

11. SPLIT PLOT AND REPEATED MEASURES DESIGNS: PARTLY NESTED ANOVAS

In Chapter 9, we described multifactor ANOVA models that can involve crossed or nested factors, or a combination of both, and in Chapter 10, we introduced designs that incorporate either blocks or repeated measures. One particular class of experimental designs with both crossed and nested factors, and either blocks or repeated measures, includes split-plot designs (from an agricultural origin), and repeated measures designs (from psychology). These designs can be complex but are particularly common in biological research, so we have devoted a chapter to their analysis. We will use the term partly nested or partly hierarchical for the linear model we fit with these designs, and the least ambiguous name for these designs might also be partly nested. One of the important messages from this chapter is that these repeated measures and split plot designs are basically analyzed with the same linear model, something that is often unappreciated by biologists, although some textbooks do emphasize the equivalence in models (e.g. Kirk 1995, Mead 1988). In its simplest form, this design has three factors: A and C are crossed, and B is nested within A but crossed with C, although the possible extensions of this design are almost limitless.

11.1. Partly nested designs

11.1.1. Split-plot designs

Split-plot designs were originally used in agricultural experiments and represent a randomized complete block (RCB) design, with one or more factors applied to experimental units within each block. A second factor (or set of factors) is then applied to whole blocks, with replicate blocks for each level of this factor. Note that the terms blocks and plots are interchangeable in the context of these designs.

There are many examples of classical split-plot designs in the biological literature. First we will consider a fictitious extension of the RCB experiment we described in Chapter 10 from Walter & O'Dowd (1992), examining the role of domatia (small cavities on the leaf surface where mites can live) in determining the number of mites on leaves from species with domatia. They set up pairs of leaves (blocks) on a tree where one leaf in each pair was a control and the other leaf had its domatia removed. The treatment factor was applied to experimental units (leaves) within each block (leaf pair). If we now include additional plant species (those that have domatia), we now have a second factor applied at the scale of whole blocks, i.e. a block will be one or other of the species. This new experiment has blocks as the scale of replication for comparisons of species and leaves within blocks as the scale of replication for comparisons of treatments.

As another example, consider the experiment from Wissinger *et al.* (1996) who studied the effects of competition and water regime (hydroperiod) on the ecology of two species of larval caddisflies (*Asynarchus nigriculus* and *Limnephilus externus*) in ponds (Figure 11-1). The experiment was set up as a RCB design, with a block (i.e. plot) being a single pond, chosen for having some consistency in environmental conditions. Within each pond, they set up six wood frame cages in the littoral zone and applied one of six competition treatments (low density *Asynarchus*, low density *Limnephilus*, high density *Asynarchus*, high density *Limnephilus*, high density both species, control with no caddisflies) to each cage within each pond. The role of hydroperiod (permanent or autumnal) was investigated by having four ponds in each category. The response variables were body mass and survival of each species analyzed separately, so there were only three density treatments (those containing the same species). So there are two factors: hydroperiod was “applied” (non-experimentally) to whole

ponds (plots) and is termed the between plots factor and density treatment was applied to cages within plots and is termed the within plots factor. Split plot designs are characterized by having factors applied to experimental units at different, usually spatial, scales.

There are a number of practical design issues for this experiment:

- The experimental design that would be simplest to analyze would be to have whole ponds that are subjected to levels of both factors, hydroperiod and density treatment, forming a completely randomized (CR) factorial arrangement of two hydroperiods by six density treatments with n ponds per cell. Ponds are large units and we would expect considerable variability between them, resulting in large residual variance.
- It is often difficult to install cages, especially large ones. For example, covering whole ponds with cages to maintain experimental densities would be very expensive to set up and probably require an immense amount of labour. We may find that we cannot physically deal with the required size of cages in the time available to set the experiment up, because the research grant has dried up, or we've exhausted the supply of eager volunteers in earlier experiments. We would also need a lot more ponds. The current design uses eight ponds, whereas a completely randomized design with even only two ponds per density and hydroperiod combination would need 24. That many ponds may simply not exist.
- The split-plot design chosen allows us to group our density treatments within ponds, minimizing spatial variation in environmental characteristics, and giving us a clearer test of the effects of density. It also reduces the size of cages. We have, however, linked together groups of cages, and changed our statistical model dramatically compared to the CR design. If anything happens to a pond (e.g. it dries up at the wrong time, or gets an algal bloom), we would be forced to discard all cages in that pond. If we'd used a CR design, we would lose just a single replicate in a cell.

As another example, Leonard *et al.* (1999) tested the prediction that flow had strong effects on the abundances of mussels and barnacles in an estuary but that these effects might vary with tidal height. They had a number of general design options for testing this prediction:

- They could have sampled a range of sites in the estuary. In the simplest case, they could sample replicate sites within combinations of flow regime and tidal height (*i.e.*, a completely randomized factorial design, with two factors, flow regime and tidal height). This approach would require a large number of sites, and it may be difficult to find enough in the estuary.
- It is likely that sites of a given flow regime vary widely, and the researchers would require many replicate sites to get adequate power. They might get less variability if they sampled a given site at different tidal heights, because they could get more similar physical habitats., and make a better test of the effect of height (although the variation between sites will still be a problem for assessing the effects of flow).

Leonard *et al.* (1999) used a split-plot design: plots were sites and they "applied" six tidal heights (from 0.0 to 3.6m above mean low water) within each site, and each site falls into one of two flow regimes, high and low flow. In analyzing this design, we need to keep the six height observations for each site together, so that we can compare their differences.

Split-plot designs can also be used when the plots or blocks do not obviously represent spatial units of replications. For example, Westley (1993) set up a split-plot experiment to examine the effects of inflorescence bud removal on asexual investment in the Jerusalem artichoke (*Helianthus tuberosus*). There were four populations of *H. tuberosus*, five genotypes (genotypes were actually tubers from single individuals) nested within each population and two treatments (normal flowering and inflorescence removed) applied to different tubers from each genotype. Genotypes were plots, population was the between plots factor and treatment was the within plots factor.

We will illustrate the analysis of split-plot designs in this chapter with a recent example from our own work on disturbances on rocky shores.

Effects of trampling on intertidal algae populations

These data come from a long-term experiment examining the impact of humans on the fauna and flora of rocky shores in southern Australia (Keough & Quinn 1998), and the full analysis is in Box 11-1. In this experiment, we were interested in disturbances caused by pedestrians, and whether a pattern of summer disturbance and autumn-winter-spring recovery results in a series of small disturbance-recovery cycles, or whether repeated disturbances eventually cause a major impact. We manipulated disturbance by trampling on marked intertidal areas each summer, using four different disturbance levels, which were the number of pedestrian passages. To determine the variability in results, we did the experiment on three different rock platforms, separated by 100s to 1000s of metres. On a smaller scale, at each site, we had two experimental plots, separated by 10s of metres, and each plot contained eight experimental strips. This arrangement corresponds to a nested design, with sites (i.e. platforms), plots within sites, and strips within each plot.

Within each plot, the eight strips were allocated randomly to one of the four trampling levels, with two replicates of each trampling level. With the same disturbance levels applied to all plots and sites, the factor trampling is orthogonal to sites & plots. The data used in this example are from a census of the number of limpets in each strip after three years of trampling.

11.1.2. Repeated measures designs

A simple repeated measures design, where the responses of a number of experimental units (or subjects) are recorded for a number of trials (or times), was discussed in detail in Chapter 10 and was also termed a subject by trials design. A modification of this design is a groups by trials design where the basic repeated measures design is modified to include a treatment structure between subjects, i.e., the subjects are randomly allocated to treatment groups in addition to their responses being recorded on a number of trials or times. Just as the linear model used for a subjects by trials repeated measures designs was the same as that used for a RCB design (an unreplicated two factor ANOVA model), groups by trials repeated measures designs can be analyzed in the same way as classical split-plot designs (with a partly nested ANOVA model). The term “plot” is replaced by “subject”, and we simply have “between subjects” and “within subjects” effects in the same way as we had between and within plot effects. In biology and ecology, the “subjects” are experimental or sampling units (animals, plants, quadrats etc.) and the trials are usually sequential times (von Ende 1993).

The term “repeated measures” has actually been used in a confusing manner in the literature. It really refers to repeated observations made on individual units (e.g. subjects, plots), either sequentially through time or under some treatment structure that is applied sequentially throughout time. Repeated measurements on experimental units can occur in any type of design. For example, a randomized block or split-plot design can have repeated measurements on each experimental unit within each block or plot (Gumpertz & Brownie 1993). The linear model used for repeated measures and split-plot designs are identical. The only complications are in the way the data are coded for computer analysis and which assumptions are applicable.

As an example of a group by trials repeated measures design, Schwartz *et al.* (1995) studied the effects of four temperatures (10°, 20°, 30°, 40°) on the dark respiration rate of five species of tree (four species of *Torrea* and one species of *Taxus*). Assume that it was desirable to have around five replicates to compare the five tree species, there are not large numbers of plants available, and individual plants were also likely to have different temperature profiles (leading to possibly reduced power). What are the design options for this experiment?

- Five replicate plants per cell, by four temperatures by five species the experiment would require 100 plants. We would analyze this experiment with a CR two-factor design (factors: species and temperature).
- One temperature profile per plant, so each plant would be used four times for the four temperatures, and only 20 plants are required (five for each species).

The second is a sensible option, to reduce the number of plants used and cut costs (and, if the experiment required sacrificing animals, reducing the number of animals killed). If we choose this option, we don't have a set of independent measurements for each temperature, but a group of five at one temperature, then another group of five for the same set of plants at the next temperature, and so on. Our analysis therefore needs to maintain the relationships between the measurements. Schwartz *et al.* (1995) used this repeated measures design with five or six plants of each species and each plant was subjected to the four temperatures. Individual plants were subjects, species was the between subjects factor and temperature was the within subjects factor. You can see the similarity to the diagrammatic representation of the pond experiment (Wissinger *et al.* 1996): the "plots" are individual plants (in repeated measures designs, these are termed "subjects"), the hydroperiod treatment corresponds to species, and the density treatments correspond to the temperatures. This experiment has one "between subjects" factor (species) and one "within subjects" factor (temperature).

In this example, the within subjects factor is a series of treatments (temperatures) applied sequentially through time. Repeated measures designs are often used when the within subjects factor does not represent different treatments but just a time sequence of interest. For example, Gange (1995) measured aphid abundance on twenty individual trees of two species of alder on twenty consecutive dates between May and September. The response variable was aphid abundance, the between subjects factor was tree species and the within subjects factor was date.

We will illustrate the analysis of groups by trials repeated measures designs with an example from a postgraduate project on physiology of amphibians.

Responses of cane toads to hypoxia

Mullens (1993) investigated the ways that cane toads (*Bufo marinus*) respond to conditions of hypoxia. Toads, the subjects, show two different kinds of breathing patterns, lung or buccal, and this breathing pattern was the between subjects factor. The second factor was O₂ concentration, which had eight levels (0, 5, 10, 15, 20, 30, 40, 50%), and was applied within subjects (toads). Various aspects of breathing rate were measured in each trial. The full analysis of this example is in Box 11-2.

11.1.3. Reasons for using these designs

The examples above demonstrate the two major reasons for using split-plot or repeated measures designs. First, if our experimental or sampling units (organisms, ponds, sites) are expensive or otherwise difficult to obtain, we might consider applying a number of treatments to each (or to subunits within each) or recording each through time. Second, if we expect lots of variation between these units, and are worried that this variation might obscure effects of our treatments, we can attempt to remove this variation by taking a biological "unit" and applying different treatments to it - sampling different parts of the same pond, applying a range of oxygen concentrations, etc. The basic difference between split-plot and a group by trials repeated measures design is that the former allocates the within plots treatments to subunits within each plot whereas the latter allocates the within subjects treatments sequentially to each subject.

11.2. Analyzing partly nested designs

We will first describe the analyses for a standard partly nested design that has three factors. One (the plots or subjects) is nested within the second, but both of these factors are crossed (orthogonal, factorial) with the third. In the split-plot example from Wissinger *et al.* (1996), we have hydroperiod (A), ponds within hydroperiods [B(A)], and density treatment (C). In the repeated measures example from Schwartz *et al.* (1995), we have species (A), plants within each species [B(A)], and temperature (C). In both examples, we have every combination of A and C (hydroperiod and density or species and temperature), so A and C form a factorial design. B and C are also factorial because every pond gets all density treatments and every plant gets all temperature treatments.

We could also have replicate observations from each combination of B and C (run each plant twice at each temperature, have replicate cages for each density treatment within each pond, etc.). As we will see below, in most cases, it makes little difference to how we test the effects of A and C.

We will describe the linear model and the various forms of analysis using split-plot terminology; keep in mind, however, that the plots are simply replaced by subjects in repeated measures designs. Components for fixed and random factors in expected mean squares are represented as “variances”; remember the different interpretations of variation between means of fixed treatment groups versus variance across all possible groups from which we have selected a subset at random – see Box 9-8.

11.2.1. Linear models for partly nested analyses

Linear effects model

Consider a design with p levels of factor A ($i = 1$ to p), q levels of factor B (plots or subjects) nested within each level of A ($j = 1$ to q) and r levels of factor C ($k = 1$ to r), crossed with both A and B (Table 11-1). From Keough & Quinn (1998), p equals three (the number of sites), q equals two (the number of plots) and r equals four (trampling treatments). From Mullens (1993), p equals two (the different breathing types), q equals eight or 13 (the number of toads of each breathing type) and r equals eight (O_2 levels). For completeness, we will describe the model with replicate observations ($l = 1$ to n) within each combination of A, B and C, e.g. Keough & Quinn (1998) had replicate observations (strips) within each site, plot and treatment combination. Usually, however, there is only a single observation (e.g. a single toad) of each level of C in each plot/subject.

