CHAPTER 10

EXPERIMENTAL DESIGNS

(Version 4, 14 March 2013)

Page

Measurements in ecology must be not only done accurately and precisely but also carried out within the general framework of a good experimental design. As ecologists have attempted to do more field experiments, the difficulties and pitfalls of experimental design have begun to emerge. There are many excellent textbooks on experimental design (for example, Cox 1958, Mead (1988), Hinkelmann and Kempthorne 2008) and I shall not attempt to summarize the detailed discussion you may obtain by reading one of these statistical texts. Rather I shall concentrate on the simple principles of experimental design as they are applied to ecological studies. The aim of this chapter is to teach you about the general principles of ecological experimental design so that you can talk intelligently to a statistician or take up one of the statistical texts on the subject without a grimace.

10.1 GENERAL PRINCIPLES OF EXPERIMENTAL DESIGN

Experimental design is a term describing the logical structure of an experiment. Let us begin by defining some of the terms that are commonly used in discussions of experimental design. The important point is that if you are talking to a statistician, you should carefully adopt their language or confusion may well result. There are three aspects of the structure of an experimental design.

1. Treatment structure: This defines the set of treatments selected for comparison. Statisticians use the term *treatment* as a general term for any set of comparisons. For example, if you wished to compare the body size of a lizard species on two islands, the two islands would be called 'treatments'. The term is somewhat stilted for many ecological studies but it fits quite well for studies in which a manipulation is applied to a set of plots or chambers, e.g. increased CO_2, normal CO_2, reduced CO_2 affecting grass growth. Treatment factors can be qualitative or quantitative. The investigator selects the treatment structure to reflect the ecological hypotheses under investigation.

2. Design structure: This specifies the rules by which the treatments are to be allocated to the experimental units. This is the central issue of experimental design and covers much of the material this chapter deals with, and is discussed much more fully in Mead (1988). The important scientific point is that the design structure depends on the scientific questions being asked, and most importantly it dictates what type of statistical analysis can be carried out on the resulting data.

3. Response structure: This specifies the measurements to be made on each experimental unit. In particular it gives the list of response variables that are measured and when and where the measurements are to be made. These are the key decisions an ecologist must make to gather the data needed to answer the question under investigation. Even though the response structure is a critical element in experimental design, it is rarely discussed. Most fields of ecology have developed paradigms of response structure, but these must be explicitly recognized. For example, if you wish to measure body size of lizards only once each year, you could not investigate important seasonal variations in growth. Again the key is to have the design coordinated with the ecological questions.

An *experiment* is an attempt to test a hypothesis, and for ecologists hypotheses are typically suggested explanations for an ecological pattern or process. There are two broad types of experiments (Hurlbert 1984):

1. *Mensurative experiments* or *observational studies*: These experiments involve making some measurements on ecological units. The ecologist does not apply some treatment to the organisms or the plots but measures what currently exists. For example, you might measure the density of deer in forest and in grassland habitats to test the hypothesis that deer prefer forest habitats. These types of studies are sometimes referred to as *quasi-experiments* to emphasize that observational studies concentrate on a set of factors of interest but cannot control all the factors that may affect the data.

2. *Manipulative experiments*: these experiments involve assigning some treatment to the experimental units or plots. At least two treatments or manipulations are required. For example, an ecologist might burn one plot of grassland and leave a second plot unburned to test the hypothesis that burning reduces grasshopper abundance.

Most ecologists use the word "experiment" to apply only to manipulative experiments and the word “quasi-experiment” to apply to observational studies. This chapter concentrates on manipulative experiments, and the following chapter will discuss quasi-experimental studies (mensurative experiments or observational studies).

Experiments always involve some measurements being taken on the units of study, termed more precisely as *experimental units*. The concept of an *experimental unit* is critical for understanding the design of ecological experiments. An experimental unit is defined as *the smallest division of the experimental material such that any two units may receive different treatments (or treatment combinations) in a manipulative experiment or have different ecological conditions for a mensurative experiment*. The experimental units must be independent of one another and this independence requires that the experimental units are bounded or physically defined in such a way that what transpires in one experimental unit can have no effect on other experimental units. Consider two manipulative experiments and one mensurative experiment.

1. In a fire study, one 10-ha block of grassland may be burned, and one 10-ha block of grassland left unmanipulated. An ecologist might sample 50 one-square-meter plots on each of these two areas. In this case there are 2

experimental units (the two 10-ha blocks), and the 50 small plots are *subsamples* of each experimental unit. Ecologists at times talk of the 50 small plots as the 'unit of study' but the statistician more precisely uses the term experimental unit for the 10 ha blocks.

2. In a plant growth study, four fertilizer treatments (none, N, N+P, N+P+K) may be applied at random to 50 one-square-meter plots on each of two areas. In this case there are 50 experimental units on each area because any single one-square-meter plot might be treated with any one of the four fertilizers. Note that the experimental units must be separated by enough space to be independent of one another, so that for example the spraying of fertilizer on one plot does not blow or leak over to adjacent plots.
3. In a study of tree growth up an altitudinal gradient, a plant ecologist wishes to determine if growth rates decrease with altitude. The unit of study or experimental unit is a single tree, and a random sample of 100 trees is selected and the growth rate and altitude of each tree is recorded. The ecological condition or variable of interest is the altitude of each tree.

So the first step in specifying your experimental design is to determine the experimental units. When a statistician asks you about the number of *replicates*, he or she usually wants to know how many experimental units you have in each treatment. Most of the difficulty which Hurlbert (1984) has described as *pseudoreplication*[1] arises from a failure to define exactly what the experimental unit is.

