the desired probability estimate:

$$\widehat{\text{Prob}}(H|D) = \frac{\widehat{\text{Prob}}(D|H)\widehat{\text{Prob}}(H)}{\widehat{\text{Prob}}(D)}$$

Mixed Effects Models

Mixed Effects as a Conditional Distribution

Going back to our example:

Two levels of a treatment factor with means $\mu_j, j = 1, 2$

K levels of a blocking factor $B_k, k = 1, 2, \dots, K$

$$Y_{ijk} = \mu + \mu_j + B_k + \varepsilon_{ijk}$$

We can incorporate the random effect via a conditional distribution

$$Y_{ijk}|B_k \sim N(\mu + \mu_j + B_k, \sigma^2)$$

(Note that we still need to the marginal distribution of B_k , which is still normal)

Linear Mixed Effects Model

Suppose we have:

- Fixed effect explanatory variables $x_j, j = 1, 2, \dots, p$
- Random effect explanatory variables $u_q, q = 1, 2, \dots, Q$

Then we write the conditional mixed effects model as: $i = 1, 2, \dots, n$

$$\mu\{Y_i | u_1, u_2, \dots, u_Q\} = \mu + \sum_{j=1}^p \beta_j X_{ij} + \sum_{q=1}^Q z_{iq} u_q$$

Here z_{iq} is just either 0 or 1

(depending on which random effect occurs with which observations)

(Note: If you specify an interaction between a fixed and random effect, that interaction is a random effect)

Linear Mixed Effects Model

It is convenient to write this in matrix form as

$$\mu\{Y|U\} = \mu + \mathbb{X}\beta + \mathbb{Z}U$$

There is also a conditional variance:

$$\text{Var}\{Y|U\} = R, \text{ where } R \text{ is an } n \text{ by } n \text{ covariance matrix}$$

(This is like writing $Y = \mu + \mathbb{X}\beta + \mathbb{Z}U + \varepsilon$, with $\text{Var}\{\varepsilon\} = R$. So far, $R = \sigma^2 I$. But this won't always be the case)

Both of these together plus the normality assumption yields:

$$Y|U \sim N(\mu + \mathbb{X}\beta + \mathbb{Z}U, R) \quad \rightarrow Y \text{ has a conditional normal distribution}$$

To complete the specification, write the marginal distribution of U as

$$U \sim N(0, G), \text{ where } G \text{ is an } Q \text{ by } Q \text{ covariance matrix}$$

(The notation “R” and “G” are so common that they lead to formal terms later)

Linear Mixed Effects Model: Estimation

The parameters that need to be estimated are μ , β , R , and G

Additionally, we need to make predictions about the random effects U

Generally speaking, this is done via (restricted) maximum likelihood (although, the implementation and details are a bit involved)

Linear Mixed Effects Model: Conditional vs. Marginal

Conditional Model:

$Y|U \sim N(\mu + \mathbb{X}\beta + \mathbb{Z}U, R)$, where R is an n by n covariance matrix

$U \sim N(0, G)$, where G is a Q by Q covariance matrix

(Some mean term has been converted
to variance/covariance)

We can alternatively write down the **Marginal Model**:

$$\mu\{Y\} = \mu + \mathbb{X}\beta$$

$$\text{Var}\{Y\} = \text{Var}\{\mu + \mathbb{X}\beta + \mathbb{Z}U + \varepsilon\} = \text{Var}\{\mathbb{Z}U\} + \text{Var}\{\varepsilon\} = \mathbb{Z}G\mathbb{Z}^T + R$$

Here is the “Equivalence Law of Statistics” at work!