The formal linear model for a split-plot design is:

$$y_{ijkl} = \mathbf{m} + \mathbf{a}_i + \mathbf{b}_{j(i)} + \mathbf{g}_k + \mathbf{ag}_{ik} + \mathbf{bg}_{j(i)k} + \mathbf{e}_{ijkl} \quad (11.1)$$

Details of this linear model, including estimation of its parameters, are provided in Box 11-3.

From Keough & Quinn (1998):

$$(\text{number of limpets})_{ijkl} = \mathbf{m} + (\text{site})_i + (\text{plot within site})_{j(i)} + (\text{trampling})_k + (\text{interaction between site and trampling})_{ik} + (\text{interaction between plot within site and trampling})_{j(i)k} + \mathbf{e}_{ijkl} \quad (11.2)$$

From Mullens (1993):

$$(\text{breathing rate})_{ijkl} = \mathbf{m} + (\text{breathing type})_i + (\text{toad within breathing type})_{j(i)} + (O_2 \text{ level})_k + (\text{interaction between breathing type and } O_2 \text{ level})_{ik} + (\text{interaction between toad within breathing type and } O_2 \text{ level})_{j(i)k} + \mathbf{e}_{ijkl} \quad (11.3)$$

In models 11.1 and 11.2:

y_{ijkl} is the number of limpets in the l th strip in the k th level of trampling treatment for the j th plot at the i th site. Commonly in these designs, n equals one, although Keough & Quinn (1998) had n equals two.

m is the overall (constant) population mean number of limpets per strip for all levels of trampling in all plots across all sites.

Factor A is fixed, so a_i is the effect of i th site on the number of limpets per strip, pooling over plots and trampling treatment. If factor A was random, e.g. sites were chosen at random along the shore, then a_i is a random variable with a mean of zero and a variance of S_a^2 , measuring the variance in mean number of limpets across all possible sites that could have been used.

Plots are random so $b_{j(i)}$ is a random variable with a mean of zero and a variance of S_b^2 , measuring the variance in mean number of limpets across all possible plots that could have been used within any site, pooling trampling treatments.

Factor C is fixed, so g_k is the effect of the k th level of trampling treatment, pooling over plots and sites. If factor C was random, e.g. trampling levels were chosen at random from the possible trampling levels that could have been used, then g_k is a random variable with a mean of zero and a variance of S_g^2 , measuring the variance in mean number of limpets of across possible levels of trampling that could have been used.

Factors A and C are both fixed, so ag_{ik} is the effect of the interaction between the i th site and k th trampling treatment. This interaction measures whether the effect of trampling is consistent at all sites used. If one factor is random, e.g. random sites, then ag_{ik} is a random variable with a mean of zero and a variance of S_{ag}^2 , measuring the variance of all the possible interaction terms between all possible sites and the fixed trampling levels. This interaction would measure whether the effect of trampling was consistent across all possible sites.

Because plots are random, the interaction between trampling treatment and plots [$bg_{(i)k}$] is a random variable with a mean of zero and a variance of S_{bg}^2 , measuring the variance of all the possible interaction terms between all the possible plots within any site and the fixed trampling treatments. This measures the variation in effects of trampling at the spatial scale of plots.

e_{ijkl} is random or unexplained error associated with the l th strip in the k th trampling treatment for the j th plot at the i th site. This is the error associated with each strip that is not due to trampling treatment, plot or site. Note that classical split-plot and repeated measures designs usually do not have replication for each combination of plot and factor C (n equals one) so e_{ijkl} usually cannot be separately estimated from S_{bg}^2 .

Predicted values and residuals

If we replace the parameters in our model by their OLS estimates (Box 11-3), it turns out that the predicted or fitted values of the response variable from our linear model 11.1 are:

$$\hat{y}_{ijkl} = \bar{y}_{ijk} \tag{11.4}$$

The error terms from model 11.1 can be estimated by the residuals, the difference between each observed and predicted value. In most applications of split-plot and groups by trials repeated measures designs, there is only a single observation per cell and these residuals all equal zero. In these circumstances, we cannot directly estimate S_e^2 , the variance of the error terms, unless we assume that S_{bg}^2 , the variance due to the B(A) x C interaction, equals zero.

Note being able to estimate \mathbf{S}_e^2 does not, however, compromise our tests of the main hypotheses of interest, those of A, C and A x C, the argument being similar to that used for analyses of RCB and simple repeated measures designs in Chapter 10 (see Section 11.2.3).

11.2.2. Analysis of variance

The partitioning of variance from the OLS fit of the linear model 11.1 is shown in Table 11-2. We do not provide computational details for sums-of-squares (SS) for each term in this ANOVA – see Winer *et al.* (1991) and Kirk (1996) for the classical formulae. In practice, however, we assume that you have access to statistical software with a general linear modeling routine when dealing with these complex designs. The SS for each source of variation are calculated by comparing the fit of the full model with the fit of an appropriate reduced model (the model including all terms except the one we wish to test in out H_0), as we described in Chapters 8, 9 and 10 for simpler ANOVA models. The general expected values of the mean squares are also provided in Table 11-2a, as well as those for the usual case of factors A and C being fixed and factor B (plots or subjects) being random.

This ANOVA is more complicated than those from previous chapters but not really that difficult - let's look at the different components. The between plots/subjects section is just a single factor ANOVA on the mean values for each plot/subject (*i.e.* averaging over the levels of Factor C) and the plots/subjects within A (*i.e.* factor B) term is the equivalent of the residual term in this single factor ANOVA. The within plots/subjects section is just a number of RCB (or simple repeated measures) designs, one for each level of A. The effects of factor C and the A x C interaction are interpreted in the same way as for a two factor crossed ANOVA (Chapter 9). The C x plots within A [*i.e.* C x B(A)] term represents the pooled error terms from the p randomized block designs which comprise the between plots/subjects component of the analysis, *i.e.* for each level of A, we have a C by plots RCB design (Kirk 1995).

These ANOVA tables are illustrated in Table 11-2b for some of the examples we described in Section 11.1. The first is the “split-plot” design from Wissinger *et al.* (1996) where the between plots factor was hydroperiod, the within plots factor was density treatment and the plots were ponds nested within each hydroperiod. The second was the groups by trials repeated measures design from Gange (1995) where the “between subjects” factor was tree type, the “within subjects” factor was date, the subjects were individual trees and the response variable was aphid abundance. The ANOVA tables for our two worked examples are also provided in Box 11-1Box 11-2.

11.2.3. Null hypotheses

There are three null hypotheses of primary interest when we fit the partly nested model 11.1.

Factor A (fixed)

$H_0: \mathbf{m}_1 = \mathbf{m}_2 = \dots = \mathbf{m}_k$. This H_0 states that there is no difference between the factor A marginal means, pooling levels of factor C. For example, no difference in the mean number of limpets per strip between sites, pooling the trampling treatments.

This is equivalent to:

$H_0: \mathbf{a}_1 = \mathbf{a}_2 = \dots = \mathbf{a}_i = 0$, *i.e.* no effect of any level of factor A. For example, no effect of site on the mean number of limpets per strip, pooling the trampling treatments.

Factor C (fixed)

$H_0: \mathbf{m}_1 = \mathbf{m}_2 = \dots = \mathbf{m}_k$. This H_0 states that there is no difference between the factor C marginal means, pooling levels of factor A. For example, no difference in the mean number of limpets per strip between trampling treatments, pooling sites.

This is equivalent to:

$H_0: \mathbf{g}_1 = \mathbf{g}_2 = \dots = \mathbf{g}_c = 0$, i.e. no effect of any level of factor C. For example, no effect of trampling treatment on the mean number of limpets per strip, pooling sites.

A x C interaction (fixed)

$H_0: \mathbf{m}_{jk} - \mathbf{m}_j - \mathbf{m}_k + \mathbf{m} = 0$, which is the same as $(\mathbf{ag})_{ik} = 0$. This H_0 states that there are no interactions between A and C, e.g. the effect of site on the mean number of limpets per strip is the same for all trampling treatments and the effect of trampling treatment on the mean number of limpets per strip is the same for all sites.

The modifications of these H_0 s for random factors are straightforward as described in Chapters 9 and 10.

Two other null hypotheses might also be tested in some circumstances

Factor B(A)

$H_0: \mathbf{S}_b^2 = 0$, i.e. the variance in mean values of the response variable across all possible plots or subjects that could have been used within any level of A equals zero. For example, there is no variance in the mean number of limpets per strip between plots in either site, pooling trampling treatments.

Factor B(A) x C

$H_0: \mathbf{S}_{bg}^2 = 0$, i.e. the variance across all the possible interaction terms between all the possible plots/subjects within any level of A and the fixed levels (if C is fixed) or all possible levels (if C is random) of factor C equals zero. For example, no variance mean number of limpets per strip across all the possible interaction terms between plots within each site and trampling treatment.

F-ratios

The F -ratios for testing these null hypotheses are based on the expected values of the relevant mean squares (Table 11-2a). When factors A and C are both fixed, the F test for factor A uses a different denominator [$MS_{B(A)}$] than those for factor C and A x C [$MS_{CB(A)}$]. This is typical for designs with both fixed and random factors and is apparent in all these partly nested designs because the plots/subjects term is nearly always random. In the classical split-plot or repeated measures design with n equals one observation for each cell, the B(A) and C x B(A) terms cannot be tested. The implications of not being able to test the C x B(A) are analogous to the implications of having no test for a block by treatment interaction in a RCB design (Chapter 10). This makes sense given that the C x B(A) interaction comprises the pooled residual terms from the p RCB designs and each of these residual terms includes the plot/subject by treatment interaction. The first implication is that if B is random, then a strong C x B(A) interaction will reduce the power of the tests of C and A x C, although these tests are still valid because the expected mean squares of both C and A x C include the variance component due to C x B(A). This is not the case if B is fixed, where C x B(A) is an inappropriate error term for testing C and A x C (see below). The second implication is that the use of C x B(A) as an error term for C and A x C can be invalid if the observations within each plot/subject are correlated, which is almost certainly the case for repeated measures designs. For the C x B(A) to be used as an error term, we must make certain assumptions about the covariances of the observations (Section 11.3).

With replicate observations in each cell, the B(A) and C x B(A) terms can be tested against the Residual. Note, however, that using many replicate observations within each cell, e.g. multiple measurements on toads or multiple strips within plots, may not be providing a much better test of the terms that you really care about. The B(A) term is rarely of much interest,

and you probably don't care much if factor C has a different effect across levels of B, i.e. a C x B(A) interaction. The effort expended in sampling at this lowest level may not be producing a more powerful statistical test of any of the biologically interesting effects, only increasing the cost of the design in terms of time and/or money. When there is no replication within plots, Underwood (1997) argued that tests of the main effects of C and the A x C interaction can only be done if we assume there is no C by B(A) interaction, i.e., the effects of C do not vary from plot to plot. However, it is clear from Table 11-2a that the expected mean squares for C and the A x C interaction include the variance due to C x B(A) interaction, so the *F* test for C and A x C is testing for these effects over and above the variation due to the C x B(A) interaction and any residual variation. Therefore, we can interpret a significant effect of C or the A x C interaction even if the effects of C do vary for different levels B(A), which is similar to the argument we made in Chapter 10 for RCB designs when the blocks factor was random. A non-significant C or A x C interaction is more difficult to interpret in the presence of a C x B(A) interaction, but interpreting non-significant tests is always problematical.

As you can see from Table 11-2a, the effects of A, C and the A x C interaction are all tested against other terms in the analysis, all featuring effects of B within A. Because you cannot control the number of levels of A or C when they are fixed factors, the only way to increase the power of these tests is to increase the degrees of freedom. This can only be achieved by using more levels of B (more plots or subjects, e.g. more toads, more plots, etc.), i.e., increasing *q*.

The expected mean squares and appropriate *F* tests for other combinations of fixed and random factors are presented in Table 11-3. When A, C and plots/subjects (i.e. B) are all fixed (Table 11-3), you can see that all terms are tested using the MS_{Residual} as the denominator. Note that you must have replicate observations in each cell if plots (B) are fixed because $MS_{\text{CB(A)}}$ is not an appropriate denominator here unless you are very sure that the C x B(A) interaction is negligible. In almost every case that we deal with, factor B will be random so this design is unlikely for biological experiments. If factors A and plots/subjects are random, but C is fixed, the tests are straightforward (Table 11-3), but note that again, they use different combinations of denominators for *F* tests for the various hypotheses. Problems occur when plots/subjects and factor C are random (Table 11-3). It does not matter whether A is fixed or random in this case. The difficulty is that the main effect of factor A (which will almost always be of central interest) cannot be tested directly, because there is no appropriate denominator. The only option is to use quasi *F*-ratios, which are combinations of mean squares that produce an approximate test of your hypothesis (see Chapter 9). Winer *et al.* (1991) discussed this option in detail, but you should be aware that the resulting *F* tests are only approximate and not necessarily robust, and you probably should avoid this situation if possible.

11.2.4. Comparing ANOVA models

The SS, df and MS for each term in the partly nested model 11.1 can be derived from comparing the fit of a full and a reduced model, where the reduced model omits the parameter specified to be zero in the H_0 . This is the same principle we have described in Chapters 9 and 10 for other multifactor models.