There is a critical distinction between the experimental units and the measurements or samples that may be taken from or made on the experimental units. Statisticians use the term *sampling unit* for these measurements or samples. In the first example above, the 50 subsamples taken on each experimental unit are the sampling units. Hurlbert (1990a) prefers the term *evaluation unit* for the older term sampling unit, so note that *evaluation unit* = *sampling unit* for statistical discussions. The precise definition of evaluation unit is *that element of an experimental unit on which an individual measurement is made*. In the second example above, each of the 50 individual plots treated with fertilizer is an evaluation unit or sampling unit as well as an experimental unit. In the first example above each of the 50 individual plots is an evaluation unit but it is not an experimental unit.

[1] *Pseudoreplication* occurs when experimental measurements are not independent. If you weigh the same fish twice you do not have two replicates

A general rule of manipulative experimentation is that one should have a "control". A *control* is usually defined as an experimental unit which has been given no treatment. Thus a control is usually the baseline against which the other treatments are to be compared (Fig. 10.1). In some cases the control unit is subjected to a sham treatment (e.g. spraying with water vs. spraying with a water + fertilizer solution). The term ‘control’ is awkward for ecologists because it has the implication that the ecologist controls something like the temperature or the salinity in a study. Perhaps the term ‘control’ must be maintained in order to talk to statisticians but ecologists should use the more general term ‘baseline ‘ for the *baseline treatment* in manipulative experiments or the *baseline comparison* in observational studies.

Mensurative experiments need a statistical ‘control’ in the sense of a set of experimental units that serve as a baseline for comparison. The exact nature of the controls will depend on the hypothesis being tested. For example, if you wish to measure the impact of competition from species A on plant growth, you can measure plants growing in natural stands in a mixture with species A and compare these with plants growing in the absence of species A (the baseline comparison plants). The baseline for comparison should be determined by the ecological questions being asked.

There is one fundamental requirement of all scientific experimentation:

Every manipulative experiment must have a control.

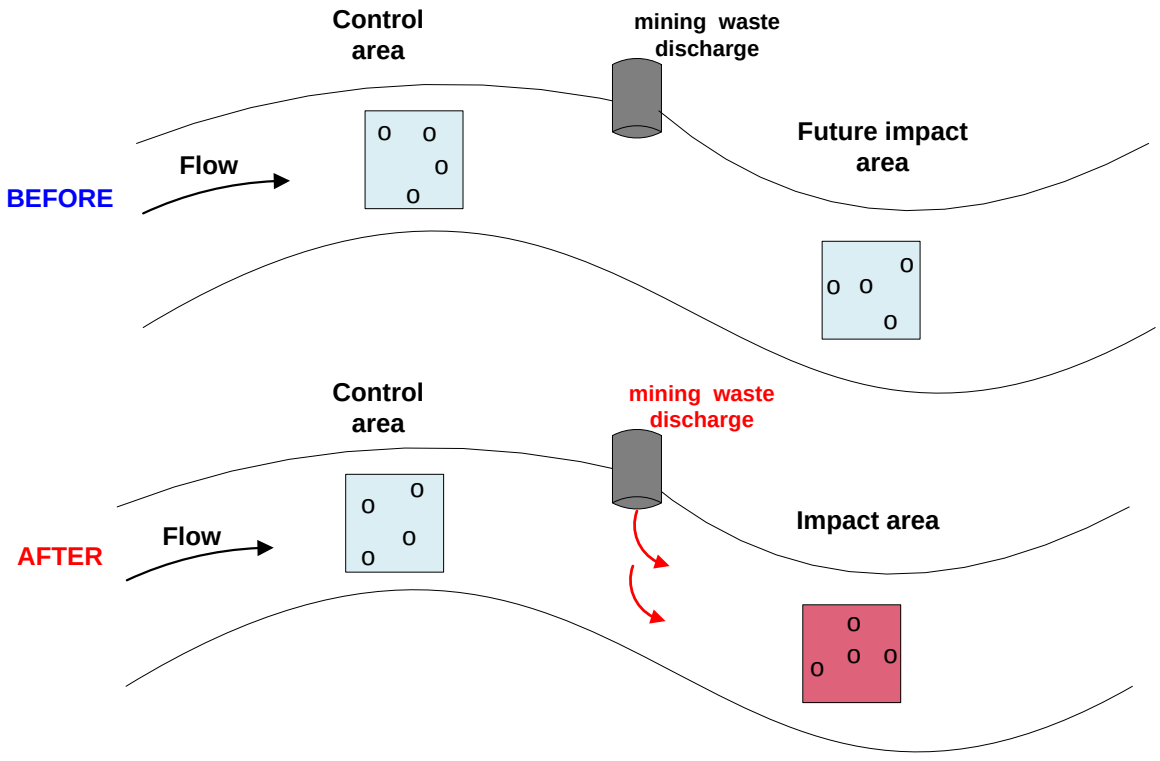
If a control is not present, it is impossible to conclude anything definite about the experiment[2].· In ecological field experiments there is so much year to year variation in communities and ecosystems that an even stronger rule should be adopted:

Every manipulative ecological field experiment must have a contemporaneous control.

Because of the need for replication this rule dictates that field experiments should utilize at least 2 controls and 2 experimental areas or units. Clearly statistical power

will increase if you have 3 replicates rather than 2, so 2 is a minimum number rather than the most desirable number.

Before-after comparisons are statistically powerful because every experimental unit can serve as its own control. For this reason most experimental design texts recommend this approach. Fisheries and wildlife management often rely on time series of data on abundance in which a management manipulation occurs at a given time. But unless there is a contemporaneous control, all before-after comparisons must assume homogeneity over time, a dubious balance-of-nature model that has been found invalid time and time again in ecological work (Green 1979). Populations and communities change over time in a way we only dimly understand and to achieve reliable statistical inference we need spatial controls for all ecological experiments. Figure 10.1 illustrates the need for both temporal and spatial controls in ecological studies.



[2] In some experiments two or more treatments are applied to determine which one is best. One treatment can act as a “control” for another.

