Random 2: Mixed Models

Outline:

0. Some Perspective
1. Blocks as random versus fixed effects (RCB, BIBD)
2. Mixed models with nested factors
3. Mixed models with crossed factors
4. Split-Plot designs
5. Strip-Plot (or Split-Block) designs
6. Split-Split Plot designs

Examples:

1. Mealybugs: RCB Random versus Fixed Blocks
2. Pillow: BIBD with Blocks as Random
3. Pharmaceutical: Two-stage nested mixed model
4. Sunscreen: Two-way (crossed) mixed model
5. Oats: Split-Plot analysis
6. Rice: Strip-Plot Analysis
7. Split-Split Plot Analysis

0. Some Perspective

In this group of notes and the next (Random 2 and 3), we will be working with **mixed models** were some terms are **fixed** and others are **random.** Mixed models commonly arise in research analysis.

Typically research questions will be focused on comparing means for **fixed factors**. We will continue to use Type 3 ANOVA table and pairwise comparisons using the emmeans function.

Random effects will be used to account for blocking structure (focus of Random 2) and/or repeated measures on the same experimental unit (focus of Random 3) where observations cannot be treated as independent.

Recall that with fixed effects models, the error term was always residual variance ($\sigma^2 = \sigma_\varepsilon^2$), estimated as $\hat{\sigma}^2$ = MSResid.
So, F = MS?/MSResid.

We will see that with mixed models, different fixed factors can have different error terms. In other words, the F statistics may have different denominators (other than MSResid).

We will use ANOVA tables with Expected Mean Squares (EMS) to illustrate this. For some examples, I will use the EMSaov package to show ANOVA table with EMS. However, I find the flexibility of this package to be limited.

In this group of notes, we will not do much model selection. The reason is that if a study was run using a particular design (ex: split plot), we will analyze the data based on that design.

We will talk about model selection for mixed models in Random 3. Early Warning: AIC approach needs to be used with caution for mixed models. More details later.

1. Blocks as random versus fixed effects

Up to this point we have modeled blocks as fixed effects. However, it is common to consider blocks as random effects.

The choice depends on the particular situation. Is it reasonable to think of the blocks as having been random selected? Or at least representative of some larger population?

Blocks as random versus fixed effects:

When comparing means, the analyses will give exactly the same results (because the block effects "cancel out") as long as there is no missing data.

When making inference about individual means, the treating blocks as random will give wider confidence intervals (and larger p-values) because we are accounting for block-to-block variability.

Mealybugs Example: RCB on Cycad plants

A zoo is testing pesticides to control mealybugs on cycad plants. They consider only treatments that are non-toxic to animals:

1. Water (control)
2. Horticultural oil
3. Fungal spores in water (the spores grow on the insects).

Five cycad plants are chosen (randomly?) and the three treatments are randomly assigned to three branches on each plant (RCB). The change (before – after) in the number of mealybugs on each branch is recorded.

Notes about mixed model analysis:

1. We will use **lmer()** from the **lme4** package to fit mixed models. (Another option is the lme() function from the nlme package. We will try this approach in Random3.)
2. The syntax (1|Block) is used to indicate a random effect associated with the intercept (1) for each unique level of block.
3. We will use REML estimation. This is the default for lmer().
4. We will continue to use the **emmeans** package and function to get pairwise comparisons for fixed factors. This works fine with mixed models. There is one catch: Depending on whether or not the lmer package has been loaded, emmeans will be expecting a mixed model (or not). In the Mealybugs example, notice that I ran the fixed effects model first, then "detached" emmeans. In order to then run the mixed model, I loaded lme4, then emmeans!

5. The **lmerTest** package is required for testing. Without loading this package, lmer() will not return any p-values for tests of fixed effects!!!!
6. To get a Type 3 ANOVA table for fixed factors, we use anova(Model, ddf="Kenward-Roger"). This is not a typo, after loading the lmerTest package, the anova() function returns Type 3 tests!
7. ddf specifies the method for computing denominator degrees of freedom. **The Kenward-Roger method is recommended as a good choice in many situations. This is the option we will use.**
8. Satterthwaite estimation is the default for lmer(). When the data are balanced the Satterthwaite method will usually give similar results. This is another common choice.