11.3. Assumptions

Irrespective of whether it is for a split-plot design or a groups by trials repeated measures design, the partly nested ANOVA model 11.1 has a number of assumptions that need to be assessed. Additionally, we should always check for outliers from our fitted model. A useful first step is to examine the residuals from the fit of the model. If we only have *n* equals one within each combination of A, B and C, then we should omit the B(A) x C term, otherwise the model is saturated (a perfect fit) and all the residuals are zero. These residuals will indicate

any obvious outliers and also indicate any strong skewness in the data. Generally, however, the assumptions, and their assessment, in these analyses are considered separately for the between plots/subjects and within plots/subjects components.

11.3.1. Between plots/subjects

The test of factor A assumes normality and homogeneity of variance and the comments about these assumptions in Chapters 8 and 9 apply here. Note that, for the usual case of B random and A and C fixed, these assumptions apply to the levels of A (pooled across C) with the mean of Y in each level of B(A) as a replicate observation. It is often useful to create a new variable that is the average across the levels of C and then use that variable in boxplots for each level of A or to examine residuals from the fit of a single factor ANOVA model with p groups to that variable.

11.3.2. Within plots/subjects and multisample sphericity

The tests for any terms including within plots/subjects factors(s), *i.e.* tests of C and A x C, have the assumption of sphericity of variances and covariance, as did RCB and simple repeated measures designs (Chapter 9). Unless this assumption holds, then the B(A) x C term is an inappropriate denominator for the test of C and A x C. Remember that these partly nested designs can be envisaged as a series of RCB (factor C by blocks, plots or subjects) experiments, one for each level of factor A, so the assumption is now multisample sphericity. Not only must the variance-covariance matrices be the same for each level of factor A, they must each show sphericity, which means that the variances of the differences between the levels of the repeated factor must be the same.

In classical repeated measures designs, the levels of the within subjects factor (C) can usually be applied in random order to each subject (Winer *et al.* 1991). Similarly, in classical split-plot designs, the levels of the within plot factor (C) should be randomly allocated to experimental units (subplots) within each plot. Under these randomization conditions, there is no reason for the sphericity assumption not to hold; in fact, the sphericity assumption is often not discussed when general statistics texts describe split-plot designs. In contrast, the sphericity assumption is unlikely to hold in repeated measures designs when subjects are recorded through time because the differences between times closer together are likely to be less variable (*i.e.* more similar) than times further apart. If sphericity is not met, the F -ratio statistics for within subjects effects (C and A x C) will be inflated, increasing the risk of a Type I error above the nominal level (e.g. 0.05) – see Keselman & Keselman (1993), Keselman *et al.* (1995) and Rasmussen (1989). There is no easy test for the null hypothesis that the variance-covariance matrices conform to multisample sphericity. Kirk (1995) recommended the W test and provided critical values, although we suggest it is safer to assume that multisample sphericity is not met in repeated measures type designs and use one or more of the following analytical strategies.

Adjusted univariate F-ratio tests

As described in Chapter 10 for RCB and simple repeated measures designs, we can make the F tests more conservative using adjusted df. An index of sphericity is the population parameter ϵ , which can be estimated by the epsilon statistic ($\hat{\epsilon}$). Two methods of estimating $\hat{\epsilon}$ were described in Chapter 10, the Greenhouse-Geisser (G-G) estimate or the Huynh-Feldt (H-F) estimate (Winer *et al.* 1991, Yandell 1997). These sample $\hat{\epsilon}$ s can be used to adjust the df for within plots/subjects tests downwards to make the tests more conservative, since non-sphericity increases the risk of Type I error. The adjustment is simple, being the original df multiplied by $\hat{\epsilon}$, although the new df will not be integers (Table 11-4). If ϵ is greater than 0.75, the correction based on the Huynh-Feldt $\hat{\epsilon}$ is better, when ϵ is less than 0.75, the correction based on the Greenhouse-Geisser $\hat{\epsilon}$ is better (Keselman & Keselman 1993). These adjusted tests are standard output from most statistical software.

Multivariate tests

An alternative solution to the sphericity assumption is to treat the levels of the within subjects or plots factor (*i.e.* the repeated measures factor) as multiple response variables in a multivariate analysis of variance (MANOVA in Chapter 16; see also Keselman & Keselman 1993, Looney & Stanley 1989, Kirk 1995, Tabachnick & Fidell 1996). The MANOVA actually uses the difference between successive repeated measurements (*i.e.* times) for each subject or plot as response variables and tests the null hypothesis that the difference scores have a population centroid (multivariate mean) equal to zero. The MANOVA approach can be useful for these designs because it doesn't assume sphericity of variances and covariances, although it does have all the usual MANOVA assumptions (Chapter 16; Johnson & Field 1993) and has fewer degrees of freedom. Also, if the n is less than the number of differences between successive repeated measurements (*i.e.* less than the number of levels of the within plots or subjects factor minus one), then the MANOVA approach cannot be used. As discussed in Chapter 16, the Pillai trace statistic is recommended for these multivariate tests.

Profile analysis

Another approach is to summarize the responses for each plot/subject as a single value and then use these values in a single factor ANOVA model comparing the levels of A. The between plots/subjects part of the partly nested univariate ANOVA does this by summarizing the responses of each plot/subject as an average across the levels of C. If factor C is quantitative, *e.g.* time, we can also summarize the responses of each plot/subject as a trend or response curve, such as a linear, quadratic etc., and analyze the coefficients of these trends in separate one factor ANOVAs (Meredith & Stehman 1991). This provides a test of whether such trends (linear, quadratic etc.) vary across factor A, *i.e.* a test of a treatment-contrast interaction (Chapter 9). Such tests are usually default output from statistical software and will be discussed in Section 11.5.2.

Which strategy is the best?

As we pointed out in Chapter 10 for RCB and simple repeated measures designs, neither the epsilon-adjusted univariate nor the multivariate approach is always more powerful, unless sphericity is met, when the traditional partly nested univariate analysis is clearly preferred. Looney & Stanley (1989) recommended using both approaches and rejecting the within subjects null hypotheses if either the adjusted univariate or multivariate tests are significant. Kirk (1995) recommended a preliminary test for multisample sphericity and using the adjusted univariate tests if the sphericity test is significant; however, his preliminary test is not straightforward and not available in statistical software. We suggest that preliminary tests for sphericity are of limited value and support the Looney & Stanley (1989) approach and the use of profile analyses if factor C is quantitative.

11.4. Robust partly nested analyses

As for other linear model analyses, the RT (rank transform) procedure has been proposed as a general method for overcoming problems of non-normality and possibly other assumptions of the partly nested analyses of variance (see discussion in Thompson 1991b). We reiterate our comments from Chapters 9 and 10. The rank transformation is nonlinear in nature (Akritas 1991) and therefore cannot effectively deal with interactions; indeed, a significant main effect may be indicated when it is simply due to an undetected interaction (Thompson 1991b). As the A x C interaction is often of considerable interest in the designs discussed in this chapter, the RT procedure seems inappropriate. RT procedures are also inappropriate for nested factors (Thompson 1991b), which are important in the models used to analyze split-plot and groups by trials repeated measures designs. Also, as discussed in the Chapter 10 for analyses of RCB designs, a rank transformation can also change the nature of variances and covariances, making the assumption of sphericity less tenable (Akritas 1991). Although

Thompson (1991b) has developed a general rank-based multivariate test statistic that is applicable to repeated measures designs, its usefulness is restricted to situations where there are no interactions.

We could also fit the models in this chapter using generalized linear models (GLMs), that allow a range of different error distributions of which normal is just one (Chapter 13). Maximum likelihood techniques are used for fitting the models and estimating the parameters and likelihood ratios are used for hypothesis tests of these parameters. Care must be taken in the choice of full and reduced models for such complex analyses because some models won't make much biological or statistical sense, e.g. a model that includes B(A) but not A. Note that GLM analyses are still sensitive to the specification of the error distribution so model diagnostics are very important, just as they are for linear models. Chapter 13 includes a more detailed discussion of GLMs.

11.5. Specific comparisons

11.5.1. Main effects

Planned contrasts for between the plots/subjects main effect are done in the same way as described in Chapter 8 and simply average across the within plots/subjects factor levels for each experimental unit. Planned contrasts for the within plots/subjects main effect assume multisample sphericity if the usual B(A) x C term is to be used as the denominator. The two alternatives are to adjust the df for these contrasts using the G-G or H-F estimates of ϵ (Section 0) or use separate error terms, e.g. $C_{\text{contrast}} \times B(A)$, for each contrast (Kirk 1995). These error terms are calculated similarly to those for analyses RCB (or simple repeated measures) designs described in Chapter 10, except that the contrasts are calculated across the levels of factor A; Kirk (1995) provides computational details but good statistical software will calculate these separate error terms. They basically represent a separate F -ratio testing for differences in the levels of C within each level of A. Keselman & Keselman (1993) suggested an approximate paired t test with separate error terms based on the two groups being compared, called the KKS test, similar to that described in Chapter 10, although Satterthwaite's adjusted df are used.

Unplanned comparisons for between plots/subjects factors are done in the same way as described in Chapter 8, and simply average across the within plots/subjects factor levels for each experimental unit or subject. The usual unplanned multiple comparison procedures may not be reliable for within plot/subjects factors because the means are probably correlated to some extent, particularly for repeated measures designs. Keselman & Keselman (1993) described some new stepwise multiple comparison procedures for within subjects/plots factors. The simplest approach might be to contrast the specific levels of C applying a Bonferroni-type adjustment of significance levels for multiple testing if required (Chapter 3). Note that these contrasts between levels of C will use the B(A) x C term as the denominator and therefore assume multisample sphericity; adjusted df should be used based on G-G or H-F estimates of ϵ .

11.5.2. Interactions

In partly nested ANOVA models, the main interaction of interest is between A and C and represents an interaction between a between plots/subjects factor and within plots/subjects factor. This interaction can be explored with "interaction" plots of means, where we might have the levels of factor C along the horizontal axis, the response variable along the vertical axis and each point represents the mean of factor A levels across plots/subjects within each A level. Deviations from parallel lines indicate some interaction between A and C.

Tests of simple main effects can also be done as described in Chapter 9, the only difficulty for the designs in this chapter is choosing the appropriate denominator for the F tests (Kirk 1995, Maxwell & Delaney 1990). In Chapter 9, we pointed out that for a two factor crossed (A, B,

A x B) linear model, the SS for simple main effect tests for factor A represent partitioning of the SS_A and SS_{AB} , whereas the simple main effects tests for B represent partitioning of the SS_B and SS_{AB} . In contrast to the two factor completely randomized design, however, the test of the A term in a partly nested model with A and C fixed and B (plots/subjects) random uses a different denominator than the tests of the C and A x C interaction terms. So what denominators do we use for the simple main effects tests in a partly nested model?

The simple effects tests for C at each level of A separately, e.g. the effect of O₂ level for each breathing type separately in the Mullens (1993) example, are relatively straightforward because both C (O₂ level) and A x C (breathing type x O₂ level) use the same denominator – C x B(A). Note that if multisample sphericity does not hold, then these tests should be based on adjusted degrees of freedom using the G-G or H-F estimates of ϵ (Section 0). Alternatively, separate denominators should be used for each simple effect, the equivalent to calculating a simple repeated measures ANOVA testing C within each level of A separately (Chapter 10).

For the simple effects tests for A at each level of C separately, e.g. the effect of breathing type for each O₂ level separately, Kirk (1995) and Maxwell & Delaney (1990) recommended using a denominator that represents the average of the B(A) and B(A) x C terms. This is sometimes called the within cells error term:

$$\frac{SS_{B(A)} + SS_{B(A) \times C}}{p(q-1) + p(q-1)(r-1)} \quad (11.5)$$

Tests using the error term in 11.5 might be biased, especially if the two terms contributing to the pooled term are very different.

11.5.3. Profile (*i.e.* trend) analysis

A useful approach, which can be used in conjunction with any experimental design where at least one factor is quantitative, is to look for trends across levels of the quantitative factor (Chapter 8). For designs in this chapter, the common approach is to test for trends across the levels of factor C (the within plots/subjects factor) if C is quantitative (e.g. time, O₂ level). The simplest trends to examine are those of a polynomial form, such as linear, quadratic, cubic etc. (see Chapter 8). Tabachnick & Fidell (1996) provide an excellent description of these methods for repeated measures designs, and therefore for partly nested models in general.

The number of polynomial contrasts that can be calculated is one less than the number of levels of the relevant factor. For example, if there were six levels of factor C, you could test for linear (X), quadratic (X^2), cubic (X^3), quartic (X^4) and quintic (X^5) polynomials, although it is often difficult to attach biological meaning to trends more complex than cubic. It is important to remember that these trend tests depend on the metric (spacing) of levels of the quantitative factor(s), as discussed in Chapter 8, and that most statistical software assumes equal spacing by default. The tests of these polynomials are statistically orthogonal (independent) of each other because each is tested using a separate component of df_C and MS_C (or A x C if trends are tested as part of the interaction), with separate components of the $df_{CB(A)}$ and $MS_{CB(A)}$ for the denominators of each trend F test.

Testing for trends as part of analyses of classical repeated measures designs is often termed profile analysis (Tabachnick & Fidell 1996). As with tests of trends in completely randomized factorial designs discussed in Chapter 9, there are two types of tests of interest in profile analysis:

- Main effect trends, usually across the within plots/subjects factor C pooling the levels of factor A. For example, is there a linear trend in breathing rate of toads across oxygen concentrations, pooling the two breathing types? Is there a quadratic trend? A cubic

trend? Trends could also be examined across factor A as part of the between plots/subjects part of the analysis.

- Treatment-contrast interactions for examining the A x C interaction term. Here we compare trends of the same form (e.g. linear) across C between different levels of factor A. For example, is the linear trend in breathing rate of toads across oxygen concentrations the same for the two breathing types? Is the quadratic trend the same? These tests are often described as tests of parallelism (Tabachnick & Fidell 1996), since testing whether the linear trends are the same across level of A is clearly a test of whether the trends are parallel.