Figure 10.1 Example of the requirements for a control in ecological studies. A stream is to be subjected to nutrient additions from a mining operation. By sampling both the control and the impact sites before and after the nutrient additions, both temporal and spatial controls are utilized. Green (1979) calls this the BACI design (Before-After, Control-Impact) and suggests that it is an optimal impact design. In other situations one cannot sample before the treatment or impact is applied, and only the spatial control of the lower diagram is present. (Modified from Green 1979)

There are at least six sources of variability that can cloud the interpretation of experiments (Table 10.1). These sources of confusion can be reduced by three statistical procedures - *randomization*, *replication*, and *design control*.

TABLE 10.1 POTENTIAL SOURCES OF CONFUSION IN AN EXPERIMENT AND MEANS FOR MINIMIZING THEIR EFFECT

Source of confusion	Features of an experimental design that reduce or eliminate confusion
1. Temporal change	Control treatments
2. Procedure effects	Control treatments
3. Experimenter bias	Randomized assignment of experimental units to treatments
	Randomization in conduct of other procedures "Blind" procedures[a]
4. Experimenter-generated variability	Replication of treatments
5. Initial or inherent variability among experimental units	Replication of treatments
	Interspersion of treatments
	Concomitant observations
6. Nondemonic intrusion[b]	Replication of treatments
	Interspersion of treatments

[a] Usually employed only where measurement involves a large subjective element.

[b] Nondemonic intrusion is defined as the impingement of chance events on an experiment in progress.

Source: After Hurlbert, 1984.

10.1.1 Randomization

Most statistical tests make the assumption that the observations are independent. As in most statistical assumptions, independence of observations is an ideal that can never be achieved. One way to help achieve the goal of independent observations is to randomize by taking a random sample from the population or by assigning treatments at random to the experimental units. If observations are not independent, we cannot utilize any of the statistical analyses that assume independence.

Randomization is also a device for reducing bias that can invade an experiment inadvertently. Randomization thus increases the accuracy of our estimates of treatment effects.

In many ecological situations complete randomization is not possible. Study sites cannot be selected at random if only because not all land areas are available for ecological research. Within areas that are available, vehicle access will often dictate the location of study sites. The rule of thumb to use is simple:

Randomize whenever possible.

Systematic sampling is normally the alternative to random sampling (see Chapter 8, page 000). While most statisticians do not approve of systematic sampling, most ecologists working with field plots use some form of systematic layout of plots. Systematic sampling achieves coverage of the entire study area, which ecologists often desire. There is so far no good evidence that systematic sampling in complex natural ecosystems leads to biased estimates or unreliable comparisons. But there is always a residue of doubt when systematic sampling is used, and hence the admonition to random sample when possible. A good compromise is to semi-systematically sample. Randomization is a kind of statistical insurance.

Randomization should always be used in manipulative experiments when assigning treatments to experimental units. If some subjective procedure is used to assign treatments, the essential touchstone of statistical analysis is lost and

probability statements cannot be assigned to the resulting data because of the possibility of bias.

10.1.2 Replication and Pseudoreplication

Replication means the *repetition* of the basic experimental units within one treatment and the main idea of replication is enshrined in the first commandment of statistics:

Let $n \geq 2$

Replication of experimental units is necessary to permit an estimate of "experimental error" which is the basic unit of measurement for assessing statistical significance or for determining confidence limits. Replication is one way of increasing the precision of a statistical estimate, and this simple fact is the basis of the graduate students' credo of "Get as many samples as possible!".

Note that replication can be applied at any level. Typically we speak of replicate experimental units, but we can also have replicate evaluation units, replicate subsamples, and replicate measurements. Consequently it is best to specify the type of replication or replicate with an adjective or noun, so that the meaning is precise.

Replication of experimental units is a type of insurance against the intrusion of chance events on ecological experiments. Chance events are one of the major sources of interference or "noise" in field ecology, and ecologists often use chance events as a good excuse when observations fail to match theory (Weatherhead 1986). Chance events are most troublesome when they impinge on one experimental unit and not the others. For example, one plot might be accidentally burned during a study, or be subject to an insect outbreak. Replication is the only way to avoid these difficulties.

Where should experimental plots be placed spatially and in what time sequence should manipulations be done? Ecologists always face these questions about the *interspersion* of treatments in space and in time (Hurlbert 1984). In field studies interspersion is more important than randomization because it deals with the critical question of how the experimental units should be distributed in

heterogeneous space. Figure 10.2 illustrates good and poor designs for interspersion of a simple 2-treatment field experiment. Let us look at each design briefly.

1. *Completely Randomized Design*: This is the simplest design recommended by many statistical tests (Fig. 10.2, A-1). Hurlbert (1984) pointed out that strict randomization can result in treatments being spatially segregated by chance, especially if only a few treatment replicates are possible. Spatial segregation will produce spurious treatment effects when there are preexisting gradients in the study area. For this reason Hurlbert (1984) recommends against this statistical design in ecological studies when treatment replicates are few, even though technically this is a perfectly acceptable statistical design to all professional statisticians.

2. *Randomized Block Design*: In this design the experimental units are grouped together in *blocks*. In ecological use the blocks may be areas of habitat, or time periods, or rooms within a greenhouse. The main point is that the blocks are relatively uniform internally, and the differences between blocks may be large or small. This is an excellent design for most field experiments because it automatically produces an interspersion of treatments (c.f. Fig. 10.2, A-2) and