ANOVA Tables with Expected Mean Squares (EMS)

The EMS is a theoretical quantity. It is the average of the values of the MS, over repeated replications of the experiment. We don't know what this value is (because it contains unknown parameters). But we use it to determine **F-tests.**

How to determine F-tests using the EMS:
Form a ratio of MS's that have the same expectation under H_0, but the numerator has larger expectation under H_A.

In this group of notes, I will focus on the tests of FIXED effects. In R, tests of random effects are based on a likelihood ratio test and obtained using rand().

ANOVA Table for RCB Analysis with Fixed Blocks

Source	df	MS	E(MS)
Trt	t-1	MSTrt	$\sigma^2 + b\theta^2_{Trt}$
Block	b-1	MSBlock	$\sigma^2 + t\theta^2_{Block}$
Resid	(t-1)(b-1)	MSResid	σ^2
Total	tb-1		

$$b\theta^2_{Trt} = b\frac{\sum(\propto_i - \bar{\propto}_.)^2}{t-1}$$

Consider the **F-test for Trt:** H0: $\alpha_1=\alpha_2= \ldots = \alpha_t$
Hence, if H0 is true, then $b\theta^2_{Trt} = 0$.
So, F = MSTrt/MSResid.

Similarly the **F-test for Block** is F = MSBlock/MSResid.

ANOVA Table for RCB Analysis with Random Blocks

Source	df	MS	E(MS)
Trt	t-1	MSTrt	$\sigma^2 + b\theta^2_{Trt}$
Block	b-1	MSBlock	$\sigma^2 + t\sigma^2_{Block}$
Resid	(t-1)(b-1)	MSResid	σ^2
Total	tb-1		

The test for Trt is the same as in the Fixed Block analysis:

F-test for Trt: F = MSTrt/MSResid

Mealybug Example: ANOVA Tables

Fixed Blocks:

```
               Sum Sq Df F value   Pr(>F)
Block        2745.60  4  9.6812 0.003708 **
Trt          1728.13  2 12.1871 0.003729 **
Residuals     567.20  8
```

Random Blocks:

```
    Sum Sq Mean Sq NumDF DenDF F.value   Pr(>F)
Trt 1728.1  864.07     2     8  12.187 0.003729 **
```

Fixed versus random blocks when blocks are incomplete

When blocked designs are incomplete (i.e. each block does not contain all treatments) the fixed and random block models give different results for comparisons of means.

1. If the incompleteness is the result of missing data in an RCB, the researcher must decide which model to use. (If there are only one or two missing data points, the results of the two analyses will likely be close.)

2. If the incompleteness is planned, as in a BIBD, the decision is harder. In stat books an analysis with fixed blocks is usually called the "intra-block" analysis. An analysis with random blocks is called "recovery of interblock information."

Example: Fixed versus random blocks in a BIBD (Pillows)

Nine pillows (A, B, C, D, E, F, G, H, I) were tested ($t=9$) for firmness in groups of three pillows at a time ($k=3$). The assignment of pillows to test groups was a BIBD ($b=12$). See previous notes for design layout.

1. This experiment is incomplete by design. If it were complete, there would be $t \times b$ ($9 \times 12 = 108$) observations. We have $k \times b$ ($3 \times 12 = 36$). The design is only 1/3 complete.

2. If blocks are randomly sampled from some population there is a substantial amount of information about the treatments in the comparisons of blocks (i.e. blocks with firm pillows will tend to have high block means.) The mixed model with random blocks will "recover" that information. Recovered information will change emmeans and perhaps make the CI's for differences narrower (depending on blocks variability.)

2. Mixed Models with Nested Factors

Example: A two-factor mixed model with nested factors.

