DESIGN AND MIXED-MODEL ANALYSIS OF EXPERIMENTS

XIII. Reflecting on the design and analysis of experiments

XIII.A A summary of the content

Arrangement of units — unrandomized factors

Completely randomized design	–	No allowance made for patterns in experimental material
Randomized complete block design	–	Units are grouped so as to be alike as possible, with no. plots/block = no. treats
Latin squares	–	Units grouped in 2 directions, with no. rows = no. cols = no. treats
Split-plot experiments	–	Units arranged in one of the above designs are split into subunits
Repeated measurements experiments	–	each unit is measured repeatedly over time

Note that the differences between these designs are the restrictions placed on randomization and whether or not a complete set of treatments is observed within blocking units.

Determining the treatments — randomized factors

Single treatment factor	–	Have only one factor to investigate
Factorial Experiments	–	Used to investigate more than one factor; factors are often crossed but may be nested
Factorial Designs at 2 levels	–	Used when have a large number of crossed factors and want to determine which affect response
Unreplicated 2^k experiments	–	Used when have a large number of crossed actors and replication expensive
Confounded 2^k experiments	–	Used when cannot fit all treatment combinations in a single block
Fractional 2^k experiments	–	Used when have at least 3 crossed factors and cannot afford all treatment combinations

Estimation of expectation parameters

For $\mathbf{V} = \sigma^2\mathbf{I}$, $\mathbf{X}$ of full rank	–	least squares and maximum likelihood estimators are $\hat{\theta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ with $\hat{\psi} = \mathbf{X}\hat{\theta}$
For $\mathbf{V} = \sigma^2\mathbf{I}$, $\mathbf{X}$ not of full rank	–	the fitted values are estimable with the least squares and maximum likelihood estimators being $\hat{\psi} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}$

For $\mathbf{V} \neq \sigma^2\mathbf{I}$, $\mathbf{X}$ not of full rank – generalized least squares and maximum likelihood estimators are

$$\hat{\psi} = \mathbf{X}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$$

For indicator variables – estimators are linear combinations of means

Hypothesis testing for model selection

ANOVA method of hypothesis testing
- SSq of effects for each term in the model and for residuals
- degrees of freedom for each source
- expected means squares for each source to justify test statistic
- distribution of test statistic
- summarize in ANOVA table