We do not provide computational details for calculating these trend tests because there is nothing additional to the information we included in Chapter 9 and these trend tests are usually default output from statistical software if the data are coded, and analysis run, as a classical repeated measures design. Note that some software will automatically test each trend MS against a separate error term so that multisample sphericity is not assumed. Alternatively, the B(A) x C term could be used, with adjustments to the df based on the G-G or H-F estimates of ϵ . Growth curve analysis can also be useful for ecophysiological studies and involves comparisons of non-linear regressions of a more complicated form than simple polynomials (Potvin *et al.* 1990).

An example of these trend analyses was provided by Sharpe & Keough (1998), who examined temporal trends in chlorophyll-*a* and in the density of herbivorous snails following the removal of dominant grazers from the intertidal zone of a rocky shore. The removal treatments were the between subjects/plots factor, and time was the repeated factor. Individual boulders were the plots/subjects, so different boulders received different removal treatments. They recorded chlorophyll-*a* from randomly selected areas on each boulder, and censused a range of herbivores once a month. They contrasted the linear temporal trend in abundance of each species (as a measure of recolonization rate) between particular combinations of treatments. We also illustrate these trend analyses in the worked examples in Box 11-2 & Box 11-4.

11.6. Analysis of unbalanced partly nested designs

Unequal sample sizes can arise in partly nested (split plot or repeated measures) designs in two ways. First, the number of plots or subjects in each level of the between plots factor might vary. Since the between plots tests average over the within plots factors, this type of unequal sample size is no different to unequal sample sizes in the usual factorial ANOVAs described in Chapter 9, and our recommendations are the same. Remember that checking assumptions becomes much more important when sample sizes are unequal and that even tests of within plots/subjects factors can be more sensitive to assumptions (e.g. sphericity) when the between plots/subjects part of the design is unbalanced (Keselman & Keselman 1990). Second, when we have no replication within each cell (the classical split-plot or repeated measures design), then missing observations equate to missing cells. If you have a reasonable number plots/subjects, then a simple approach is to delete the plot(s) or subject(s) with the missing observations; this causes problems if sample sizes (number of plots or subjects) are small because the between subjects/plots part of the analysis may become severely unbalanced. Basically, most statistical software will use this approach by default if the data are set up for a classical repeated measures analysis. If you don't have many plots/subjects but lots of levels of the within subjects/plots factor(s), then it might be better to omit the level of factor C (or the combination of levels if you have more than one within subjects/plots factor) with the missing observations. This approach changes the null hypotheses being tested, of course, but if the hypotheses are general ones about trends through time and you have a long time sequence, then omitting one or two times may not have much effect.

An alternative solution is to simply fit the partly nested linear model (Berk 1987) and compare this full model with appropriate reduced models for specific hypotheses, as described in Chapter 10 for RCB designs. Unfortunately, the F tests are more sensitive to the sphericity assumption when observations are missing and most statistical software doesn't provide epsilon estimates, nor adjusted univariate tests, when the analysis is run this way, so be careful. As we recommended in Chapter 10 for RCB and simple repeated measures designs, a practical strategy may be to delete the subject(s)/plot(s) with the missing observation(s), running the analysis as a classical repeated measures design to check sphericity and then only fit a partly nested linear models to the data with all subjects/plots if that assumption is tenable. This is messy but there are not many practical options when dealing with missing observations in these designs.

More complicated solutions are provided by Berk (1987), who suggested ML and REML estimation procedures that weight the observations, by Kirk (1995), who described using the cell means model and testing a subset of hypotheses using contrasts, and by Rovine & Delaney (1990). All these methods will be difficult for practicing biologists, at least until they are standard components of statistical software.

11.7. Power for partly nested designs

As expected, power calculations become more complicated with these complex designs, with the possibility of separate power calculations for a series of main effects, and interactions. We can divide these tests into those involving only between plot/subject terms, only within plot/subject terms, and interactions between the two groups. For between subjects factors, power calculations are similar to those described for Chapters 8 & 9. They are routine when main effects are of interest, and they can be made easier by recoding the data file as means, averaging across the repeated or within-plots factor levels. For the more complex within-subjects/plots effects, the power calculations can be done, with two important steps. First, specifying an effect size can be very difficult, as for all complex interactions. Second, in computing power for a particular effect, we must identify the denominator used to test that effect, and use that MS to generate the variance estimate needed to calculate power.

One special case in which the power calculation is relatively simple is the family of BACI (Before-After-Control-Impact) designs used in environmental monitoring. The test for an environmental impact is an interaction between Before-After and Control-Impact, tested using, for example, changes Before-After at replicate locations within Control and Impact categories. In the original formulation of this design, with two locations (C & I), two periods (B & A), and multiple sampling times within each period, we could use a partly nested analysis, with Periods, Times within Periods, and samples at C & I at each time. An impact would be revealed as a change in the difference between C & I, from the Before to the After period. Stewart Oaten *et al.* (1985) pointed out that this design can be analyzed as a t test, simply by calculating the difference, C-I, and comparing that difference between Periods. As a consequence, rather than formulating an effect size based on the interaction, we can specify an effect size as the divergence or convergence of these C-I differences. More complex formulations of this design (e.g. Downes *et al.* 2001, Keough & Mapstone 1997) can also be simplified in this way, because the interaction of interest is between the main between and within plots factors, and each of them has only two levels.

11.8. More complex designs

So far we have considered partly nested designs involving one between subjects/plots factor (A), one within subjects/plots factor (C) and one factor representing subjects/plots (B). These experimental designs can be extended to include more than one between subjects/plots factor and/or more than one within subjects/plots factor.

11.8.1. Additional between plots/subjects factors

There is nothing difficult about additional between subjects/plots factors, because this part of the analysis is just an ANOVA on the average of the response variable for each plot/subject. For example, a four factor design might have two between subjects (plots) factors (A and C), one within subjects (plots) factor (D), and factor B representing plots nested within A and C. For example, McGoldrick & Mac Nally (1998) studied the impact of eucalypt flowering on the dynamics of bird communities in forests of southeastern Australia. They had eight sites (i.e. plots) arranged in a two factor crossed design with factor A being habitat (two levels: dominated by ironbark eucalypts vs dominated by stringybark eucalypts) and factor B being region (two levels: north of Great Dividing Range and south of Great Dividing Range) with two sites within each combination. Each site was censused monthly for twelve months, so month was the within plots/subjects factor. The response variables included flowering index, density of nectarivorous birds, species richness of nectarivorous birds etc. The analysis for this example is in Box 11-4, where we analyze the density of nectarivorous birds, transformed to logs after adding one to each observation to account for zero values.

The appropriate linear model for a split-plot or repeated measures design with two crossed between plots/subjects factors is:

$$y_{ijklm} = \mathbf{m} + \mathbf{a}_i + \mathbf{g}_k + \mathbf{ag}_{ik} + \mathbf{b}_{j(ik)} + \mathbf{d}_l + \mathbf{ad}_{il} + \mathbf{gd}_{kl} + \mathbf{agd}_{ikl} + \mathbf{bd}_{j(ik)l} + \mathbf{e}_{ijklm} \quad (11.6)$$

From McGoldrick & Mac Nally (1998):

(log density of nectarivorous birds plus one) $_{ijklm} = \mathbf{m} + (\text{habitat})_i + (\text{region})_k + (\text{interaction between habitat and region})_{ik} + (\text{site within habitat and region})_{j(ik)} + (\text{month})_l + (\text{interaction between habitat and month})_{il} + (\text{interaction between region and month})_{kl} + (\text{interaction between habitat, region and month})_{ikl} + (\text{interaction between site within habitat and region and month})_{j(ik)l} + \mathbf{e}_{ijklm} \quad (11.7)$

In models 11.6 and 11.7:

\mathbf{m} is the overall (constant) population mean log density of nectarivorous birds plus one

\mathbf{a}_i is the effect of i th level of the first between plots factor A (effect of habitat), pooling regions and months

\mathbf{g}_k is the effect of the k th level of the second between plots factor C (effect of region), pooling habitats and months

\mathbf{ag}_{ik} is the effect of the interaction between the i th level of A and k th level of C (interaction between habitat and region), pooling months

$\mathbf{b}_{j(ik)}$ is the effect of the j th plot (factor B, site) within the ik th combination A and C

\mathbf{d}_l is the effect of the l th level of factor the within plots factor D (effect of month)

\mathbf{ad}_{il} is the effect of the two way interaction between the i th level of A and the l th level of D (interaction between habitat and month)

\mathbf{gd}_{kl} is the effect of the two way interaction between the k th level of C and the l th level of D (interaction between region and month)

\mathbf{agd}_{ikl} is the effect of the three way interaction between the i th level of A, the k th level of C and the l th level of D (interaction between habitat and region and month)

$\mathbf{bd}_{j(ik)l}$ is the effect of the interaction between the j th plot (factor B) within the ik th combination A and C and the l th level of D [interaction between site (within habitat and region) and month]

e_{ijklm} is the error term. Note that e_{ijklm} cannot be estimated separately from $bd_{j(ik)l}$ in this model unless there is replication within each cell, which is unusual. By recording the same sites once at each time, McGoldrick & Mac Nally (1998) did not have replicates within each combination of habitat, region and month and so could not estimate e_{ijklm} .

The general expected mean squares are in Table 11-5, as well as those for the common case whereby A, C and D are fixed and B (plots or subjects) is random. The between plots/subjects terms are tested against $MS_{B(AC)}$ and the within plots/subjects terms are tested against $MS_{DB(AC)}$. Error terms for other combinations can be determined from the expected mean squares are provided in Table 11-5 following the rules in Box 9-8 – see also Kirk (1995) and Winer *et al.* (1991).

To further illustrate this design, consider the study of Morris (1996) who examined factors affecting the density of rodents in the Rocky Mountains of the USA. He had nine locations, with two habitats (xeric and mesic) at each location (*i.e.*, a 9 x 2 factorial design), with two replicate grids for each combination of location and habitat. Grids were thus the plots or subjects and location and habitat were the between plots/subjects factors. He sampled each grid at three times (early, mid, late summer), so sampling time was the within plots/subjects factor. The ANOVA for this study is in Table 11-6, illustrating the appropriate error terms for each effect in the model, based on all factors except grids (*i.e.* plots) being fixed.

A more complicated version of this design was used by Letourneau & Dyer (1998), who examined the effects of top predators (beetle larvae present or absent), soil type (nutrient rich or poor) and light level (high or low) on colony size of an ant species on seedlings planted in three replicate pots (*i.e.* plots) in a three factor crossed design. Each pot was recorded on five occasions over 18 months, with time as the within plots/subjects factor. The ANOVA for this study is in Table 11-7 with error terms based on all factors except plots being fixed.

A further modification of the between plots/subjects part of the design is where A (the between plots factor) and plots are arranged as a RCB design (Table 11-8). The appropriate linear model for this design is:

$$y_{ijkl} = m + a_i + b_j + ab_{ij} + g_k + ag_{ik} + bg_{jk} + abg_{ijk} + e_{ijkl} \quad (11.8)$$

In model 11.8:

m is the overall (constant) population mean

a_i is the effect of factor A

g_k is the effect of factor C

ag_{ik} is the interaction between factors A and C

b_j is the effect of plots/subjects (*i.e.* blocks)

ab_{ij} , bg_{jk} , and abg_{ijk} are the interactions between A, C, AC and plots/subjects

e_{ijkl} is the error effect. Note that e_{ijkl} cannot be estimated separately from abg_{ijk} in this model unless there is replication within each cell, which is unusual.

This is basically a three factor unreplicated ANOVA, identical to a factorial RCB design (Chapter 10). If A and C are fixed and B (plots/subjects or blocks) is random, then A is tested against A x B (plot), as in all RCB designs, C is tested against C x B and A x C is tested against A x B x C (Table 11-8). There are no tests for plot/subject (*i.e.* block) or its interactions with A and C, unless quasi *F*-ratios are used.

Aguiar & Sala (1996) used such an analysis in their investigation of seed movement in the Patagonia steppe. They had three sites recorded on three dates and they measured seed availability in four microsites (bare ground, grass, shrub, litter) in each site on each date. Site and date were between plots factors (although there was only one “plot” for each combination

of site and date) and microsite was a within-plot factor (Table 11-9). Although not stated in their paper, they treated site as a random block effect and assumed there was no site by date interaction since they tested the random site effect against this interaction term. A second example comes from Evans & England (1996) who looked at the effect of artificial honeydew on the numbers of adult weevil parasitoids on alfalfa plants. They had three treatments (early application of artificial honeydew followed by water, early application of water followed by artificial honeydew, two applications of water only), each allocated to one of three “subplots” in each of ten rows (plots or blocks). The numbers of parasitoids were recorded from each subplot on two separate dates about ten days apart. The ANOVA for this design is also in Table 11-9 and Evans & England (1996) fitted an additive model with no treatment x row interactions, allowing tests for row (i.e. plots) and row x date.