Linear Mixed Effects Model: Conditional vs. Marginal

Conditional Model:

$$Y|U \sim N(\mu + \mathbb{X}\beta + \mathbb{Z}U, R)$$

$U \sim N(0, G)$, where G is an Q by Q covariance matrix

Marginal Model:

$$Y \sim N(\mu + \mathbb{X}\beta, \mathbb{Z}G\mathbb{Z}^T + R)$$

Although these models are equivalent when Y is modeled as normal:

- Equivalent doesn't mean the same (especially in terms of algorithms)
- the conditional and marginal models won't be equivalent under other distributional assumptions e.g. Bernoulli

Models can be fit with a very wide range of \mathbb{X} , R , and G

(Multiway ANOVA. \mathbb{X} as 0s and 1s (depending on design), $R = \sigma^2 I$, and $G = 0$)

Linear Mixed Effects Model: Conditional vs. Marginal

Conditional Model:

$$Y|U \sim N(\mu + \mathbb{X}\beta + \mathbb{Z}U, R)$$

$U \sim N(0, G)$, where G is an Q by Q covariance matrix

Marginal Model:

$$Y \sim N(\mu + \mathbb{X}\beta, \mathbb{Z}G\mathbb{Z}^T + R)$$

Note that the marginal model specification won't have G-side effects

(Reminder: the notation "R" and "G" are so common that they lead to formal terms)

Common Terminology:

- Parameters encoded in R are called "R-side effects" (for **residual**)
- Parameters encoded in G are called "G-side effects" (for the random effect)

Mixed Effects Models: Example

An Agricultural Experiment

Suppose we are testing the effectiveness of two cultivars of a crop at a large, commercial farm

The farm is naturally divided into a large number of pairs of plots, so we decide to create a blocking factor that contains two plots

These plots will be the experimental unit with respect to the treatment

We choose to model the blocking variable as random due to:

- As the farm contains many such pairs of plots, we subsample from them randomly (Let's choose 10 pairs of plots)
- We are not directly interested in the blocking effect

(Generally, never treat a factor you are interested in as random)

Formally, this is a split-plot design, but it is more common to call it a "Completely Randomized Block Design" (RCB)

Testing for Treatment Effect with a Blocked Design

Let's represent the two varieties of cultivars as a treatment factor with means $\mu_j, j = 1, 2$

The blocking factor will be represented as $K = 10$ levels of a random effect $B_k, k = 1, 2, \dots, 10$

Then the model can be written

$Y_{jk} = \mu + \mu_j + B_k + \varepsilon_{jk}$ (it is worth thinking about why there is no "i")

We have two experimental units per block assigned to treatment (n=20)

For any observation, $\varepsilon_{jk} \sim N(0, \sigma^2)$ (all independent), $R = \sigma^2 I$

For this experiment, it makes sense that the blocks are uncorrelated:
 $G = \sigma_B^2 I$

Also, the B_k and ε_{jk} are independent

Testing for Treatment Effect with a Blocked Design

Let's look at the marginal model and compute the form for $\text{Var}\{Y\}$

(These will be R-side effects. Reminder: R-side comes from “residual”)

$$Y_{jk} = \mu + \mu_j + B_k + \varepsilon_{jk}$$

- Every observation has variance $\sigma^2 + \sigma_B^2$
- Any two observations in the same block will additionally have a covariance σ_B^2
- Any two observations in different blocks will have zero covariance

This can all be summarized as

$$\text{Cov}(Y_{jk}, Y_{j'k'}) = \begin{cases} \sigma^2 + \sigma_B^2 & \text{if } j = j', k = k' \\ \sigma_B^2 & \text{if } j \neq j', k = k' \\ 0 & \text{otherwise} \end{cases}$$

The Marginal Covariance Matrix

Any covariance structure that look like:

$$\text{Cov}(Y_{jk}, Y_{j'k'}) = \begin{cases} \sigma^2 + \sigma_B^2 & \text{if } j = j', k = k' \\ \sigma_B^2 & \text{if } j \neq j', k = k' \\ 0 & \text{otherwise} \end{cases}$$

As a matrix (which is $n \times n$, remember), this looks like:

$$\text{Var}\{Y\} = \begin{bmatrix} \sigma^2 + \sigma_B^2 & \sigma_B^2 & 0 & \dots & 0 \\ \sigma_B^2 & & & & \vdots \\ 0 & & \ddots & & 0 \\ \vdots & & & \sigma_B^2 & \\ 0 & \dots & \sigma_B^2 & \sigma^2 + \sigma_B^2 \end{bmatrix} \text{ as long as } Y = \begin{bmatrix} Y_{1,1} \\ Y_{2,1} \\ Y_{1,2} \\ Y_{2,2} \\ \vdots \\ Y_{2,10} \end{bmatrix}$$