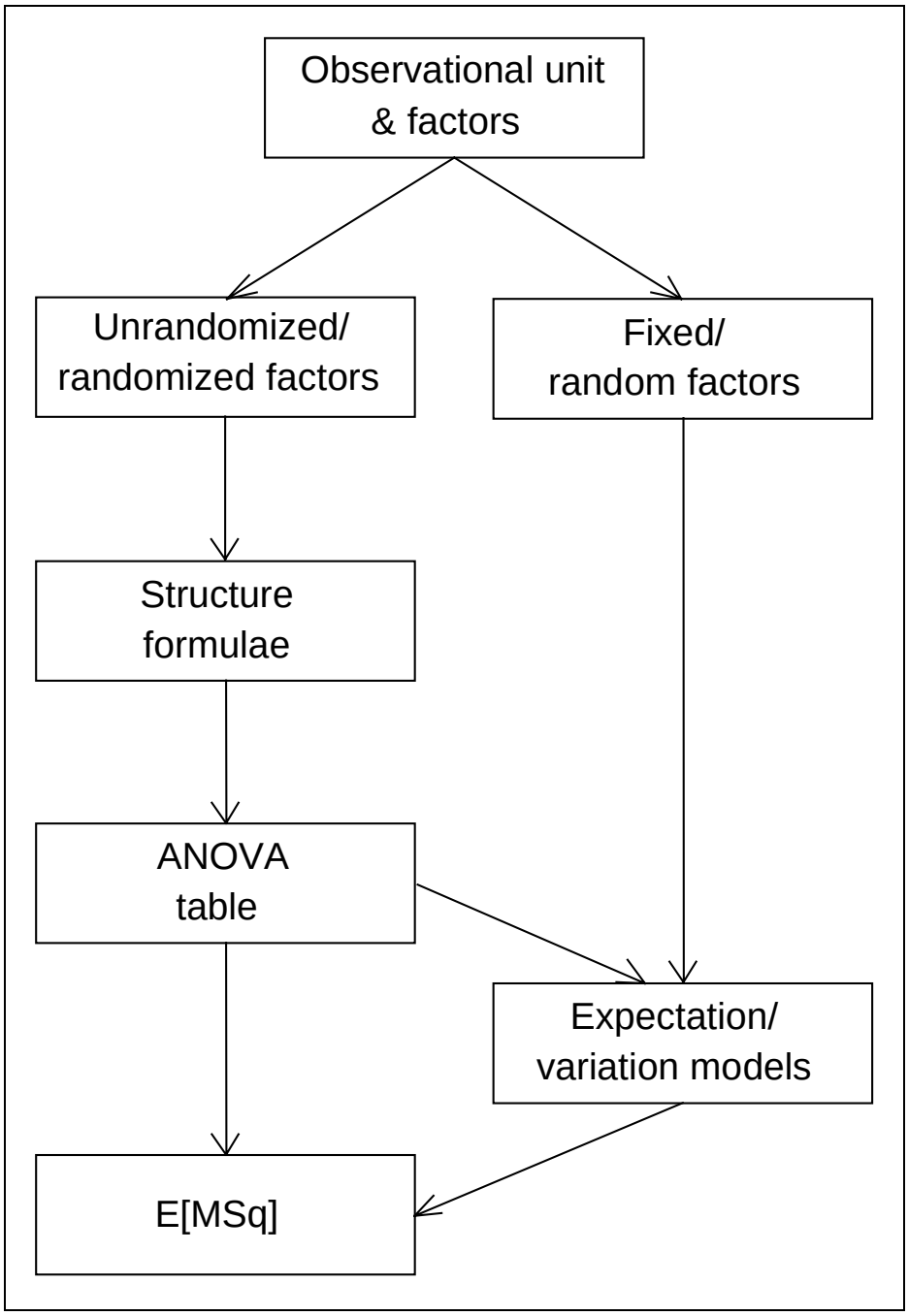
Analysis elements

Test for main effect and interactions	– Used to determine how factors, in general, affect the response.
Multiple comparisons procedures	– Used with terms involving only qualitative factors to examine, in detail, the effects of the factors
Polynomial submodels	– Used with terms involving at least one quantitative factor to characterise the response
Orthogonal contrasts	– To examine specific differences between treatments
Diagnostic checking of the residuals	– Normal probability and residuals-versus-fitted-values plots and Tukey's one-degree-of-freedom-for-nonadditivity
Normal plot of Yate's effects	– Used to determine significance of main effects and interactions in 2^k experiments (preferred to hypothesis testing)

XIII.B Distinctive features of the approach to mixed-model analysis

As can be seen in the following figure, the approach is based on identifying the observational unit and the factors indexing them. Then the factors are divided into both unrandomized and randomized factors and into fixed and random factors. Structure formulae are derived from the unrandomized factors and randomized factors and these expanded to yield terms from which the analysis of variance table is derived. Using the classification into fixed and random factors, terms from the analysis of variance table are designated as expectation and variation terms and the expectation and variation models formulated from these. Finally, the expected mean squares under the hypothesised expectation and variation models are derived and added to the analysis of variance table.

Outline of the approach



The distinctive features of this approach include:

1. Use of observational unit, not the experimental unit
2. Inclusion of all factors from the experiment
3. Division of factors into unrandomized and randomized factors, in addition to the usual division into fixed and random factors, and the use of this in deriving the models and the analysis of variance table
4. Division into fixed and random factors based on model considerations
5. Models do not involve constraints on the parameters and model comparison is emphasized
6. Inference is model-based not randomization-based
7. Role of randomization

The consequences of each of these will now be discussed.

a) Use of observational unit

The observational unit has been defined as the unit for which a single value of a response variable is obtained. This is usually not difficult to determine. On the other hand, an experimental unit is usually defined to be a unit to which a treatment is applied. While in many designs this is easy to identify, it is not so easy for some designs. It is more difficult for experiments involving more than one type of experimental unit, such as complex split-plot experiments. Possibly the most difficult situation in which to determine the experimental unit is grazing trials and we will examine them in more detail later.

b) Inclusion of all factors from the experiment

It is an important aspect of this approach that all the factors involved in the experiment are identified and included in the specification. For example, for the RCBD, the factors are Blocks, Plots and Treatments. Traditionally, the factor Plots is overlooked. It is essential not to do this if the confounding that results from randomization is to be displayed in the model and the sources of variation occurring in the experiment are to be correctly identified — these aspects are further discussed in the section c, *Division of the factors into unrandomized and randomized factors*.