We will return to the same nested design example we used in Random1. Previously we considered sites as random. We now consider sites as a fixed. They are not randomly selected; they are the only two sites of interest. As before, three batches are randomly selected from each site, and five samples randomly selected from each batch.
This is still a two-stage nested model, but it is now a "mixed model" because we are treating site as fixed, rather than randomly selected from a normal population.

Site (Factor A): Fixed with 2 levels.

Batch nested within Site (Factor B(A)): Random with 3 levels.

Sample is the random error term, nested within Site and Batch.

Nested Mixed Model Analysis:

1. Because sites are fixed, we are now interested in estimating and comparing means for sites. We can do this using the ANOVA table and emmeans.

2. Tests of variance components could be done using rand(), but not usually of interest.

ANOVA Table for Mixed Nested Design

Source	df	MS	EMS
A	a-1	MSA	$\sigma^2 + n\sigma^2_{\beta(\alpha)} + bn\theta^2_\alpha$
B(A)	(b-1)(a)	MSB(A)	$\sigma^2 + n\sigma^2_{\beta(\alpha)}$
Resid	ab(n-1)	MSResid	σ^2
Total	abn-1		

Consider the **F-test for A:** H0: $\alpha_1=\alpha_2= \ldots = \alpha_t$
Hence, if H0 is true, then $bn\theta^2_\alpha = 0$.
So, F = MSA/MSB(A).

Notice that in this case, the error term is not MSResid, instead it is MSB(A).

Pharmaceutical Example: ANOVA Table

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr(>F)
site	0.0019445	0.0019445	1	4	0.16082	0.7089

3. Mixed Models with Crossed Factors

Sunscreen Example: (Ex 17.3 from O & L) Two sunscreens (Factor A) are tested on 10 randomly selected subjects (Factor B). Four 1" square patches are identified on each subject and the two sunscreens are randomly assigned the four patches (2 reps). Color is measured before and after sun exposure.

Notes:

1. The primary interest is in comparing sunscreens. Estimating variance components would be of secondary (or no) interest.

2. This an RCB design (subject = block), but now with replicates. In the "standard" RCB design where each trt appears once in each block we could not estimate interaction. In this case, since we have 2 replicates of each sunscreen we can include an interaction term.

Model for the Sunscreen Example:

y_{ijk} = Change in skin color after exposure for the

i^{th} sun screen, $i = 1,\ldots,a = 2$

j^{th} subject, $j = 1,\ldots, b = 10$

k^{th} patch, $k = 1,\ldots, n = 2$

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

μ and α_i are fixed effects

$\beta_j \sim N(0, \sigma_\beta^2)$

$(\alpha\beta)_{ij} \sim N(0, \sigma_{\alpha\beta}^2)$

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

Rules for the Classification of Interactions

1. If a fixed effect interacts with another fixed effect, the resulting interaction term is a fixed effect.
2. If a random effect interacts with another effect (fixed or random), the resulting interaction term is a random component.

Note: Nested effects are usually random effects. But there are occasional exceptions.

ANOVA Table for Mixed Crossed Design

Source	df	MS	E(MS)
A	a-1	MSA	$\sigma^2 + n\sigma^2_{\alpha\beta} + bn\theta^2_{\alpha}$
B	b-1	MSB	$\sigma^2 + n\sigma^2_{\alpha\beta} + an\sigma^2_{\beta}$
A*B	(a-1)(b-1)	MSAB	$\sigma^2 + n\sigma^2_{\alpha\beta}$
Error	ab(n-1)	MSE	σ^2
Total	abn-1		

Consider the **F-test for A:** H0: $\alpha_1=\alpha_2= \ldots = \alpha_t$
Hence, if H0 is true, then $bn\theta^2_{\alpha} = 0$.
So, F = MSA/MSAB.

Notice that in this case, the error term is not MSResid, instead it is MSAB.