Covariance Structures

A General Covariance Matrix

A general covariance matrix would look like

$$\text{Var}\{Y\} = \begin{bmatrix} \sigma_{1,1}^2 & \sigma_{1,2}^2 & \cdots & \sigma_{1,n}^2 \\ \vdots & \ddots & & \vdots \\ \sigma_{n,1}^2 & \cdots & \sigma_{n,(n-1)}^2 & \sigma_{n,n}^2 \end{bmatrix} = R$$

(This is an R-side effect. There is no G-side effect in a marginal model as all variability is subsumed into ε)

This is the maximally flexible model & is known as the **unstructured (UN) covariance**

It requires the most possible number of parameters: $(n-1)n/2$

(due to the symmetry in a covariance matrix)

Often there is more structure and hence we can use less complex models:

- Variance Components (VC)
- Compound Symmetry (CS)
- Autoregressive(1) (AR(1))

(There are many others. See the SAS document: <http://www2.sas.com/proceedings/sugi30/198-30.pdf>)

Variance Components

The **variance components (VC)** model is the SAS default and looks like

$$\text{Var}\{Y\} = \begin{bmatrix} \sigma_{1,1}^2 & \sigma_{1,2}^2 & \cdots & \sigma_{1,n}^2 \\ \vdots & \ddots & & \vdots \\ \sigma_{n,1}^2 & \cdots & \sigma_{n,(n-1)}^2 & \sigma_{n,n}^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix},$$

This encodes no covariances, but encodes different variances, usually at different combinations of treatments

Example: We might have an interaction model with effect A and B:

$$\text{Var}\{Y\} = \begin{bmatrix} \sigma_A^2 & 0 & 0 & \cdots & 0 \\ 0 & & & & \vdots \\ 0 & & \ddots & & 0 \\ \vdots & & & & 0 \\ 0 & \cdots & 0 & & \sigma_{AB}^2 \end{bmatrix} \quad (\text{there are 3 variance parameters})$$

Compound Symmetry

The **compound symmetry (CS)** model looks like

$$\text{Var}\{Y\} = \begin{bmatrix} \sigma_{1,1}^2 & \sigma_{1,2}^2 & \cdots & \sigma_{1,n}^2 \\ \vdots & \ddots & & \vdots \\ \sigma_{n,1}^2 & \cdots & \sigma_{n,(n-1)}^2 & \sigma_{n,n}^2 \end{bmatrix} = \begin{bmatrix} \Sigma & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Sigma \end{bmatrix},$$

These Σ are known as “**blocks**” of the matrix. Not to be confused with “blocking” in experiment design, though they often overlap

$$\text{Where } \Sigma = \begin{bmatrix} \sigma^2 & \sigma_c^2 & \cdots & \sigma_c^2 \\ \sigma_c^2 & & \ddots & \vdots \\ \vdots & & & \sigma_c^2 \\ \sigma_c^2 & \cdots & \sigma_c^2 & \sigma^2 \end{bmatrix} \quad (\text{Note that the variances/covariances are all equal})$$

Example: The covariance model we specified with our RCB is a CS model:

$$\text{Var}\{Y\} = \begin{bmatrix} \sigma^2 + \sigma_B^2 & \sigma_B^2 & 0 & \cdots & 0 \\ \sigma_B^2 & & & & \vdots \\ 0 & & \ddots & & 0 \\ \vdots & & & & \sigma_B^2 \\ 0 & \cdots & \sigma_B^2 & & \sigma^2 + \sigma_B^2 \end{bmatrix} \quad (\text{there are 2 variance parameters})$$

(We will return to AR(1) with time series and with repeated measures)

Autoregressive (1)