c) Division of the factors into unrandomized and randomized factors

The most important consequence of dividing the factors into unrandomized and randomized factors and using this in deriving the models and the analysis of variance table is that the confounding that occurs in the experiment as a result of randomization is displayed in the analysis of variance table. In this it is recognized that confounding occurs in all experiments and that there is a difference between confounding and aliasing. Confounding results from the way the randomization was performed and aliasing results from the choice of treatment combinations to observe in the experiment. So the analysis of variance table derived using the approach taken in this subject is of a different form to the traditional analysis of variance table. For example a comparison of the two analyses for the RCBD is given in the following table:

Source	df	Source in two-way ANOVA
Blocks	$b-1$	Between Blocks
Blocks.Plots	$b(t-1)$	
Treatments	$t-1$	Between Treatments
Residual	$(b-1)(t-1)$	Error
Total	$bt-1$	Total

The degrees of freedom, sums of squares, mean squares and F statistics for the two analyses are the same. The differences are in the display of the confounding and in the sources of variation associated with terms, in particular the Residual/Error terms. In the analysis derived using the approach presented in this subject, it can be seen that Treatments is confounded with Blocks.Plots — this is not evident for the two-way ANOVA which does not even include the factor Plots in the analysis. Secondly, the Residual term, and hence the corresponding term in the model, represents differences between plots within blocks. On the other hand, the Error term of the two-way ANOVA represents the interaction of Blocks and Treatments. These two sources are quite different. One represents differences arising from inherent variability between the units in the experiment. The other is the way the treatments react to the units or a block-treatment interaction. Samuels, Casella and McCabe (1991) characterise one as inherent unit variability and the other as random interaction. In our analysis we have assumed that there is no Blocks.Treatments interaction. However, it is possible to include both in the analysis if it is thought that there is such an interaction. The analysis incorporating the Blocks.Treatments interaction would be:

Source	df	Source in two-way ANOVA
Blocks	$b-1$	Between Blocks
Blocks.Plots	$b(t-1)$	
Treatments	$t-1$	Between Treatments
Blocks.Treatments	$(b-1)(t-1)$	Error
Total	$bt-1$	Total

In this case we have not gained much because we cannot separate out the plot differences from the interaction. However, there are experiments where it is possible to perform tests of no block-treatment interaction. In particular, it is possible for the generalized RCBD, in which each treatment is replicated r times within each block.

Because the confounding is displayed in the analysis of variance table, experiments that differ in their randomization procedures must also differ in their analysis of variance tables and models. For example, the CRD and a stratified random sample that would traditionally both have a one-way ANOVA differ as shown in the following table:

Sources for CRD	df	Sources for Stratified R.S.
Units	$n-1$	
Treatments	$t-1$	Strata
Residual	$n-t$	Strata.Individuals

In this case we see that the *n-t* degrees of freedom term from the CRD reflects the variability between the Units, whereas that from the stratified random sample represents differences between the individuals within the strata. Other examples are the RCBD compared to the two-way factorial with no interaction and the split-plot experiment compared to the repeated measurements experiment.

d) Division into fixed and random factors based on model considerations

We have emphasized that random factors are those for which an appropriate model for effects involving it is a probability distribution with some variance. On the other hand, fixed effects involve allowing for arbitrary differences such that a probability distribution may not be appropriate as a model for effects involving them. Fixed effects involve less stringent assumptions than random effects.

This basis for distinguishing between fixed and random effects differs from others given in the literature. For example we have argued that random sampling of levels versus no sampling is not an appropriate basis. Nor is an infinite population of levels versus a finite population of levels.

e) Models do not involve constraints on the parameters and model comparison is emphasized

For example, take the models considered for a two-factor experiment arranged in a CRD. The models considered were:

$$E\left[y_{ijk}\right]=\mu \qquad (\text{no factors affect response})$$
$$E\left[y_{ijk}\right]=\alpha_i \qquad (\text{A only affects response})$$
$$E\left[y_{ijk}\right]=\beta_j \qquad (\text{B only affects response})$$
$$E\left[y_{ijk}\right]=\alpha_i+\beta_j \qquad (\text{A and B independently affect response})$$
$$E\left[y_{ijk}\right]=(\alpha\beta)_{ij} \qquad (\text{A and B interact in effect on response})$$

We notice that all these models except the additive model are off full rank and so there is no need for constraints to obtain estimates. Even in the case of the additive model, which is not of full rank, constraints are not necessary. Rather a generalized inverse is used to obtain the estimates under the model.