Sunscreen Example: ANOVA Table

```
         Sum Sq Mean Sq NumDF DenDF F.value  Pr(>F)
screen 0.89239 0.89239     1     9  6.7605 0.02873 *
```

4. The Split-Plot Design

Oats Example: (Steel and Torrie p.384) Yield of oats from four Seed lots (Factor A) treated with four chemical seed treatments (Factor B). A total of 4 blocks were used for the experiment.

Lots (Factor A) = Whole Plot Factor

(1) Vickland (infected with *H.Victoriae)*
(2) Vickland (not infected)
(3) Clinton
(4) Branch

Trts (Factor B) = Sub Plot Factor

(1) Check
(2) Ceresan M
(3) Panogen
(4) Agrox

Two-stage randomization: In each of four blocks, seed lots were randomly assigned to large whole plots. Whole plots were subdivided, then chemical seed treatments were randomly assigned to subplots.

Blocking Design:

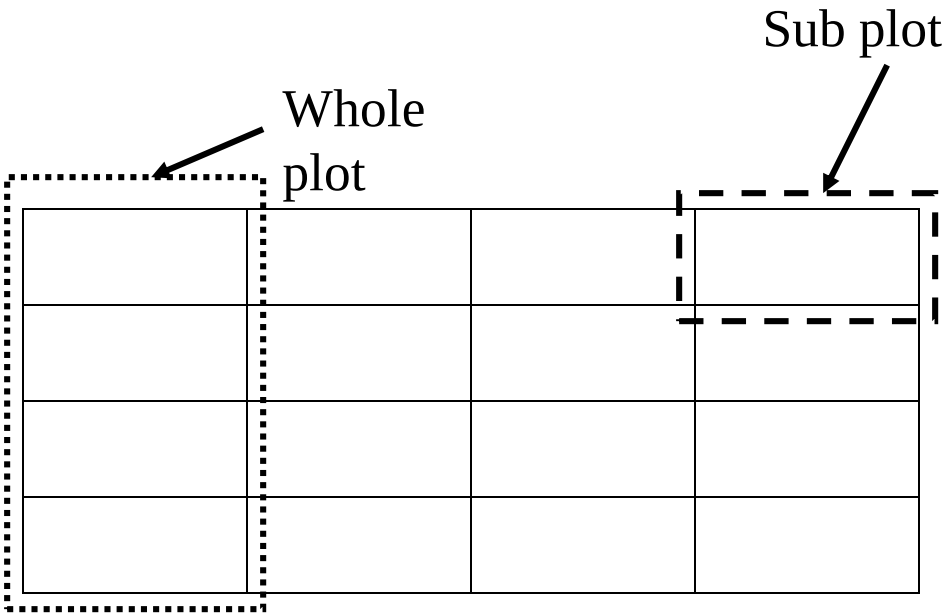
Split Plot

Whole Plot Factor = Lots (Factor A)

Sub Plot Factor = Trt (Factor B)

Treatment Design:

4x4 factorial



Split Plot Model:

Let y_{ijk} = response (yield) for the

i^{th} lot (Whole Plot Factor A), $i = 1,\ldots,a = 4$

j^{th} block, $j = 1,\ldots, r = 4$

k^{th} trt (Sub Plot Factor B), $k = 1,\ldots, b = 4$

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \tau_k + (\alpha\tau)_{ik} + \varepsilon_{ijk}$$

$\mu.\ \alpha_{i.}, \tau_k, (\alpha\tau)_{ik}$ are fixed effects

$\beta_j \sim N(0, \sigma_\beta^2)$

$(\alpha\beta)_{ij} \sim N(0, \sigma_{\alpha\beta}^2)$

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

Primary interest: Lot, Trt and Lot*Trt interaction.

Variance components are typically of little (or no) interest.