The **autoregressive(1) (AR(1))** model looks like

$$\text{Var}\{Y\} = \begin{bmatrix} \sigma_{1,1}^2 & \sigma_{1,2}^2 & \cdots & \sigma_{1,n}^2 \\ \vdots & \ddots & & \vdots \\ \sigma_{n,1}^2 & \cdots & \sigma_{n,(n-1)}^2 & \sigma_{n,n}^2 \end{bmatrix} = \begin{bmatrix} \Sigma & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Sigma \end{bmatrix},$$

$$\text{Where } \Sigma = \sigma^2 \begin{bmatrix} 1 & (\sigma_{AR}^2)^1 & \cdots & (\sigma_{AR}^2)^m \\ (\sigma_{AR}^2)^1 & & \ddots & \vdots \\ \vdots & & & \sigma_c^2 \\ (\sigma_{AR}^2)^m & \cdots & (\sigma_{AR}^2)^1 & \sigma^2 \end{bmatrix}$$

Example: We make 3 measurements of same person at equal time intervals

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & (\sigma_{AR}^2)^1 & (\sigma_{AR}^2)^2 & (\sigma_{AR}^2)^3 \\ (\sigma_{AR}^2)^1 & 1 & (\sigma_{AR}^2)^1 & (\sigma_{AR}^2)^2 \\ (\sigma_{AR}^2)^2 & (\sigma_{AR}^2)^1 & 1 & (\sigma_{AR}^2)^1 \\ (\sigma_{AR}^2)^3 & (\sigma_{AR}^2)^2 & (\sigma_{AR}^2)^1 & \sigma^2 \end{bmatrix} \text{ (there are 2 variance parameters)}$$

Back to the Example

Reminder of Our Model

Let's look at the marginal model and compute the form for $\text{Var}\{Y\}$

(These will be R-side effects. Reminder: R-side comes from “residual”)

$$Y_{jk} = \mu + \mu_j + B_k + \varepsilon_{jk}$$

- Every observation has variance $\sigma^2 + \sigma_B^2$
- Any two observations in the same block will additionally have a covariance σ_B^2
- Any two observations in different blocks will have zero covariance

This can all be summarized as

$$\text{Cov}(Y_{jk}, Y_{j'k'}) = \begin{cases} \sigma^2 + \sigma_B^2 & \text{if } j = j', k = k' \\ \sigma_B^2 & \text{if } j \neq j', k = k' \\ 0 & \text{otherwise} \end{cases}$$

Using SAS to Analyze Mixed Models

Like usual there are (at least) two main ways to analyze mixed models in SAS

- Proc Glimmix
- Proc Mixed

Within each procedure, there are many ways of specifying an equivalent model (e.g. Conditional or Marginal)

Though these specify the equivalent models, that doesn't mean they are the same!

We will focus on Proc Glimmix for now

Let's look at this in our current example...

The Data

```
DATA example;  
  INPUT block treatment F N Y1 Y2;  
  DATALINES;  
1 0 86 100 21.3 39.5  
1 1 98 100 10.0 30.3  
2 0 48 100 19.7 32.6  
2 1 93 100 22.0 33.6  
3 0 87 100 20.2 32.1  
3 1 43 100 17.6 38.8  
4 0 64 100 20.0 41.9  
4 1 89 100 16.0 35.1  
5 0 99 100 17.4 29.1  
5 1 100 100 20.3 34.1  
6 0 52 100 26.0 42.9  
6 1 49 100 25.6 31.8  
7 0 89 100 19.7 40.8  
7 1 96 100 16.2 28.2  
8 0 63 100 18.5 40.9  
8 1 98 100 12.4 38.1  
9 0 48 100 20.9 39.5  
9 1 83 100 13.5 34.7  
10 0 85 100 23.0 42.0  
10 1 97 100 23.2 32.8  
;
```

Here, we are interested in the
block, treatment, and Y1 columns

PROC GLIMMIX

This is a DF approximation. You should use it (I won't always in the code for simplicity, however)

```
PROC GLIMMIX DATA = example;  
CLASS treatment block;  
MODEL *SPECIFY FIXED EFFECTS* / DDFM=KENWARDROGER;  
RANDOM *DEFINE A RANDOM EFFECT (EITHER G OR R SIDE)* / TYPE = *COVARIANCE*;  
RUN;
```