Further it is asserted that what the model represents is clearer. Consider the maximal model in both forms:

$$E\left[y_{ijk}\right]=(\alpha\beta)_{ij}$$

versus

$$E\left[y_{ijk}\right]=\mu+\alpha'_i+\beta'_j+(\alpha\beta)'_{ij} \text{ with } \sum_{i=1}^{a}\alpha'_i=\sum_{j=1}^{b}\beta'_j=\sum_{i=1}^{a}(\alpha\beta)'_{ij}=\sum_{j=1}^{b}(\alpha\beta)'_{ij}=0$$

It is clear in the first form that this model represents arbitrary differences between expected (average) values of observations for the different combinations of *i* and *j*.

But what does the second model tell us about the expected values of our observations? It difficult to see that it is telling us exactly the same thing as the first model: that there are arbitrary differences between the expected values for different combinations of *i* and *j*. In any case what interest is there in the main effect of A or B when the maximal model is appropriate? I suggest there is none and so it is misleading to include them in the model. It can be shown that the second model is equivalent to a reparametrization of the first in that we can rewrite the first model as

$$E\left[y_{ijk}\right] = \left(\alpha\beta\right)_{ij}$$
$$= \overline{\left(\alpha\beta\right)}_{..} + \left[\overline{\left(\alpha\beta\right)}_{i.} - \overline{\left(\alpha\beta\right)}_{..}\right] + \left[\overline{\left(\alpha\beta\right)}_{.j} - \overline{\left(\alpha\beta\right)}_{..}\right] + \left[\left(\alpha\beta\right)_{ij} - \overline{\left(\alpha\beta\right)}_{i.} - \overline{\left(\alpha\beta\right)}_{.j} + \overline{\left(\alpha\beta\right)}_{..}\right]$$

so that terms in the second model are equivalent to terms in the reparametrized model as follows:

$$\mu = \overline{\left(\alpha\beta\right)}_{..},\ \alpha'_i = \overline{\left(\alpha\beta\right)}_{i.} - \overline{\left(\alpha\beta\right)}_{..},\ \beta'_j = \overline{\left(\alpha\beta\right)}_{.j} - \overline{\left(\alpha\beta\right)}_{..} \text{ and}$$

$$\left(\alpha\beta\right)'_{ij} = \left(\alpha\beta\right)_{ij} - \overline{\left(\alpha\beta\right)}_{i.} - \overline{\left(\alpha\beta\right)}_{.j} + \overline{\left(\alpha\beta\right)}_{..}$$

However, this reparametrization leads to the need to solve the normal equations subject to constraints and is a device for getting a set of unique estimators. It is unnecessary because we have shown how we can obtain estimators without resorting to it. It is confusing in that the model does not relate directly to the mechanisms hypothesized to generate the data.

Another aspect of using the first model is that it emphasizes that the ANOVA is concerned with choosing between very different models that correspond to different mechanisms by which the factors affect the pattern in the expected values, as opposed to deciding which terms to include in the model.

f) Inference is model-based not randomization-based

The approach is not randomization based because it is not uncommon for there to be scientifically interesting questions involve terms that cannot be tested using a randomization test (as opposed to a permutation test).

A very common situation is that of experiments in which block-treatment interactions of interest. Now, block-treatment interactions involve at least one unrandomized and at least one randomized factor. For example, repeated measurements experiments include the factor Time, which is not randomized. The researcher is usually interested in whether there is a Treatment.Time interaction and there is not a randomization-based test for this. Another case where a block-treatment interaction is of interest is example VII.11, *Plant rehabilitation study*, where the interaction of Regimes and Sites is of interest. Also, in the absence of interaction, one may well be interested in the Site main effects. Another example would be an experiment involving a number of animals of each Sex where treatments were randomized to the animals within each sex. The interaction of Treatments with Sex would be of interest and is a block-treatment interaction with no randomization-based test available.