ANOVA Table for a Split-Plot Design:

Source	df	EMS
Blk	r-1	$\sigma^2 + b\sigma^2_{\alpha\beta} + ab\sigma^2_{\beta}$
Lots=A	a-1	$\sigma^2 + b\sigma^2_{\alpha\beta} + rb\theta^2_{\alpha}$
Blk*A=Error(a)	(r-1)(a-1)	$\sigma^2 + b\sigma^2_{\alpha\beta}$
Trts=B	b-1	$\sigma^2 + ra\theta^2_{\tau}$
A*B	(a-1)(b-1)	$\sigma^2 + r\theta^2_{\alpha\tau}$
Resid=Error(b)	a(b-1)(r-1)	σ^2
Total	rab-1	

Note: Whole plot factor = A = Lots

Sub plot Factor = B = Trt

F-tests for Split-Plot Example:

F-test for Lots=MSA/MSE(a) =MSA/MS(Blk*A)

F-test for Trts=MSB/MSE(b) = MSB/MSResid

F-test for Lots*Trts=MS(A*B)/MSE(b) = MSB/MSResid

Oats Split Plot ANOVA Table:

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr(>F)
lot	842.03	280.675	3	9	13.8188	0.001022
trt	170.54	56.846	3	36	2.7987	0.053859
lot:trt	586.47	65.163	9	36	3.2082	0.005945

Some comments about the split-plot design:

1. A good characterization of a split plot design is "an experiment within an experiment."

   A. If we ignore factor B, and look just at factor A, it is an RCB design on the whole plots. (The top half of the ANOVA table reflects this structure.)
   B. If we ignore factor A, and look just at factor B, it is also an RCB design, where a whole plot is a block. (The bottom half of the ANOVA table reflects this structure, except for the addition of the A*B interaction.)

2. The accuracy of comparisons on factor B (subplots) and the interaction is better than the accuracy of comparisons on factor A (whole plots). We see this:

   A. Technically in the ANOVA table, because the error terms for B and A*B have smaller expectations.

B. Intuitively, because the sub-plot factor B uses whole plots as blocks. These are small blocks and likely to be more homogeneous.

The split-plot design is sometimes chosen because the experimenter wants to have more accuracy on sub-plot factor B, at the expense of giving up accuracy on whole-plot factor A.

3. Experimenters more often choose the split-plot design for convenience, so that one factor (factor A) can be applied in larger areas. Sometimes, this is because factor A can not be applied to small areas (e.g. the tillage or harvesting equipment is too large). Sometimes factor A can be applied to small areas, but it is expensive (e.g. irrigation treatments need too much pipe when applied to small areas).

4. One of the most important mistakes in experimental design is to designate the factor needing the most accuracy as the wholeplot treatment.

5. The split-plot design is very popular, but the split-plot analysis is even more common because many experiments that do not actually involve splitting plots are analyzed using the same method:

   A. RCB design at multiple locations.
   B. Some types of replicated Latin Squares.
   C. Many repeated measures designs.
   D. There are many variations on the split-plot structure. We will see some of these.

Example write-up for Oats Split Plot Analysis:

Analysis was done using R and the lme4, lmerTest and emmeans packages (provide references!). A mixed model was fit using yield as the response. Fixed effects included seed lot (Bran, Clin, Vic1, Vic2), treatment (Agrox, Ceres, Check, Panong) plus lot*treatment interaction. Block and block*lot were included as random effects to account for the split plot design (with lot as the whole plot factor). For each lot, treatments were compared using Tukey adjusted pairwise comparisons. Model assumptions were checked using residual diagnostic plots.

Also include:

1. A summary table with means (or emmeans), SE and sample sizes.
2. Interaction plot or grouped bar chart.

Summary of Results:

There were no significant differences between treatments for lots Bran, Clin and Vic2.

For the Vic1 lot, Ceres had higher mean yield than both Agrox (estimated difference = 13.325 , p = 0.001) and Check (estimated difference =14.575 , p = 0.0003). For Vic1 lot, Panong had higher mean yield than Check (estimated difference = 9.800 , p = 0.020)

5. The Strip-Plot (or Split-Block) Design

Rice Example: (Gomez and Gomez, p. 110) Yield of six varieties of rice (Factor A) treated with three Nitrogen rates (Factor B).