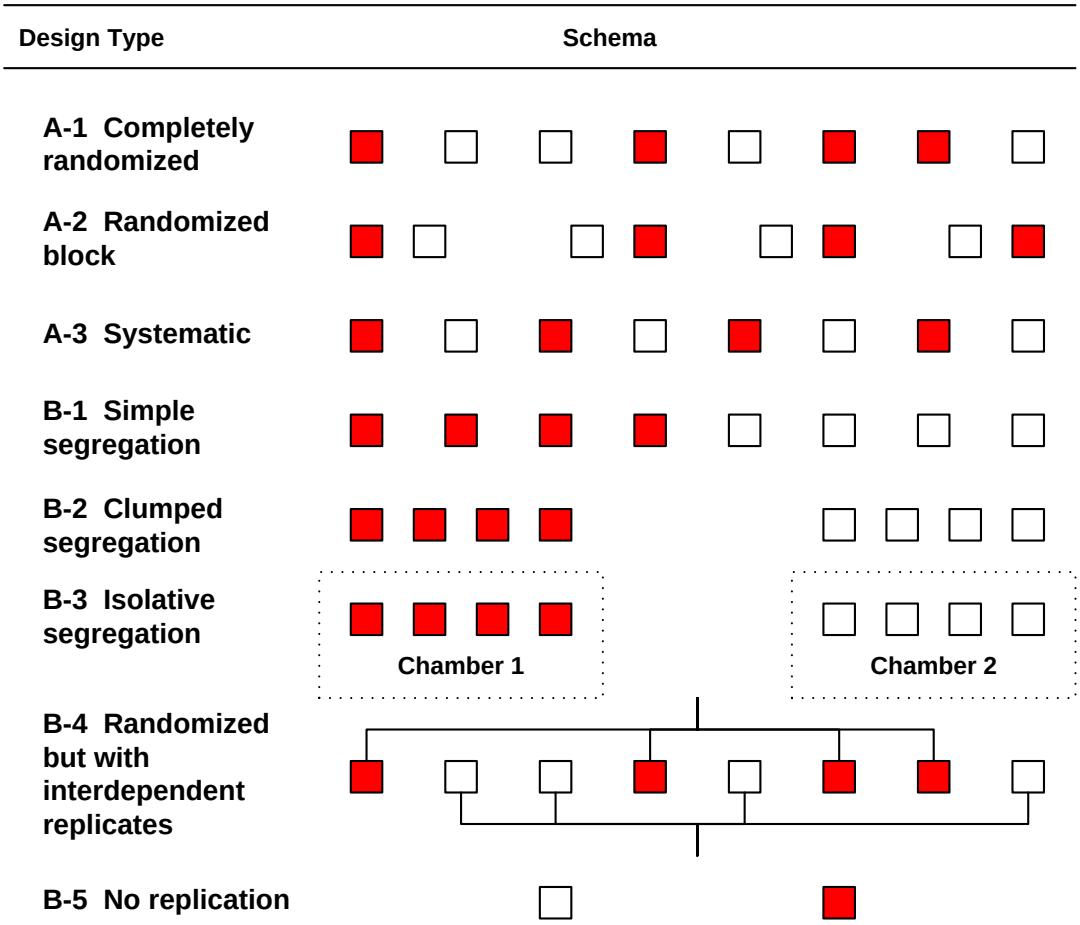


Figure 10.2 Schematic representation of various acceptable modes (A) of interspersing the replicates of two treatments (shaded red, unshaded) and various ways (B) in which the principle of interspersion can be violated. (From Hurlbert, 1984.)

thus reduces the effect of chance events on the results of the experiment. One additional advantage of the randomized block design is that whole blocks may be lost without compromising the experiment. If a bulldozer destroys one set of plots, all is not lost.

3. *Systematic Design*: This design achieves maximum interspersion of treatments at the statistical risk of errors arising from a periodic environment. Since spatially periodic environments are almost unknown in natural ecosystems, this problem is non-existent for most ecological work. Temporal periodicities are however quite common and when the treatments being applied have a time component one must be more careful to avoid systematic designs.

Field experiments can be assigned systematically or in random blocks on the basis of some measured property of the experimental plots. For example, we might know the density of deer on 8 study plots and block these into 4 low density plots and 4 high density plots. The danger of assigning randomized blocks in this way is that spatial segregation of plots may occur as a side effect. Figure 10.3 illustrates some examples of field studies with inadequate interspersion of treatments. Hurlbert (1984) recommends a hybrid approach in these cases to assign treatments to achieve the two goals of maximum interspersion and maximum similarity within the blocks. In practice this is the way many large-scale ecological experiments must be set up in the field, even though a strict statistician would not approve of this subjective way of assigning treatments.

4. Segregated Designs: The simple segregated designs shown in Figure 10.2 (B-1 to B-3) are rarely found in ecological field studies but may occur in laboratory experiments. Even if laboratory or greenhouse experiments are set up with the same initial conditions, subsequent chance events may have uneven effects if the treatments are isolated. Hurlbert (1984) gives several examples from laboratory work.

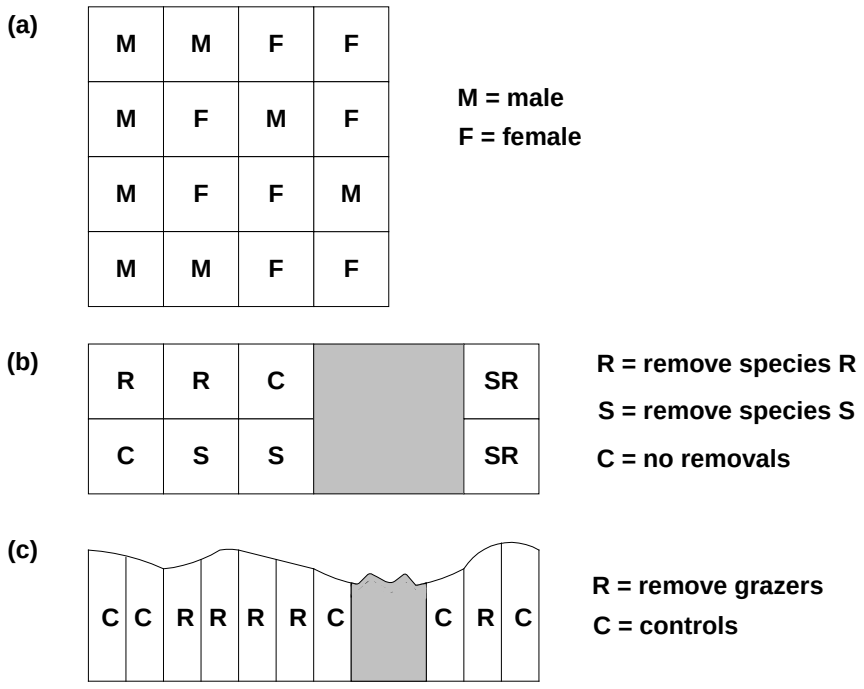


Figure 10.3 Three experimental layouts that show partial but inadequate interspersion of treatments. (a) An experiment to compare predation rates on male vs. female floral parts (Cox 1982). (b) Impact of removals on dispersal rates of two rodent species into field plots (Joule and Cameron (1975). (c) Effects on algal growth of grazer removals in rocky intertidal areas (Slocum 1980). Shaded areas of the diagrams represent unused areas. In all these cases Hurlbert (1984) recommends using subjective assignments of treatments to even out the interspersion. (After Hurlbert 1984)