11.2.4. Additional within plots/subjects factors

Extra within plots/subjects factors can also be included in these designs, although this complicates the analysis because multiple denominators now must be used for the F tests for the within subjects/plots terms. With one between plots factor (A), two within plots factors (C and D) and plots (factor B) nested within A, the appropriate linear model is:

$$y_{ijklm} = m + a_i + b_{j(i)} + g_k + ag_{ik} + bg_{j(i)k} + d_l + ad_{il} + bd_{j(i)l} + gd_{kl} + agd_{ikl} + bgd_{j(i)kl} + e_{ijklm} \quad (11.8)$$

In model 11.8:

m is the overall (constant) population mean

a_i is the effect of i th level of factor A (the between plots factor)

$b_{j(i)}$ is the effect of the j th plot (factor B) within the i th level of factor A

g_k is the effect of the k th level of factor C (the first within plots factor)

ag_{ik} is the effect of the two way interaction between the i th level of A and k th level of C (i.e. A x C interaction)

$bg_{j(i)k}$ is the interaction between the k th level of C and the j th plot (B) within the i th level of A (C x B within A interaction)

d_l is the effect of the l th level of factor D (the second within plots factor)

ad_{il} is the effect of the two way interaction between the i th level of A and the l th level of D (A x D interaction)

$bd_{j(i)l}$ is the interaction between the l th level of D and the j th plot (B) within the i th level of A (D x B within A interaction)

gd_{kl} is the effect of the two way interaction between the k th level of C and the l th level of D (C x D interaction)

agd_{ikl} is the effect of the three way interaction between the i th level of A, the k th level of C and the l th level of D (A x C x D interaction)

$bgd_{j(i)kl}$ is the effect of the interaction between the k th level of C and the l th level of D and j th plot (B) within the i th level of A (C x D x B within A interaction)

e_{ijklm} is the error effect. Note that e_{ijklm} cannot be estimated separately from $bgd_{j(i)kl}$ in this model unless there is replication within each cell, which is unusual.

The general expected mean squares, and those when factors A, C and D are fixed and plots/subjects random, are provided in Table 11-10. Note that when A, C and D are fixed, C and A x C are tested against C x plots within A, D and A x D against D x plots within A and C x D and A x C x D against C x D x plots within A.

These designs are sometimes termed split-split-plot designs because we can have a main between plots factor and two within plots factors, one applied to sub-plots within each plot and one applied to sub-sub-plots within each sub-plot. More commonly, however, these experimental designs include a single within plots factor with repeated measurements through time or two within-subjects time factors. For example:

- Vasquez (1996) looked at the effect of illumination (two fixed levels: bright and dark) and seed distribution (two fixed levels: dispersed and clumped) on seed consumption in experimental arenas for three species of rodents. Species was the between subjects factor and there were approximately 17 individuals/subjects for each species. Each individual was tested under each illumination level and each seed distribution in a crossed arrangement (four combinations), so illumination and seed distribution were separate within subjects factors (Table 11-11).
- Green *et al.* (1997) studied the effects of land crabs on recruitment of rainforest seedlings on Christmas Island. They used two habitats (understory and gap) in the rainforest as the between plots factor with seven paired plots in understory habitat and three paired plots in gap habitat. These pairs were the “plots” or “subjects”. One plot (i.e. “subplot”) in each pair allowed access to crabs and one (sub)plot excluded crabs, so exclusion was one within plots factor. Additionally, each plot and (sub)plot was recorded monthly for 23 months (although only 22 months were analyzed) so time was a second within plots factor. This example includes a factor whose levels are allocated to (sub)plots within plots (pairs) plus a factor representing the whole plots recorded through time (Table 11-11).

Other designs can be termed doubly repeated measures designs because the within plots factors both represent repeated measurements through time. Meserve *et al.* (1996) set up an experiment to examine the effect of predation on the survivorship of degus, a species of rodent. One factor was predation (two fixed levels: predators excluded using fencing and netting and control), with four plots within each level. The number of rodents alive was recorded on each plot at six monthly censuses over four years - year (four fixed levels) and month (six fixed levels) were within plots factors and were crossed (Table 11-11). In all these examples, there were four different denominators used for testing hypotheses in the ANOVA.

11.2.5. Additional between plots/subjects and within plots/subjects factors

These partly nested analyses of variance can be applied to a variety of complex split-plot (repeated measures) experimental designs that include multiple between plots/subjects factors and multiple within/subjects plots factors. We will use the study of Gough & Grace (1998) on the effects of herbivores and productivity levels on plant species densities to illustrate such a complex design. They chose two freshwater marshes on a river near the Louisiana/Mississippi border in eastern USA. In each marsh, they established eight fenced areas (plots), to exclude herbivores like rabbit, muskrat etc., and eight unfenced areas. So the between plots component of the design had two fixed factors (marsh and fence) in a crossed arrangement with replicate areas (i.e. plots). There were three subplots within each fenced or unfenced area and each subplot received one of three nutrient enrichment treatments (no addition, nutrient addition, and natural soil addition). So enrichment was the first within plots factor. Additionally, each subplot was also censused seven times over two years, so time was a second within plots factor. All factors were considered fixed except for area (i.e. plot). The resulting ANOVA model (Table 11-12) had 19 terms and four different denominators for testing hypotheses.

11.2.6. General comments about complex designs

Gumpertz & Brownie (1993) discussed split-plot designs that include repeated measures (usually multiple times) in some detail. They recommended using trend analyses to examine patterns in the repeated factor and its interactions with the other factors in the design (Section

11.5.2). They also recommended against analyzing such designs as univariate split-split-plot designs, *i.e.* using the partly nested models we have described, because of the assumption of sphericity of variance-covariance matrices across times, and preferred a multivariate approach. We agree that the sphericity assumption may be important but instead recommend epsilon-adjusted univariate tests in addition to the multivariate tests. Winer *et al.* (1991) and Kirk (1995) provide details of these complex designs and approaches to analyses; they also provide general formula for determining EMS for any combination of fixed and random factors. Kirk's (1995) unique terminology adapts well to these designs.

11.3. Partly nested designs and statistical software

Data files for these partly nested analyses can be set up in two ways. First, we could create a file for a classical "split-plot analysis" with each factor in a separate column (Table 11-13). A partly nested linear model is then fitted and most software requires that all terms are specified in the model and each term specifically tested against the appropriate denominator. Only unadjusted univariate tests are usually provided but this approach provides great flexibility in structuring the model and choosing denominators for F tests. Second, the data can be coded for a "repeated measures analysis", with between subjects factors coded as usual but the different levels of the within subjects factors are in individual columns (Table 11-13). If you have replicate observations within each cell, it can be difficult to code the data file for "repeated measures" analysis and you must either just use cell means or switch to the "split-plot" set up. Software using the "repeated measures" approach nearly always assumes B(A) is random and all other factors are fixed but provides additional output, including estimates of \mathbf{e} (for multisample sphericity), adjusted and unadjusted univariate tests, multivariate tests, and polynomial trend analyses; it also explicitly distinguishes "between subjects" and "within subjects" components of the ANOVA. Note that the unadjusted univariate tests will be identical to those provided by the first analysis. The important point is that although the two univariate analyses are functionally identical, the alternative analyses (adjusted univariate, MANOVA) and automatic extras (profile analyses) will often only be provided when the data are coded for a classical repeated measures design, not for a split-plot. The profile analyses can usually also be obtained by including contrasts as part of a split-plot analysis.

General Issues and Hints for analysis

General issues

- These designs are very commonly used in biology, as ways to use resources more economically - save money, kill fewer organisms, etc. There is a cost to this rationalization, as the statistical models have more assumptions than completely randomized factorial designs.
- Although they are treated differently in many textbooks, unreplicated partly nested, split-plot and repeated measures (“groups x trials”) designs are analyzed with an identical linear model. For repeated measures designs, this model is usually described with a larger set of assumptions, which imposes more restrictions on the analysis. We recommend that, because the two designs (split-plot and repeated measures) require identical models, you should examine the larger set of assumptions for all partly nested designs.
- Unreplicated partly nested designs, i.e. those with only a single observation of each level of C for each plot/subject within each level of A, prevent you from testing one higher-order interaction or require that you assume that interaction to be zero, depending on the exact design. In the usual situation of all factors being fixed except B (i.e. plots or subjects), this does not preclude tests of the fixed factors or their interactions.
- These designs can include additional factors, both between plots/subject and within plots/subjects. Once the model is decided, the analysis is straightforward except that care must be taken to determine the correct F -ratios depending on which factors are fixed and which are random.

Hints for individual analyses

- These designs are complex, and generally have mixtures of fixed and random factors. As a first step, before doing the experiment, write out the linear model, the ANOVA table, and include details of the F -ratios
- The different designs will change the df, and hence the power, of many of your tests of hypotheses. Before doing the experiment, look at all of the relevant degrees of freedom, and decide whether this arrangement of your experimental units and resources will give you the best compromise between power and cost.
- The assumption of normality is less a problem for the between plots/subjects factor, as those analyses effectively use means of other data, allowing the Central Limit Theorem to be invoked.
- Tests of the between plots/subjects factors assume homogeneity of between group variances
- The assumption of sphericity is important for the tests within plots/subjects factors and incorporates the homogeneity of variance assumption for this component of the analysis. Examine the various measures of the validity of this assumption (particularly the Greenhouse-Geiser and Huynh-Feldt estimates of ϵ), and, if $\hat{\epsilon}$ values are low, use the conservative corrections to the F tests or the MANOVA approach. There is no agreed-upon test for the assumption of multisample sphericity.
- If the design is unreplicated, consider coding the data file up as repeated measures, allowing you to routinely get the $\hat{\epsilon}$ values, corrected univariate F -ratios, and the multivariate equivalents. We follow Looney & Stanley (1989), and suggest you look for a significant result in either the univariate or multivariate analyses.

Box 11-1. Worked example of split-plot design: effects of trampling on intertidal limpet populations

Keough & Quinn (1998) examined the effect of pedestrian traffic (trampling) on the abundance of macroalgae and gastropods on rocky intertidal shores. They used three sites, representing different rock platforms separated by 100s to 1000s of metres. Within each site, there were two experimental plots separated by 10s of metres and four levels of trampling intensity (0, 5, 10, 25 pedestrian passages per low tide on 6-8 days each summer) were allocated to each of two strips within each plot in each site. The response variable is the number of limpets per 0.25m² quadrat per strip. With only two replicates of each plot-trampling combination, it is not worth producing boxplots, and the number of limpets did not vary widely, with numbers generally less than ten, and no extreme values. We did not transform the response variable, a decision that seems reasonable, as the model fitted the data very well (Keough & Quinn 1998, their Table 3). Model 11.2 was fitted, and includes a term for plots within sites x trampling, because we had replicate strips for each trampling treatment in every plot.

In this design, sites and plots are random factors, so you need to be sure that you use correct *F*-ratios (Table 11-3). You might need to recalculate the sites, trampling, and sites x trampling *F*-ratios from the default statistical output if your statistical software does not allow you to specify fixed and random factors.

The specific null hypotheses of interest were:

No difference between sites in the mean number of limpets per strip, pooling trampling treatments

No difference between trampling treatments in the mean number of limpets per strip, pooling sites.

No interaction between site and trampling treatment on the mean number of limpets per strip, i.e. the effect of trampling on the mean the number of limpets per strip was the same at the three sites.

Because we had replicate strips for each trampling treatment within each plot at each site, we could also test two additional null hypotheses:

No added variance in mean number of limpets per strip due to all possible plots within each site

No interaction between trampling treatment and all possible plots within each site on the mean number of limpets per strip, i.e. the effect of trampling on the mean the number of limpets per strip was the same on all possible plots within each site.

The final ANOVA table is shown below.

Source	SS	df	MS	<i>F</i>	<i>P</i>	Denominator
Sites	8.719	2	4.359	0.521	0.639	Plots(Sites)
Plots(Sites)	25.094	3	8.365	5.214	0.006	Residual
Trampling	18.354	3	6.118	5.071	0.044	Site x Trampling
Site x Trampling	7.240	6	1.207	0.485	0.805	Plots(Sites) x Trampling
Plots(Sites) x Trampling	22.406	9	2.490	1.552	0.187	Residual
Residual	38.500	24	1.604			

We would conclude that there is a significant main effect of trampling, and that the effect of trampling on the number of limpets does not vary between sites or plots. There is also

significant spatial variation at the scale of plots. The number of limpets rose with the intensity of trampling and Figure 11-3a shows similar increases at all three platforms (with data averaged across plots). Trampling appears to benefit limpets! This effect occurred because the species most affected by trampling is a brown alga, *Hormosira banksii*, which forms dense mats on these rock platforms. Dense mats provide a poor habitat for the herbivorous limpets, with little food, so the destruction of these mats generates new, usable habitat for limpets. At the level of plots, we found wide variation in overall abundance of limpets (averaged across trampling levels) (Figure 11-3b). The plots with higher numbers were on different platforms (sites), as were those with low numbers, accounting for significant variation among plots, but not sites.

Box 11-2. Worked example of groups by trials repeated measures design: responses of cane toads to hypoxia

Mullens (1993) investigated how the breathing rates of cane toads (*Bufo marinus*) respond to conditions of hypoxia. Toads, the subjects, show two different kinds of breathing patterns, lung or buccal, and this breathing pattern was the between subjects factor. The second factor was O₂ concentration, which had eight levels (0, 5, 10, 15, 20, 30, 40, 50%), and was applied within subjects (toads). The response variable was the frequency of buccal breathing and was transformed to square-roots to reduced positive skewness (based on boxplots of the data for each O₂ concentration) and improved variance homogeneity (based on residual plots).

The specific null hypotheses of interest were:

No difference between breathing types in the mean square root rate of breathing, pooling O₂ levels.

No difference between O₂ levels types in the mean square root rate of breathing, pooling, pooling breathing types.

No interaction between breathing type and O₂ level on the mean square root rate of breathing, i.e. the effect of O₂ level on the mean square root rate of breathing was the same for both breathing types.

With no replicates within each combination of breathing type, toad and O₂ level, we could not test hypotheses about the random factor toads within breathing type or O₂ levels by toads within breathing type.