Another situation that arises is when some randomized factors are designated as random. It then happens that tests of some terms may involve a randomized terms as a denominator. Such tests do not have an equivalent randomization test.

g) Role of randomization

While the approach is model-based, not randomization-based, randomization does have a role to play in it. In the approach presented here, randomization:

- Influences the terms to be included in the model and determines the confounding relationships between them
- provides an insurance against biased allocation of treatments and allows for causal inference

The randomization influences the terms to be included in the model through its impact on the crossing and nesting relationships between terms. For example, in the RCBD the plots are nested within blocks, even if Blocks and Plots are intrinsically crossed, because the treatments are randomized to the plots within a block. Consequently, the main effect for Plots does not occur in the model and analysis for this experiment.

We have discussed how the confounding is displayed in the analysis of variance table above. For this to happen requires that ALL factors involved in the experiment be identified and that the factors be divided into unrandomized and randomized factors.

Randomization also provides a way of allocating treatments to units so that, possibly unknown, systematic effects associated with the units do not lead to differences between the treatment groups. That it does this means that we are able to draw the conclusion that our treatments caused the difference. To illustrate, suppose that we are interested in the influence of smoking on blood cholesterol. We draw of random sample of p patients known to smoke and p patients known not to have smoked. This is a stratified random sample and the experimental structure for it is *2* Smoking/p Patients. The analysis of variance appropriate to it is the one-way analysis of variance shown in the table below. In this analysis the hypothesis test answers the question “Is there a difference in blood cholesterol between the two groups that is greater than the variability within the groups?”. If the answer is yes, we cannot be sure that the difference is due to smoking — it may be different due to some other group difference. All we can conclude is that the two groups differ in their blood cholesterol.

Sources for Stratified R.S.	Sources for CRD	df
	Patients	$2p–1$
Smoking	Smoking	1
Smoking. Patients	Residual	$2p–2$

On the other hand, suppose that it was possible to randomize to patients who smoked and who did not. Then we have a completely randomized design and its analysis is also as shown in the table above. Now the hypothesis test asks the

question “Is the difference in blood cholesterol between patients who smoked and those who did not greater than Patient variability?”. The subtle difference is that in the second case there is no natural grouping of patients and the assignment of patients to smoking is arbitrary. This is symbolized in the analysis in that the factor Patients occurs in a term on its own, that is without any nesting factor to indicate groups. In this case we are able to conclude that the smoking differences are due to smoking and not to differences between patients that smoked and those who did not.

XIII.C Obtaining the analysis in packages other than Genstat

The approach presented in this course requires two structure formulae, or equivalently two sets of terms, to be specified to obtain the analyses that we have presented. Genstat, and possibly p-Stat, are the only packages that I know of that allow you to do this. Other packages such as Minitab, SAS, SPSS and S-plus only allow a single set of terms or a single structure formula. Consequently, it is not possible to obtain the analyses as presented here. However, for orthogonal experiments, and we have only looked at orthogonal experiments, it is possible to obtain the correct analysis of variance using the single formula/set of terms. It is just that the insights described in section XIII.B, *Distinctive features of the approach to mixed-model analysis*, will be lost.

For example, for the RCBD, the analysis can be achieved with the single structure/set of terms Blocks + Treatments. This will produce the two-way ANOVA given in section c, *Division of the factors into unrandomized and randomized factors*. It will have the disadvantage of not exhibiting the confounding as we described. However, as a convenient method of getting the analysis, it is fine.

Similarly, the analysis for a factorial experiment arranged in a randomized complete block design could be obtained with the formula Blocks + A + B + A.B. The analysis for a split-plot experiment would require Blocks * A + A * B.

XIII.D Generalization to more than one randomization

So far in this subject the experiments presented have involved just the randomization of one set of factors onto a second set. There are many experiments whose randomizations involve more than two sets of factors. For example, grazing experiments and multiphase experiments.

a) Grazing experiments

Grazing experiments generally involve two randomizations: the randomization of treatments to field units and the randomization of field units to animals. One can then recognise three sets of factors based on the randomization for such a grazing experiment: treatment factors, field factors and animal factors. Of these groups the animal factors are unrandomized, the field factors involved in one randomization and the treatment factors in a second randomization. The analysis of variance can be derived using these three sets of factors in a similar way to that derived from two sets of factors.