Another type of segregated experimental design is less easily seen - one in which the replicates are physically interdependent (Fig. 10.2, B-4). In this case the replicates may share a common heating duct, or a common filtration or water system, even though they are spatially separated. It is essential that all replicates of all treatments are hooked up together to the same heating, water, etc. systems, or else that each replicate has its own system.

Randomization and interspersion often conflict in field experiments. Some statistically valid designs will produce on occasion very spatially segregated treatment plots. From an ecological view the best approach is to reject these segregated layouts and go back and re-randomize until you get a layout with an acceptable amount of interspersion (Cox 1958 pp. 86-87). Segregated layouts are not usually a problem when there are many treatment replications.

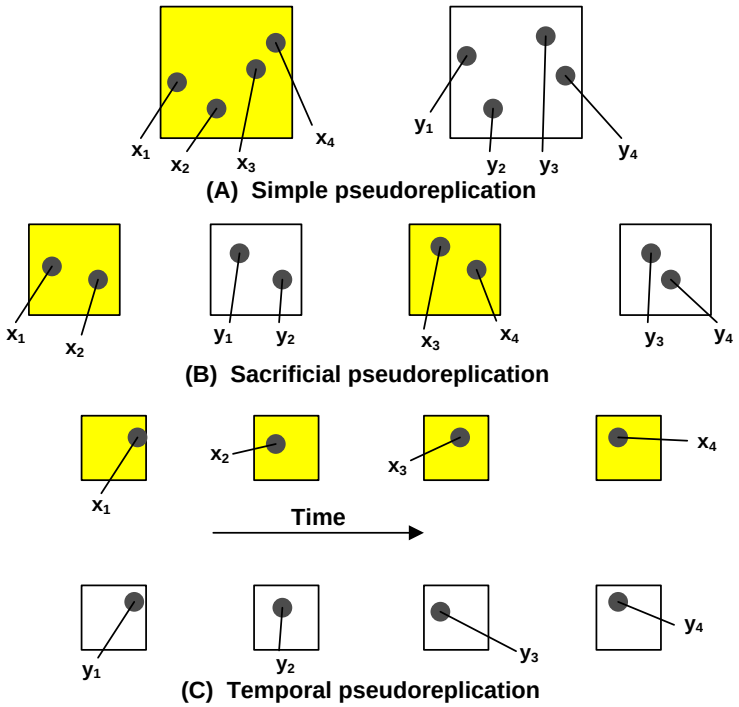


Figure 10.4 The three most common types of pseudoreplication. Yellow and white boxes represent experimental units receiving two different treatments. Each dot represents an evaluation unit or sample measurement. Pseudoreplication is a consequence, in each example, of statistically testing for a treatment effect by means of procedures (e.g., *t*-test, *U*-test) which assume, implicitly, that the four data for each treatment have come from four independent experimental units (=treatment replicates). (From Hurlbert, 1984.)

Hurlbert (1984) introduced the useful idea of *pseudoreplication* to describe a statistical error of using treatment replicates from experiments which violate the principle of interspersion (Fig. 10.2, B-1 to B-5). The basic statistical problem is that in these cases "replicates" are not independent, and the first assumption of statistical inference is violated. Hurlbert (1984) reported that in two separate surveys 26% and 48% of the ecological papers surveyed showed the statistical error of pseudoreplication. Underwood (1981) found statistical errors in 78% of the papers he surveyed in marine ecology. Clearly there is a need to improve the statistical design of ecological studies.

Three types of pseudoreplication can be recognized (Fig. 10.4). The simplest and most common type of pseudoreplication occurs when there is only one replicate per treatment. For example, there may be one large burned area and one unburned area. If several 1 m^2 plots are measured within each area, these 1 m^2 plots are *not* replicates (they are subsamples) and they should not be used in a *t*-test to compare burned vs. unburned areas in general. A *t*-test would only answer the specific question of whether this particular burned area differed from this particular unburned area and we would not know if we were measuring an area effect or a burn effect. Sacrificial pseudoreplication occurs when there is a proper, replicated experimental design but the data for the replicates are pooled together prior to measurement or prior to statistical analysis. This is a simple statistical error and should be a recoverable problem in data analysis unless the samples were physically pooled before measurement[3]. Temporal pseudoreplication is also common in ecological experiments in which a time series of data are accumulated. For example, with

[3] In some cases, for example plant clip plots, all the plants from the two plots 4might be put together in a single sample bag before taking them back to the laboratory for sorting.

burned and unburned plots one might return to sample quadrats each week for two months after the fire. Successive samples over time from a single experimental unit are clearly not independent samples and can be analysed as a series of date-by-date *t*-tests or ANOVAs or as a repeated measures ANOVA (see page 441).

10.1.3 Design Control

One objective of a good experimental design is to reduce the size of experimental error, or in other words to make the conclusions more precise. There are four general ways to increase precision in any statistical comparison.