General Linear Mixed Model:

$$Y = \mu + X\beta + ZU + \varepsilon$$

with variance parameters R and G

(You can specify multiple random statements for different random effects)

The default in SAS is “variance components”, which specifies that each effect gets its own variance parameter

Overall idea:

- Specify any random effect (a random ε is included by default)
- Specify any variance structure other than “VC”

PROC GLIMMIX: Example

“blocks”

The Model: $Y_{jk} = \mu + \mu_j + B_k + \varepsilon_{jk}$, $\varepsilon_{jk} \sim N(0, \sigma^2)$, $B_k \sim N(0, \sigma_B^2)$,

Taken as the Conditional Model, this defines $R = \sigma^2 I$, $G = \sigma_B^2 I$

Since there is only 1 random effect, we can encode the single variance parameter for B_k using the default “VC” (hence no “Type” parameter)

```
TITLE 'conditional model';
PROC GLIMMIX DATA = example;
  CLASS treatment block;
  MODEL Y1 = treatment;
  RANDOM block;
RUN;
```

The treatment is a fixed effect

We are treating block as random. No covariance is specified as VC is the default

Dimensions	
G-side Cov. Parameters	1
R-side Cov. Parameters	1
Columns in X	3
Columns in Z	10
Subjects (Blocks in V)	1
Max Obs per Subject	20

Covariance Parameter Estimates		
Cov Parm	Estimate	Standard Error
block	5.6193	5.5187
Residual	9.9538	4.6923

$\hat{\sigma}_B^2$

$\hat{\sigma}^2$

$$H_0: \mu_1 = \mu_2$$

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
treatment	1	9	4.49	0.0631

PROC GLIMMIX: Example

Recall that we can think about the covariance matrix in terms of “blocks”

```
TITLE 'conditional model: output X and Z';
PROC GLIMMIX DATA = example outdesign=matrix;
  CLASS treatment block;
  MODEL Y1 = treatment / DDFM=KENWARDROGER;
  RANDOM block;
RUN;
```

Dimensions	
G-side Cov. Parameters	1
R-side Cov. Parameters	1
Columns in X	3
Columns in Z	10
Subjects (Blocks in V)	1
Max Obs per Subject	20

```
PROC PRINT DATA=matrix;
RUN;
```

```
TITLE 'conditional model: output X and Z';
PROC GLIMMIX DATA = example outdesign=matrix;
  CLASS treatment block;
  MODEL Y1 = treatment / DDFM=KENWARDROGER;
  RANDOM intercept / SUBJECT = block;
RUN;
```

```
PROC PRINT DATA=matrix;
RUN;
```

The random effect is an

adjustment to the intercept

Dimensions	
G-side Cov. Parameters	1
R-side Cov. Parameters	1
Columns in X	3
Columns in Z per Subject	1
Subjects (Blocks in V)	10
Max Obs per Subject	2

PROC GLIMMIX: Example

```
TITLE 'conditional model: output X and Z';
PROC GLIMMIX DATA = example outdesign=matrix;
  CLASS treatment block;
  MODEL Y1 = treatment / DDFM=KENWARDROGER;
  RANDOM block;
RUN;
```

```
PROC PRINT DATA=matrix;
RUN;
```

```
TITLE 'conditional model: output X and Z';
PROC GLIMMIX DATA = example outdesign=matrix;
  CLASS treatment block;
  MODEL Y1 = treatment / DDFM=KENWARDROGER;
  RANDOM intercept / SUBJECT = block;
RUN;
```

```
PROC PRINT DATA=matrix;
RUN;
```