The data were initially coded in classical split-plot form, where toads were plots, and the model in 11.3 was fitted. Because there is only one replicate observation for each toad for each O₂ concentration, this model is fully saturated, i.e. it fits the data perfectly because all sources of variation have been accounted for. The output from your statistical software usually won't include *F* tests or *P* values. You might just need to specify each effect in the model and its appropriate denominator to get these. In this example, breathing type and oxygen are clearly fixed factors, but toad is random, so breathing type is tested against toad within breathing type and O₂ level and interaction between breathing type and O₂ level against toad within breathing type by O₂ level interaction. We could also achieve these latter tests by fitting a model without the toad within breathing type by O₂ level interaction, which would then become the residual term. Note that many statistical programs assume all factors are fixed and default to using this as the denominator for all tests, which is incorrect if B(A) is random.

Source	SS	df	MS	<i>F</i>	<i>P</i>
Breathing type	39.921	1	39.921	5.762	0.027
Toad(Breathing type)	131.634	19	6.928		
O ₂ level	25.748	7	3.678	4.884	<0.001
Breathing type x O ₂ level	56.372	7	8.053	10.693	<0.001
Toad(Breathing type) x O ₂ level	100.166	133	0.753		

We would conclude that there is a significant difference between toads with the different breathing types, but this depends on O₂ level (significant breathing type x O₂ level interaction).

We then re-analyzed the data after recoding them as a “repeated measures” design. For most software, , we get even more extensive output:

BETWEEN SUBJECTS

Source	SS	df	MS	F	P
Breathing type	39.921	1	39.921	5.762	0.027
Residual	131.634	19	6.928		

Note: this residual is actually toads nested within breathing type.

WITHIN SUBJECTS

Source	SS	df	MS	F	P	GG P	HF P
O ₂ level	25.748	7	3.678	4.884	<0.001	0.004	0.002
Breathing type x O ₂ level	56.372	7	8.053	10.693	<0.001	<0.001	<0.001
Residual	100.166	133	0.753				

Note: this residual is actually toads within breathing type by O₂ level.

Greenhouse-Geisser Epsilon: 0.4282
 Huynh-Feldt Epsilon: 0.5440

The “between subjects” and “within-subjects” parts of the ANOVA are distinguished and B(A), in this example toads within breathing type, is assumed to be random and all other factors fixed. The ANOVA output is, however, identical to the partly nested ANOVA above. Estimates of ϵ are also provided – the Greenhouse-Geisser is more conservative than the Huynh-Feldt estimate and neither is close to one, suggesting that the sphericity assumption is not met. Because both estimates of epsilon are less than 0.75, the Greenhouse-Geisser adjustment is preferred. Our conclusions would not be affected by these more conservative tests; there is a significant interaction between oxygen concentration and breathing type. Both main effects are also significant, although it is more sensible to base further interpretation on the interaction. It is clear from Fig Figure 11-4 that breathing rate decreases with increasing O₂ level for buccal breathing toads but increases with O₂ level for lung breathing toads.

Because of the interaction, simple main effects tests for O₂ level at each breathing type separately might be of interest. We adjust the df for these tests based on the Greenhouse-Geisser estimate of epsilon.

Buccal:

Source	SS	df	MS	F	P	GG df	GG P
O ₂ level	75.433	7	10.776	14.311	<0.001	2.997	<0.001
Residual	100.166	133	0.753			56.951	

Lung:

Source	SS	df	MS	F	P	GG df	GG P
O ₂ level	19.907	7	2.844	3.777	0.001	2.997	0.015
Residual	100.166	133	0.753			56.951	

There is a significant effect of O₂ level for both breathing types, although the effect seems stronger for buccal breathing toads than lung breathing toads.

For most statistical software, “repeated measures” output will include polynomial trend analyses. With eight O₂ levels, up to 7th order polynomials could be examined, although we will just look at the first three. The interaction test of these polynomials is testing whether the trend (linear, quadratic etc.) through O₂ level differs between breathing types; the main effect test is examining whether there is a trend through O₂ level pooling breathing types.

Polynomial Test of Order 1 (Linear).

Source	SS	df	MS	F	P
O ₂ level	17.010	1	17.010	8.255	0.010
Breathing type x O ₂ level	40.065	1	40.065	19.444	<0.001
Residual	39.149	19	2.060		

Polynomial Test of Order 2 (Quadratic)

Source	SS	df	MS	F	P
O ₂ level	5.007	1	5.007	6.967	0.016
Breathing type x O ₂ level	12.326	1	12.326	17.150	0.001
Residual	13.655	19	0.719		

Polynomial Test of Order 3 (Cubic)

Source	SS	df	MS	F	P
O ₂ level	1.747	1	1.747	3.263	0.087
Breathing type x O ₂ level	1.784	1	1.784	3.331	0.084
Residual	10.174	19	0.535		

Both linear and quadratic trends are different between the two breathing types; there is no evidence of a cubic trend. It is clear from Fig Figure 11-4 that the linear trends are in different directions for the two breathing types. Note that separate error terms are used for each trend test, a requirement if there is a chance that sphericity of variances and covariances does not hold.

Finally, we get the multivariate tests of the “within subjects” hypotheses.

O₂ level

		Hypoth. df	Error df	F	P
Wilks' Lambda	0.115	7	13	14.277	<0.001
Pillai Trace	0.885	7	13	14.277	<0.001
Hotelling-Lawley Trace	7.688	7	13	14.277	<0.001

Breathing type x O₂ level

		Hypoth. df	Error df	F	P
--	--	------------	----------	---	---

Wilks' Lambda	0.325	7	13	3.853	0.017
Pillai Trace	0.675	7	13	3.853	0.017
Hotelling-Lawley Trace	2.075	7	13	3.853	0.017

The conclusions from the Pillai statistic agree with the univariate analysis – a significant interaction between O₂ level and breathing type.

Box 11-3 The partly nested linear model and its parameters

Consider a design with p levels of factor A ($i = 1$ to p), q levels of factor B (plots or subjects) nested within each level of A ($j = 1$ to q) and r levels of factor C ($k = 1$ to r), crossed with both A and B (Table 11-1) and replicate observations ($l = 1$ to n) within each combination of A, B and C.

The formal linear model for a split-plot design is:

$$y_{ijkl} = \mathbf{m} + \mathbf{a}_i + \mathbf{b}_{j(i)} + \mathbf{g}_k + \mathbf{ag}_{ik} + \mathbf{bg}_{j(i)k} + \mathbf{e}_{ijkl}$$

In this model:

y_{ijkl} is the value of the response variable for the l th observation in the k th level of factor C for the j th plot/subject in the i th level of factor A. Commonly in these designs, $l = 1$, but our two worked examples in this chapter include one in which $l = 2$ (Box 11-1) and one with $l = 1$ (Box 11-2).

\mathbf{m} is the overall (constant) population mean of the response variable.

If factor A is fixed, \mathbf{a}_i is the effect of i th level of factor A ($\mathbf{m}_i - \mathbf{m}$), pooling over factor C. If factor A is random, \mathbf{a}_i is a random variable with a mean of zero and a variance of \mathbf{S}_a^2 , measuring the variance in mean values of the response variable across all possible levels of factor A that could have been used.

Plots or subjects are nearly always random so $\mathbf{b}_{j(i)}$ is a random variable with a mean of zero and a variance of \mathbf{S}_b^2 , measuring the variance in mean values of the response variable across all possible plots or subjects that could have been used within any level of A.

If factor C is fixed, \mathbf{g}_k is the effect of the k th level of factor C ($\mathbf{m}_k - \mathbf{m}$), pooling over factor A. If factor C is random, \mathbf{g}_k is a random variable with a mean of zero and a variance of \mathbf{S}_g^2 , measuring the variance in mean values of the response variable across all possible levels of factor C that could have been used.

If factors A and C are both fixed, \mathbf{ag}_{ik} is the effect of the interaction between the i th level of A and k th level of C. If either factor is random, then \mathbf{ag}_{ik} is a random variable with a mean of zero and a variance of \mathbf{S}_{ag}^2 , measuring the variance across all the possible interaction terms between the fixed levels (if A is fixed) or all possible levels (if A is random) of factor A and the fixed levels (if C is fixed).

Because plots/subjects are nearly always random, the interaction between factor C and plots/subjects $\mathbf{bg}_{j(i)k}$ is a random variable with a mean of zero and a variance of \mathbf{S}_{bg}^2 , measuring the variance across all the possible interaction terms between all the possible plots/subjects within any level of A and the fixed levels (if C is fixed) or all possible levels (if C is random) of factor C.

\mathbf{e}_{ijkl} is random or unexplained error associated with the l th observation in the k th level of factor C for the j th plot/subject in the i th level of factor A. These error terms are assumed to be normally distributed in each combination of A, B and C with a mean of zero and a variance of \mathbf{S}_e^2 . Note that classical split-plot and repeated measures designs usually do not have replication for each combination of plot and factor C ($n = 1$) so \mathbf{e}_{ijkl} usually cannot be separately estimated.

This effects model is over-parameterized (Box 8.3) so the usual sum-to-zero constraints are imposed on the fixed effects. We can also use a cell means model (Kirk 1995) which may be useful when there are missing observations (Section 11.6).

Estimating the parameters of the partly nested model follows the procedures outlined in the previous three chapters for single factor, multifactor and randomized block models. The cell means (\mathbf{m}_{jk}) are estimated by means of the observations in each cell, although there is often only a single observation for each A, B and C combination. The marginal means are shown in Table 11-1 and represent averages across the appropriate cell means (or single observations). Standard errors for these means must be based on the appropriate variance estimate (mean square), the one that is used as the denominator of the F -ratio for testing the H_0 that the means are equal (see Boxes 9-2 and 9-6).

Box 11-4. Impact of flowering on forest bird communities

As described in Section 11.8.1, McGoldrick & Mac Nally (1998) studied the impact of eucalypt flowering on the dynamics of bird communities in forests of S.E. Australia. They used a partly nested design with two between plots/subject factors (habitat and region) with two sites within each combination. Each site was censused monthly for twelve months, so time was the within plots/subjects factor. The response variable we will analyze is natural log transformed (density of nectarivorous birds + 1).

The specific null hypotheses of interest were:

No difference between habitat in the mean log (density of nectarivorous birds + 1), pooling regions and months.

No difference between regions in the mean log (density of nectarivorous birds + 1), pooling habitat and months.

No interaction between habitat and region on the mean log (density of nectarivorous birds + 1), pooling months. Rephrased, the effect of habitat on the mean log (density of nectarivorous birds + 1) was the same for both regions and vice-versa, pooling months.

No difference between months in the mean log (density of nectarivorous birds + 1), pooling habitats and regions.

No interactions between habitat and month, region and month, or habitat and region and month on the mean log (density of nectarivorous birds + 1). Re-phrased, the effect of habitat, pooling regions, was the same in all months, the effect of region, pooling habitats, was the same in all months, and the interaction between habitat and region was the same in all months.

With no replicates within each combination of habitat, region, site and month, we could not test hypotheses about the random factor sites within habitat and region or months by sites within habitat and region.

Between plots/subjects

Source	SS	df	MS	F	P
Habitat	88.313	1	88.313	48.975	0.002
Region	0.106	1	0.106	0.059	0.821
Habitat x Region	1.334	1	1.334	0.740	0.438
Site(Habitat, Region)	7.213	4	1.803		

Within plots/subjects

Source	SS	df	MS	F	P	GG
Month	48.676	11	4.425	5.941	<0.001	0.019
Habitat x Month	75.152	11	6.559	8.806	<0.001	<0.006
Region x Month	11.436	11	1.040	1.396	0.209	0.299
Habitat x Region x Month	3.858	11	0.351	0.471	0.911	0.665
Site(Habitat, Region) x Month	32.774	44	0.745			

Greenhouse-Geisser Epsilon: 0.2104

Huynh-Feldt Epsilon: 0.8907

Our analysis agrees with that published by McGoldrick & Mac Nally (1998), although they did not present adjusted tests for within plots/subjects tests. The adjusted df did not change our conclusions. The month effect varied between habitats and there were neither effects of region nor any interactions between habitat and region or region and month. Note that the epsilon estimates differ greatly and for the three factor interaction, the adjusted test is more liberal than the unadjusted tests.

Linear trends:

Source	SS	df	MS	F	P
Time	16.056	1	16.056	12.231	0.025
Habitat x Month	24.532	1	24.532	18.689	0.012
Region x Month	3.028	1	3.028	2.307	0.203
Habitat x Region x Month	0.717	1	0.717	0.546	0.501
Site(Habitat, Region) x Month	5.251	4	1.313		

Quadratic trends:

Source	SS	df	MS	F	P
Time	13.099	1	13.099	8.897	0.041
Habitat x Month	17.935	1	17.935	12.182	0.025
Region x Month	1.574	1	1.574	1.069	0.360
Habitat x Region x Month	0.822	1	0.822	0.558	0.496
Site(Habitat, Region) x Month	5.889	4	1.472		

Cubic trends:

Source	SS	df	MS	F	P
Time	1.696	1	1.696	2.943	0.161
Habitat x Month	22.695	1	22.695	39.375	0.003
Region x Month	1.401	1	1.401	2.432	0.194
Habitat x Region x Month	0.167	1	0.157	0.290	0.619
Site(Habitat, Region) x Month	2.306	4	0.576		

The trend analyses indicate that any linear, quadratic or cubic trends through time differ between the two habitats. It is clear from Figure 11-5 that there is little change through time in stringybark habitats but marked declines from the austral autumn and winter through to spring and summer for ironbrak habitat.