1. Use more homogenous experimental units. This advice is useful to laboratory experimenters but is difficult to use in many field experiments.
2. Use information provided by related variables that can be measured in each experimental unit. The analysis of covariance is the simplest example of this approach (Green 1979).
3. Use more replicates. This is the bulldozer approach to statistical precision. It always works, if you have the time, money, and enough space.
4. Use a more efficient experimental design. By this statisticians mean the amount of balancing and blocking in the layout of the experiment.

We shall discuss the general idea of grouping the experimental units into homogeneous blocks on page 438. Balanced designs, in which an equal number of replicates are used for each treatment, are always more efficient than unbalanced designs of the same type. One of the unfortunate statistical problems of field ecologists is to have unbalanced designs routinely. Often one cannot avoid unequal replication, but as a general rule one should try to achieve balance even in field studies.

10.2 TYPES OF EXPERIMENTAL DESIGNS

There is a wide and ever increasing range of experimental designs but fortunately for ecologists they fall into relatively few classes and many are of limited use for ecological research. I shall describe very briefly and non-technically five types of designs that are useful in ecology. For technical details please see Mead (1988) or Hinkelmann and Kempthorne (2008).

Before we discuss experimental designs, we must define fixed and random classifications. The decision about whether a treatment[4] in ANOVA is fixed or random is crucial for all hypothesis testing (see Mead 1988).

Fixed Factors: **1.** All levels of the classification are in the experiment, or
2. The only levels of interest to the experimenter are in the experiment, or
3. The levels in the experiment were deliberately and not randomly chosen

Random Factors: **1.** All levels in the experiment are a random sample from all possible levels

Thus *sex* is a fixed factor because both sexes are studied, and *temperature* (10^o, 16^0, 27^0) could be a fixed factor if these are the only temperatures the experimenter is interested about or a random factor if these are a random sample of all possible temperature levels. It is important as a first step in experimental design to decide whether the factors you wish to study will be fixed or random factors since the details of statistical tests differ between fixed and random factor designs.

10.2.1 Linear Additive Models

All of the complex designs used in the analysis of variance can be described very simply by the use of linear additive models. The basic assumption underlying all of these models is *additivity*. The measurement obtained when a particular treatment is applied to one of the experimental units is assumed to be:

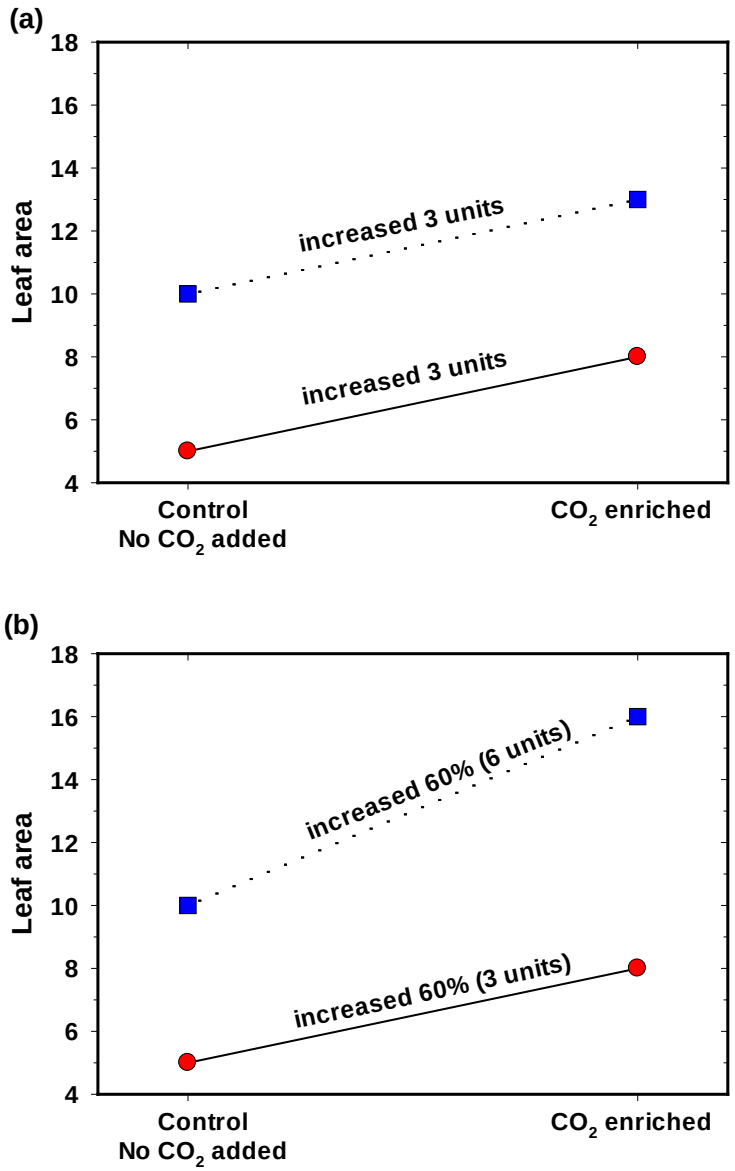
$$\left\{\begin{array}{l}\text{A quantity depending}\\ \text{only on the particular}\\ \text{experimental unit}\end{array}\right\} + \left\{\begin{array}{l}\text{A quantity depending}\\ \text{on the treatment}\\ \text{applied}\end{array}\right\}$$

The essential feature is that the treatment effect *adds* on to the unit term, rather than multiplying. Figure 10.5 illustrates this idea graphically. A second critical assumption is that the treatment effects are constant for all experimental units.

[4] Treatments are called *factors* in most statistics discussions of ANOVA, and each factor has several *levels*. For example, *sex* can be a factor with 2 levels, *males* and

Finally, you must assume that the experimental units operate independently so that treatment effects do not spill over from one unit to another.

These are the essential features of *linear additive models* which form the core of modern parametric statistics. Consider one simple example of a linear additive model. The density of oak seedlings was measured on a series of 6 burned and 6 unburned plots. The linear additive model is:



females.