Obs	block	treatment	F	N	Y1	Y2	_SubjectID_	_X1	_X2	_X3	_Z1
1	1	0	86	100	21.3	39.5	1	1	1	0	1
2	1	1	98	100	10.0	30.3	1	1	0	1	1
3	2	0	48	100	19.7	32.6	2	1	1	0	1
4	2	1	93	100	22.0	33.6	2	1	0	1	1
5	3	0	87	100	20.2	32.1	3	1	1	0	1
6	3	1	43	100	17.6	38.8	3	1	0	1	1
7	4	0	64	100	20.0	41.9	4	1	1	0	1
8	4	1	89	100	16.0	35.1	4	1	0	1	1
9	5	0	99	100	17.4	29.1	5	1	1	0	1
10	5	1	100	100	20.3	34.1	5	1	0	1	1
11	6	0	52	100	26.0	42.9	6	1	1	0	1
12	6	1	49	100	25.6	31.8	6	1	0	1	1
13	7	0	89	100	19.7	40.8	7	1	1	0	1
14	7	1	96	100	16.2	28.2	7	1	0	1	1
15	8	0	63	100	18.5	40.9	8	1	1	0	1
16	8	1	98	100	12.4	38.1	8	1	0	1	1
17	9	0	48	100	20.9	39.5	9	1	1	0	1
18	9	1	83	100	13.5	34.7	9	1	0	1	1
19	10	0	85	100	23.0	42.0	10	1	1	0	1
20	10	1	97	100	23.2	32.8	10	1	0	1	1

Obs	block	treatment	F	N	Y1	Y2	_X1	_X2	_X3	_Z1	_Z2	_Z3	_Z4	_Z5	_Z6	_Z7	_Z8	_Z9	_Z10
1	1	0	86	100	21.3	39.5	1	1	0	1	0	0	0	0	0	0	0	0	0
2	1	1	98	100	10.0	30.3	1	0	1	1	0	0	0	0	0	0	0	0	0
3	2	0	48	100	19.7	32.6	1	1	0	0	1	0	0	0	0	0	0	0	0
4	2	1	93	100	22.0	33.6	1	0	1	0	1	0	0	0	0	0	0	0	0
5	3	0	87	100	20.2	32.1	1	1	0	0	0	1	0	0	0	0	0	0	0
6	3	1	43	100	17.6	38.8	1	0	1	0	0	1	0	0	0	0	0	0	0
7	4	0	64	100	20.0	41.9	1	1	0	0	0	0	1	0	0	0	0	0	0
8	4	1	89	100	16.0	35.1	1	0	1	0	0	0	1	0	0	0	0	0	0
9	5	0	99	100	17.4	29.1	1	1	0	0	0	0	0	1	0	0	0	0	0
10	5	1	100	100	20.3	34.1	1	0	1	0	0	0	0	0	1	0	0	0	0
11	6	0	52	100	26.0	42.9	1	1	0	0	0	0	0	0	1	0	0	0	0
12	6	1	49	100	25.6	31.8	1	0	1	0	0	0	0	0	1	0	0	0	0
13	7	0	89	100	19.7	40.8	1	1	0	0	0	0	0	0	0	1	0	0	0
14	7	1	96	100	16.2	28.2	1	0	1	0	0	0	0	0	0	1	0	0	0
15	8	0	63	100	18.5	40.9	1	1	0	0	0	0	0	0	0	0	1	0	0
16	8	1	98	100	12.4	38.1	1	0	1	0	0	0	0	0	0	0	1	0	0
17	9	0	48	100	20.9	39.5	1	1	0	0	0	0	0	0	0	0	0	1	0
18	9	1	83	100	13.5	34.7	1	0	1	0	0	0	0	0	0	0	0	1	0
19	10	0	85	100	23.0	42.0	1	1	0	0	0	0	0	0	0	0	0	0	1
20	10	1	97	100	23.2	32.8	1	0	1	0	0	0	0	0	0	0	0	0	1

PROC GLIMMIX: Example

The Model: $Y_{jk} = \mu + \mu_j + B_k + \varepsilon_{jk}$, $\varepsilon_{jk} \sim N(0, \sigma^2)$, $B_k \sim N(0, \sigma_B^2)$,

Taken as the Marginal Model, this defines $\text{Var}\{Y|B\} = R$:

$$R = \begin{bmatrix} \sigma^2 + \sigma_B^2 & \sigma_B^2 & 0 & \dots & 0 \\ \sigma_B^2 & \sigma^2 + \sigma_B^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma^2 + \sigma_B^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma^2 + \sigma_B^2 \end{bmatrix}, \text{ which is compound symmetry (CS)}$$

```
TITLE 'marginal model';
```

```
PROC GLIMMIX DATA = example;
```

```
CLASS treatment block;
```

```
MODEL Y1 = treatment;
```

```
RANDOM _residual_ / TYPE = cs SUBJECT = block V;
```

```
RUN;
```