		C_1	C_2	C_k	B marginal means	A marginal means
A_1	$B_{1(1)}$	y_{111}	y_{112}	y_{11k}	$\bar{y}_{j=1(1)}$	
	$B_{2(1)}$	y_{121}	y_{122}	y_{12k}	$\bar{y}_{j=2(1)}$	$\bar{y}_{i=1}$
	$B_{j(1)}$	y_{1j1}	y_{1j2}	y_{1jk}	$\bar{y}_{j(1)}$	
A_2						
	$B_{j(2)}$	y_{2j1}	y_{2j2}	y_{2jk}	$\bar{y}_{j(2)}$	$\bar{y}_{i=2}$
A_i	$B_{j(i)}$	y_{ij1}	y_{ij2}	y_{ijk}	$\bar{y}_{j(i)}$	\bar{y}_i
C marginal means		$\bar{y}_{k=1}$	$\bar{y}_{k=2}$	\bar{y}_k		

Table 11-1 Marginal means for a partly nested design with i levels of factor A, j levels of factor B within each level of factor A and k levels of factor C crossed with both A and B.

(a)

Source	df	EMS	EMS (A, C fixed, B random)	Test (A, C fixed, B random)
<i>Between plots/subjects</i>				
A	$p-1$	$\mathbf{s}_e^2 + nD_q D_r \mathbf{s}_{bg}^2 + nqD_r \mathbf{s}_{ag}^2 + nrD_q \mathbf{s}_b^2 + nqr\mathbf{s}_a^2$	$\mathbf{s}_e^2 + nr\mathbf{s}_b^2 + nqr\mathbf{s}_a^2$	$MS_A/MS_{B(A)}$
B(A)	$p(q-1)$	$\mathbf{s}_e^2 + nD_r \mathbf{s}_{bg}^2 + nr\mathbf{s}_b^2$	$\mathbf{s}_e^2 + nr\mathbf{s}_b^2$	$MS_B/MS_{Residual}$
<i>Within plots/subjects</i>				
C	$r-1$	$\mathbf{s}_e^2 + nD_q \mathbf{s}_{bg}^2 + nqD_p \mathbf{s}_{ag}^2 + npq\mathbf{s}_g^2$	$\mathbf{s}_e^2 + n\mathbf{s}_{bg}^2 + npq\mathbf{s}_g^2$	$MS_C/MS_{B(A)C}$
A x C	$(p-1)(r-1)$	$\mathbf{s}_e^2 + nD_q \mathbf{s}_{bg}^2 + nq\mathbf{s}_{ag}^2$	$\mathbf{s}_e^2 + n\mathbf{s}_{bg}^2 + nq\mathbf{s}_{ag}^2$	$MS_{AC}/MS_{B(A)C}$
B(A) x C	$p(q-1)(r-1)$	$\mathbf{s}_e^2 + n\mathbf{s}_{bg}^2$	$\mathbf{s}_e^2 + n\mathbf{s}_{bg}^2$	$MS_{BC}/MS_{Residual}$
Residual	$Pqr(n-1)$	\mathbf{s}_e^2	\mathbf{s}_e^2	

(b)

Source	df	Source	df
<i>Between plots (i.e. ponds)</i>		<i>Between subjects (i.e. trees)</i>	
Hydroperiod	1	Species	1
Ponds within Hydroperiod	6	Trees within Species	38
<i>Within plots (i.e. ponds)</i>		<i>Within subjects (i.e. trees)</i>	
Treatment	2	Date	19
Hydroperiod x Treatment	2	Tree species x Date	19
Ponds within Hydroperiod x Treatment	12	Trees within Species x Date	799

Table 11-2 (a) Classical split-plot or repeated measures design with general EMS and those for the specific case of A & C fixed, but B (plots or subjects) random, showing F-ratios used to test all hypotheses. For explanation of the conversion of the general model to particular combinations of fixed and random factors, see Box 9-8. (b) ANOVA for split-plot design from Wissinger et al. (1996), where Hydroperiod and Treatment are fixed factors, Ponds is random and nested within Hydroperiod, and from repeated measures design from Gange (1995), where Species and Date are fixed factor; Trees is random and nested within Species.

Source	Df	A, B, C fixed		A, B random, C fixed		A fixed, B, C random	
		EMS	Test	EMS	Test	EMS	Test
A	$p-1$	$\mathbf{s}_e^2 + nqr\mathbf{s}_a^2$	$MS_A/MS_{Residual}$	$\mathbf{s}_e^2 + nr\mathbf{s}_b^2 + nqr\mathbf{s}_a^2$	$MS_A/MS_{B(A)}$	$\mathbf{s}_e^2 + n\mathbf{s}_{bg}^2 + nq\mathbf{s}_{ag}^2 + nr\mathbf{s}_b^2 + nqr\mathbf{s}_a^2$	No test
B(A)	$p(q-1)$	$\mathbf{s}_e^2 + nr\mathbf{s}_b^2$	$MS_B/MS_{Residual}$	$\mathbf{s}_e^2 + nr\mathbf{s}_b^2$	$MS_{B(A)}/MS_{Residual}$	$\mathbf{s}_e^2 + n\mathbf{s}_{bg}^2 + nr\mathbf{s}_b^2$	$MS_{B(A)}/MS_{B(A)C}$
C	$r-1$	$\mathbf{s}_e^2 + npq\mathbf{s}_g^2$	$MS_C/MS_{Residual}$	$\mathbf{s}_e^2 + n\mathbf{s}_{bg}^2 + nq\mathbf{s}_{ag}^2 + npq\mathbf{s}_g^2$	MS_C/MS_{AC}	$\mathbf{s}_e^2 + n\mathbf{s}_{bg}^2 + nq\mathbf{s}_{ag}^2 + npq\mathbf{s}_g^2$	MS_C/MS_{AC}
A x C	$(p-1)(r-1)$	$\mathbf{s}_e^2 + nq\mathbf{s}_{ag}^2$	$MS_{AC}/MS_{Residual}$	$\mathbf{s}_e^2 + n\mathbf{s}_{bg}^2 + nq\mathbf{s}_{ag}^2$	$MS_{AC}/MS_{B(A)C}$	$\mathbf{s}_e^2 + n\mathbf{s}_{bg}^2 + nq\mathbf{s}_{ag}^2$	$MS_{AC}/MS_{B(A)C}$
B(A) x C	$p(q-1)(r-1)$	$\mathbf{s}_e^2 + n\mathbf{s}_{bg}^2$	$MS_{BC}/MS_{Residual}$	$\mathbf{s}_e^2 + n\mathbf{s}_{bg}^2$	$MS_{B(A)C}/MS_{Residual}$	$\mathbf{s}_e^2 + n\mathbf{s}_{bg}^2$	$MS_{BC}/MS_{Residual}$
Residual	$pqr(n-1)$	\mathbf{s}_e^2		\mathbf{s}_e^2		\mathbf{s}_e^2	

Table 11-3 ANOVA tables with expected mean squares (EMS) for partly nested models, showing F-ratios used to test all hypotheses.

Source	df	adjusted df
<i>Within plots/subjects</i>		
C	$(r-1)$	$(r-1)\hat{\epsilon}$
A x C	$(p-1)(r-1)$	$(p-1)(r-1)\hat{\epsilon}$
B(A) x C	$p(q-1)(r-1)$	$p(q-1)(r-1)\hat{\epsilon}$

Table 11-4 Degrees of freedom for within plots or subjects components of partly nested ANOVA. Adjustment based on estimate of ϵ indicating how far variance-covariance matrix is from sphericity.

Source	df	General expected mean squares (EMS)	EMS (A, C, D fixed; B random)	Test
<i>Between plots/subjects</i>				
A	(p-1)	$\mathbf{s}_e^2 + D_q D_t \mathbf{s}_{bd}^2 + q D_r D_t \mathbf{s}_{agd}^2 + qr D_t \mathbf{s}_{ad}^2 + t D_q \mathbf{s}_b^2 + qt D_r \mathbf{s}_{ag}^2 + qrt \mathbf{s}_a^2$	$\mathbf{s}_e^2 + t \mathbf{s}_b^2 + qrt \mathbf{s}_a^2$	$MS_A / MS_{B(AC)}$
C	(r-1)	$\mathbf{s}_e^2 + D_q D_t \mathbf{s}_{bd}^2 + q D_p D_t \mathbf{s}_{agd}^2 + pq D_t \mathbf{s}_{gd}^2 + t D_q \mathbf{s}_b^2 + qt D_p \mathbf{s}_{ag}^2 + pqt \mathbf{s}_g^2$	$\mathbf{s}_e^2 + t \mathbf{s}_b^2 + pqt \mathbf{s}_g^2$	$MS_B / MS_{B(AC)}$
A x C	(p-1)(r-1)	$\mathbf{s}_e^2 + D_q D_t \mathbf{s}_{bd}^2 + q D_t \mathbf{s}_{agd}^2 + t D_q \mathbf{s}_b^2 + qt \mathbf{s}_{ag}^2$	$\mathbf{s}_e^2 + t \mathbf{s}_b^2 + qt \mathbf{s}_{ag}^2$	$MS_{AB} / MS_{B(AC)}$
B(AC)	pr(q-1)	$\mathbf{s}_e^2 + D_t \mathbf{s}_{bd}^2 + t \mathbf{s}_b^2$	$\mathbf{s}_e^2 + t \mathbf{s}_b^2$	No test
<i>Within plots/subjects</i>				
D	(t-1)	$\mathbf{s}_e^2 + D_q \mathbf{s}_{bd}^2 + q D_p D_r \mathbf{s}_{agd}^2 + pq D_r \mathbf{s}_{gd}^2 + qr D_p \mathbf{s}_{ad}^2 + pqr \mathbf{s}_d^2$	$\mathbf{s}_e^2 + \mathbf{s}_{bd}^2 + pqr \mathbf{s}_d^2$	$MS_D / MS_{DB(AC)}$
A x D	(p-1)(t-1)	$\mathbf{s}_e^2 + D_q \mathbf{s}_{bd}^2 + q D_r \mathbf{s}_{agd}^2 + qr \mathbf{s}_{ad}^2$	$\mathbf{s}_e^2 + \mathbf{s}_{bd}^2 + qr \mathbf{s}_{ad}^2$	$MS_{AD} / MS_{DB(AC)}$
C x D	(r-1)(t-1)	$\mathbf{s}_e^2 + D_q \mathbf{s}_{bd}^2 + q D_p \mathbf{s}_{agd}^2 + pq \mathbf{s}_{gd}^2$	$\mathbf{s}_e^2 + \mathbf{s}_{bd}^2 + pq \mathbf{s}_{gd}^2$	$MS_{BD} / MS_{DB(AC)}$
A x C x D	(p-1)(r-1)(t-1)	$\mathbf{s}_e^2 + D_q \mathbf{s}_{bd}^2 + q \mathbf{s}_{agd}^2$	$\mathbf{s}_e^2 + \mathbf{s}_{bd}^2 + q \mathbf{s}_{agd}^2$	$MS_{ABD} / MS_{DB(AC)}$
D x B(AC)	(t-1)pr(q-1)	$\mathbf{s}_e^2 + \mathbf{s}_{bd}^2$	$\mathbf{s}_e^2 + \mathbf{s}_{bd}^2$	No test

Table 11-5 ANOVA table for partly nested design with two crossed between plots factors and one within plots factor. Expected mean squares are provided for the general case (see Box 9-8) and for the usual case of A, C and D fixed with B (plots or subjects) random. There is only one observation within each combination of A, B, C and D.

Source	Source	df	F-ratio denominator
<i>Between plots/subjects</i>			
	<i>Between grids</i>		
A	Habitat	1	Grid within (Habitat, Location)
C	Location	8	Grid within (Habitat, Location)
A x C	Habitat x location	8	Grid within (Habitat, Location)
B within (AC)	Grid within (Habitat, Location)	18	
<i>Within plots/subjects</i>			
	<i>Within grids</i>		
D	Time	2	Time x Grid within (Habitat, Location)
A x D	Habitat x Time	2	Time x Grid within (Habitat, Location)
C x D	Location x Time	16	Time x Grid within (Habitat, Location)
A x C x D	Habitat x Location x Time	16	Time x Grid within (Habitat, Location)
D x B within (AC)	Time x Grid within (Habitat, Location)	36	

Table 11-6 ANOVA table for partly-nested design from Morris (1996) with two crossed between plots factors (Habitat and Location, both fixed), one within plots factor (Time, fixed) and Grids as random plots. There is only one observation within each combination of A, B, C and D.

Source	Source	df	F-ratio denominator
<i>Between plots/subjects</i>			
<i>Between plots</i>			
A	Predators	1	Plot within (Predator, Soil, Light)
C	Soil	1	Plot within (Predator, Soil, Light)
D	Light	1	Plot within (Predator, Soil, Light)
A x C	Predator x Soil	1	Plot within (Predator, Soil, Light)
A x D	Predator x Light	1	Plot within (Predator, Soil, Light)
C x D	Soil x Light	1	Plot within (Predator, Soil, Light)
A x C x D	Predator x Soil x Light	1	Plot within (Predator, Soil, Light)
B within (ACD)	Plot within (Predator, Soil, Light)	16	
<i>Within plots/subjects</i>			
<i>Within plots</i>			
E	Time	4	Time x Plot within (Predator, Soil, Light)
A x E	Predator x Time	4	Time x Plot within (Predator, Soil, Light)
C x E	Soil x Time	4	Time x Plot within (Predator, Soil, Light)
D x E	Light x Time	4	Time x Plot within (Predator, Soil, Light)
A x C x E	Predator x Soil x Time	4	Time x Plot within (Predator, Soil, Light)
A x D x E	Predator x Light x Time	4	Time x Plot within (Predator, Soil, Light)
C x D x E	Soil x Light x Time	4	Time x Plot within (Predator, Soil, Light)
A x C x D x E	Predator x Soil x Light x Time	4	Time x Plot within (Predator, Soil, Light)
E x B within (ACD)	Time x Plot within (Predator, Soil, Light)	64	

Table 11-7 ANOVA table for partly-nested design from Letourneau & Dyer (1998) with three crossed between plots factors (Predators, Light level, Soil; all fixed), one within plots factor (Time, fixed) and Plots as random plots.