Figure 10.5 Schematic illustration of the concept of additivity that is critical to all linear additive models in the analysis of variance. (a) Addition of CO_2 increases leaf area by the same amount no matter what the initial size of the plant. This would be highly unlikely in any growth experiment. (b) Addition of CO_2 increases leaf area proportionally by 60%. Thus the treatment effect is not additive but is multiplicative.

TABLE 10.2 HYPOTHETICAL DATA FOR OAK SEEDLING DENSITY[B] ON EACH OF SIX BURNED AND SIX UNBURNED PLOTS

Plot no.	Unburned plots	Plot no.	Burned plots
1	6	7	1
2	9	8	2
3	5	9	2
4	8	10	1
5	11	11	4
6	9	12	2
n = 6		n = 6	
Mean =	8.0		2.0
	Total n = 12		
	Grand mean = 5.0		

[a] Number of trees per square meter.

$$\left\{\begin{matrix}\text{Density of}\\ \text{oak seedlings}\\ \text{on plot}\end{matrix}\right\} = \left\{\begin{matrix}\text{Average density}\\ \text{of oak seedlings}\\ \text{on burned and}\\ \text{unburned plots}\end{matrix}\right\} + \left\{\begin{matrix}\text{Effect of}\\ \text{burning or}\\ \text{not burning}\end{matrix}\right\} + \left\{\begin{matrix}\text{Experimental}\\ \text{error}\end{matrix}\right\}$$

Or more abstractly:

$$Y_{ij} = \mu + T_i + e_{ij}$$

where:

Y = Variable being measured
μ = Mean value of the Y variable
T = Treatment effect
e = Experimental error
i = Treatment number (1= burning, 2= not burning)
j = Replicate number (1, 2, …, 6)

Linear additive models are often written as deviations:

$$Y_{ij} - \mu = T_i + e_{ij}$$

Interest usually centers on the treatment effects which can be estimated from the observed means:

$$\begin{Bmatrix}\text{Effect of burning}\\ \text{on density}\end{Bmatrix} = \begin{Bmatrix}\text{Average density}\\ \text{in burned plots}\end{Bmatrix} - \begin{Bmatrix}\text{Average density}\\ \text{in all plots}\end{Bmatrix}$$

Note that the effects of burning in this case are being related to a hypothetical world which is half burnt and half unburnt. From the data in Table 10.2:

$$\begin{aligned}\text{Effect of burning on density } (T_1) &= 2.0 - 5.0\\ &= -3.0\end{aligned}$$

Thus burning reduces density by 3.0 trees/m^2. The effect of not burning is similarly:

$$\begin{aligned}\begin{Bmatrix}\text{Effect of not}\\ \text{burning } (T_2)\end{Bmatrix} &= \begin{Bmatrix}\text{Average density on}\\ \text{unburned plots}\end{Bmatrix} - \begin{Bmatrix}\text{Average density}\\ \text{on all plots}\end{Bmatrix}\\ &= 8.0 - 5.0\\ &= +3.0\end{aligned}$$

Note that for designs like this with two levels of treatment, the measured effects are identical in absolute value but opposite in sign.

Note that treatment effects are always relative and we need to estimate effects of one treatment in comparison with the other so that we can determine the ecological significance of the treatment. For example, in this burning experiment an ecologist wants to know the difference between burnt and unburnt:

$$\begin{Bmatrix}\text{Difference between}\\ \text{two treatments}\end{Bmatrix} = T_1 - T_2$$

$$\begin{Bmatrix}\text{Difference between}\\ \text{burned and unburned}\end{Bmatrix} = -3.0 - 3.0 = 6.0 \text{ trees/m}^2$$

You can also use these treatment effects to decompose the data from each individual quadrat. For example, for plot # 5:

$$\begin{aligned}Y_{ij} &= \mu + t_i + e_{ij}\\ 11 &= 5.0 + 3.0 + 3.0\end{aligned}$$

Since you know there were 11 trees/m^2 in this plot, and the overall density μ is 5.0 (the grand mean), and the effect of not burning (T_2) estimated above is +3.0, the experimental error term must be 3.0 to balance the equation. Note again that the "error" term measures inherent biological variability among plots, and not "error" in the sense of "mistake".

Linear additive models may be made as complex as you wish, subject to the constraint of still being linear and additive. Many statistical computer packages will compute an ANOVA for any linear additive model you can specify, assuming you have adequate replication. Linear additive models are a convenient shorthand for describing many experimental designs.

10.2.2 Factorial Designs

When only one factor is of interest, the resulting statistical analysis is simple. But typically ecologists need to worry about several factors at the same time. For example, plankton samples may be collected in several lakes at different months of the year. Or rates of egg deposition may be measured at three levels of salinity and two temperatures. Two new concepts arise when one has to deal with several factors - *factorials* and *interaction*.

The concept of factorials is just the commonsense notion that all treatments of one factor should be tried with all treatments of the other factors. Thus if egg laying rates for a particular species of amphipod are measured in a laboratory study at 3 salinities and 2 temperatures, the investigator should do all 3 salinities at each of the two temperatures. The test for a factorial arrangement of treatments is simple - make a table!

		Salinity			
		None	Low	Medium	High
Temperature	Low				
	High				

Statisticians refer to these eight boxes as *cells* in an ANOVA. You must have observations in each box or cell of this table to have a factorial design. In the best of all worlds, you will have equal sample sizes in each box (or *cell*) and thus a

balanced design. But in many ecological situations sample sizes are unequal and you will have an *unbalanced* design.

In an ideal world all factors operate independently. Salinity will raise or lower egg production, and temperature will independently change it as well. In the real world, factors enhance or interfere with one another and thus are not independent in their effects. Statisticians say factors *interact*. The simplest way to look at and understand interactions is graphically. Figure 10.6 illustrates a hypothetical set of data for this experimental design with and without interactions. When there are no interactions in the data, a graph will show only sets of parallel lines, as in Figure 10.6(a). In this example level 2 of treatment A (high temperature, A_2) always has a higher mean egg productions than level 1 (low temperature, A_1), regardless of what the salinity is. When interaction is present, the lines diverge (Fig. 10.6(b)) or cross (Fig. 10.6 (c)). In this example high temperature A_2 stimulates more egg laying in the first two low salinities but lowers egg production in the second two high salinities, relative to low temperature treatment A_1.