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
CS	block	5.6193	5.5187
Residual		9.9538	4.6923

Estimated V Matrix for block 1		
Row	Col1	Col2
1	15.5732	5.6193
2	5.6193	15.5732

The “blocks” of the covariance matrix get specified via the “SUBJECT” statement

$$H_0: \mu_1 = \mu_2$$

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
treatment	1	9	4.49	0.0631

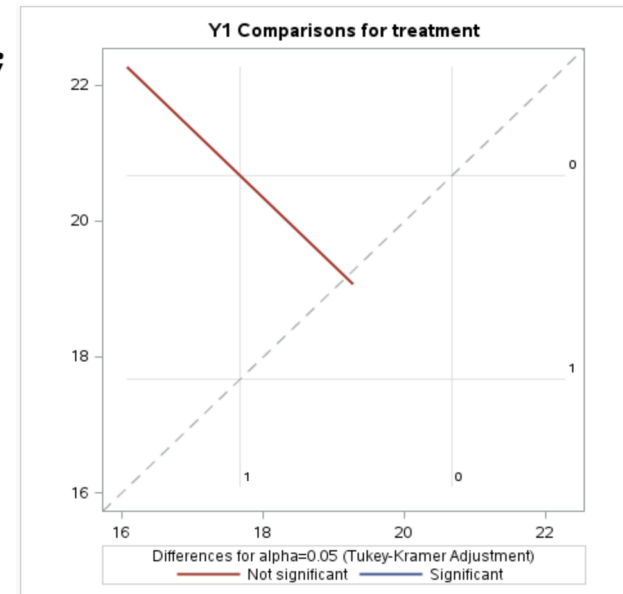
PROC GLIMMIX: Estimating the Treatment Effect

```

TITLE 'conditional model';
PROC GLIMMIX DATA = example;
  CLASS treatment block;
  MODEL Y1 = treatment;
  RANDOM block;
  LSMEANS treatment / adjust = TUKEY DIFF CL PLOTS=DIFF;
RUN;
  
```

This isn't needed in this problem. Why?

(These have the same information but it is displayed differently)



Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
treatment	1	9	4.49	0.0631

Differences of treatment Least Squares Means Adjustment for Multiple Comparisons: Tukey-Kramer												
treatment	_treatment	Estimate	Standard Error	DF	t Value	Pr > t	Adj P	Alpha	Lower	Upper	Adj Lower	Adj Upper
0	1	2.9900	1.4109	9	2.12	0.0631	0.0631	0.05	-0.2018	6.1818	-0.2018	6.1818

When Conditional vs. Marginal Matters

When Conditional vs. Marginal Matters

Notice that there is a zero variance estimate

```
TITLE 'conditional model';
PROC GLIMMIX DATA = example;
  CLASS treatment block;
  MODEL Y2 = treatment / DDFM=KENWARDROGER;
  RANDOM intercept / SUBJECT = block V;
RUN;
```

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
Intercept	block	0	.
Residual		17.4270	5.8090

Estimated V Matrix for block 1		
Row	Col1	Col2
1	17.4270	
2		17.4270

```
TITLE 'marginal model';
PROC GLIMMIX DATA = example;
  CLASS treatment block;
  MODEL Y2 = treatment;
  RANDOM _residual_ / TYPE = cs SUBJECT = block V;
RUN;
```

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
CS	block	-5.2072	6.0628
Residual		22.6342	10.6699

Estimated V Matrix for block 1		
Row	Col1	Col2
1	17.4270	-5.2072
2	-5.2072	17.4270

The REML solution in the conditional model yields a negative variance, which gets thresholded at 0 (this occurs when the $MS(\text{block}) < MS(\text{error})$). If you use the “nobound” option, you will get the same inference in both models but a negative variance in the conditional model