Source	df	EMS general	EMS (A, C fixed, B random)	Test
<i>Between plots/subjects</i>				
A	(p-1)	$\mathbf{s}_e^2 + D_q D_r \mathbf{s}_{agb}^2 + D_r q \mathbf{s}_{ag}^2 + D_q r \mathbf{s}_{ab}^2 + q r \mathbf{s}_a^2$	$\mathbf{s}_e^2 + r \mathbf{s}_{ab}^2 + q r \mathbf{s}_a^2$	MS _A / MS _{AB}
B (= block)	(q-1)	$\mathbf{s}_e^2 + D_p D_r \mathbf{s}_{agb}^2 + D_r p \mathbf{s}_{gb}^2 + D_p r \mathbf{s}_{ab}^2 + p r \mathbf{s}_b^2$	$\mathbf{s}_e^2 + r \mathbf{s}_{ab}^2 + p r \mathbf{s}_b^2$	No test
A x B	(p-1)(q-1)	$\mathbf{s}_e^2 + D_r \mathbf{s}_{agb}^2 + r \mathbf{s}_{ab}^2$	$\mathbf{s}_e^2 + r \mathbf{s}_{ab}^2$	No test
<i>Within plots/subjects</i>				
C	(r-1)	$\mathbf{s}_e^2 + D_q D_p \mathbf{s}_{agb}^2 + D_p q \mathbf{s}_{ag}^2 + D_q p \mathbf{s}_{gb}^2 + p q r \mathbf{s}_g^2$	$\mathbf{s}_e^2 + p \mathbf{s}_{gb}^2 + p q r \mathbf{s}_g^2$	MS _C / MS _{BC}
A x C	(p-1)(r-1)	$\mathbf{s}_e^2 + D_q \mathbf{s}_{agb}^2 + q \mathbf{s}_{ag}^2$	$\mathbf{s}_e^2 + \mathbf{s}_{agb}^2 + q \mathbf{s}_{ag}^2$	MS _{AC} / MS _{ABC}
B x C	(q-1)(r-1)	$\mathbf{s}_e^2 + D_p \mathbf{s}_{agb}^2 + p \mathbf{s}_{gb}^2$	$\mathbf{s}_e^2 + p \mathbf{s}_{gb}^2$	No test
A x B x C	(p-1)(q-1)(r-1)	$\mathbf{s}_e^2 + \mathbf{s}_{agb}^2$	$\mathbf{s}_e^2 + \mathbf{s}_{agb}^2$	No test

Table 11-8 ANOVA table for partly-nested design with a RCB between-plots component (A fixed & B=blocks random) and within-plots factor C fixed. The fitted model is non-additive, including A x Blocks interactions. An additive model means that all the (**ab**) terms would disappear from the EMS, allowing tests for B (blocks) and B (blocks) x C.

Aguiar & Sala (1997)			Evans & England (1996)		
Source	df	F-ratio denominator	Source	df	F-ratio denominator
<i>Between plots/subjects</i>			<i>Between plots/subjects</i>		
Date	2	Residual (Site x Date)	Treatment	2	Residual
Site (= block)	2		Row (block)	9	Residual
Residual (Date x Site)	4		Residual	18	
<i>Within plots/subjects</i>			<i>Within plots/subjects</i>		
Microsite	3	Date x Microsite	Date	1	Treatment x Row x Date
Date x Microsite	6	Date x Site x Microsite	Treatment x Date	2	Treatment x Row x Date
Site x Microsite	6		Row x Date	9	Treatment x Row x Date
Date x Site x Microsite	12		Treatment x Row x Date	18	

Table 11-9 ANOVA table for partly-nested designs with a RCB between-plots component. The example from Aguiar & Sala (1997) has Date (fixed) and Site (a random blocking factor) as between plots (blocks) factors and Microsite (fixed) as a within plots (blocks) factor. There is only one observation for each Site and Date combination and the four microsites were located within each plot (block). The example from Evans & England (1996) has Treatment (fixed) and Row (a random blocking factor) as between plots (blocks) factors and Date as a within plots (blocks) factor. Evans & England (1996) fitted an additive model assuming no Treatment x Row and no Treatment x Row x Date interactions, allowing tests for Row and Row x Date. There is only one observation for each Treatment and Row combination and each combination was recorded on two dates.

Source	df	EMS general	EMS (A, C, D fixed, B random)	Test
<i>Between plots/subjects</i>				
A	(p-1)	$s_e^2 + D_q D_r D_t s_{bgd}^2 + q D_r D_t s_{agd}^2 + r D_q D_t s_{bd}^2 + q r D_t s_{ad}^2 +$ $t D_q D_r s_{bg}^2 + q t D_r s_{ag}^2 + r t D_q s_b^2 + q r t s_a^2$	$s_e^2 + r t s_b^2 + q r t s_a^2$	MS _A / MA _{B(A)}
B(A)	p(q-1)	$s_e^2 + D_r D_t s_{bgd}^2 + r D_t s_{bd}^2 + t D_r s_{bg}^2 + r t s_b^2$	$s_e^2 + r t s_b^2$	No test
<i>Within plots/subjects</i>				
C	(r-1)	$s_e^2 + D_q D_t s_{bgd}^2 + q D_p D_r s_{agd}^2 + p q D_t s_{gd}^2 + t D_q s_{bg}^2 + q t D_p s_{ag}^2 + p q t s_g^2$	$s_e^2 + t s_{bg}^2 + p q t s_g^2$	MA _C / MS _{B(A)C}
A x C	(p-1)(r-1)	$s_e^2 + D_q D_t s_{bgd}^2 + q D_t s_{agd}^2 + t D_q s_{bg}^2 + q t s_{ag}^2$	$s_e^2 + t s_{bg}^2 + q t s_{ag}^2$	MS _{AC} / MS _{B(A)C}
B(A) x C	p(q-1)(r-1)	$s_e^2 + D_t s_{bgd}^2 + t s_{bg}^2$	$s_e^2 + t s_{bg}^2$	No test
D	(t-1)	$s_e^2 + D_q D_r s_{bgd}^2 + q D_p D_r s_{agd}^2 + p q D_r s_{gd}^2 + r D_q s_{bd}^2 + q r D_p s_{ad}^2 + p q r s_d^2$	$s_e^2 + r s_{bd}^2 + p q r s_d^2$	MS _D / MS _{B(A)D}
A x D	(p-1)(t-1)	$s_e^2 + D_q D_r s_{bgd}^2 + q D_r s_{agd}^2 + r D_q s_{bd}^2 + q r s_{ad}^2$	$s_e^2 + r s_{bd}^2 + q r s_{ad}^2$	MS _{AD} / MS _{B(A)D}
B(A) x D	p(q-1)(t-1)	$s_e^2 + D_r s_{bgd}^2 + r s_{bd}^2$	$s_e^2 + r s_{bd}^2$	No test
C x D	(r-1)(t-1)	$s_e^2 + D_q s_{bgd}^2 + q D_p s_{agd}^2 + p q s_{gd}^2$	$s_e^2 + s_{bgd}^2 + p q s_{gd}^2$	MS _{CD} / MS _{B(A)CD}
A x C x D	(p-1)(r-1)(t-1)	$s_e^2 + D_q s_{bgd}^2 + q s_{agd}^2$	$s_e^2 + s_{bgd}^2 + q s_{agd}^2$	MS _{ACD} / MS _{B(A)CD}
B(A) x C x D	p(q-1)(r-1)(t-1)	$s_e^2 + s_{bgd}^2$	$s_e^2 + s_{bgd}^2$	No test

Table 11-10 ANOVA with expected mean squares for partly nested design with one between plots factor (A), plots/subjects (B) and two crossed within plots factors (C & D). There is only a single observation within each combination of A, B, C and D.

General		Meserve <i>et al.</i> (1996)		Vasquez (1996)		Green <i>et al.</i> (1997)	
Source	<i>F</i> -ratio denominator	Source	df	Source	df	Source	df
Between plots/subjects							
A	B(A)	Predation	1	Species	2	Habitat	1
B(A)		Plots(Predation)	6	Subject(Species)	48	Pairs(Habitat)	8
Within plots/subjects							
C	C x B(A)	Year	3	Illumination	1	Exclusion	1
A x C	C x B(A)	Year x Predation	3	Illumination x Species	2	Exclusion x Habitat	1
C x B(A)		Year x Plots(Predation)	18	Illumination x Subject(Species)	48	Exclusion x Pairs(Habitat)	8
D	D x B(A)	Month	5	Distribution	1	Time	21
A x D	D x B(A)	Month x Predation	5	Distribution x Species	2	Time x Habitat	21
D x B(A)		Month x Plots(Predation)	30	Distribution x Subject(Species)	48	Time x Pairs(Habitat)	168
C x D	C x D x B(A)	Year x Month	15	Illumination x Distribution	1	Exclusion x Time	21
A x C x D	C x D x B(A)	Year x Month x Predation	15	Illumination x Distribution x Species	2	Exclusion x Time x Habitat	21
C x D x B(A)		Year x Month x Plots(Predation)	90	Illumination x Distribution x Subject(Species)	48	Exclusion x Time x Pairs(Habitat)	168

Table 11-11 Examples of partly-nested designs from the literature with one between plots factor and two crossed within plots (subjects) factors. There is only a single observation within each combination of A, B, C and D. See Section 11.8.2 for more details of specific examples.

Source	Source	df	F-ratio denominator
<i>Between plots</i>			
A	Marsh	1	Plots within (Marsh, Fence)
B	Fence	1	Plots within (Marsh, Fence)
A x B	Marsh x Fence	1	Plots within (Marsh, Fence)
C(AB)	Plots within (Marsh, Fence)	28	
<i>Within plots</i>			
D	Enrichment	2	Enrichment x Plots within (Marsh, Fence)
A x D	Marsh x Enrichment	2	Enrichment x Plots within (Marsh, Fence)
B x D	Fence x Enrichment	2	Enrichment x Plots within (Marsh, Fence)
A x B x D	Marsh x Fence x Enrichment	2	Enrichment x Plots within (Marsh, Fence)
D x C(AB)	Enrichment x Plots within (Marsh, Fence)	56	
E	Time	6	Time x Plots within (Marsh, Fence)
A x E	Marsh x Time	6	Time x Plots within (Marsh, Fence)
B x E	Fence x Time	6	Time x Plots within (Marsh, Fence)
A x B x E	Marsh x Fence x Time	6	Time x Plots within (Marsh, Fence)
E x C(AB)	Time x Plots within (Marsh, Fence)	168	
D x E	Enrichment x Time	12	Enrichment x Time x Plots within (Marsh, Fence)
A x D x E	Marsh x Enrichment x Time	12	Enrichment x Time x Plots within (Marsh, Fence)
B x D x E	Fence x Enrichment x Time	12	Enrichment x Time x Plots within (Marsh, Fence)
A x B x D x E	Marsh x Fence x Enrichment x Time	12	Enrichment x Time x Plots within (Marsh, Fence)
D x E x C(AB)	Enrichment x Time x Plots within (Marsh, Fence)	336	

Table 11-12 ANOVA table from complex partly design from Gough & Grace (1998) with two crossed and fixed between plots/subjects factors and two crossed and fixed within plots/subjects factors – see Section 11.8.3.

Data file for “split-plot” analysis

Factor A	Plots/subjects (B)	Factor C	Y
1	1	1	y_{111}
1	1	2	y_{112}
1	2	1	y_{121}
1	2	2	y_{122}
2	3	1	y_{231}
i	j	k	y_{ijk}

Data file for “repeated measures” analysis

Factor A	Plots/subjects (B)	C_1	C_k
1	1	y_{111}	y_{11k}
1	2	y_{121}	y_{12k}
1	3	y_{231}	y_{23k}
i	j	y_{ij1}	y_{ijk}

Table 11-13 Data coding for unreplicated “split-plot” analysis and for unreplicated “repeated measures” analysis.

Figure 11-1 Diagrammatic representation of the split-plot experiment from Wissinger et al. (1996). There are four ponds (only two shown here) in each of two hydroperiods (permanent, autumnal), the between plots factors. Within each pond, there were six cages, each containing one level of the within plots factor, competition treatment.

Figure 11-2 Part of data set for partly nested design, with p levels of factor A ($i = 1$ to p), q levels of factor B ($j = 1$ to q) nested within each level of A, r levels of factor C ($k = 1$ to r) crossed with factors A and B(A), and n replicate observations ($l = 1$ to n) within each combination (cell) of A, B(A) and C.

Figure 11-3 Variation in numbers of limpets under different trampling regimes and at different places, from Keough & Quinn (1998). Panel (a) shows number of limpets vs intensity of trampling for three rock platforms, and panel (b) shows variation among plots within platforms in overall abundance of limpets.

Figure 11-4 Mean square root transformed rate of buccal breathing for lung and buccal breathing toads for eight levels of O_2 concentration from Mullens (1993).

Figure 11-5 Mean log transformed density of birds (+1) for two habitats (ironbark and stringybark forests) and two regions (north and south) for twelve months from McGoldrick & Mac Nally (1998).