Example XIII.1 Grazing experiment

Suppose a continuous grazing trial has *t* treatments assigned to *t* paddocks in each of *b* blocks using an RCBD. Also, *bt* paddocks are randomized to *bta* animals by dividing the animals into *a* homogeneous weight classes based on initial weight and randomizing the *bt* paddocks to the *bt* animals in each class. As a result each paddock receives *a* animals. The response variable is a variable measured on each animal, say weight gain. So observational unit is an animal, the sets of factors are:

unrandomized animal factors:	Classes, Animals
randomized field factors:	Blocks, Paddocks
randomized treatment factors:	Treatments

and the structure formulae based on them are: Classes / Animals; Blocks / Paddocks; Treatments. Also, we will designate Blocks, Paddocks and Animals as random factors and Classes and Treatments as fixed factors. The analysis of variance table derived from the structure formulae is as follows:

Source	df	E[MSq]		
Classes	$(a-1)$	σ^2_{CA}		$+f_C(\psi)$
Classes.Animals	$a(bt-1)$			
Blocks	$(b-1)$	σ^2_{CA}	$+a\sigma^2_{BP}$	$+ta\sigma^2_B$
Blocks.Paddocks	$b(t-1)$			
Treatments	$(t-1)$	σ^2_{CA}	$+a\sigma^2_{BP}$	$+f_T(\psi)$
Residual	$(b-1)(t-1)$	σ^2_{CA}	$+a\sigma^2_{BP}$	
Residual	$(bt-1)(a-1)$	σ^2_{CA}		

What we can see from this analysis is that Treatments is confounded with Blocks.Paddocks which in turn is confounded with Classes.Animals. Further, a test of Treatments is provided by the Residual for Blocks.Paddocks.

Another point in relation to such experiments is that there is considerable debate in the literature about what constitutes the experimental unit. Is it the paddock, the animal or both? The problem is that it is often not recognized that the experiment involves two randomizations and it is not clear what is an appropriate definition of the experimental units in this situation (Brien and Demétrio, 1998). The importance of identifying the experimental unit is that it is used in the identification of appropriate error terms.

This debate is avoided with approach presented in this subject. We identify only the observational unit and then derive expected mean squares that tell us which terms contribute variation to an individual source. In particular, we note that both animal and plot variability contribute to the differences between treatments and so a

denominator for testing for treatment differences must have just these two sources contributing to it.

The problem of improperly replicated grazing trials is another problem that clarified by the approach presented in this subject.

Example XIII.2 Improperly replicated grazing trial

Suppose that each treatment of *t* treatments is randomly applied to just *t* paddocks and that the *t* paddocks are randomized to *ta* animals by dividing the animals into *a* homogeneous weight classes based on initial weight and randomizing the *t* paddocks to the *t* animals in each class. Again, each paddock receives *a* animals. Usually, this analysis would be analyzed using a two-way ANOVA:

Source	df	E[MSq]
Classes	$(a-1)$	$\sigma^2 + f_C(\psi)$
Treatments	$(t-1)$	$\sigma^2 + f_T(\psi)$
Residual	$(t-1)(a-1)$	σ^2

Everything seems fine. The animal scientist would be happy as he has replicates, several animals per plot, and has a Residual for testing for Treatments. However, the analysis derived using the approach presented in this subject is as follows:

Source	df	E[MSq]
Classes	$(a-1)$	$\sigma^2_{CA} + f_C(\psi)$
Classes.Animals	$a(t-1)$	
Paddocks	$(t-1)$	
Treatments	$(t-1)$	$\sigma^2_{CA} + a\sigma^2_{BP} + f_T(\psi)$
Residual	$(t-1)(a-1)$	σ^2_{CA}

Now we can see that there is a problem. Both animals and paddocks affect the variability of treatments but there is no term with just these two sources contributing to it that can be used as a denominator. The problem is that Treatments is completely confounded with Paddocks — there is no residual for Paddock left over to allow the estimation of Paddock variability. The test suggested in the two-way ANOVA is numerically the same as using the Residual for Classes.Animals. However, we see that in using this test we are actually testing whether $a\sigma^2_{BP} + f_T(\psi) = 0$. If the test indicates that it is non-zero, we can only conclude that it is due to treatments if we can be sure that $\sigma^2_{BP} \approx 0$. That is, that there is little plot variability. It is hard to justify this assumption — statistics certainly cannot help.

b) Multiphase experiments

An example of a multiphase experiment is one involving a field experiment from which material is harvested. Then the material from the field experiment has to be processed in a laboratory. So the experiment involves a field phase and a laboratory phase. In general, there should be randomization of treatments to field units in the field phase and the randomization of field units to laboratory units in the laboratory phase. It can be proven that, if each randomization is balanced, the experiment overall will be balanced.