Var: 1) IR81
2) IR127-80
3) IR305-4-12
4) IR400-2-5
5) IR665-58
6) Peta.

Nrates: 1) 0
2) 60 kg/ha
3) 120 kg/ha.

Block I

	0N	60N	120N
V6			
V3			
V5			
V1			
V2			
V4			

<u>Two-stage randomization:</u> In each of three blocks, varieties were randomly assigned to horizontal strips. N rates were randomly assigned to vertical strips.

Strip-Plot Model:

Let y_{ijk} = yield for ith Var (A), i=1,...,a=6

jth Block, j=1,...,r=3

kth N rate (B), k=1,...,b=3

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \tau_k + (\beta\tau)_{jk} + (\alpha\tau)_{ik} + \varepsilon_{ijk}$$

μ , α_i, τ_k and $(\alpha\tau)_{ik}$ are fixed effects

β_j are $N(0,\sigma^2_\beta)$ $(\alpha\beta)_{ij}$ are $N(0,\sigma^2_{\alpha\beta})$

$(\beta\tau)_{jk}$ are $N(0,\sigma^2_{\beta\tau})$ ε_{ijk} are $N(0,\sigma^2)$, all independent.

Primary interest: variety effects, N rate effects and the interactions.

(We are not interested in variance components.)

Note: The roles of Variety and N-rate are interchangeable in this notation.

Strip-Plot ANOVA table:

Source	df	EMS
Blk	r-1	.....
Var=A	a-1	$\sigma^2 + b\sigma^2_{\alpha\beta} + rb\theta^2_{\alpha}$
Blk*A=Error(a)	(r-1)(a-1)	$\sigma^2 + b\sigma^2_{\alpha\beta}$
Nrate=B	b-1	$\sigma^2 + b\sigma^2_{\beta\tau} + ra\theta^2_{\tau}$
Blk*B=Error(b)	(r-1)(b-1)	$\sigma^2 + b\sigma^2_{\beta\tau}$
A*B	(a-1)(b-1)	$\sigma^2 + r\theta^2_{\alpha\tau}$
Resid=Error(c)	(a-1)(b-1)(r-1)	σ^2
Total	rab-1	

F-tests for the Strip-Plot Design:

F-test for Variety=MSA/MSE(a) = MSA/MS(Blk*A)

F-test for N rate=MSB/MSE(b) = MSB/MS(Blk*B)

F-test for Var*N rate=MS(A*B)/MSE(c) = MS(A*B)/MSResid

Rice Example: ANOVA Table

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr(>F)
var	15751300	3150260	5	10	7.653	0.00337
nrate	28048730	14024365	2	4	34.069	0.00307
var:nrate	23877979	2387798	10	20	5.801	0.00042

Some comments about the Strip-Plot design:

1. Each factor in a Strip-Plot design looks like an RCB design:
   A. If we ignore factor B, and look just at factor A, it is an RCB design on the rows.
   B. If we ignore factor A, and look just at factor B, it is also an RCB design on columns.

2. The accuracy of the comparisons on factor A is comparable to the accuracy of the comparisons on factor B. The greatest accuracy is given to the interaction contrasts. However, comparison of levels of one factor, holding the level of the other factor fixed are not as accurate as the interaction contrasts.

3. Experimenters use strip-plot designs primarily for experiments in which it is impractical to apply treatments in small areas (e.g. irrigation, tillage, etc.).

4. Some RCB experiments in which multiple measurements are made on the same plot (e.g. multiple cuttings of the crop, or multiple depth measurements) are analyzed as if they were Strip-Plot experiments.

5. Some experiments in which a two-factor treatment structure is measured on <u>each subject</u> are analyzed using the same structure as a Strip-Plot.

6. The Split-Split Plot Design

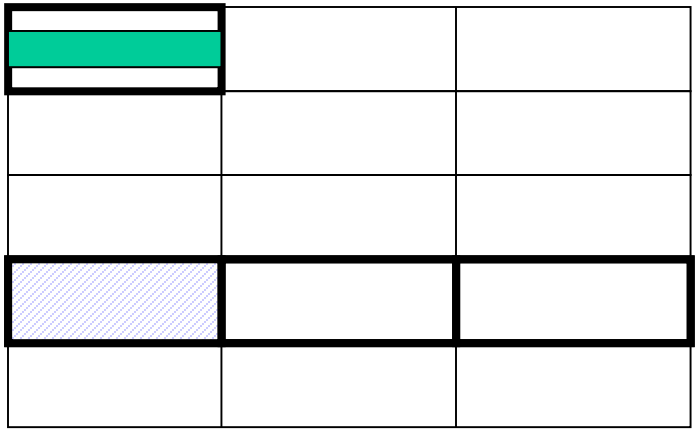
Example: (Gomez and Gomez, p. 143) Yield of rice grown with five Nitrogen rates (Factor A), three Management practices (Factor B) and three Varieties (Factor C).

Nrates (Factor A) = Whole Plot Factor: 0, 50, 80, 110,140kg/ha.

Management (Factor B) = Sub-Plot Factor: M1, M2, M3.

Varieties (Factor C)= Sub-sub-plot Factor: V1, V2, V3.

Block I



Three-stage randomization: In each of three blocks, nrates were randomly assigned to whole plots. Within each whole plot, management practices were assigned to sub-plots. Within each subplot varieties were randomly assigned to sub-sub-plots.

Split-split plot ANOVA table:

Source	df
Blk	r-1
Nrate=A	a-1
Blk*A=Error(a)	(r-1)(a-1)
Manage=B	b-1
A*B	(a-1)(b-1)
Blk*A*B=Error(b)	a(b-1)(r-1)
Var=C	c-1
A*C	(a-1)(c-1)
B*C	(b-1)(c-1)
A*B*C	(a-1)(b-1)(c-1)
Resid=Error(c)	ab(c-1)(r-1)
Total	rabc-1

Note:

Nrate = A = Whole plot factor

Manage = B = Subplot factor

Var = C = Sub-subplot factor

F-tests in the Split-Split Plot design:

A:	Denominator = MSE(a)=MS(A*Blk)
B or A*B	Denominator = MSE(b)=MS(A*B*Blk)
A*C, B*C or A*B*C	Denominator = MSE(c)=MSResid

Some comments about the Split-Split Plot design:

1. A Split-Split-Plot design is "an experiment within an experiment, within an experiment."

2. Comparisons on factor C (sub-sub-plot factor) and its interactions have the best accuracy. Comparisons on factor B (sub-plot factor) and its interaction with A have the next best accuracy, and comparisons on factor A (whole plot factor) have the least accuracy.

3. The experimenter sometimes chooses the Split-Split Plot design for convenience of treatment application, and sometimes to achieve better accuracy on the subplot and sub-sub-plot factors.

Comments on the Split-split plot Example

1. In this example, there is a significant management main effect, and no interactions with management, so the management means are compared, averaging over other factors.
2. There is also a significant nrate*var effect. Study of the nrate*var interaction could be done a variety of ways. In the example, we compare varieties, separately for each nrate. Other strategies are possible.
3. Notice that 2 of the variance components are estimated to be zero (block and block:nrate:manage).
4. When a variance component is estimated to be zero, it is effectively the same as dropping the corresponding random effect from the model (and pooling the corresponding MS and df with another term).

Model Selection for Mixed Models:

1. In many cases, tests of variance components are NOT of interest.
2. However, Random effect terms with small positive, but non-significant values (based on test from rand()) may be retained or deleted, depending on the purpose of the model. Random effects that are dropped from the model will be pooled with other terms.
3. AIC may be used to compare REML models, as long as the models being compared have the same fixed effects.
4. In order to compare mixed models with different fixed and random effects with AIC, ML estimation (REML = FALSE) can be used.