Another way to understand what interactions are is to ask a simple question about each of the factors in the experiment. For example, from Figure 10.6 in which the two factors are *temperature* and *salinity*, we can ask:

1. What effect does *temperature* have on number of eggs produced?

2. What effect does *salinity* have on number of eggs produced?

When there is no interaction the answer to these questions is straightforward - temperature 2 has twice the number of eggs of temperature 1, or high salinity increases egg production to 3 times that of low salinity, for example. But when there is interaction, these questions have no simple answer and you must reply "It all depends ....." and give a more detailed answer, like:

For low temperature A_1, salinity has no effect on egg production, but for high temperature A_2 there is a strong effect of high salinity on egg production.

Interactions produce statistical headaches but interesting ecology so there is always a conflict in factorial designs between the desire for simplicity with no interactions or for complexity with interactions that require ecological understanding.

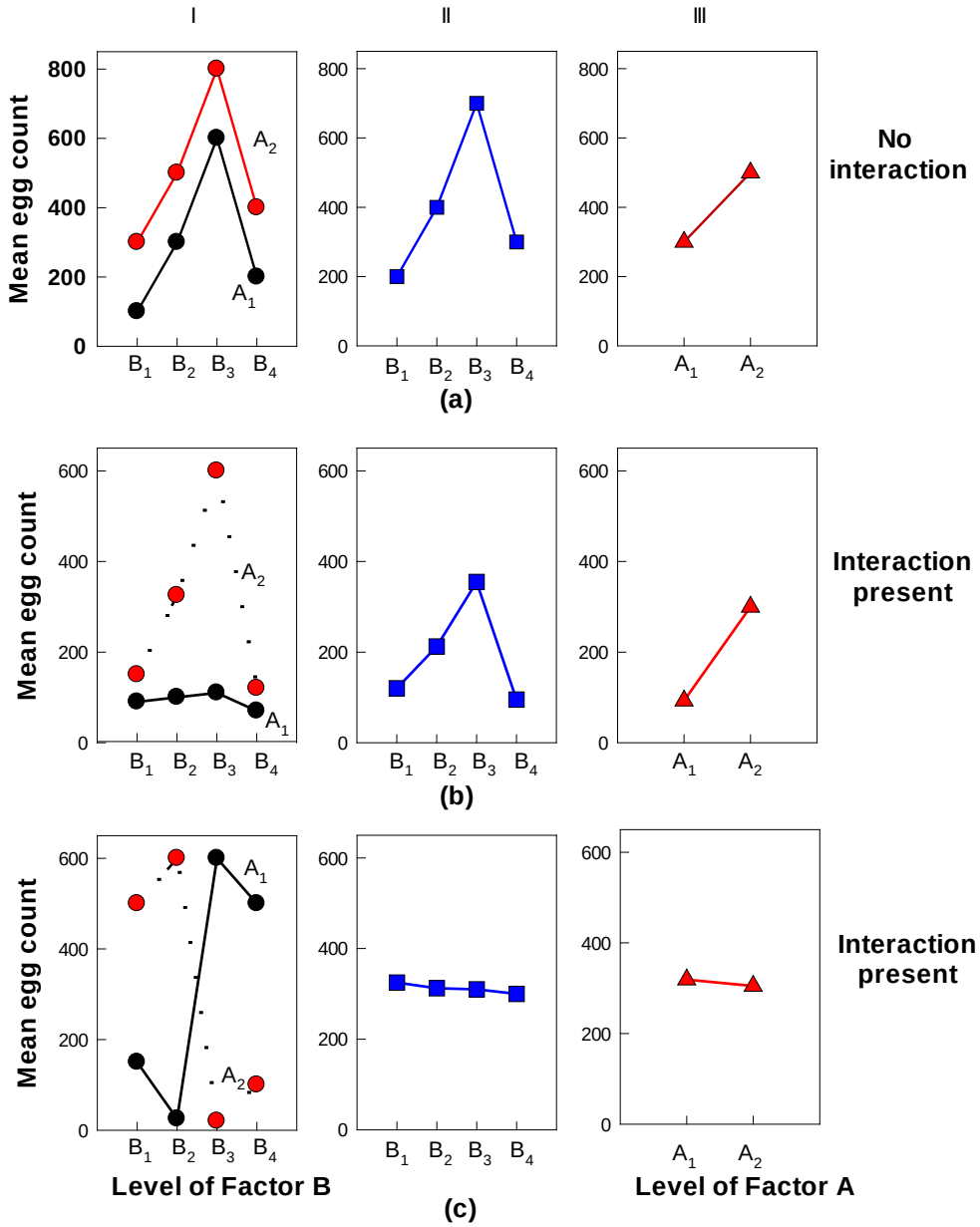


Figure 10.6 Graphical illustration of interactions between two factors. In each set of graphs, the means of hypothetical data are plotted. In column I, the means (number of eggs laid) from two temperature treatments (A_1 black and A_2 red) are plotted for four salinity levels (B_1 to B_4). In column II, the mean egg count in each salinity, averaged from the two temperatures, are plotted. In column III, the mean egg count in each temperature are averaged from the four salinities. (a) No interaction between the two factors; (b) and (c) different types of interactions present between the two factors.

Linear additive models for factorial designs must therefore have a term in them for interactions. For example, a 2-factor analysis like that in Figure 10.6 can be represented by:

$$Y_{ijk} - \mu = A_i + B_j + AB_{ij} + e_{ijk}$$

where:

$Y_{ijk} - \mu =$ Deviation of observed value from the grand mean μ
$A_i =$ Main effect of level i of factor A
$B_j =$ Main effect of level j of factor B
$AB_{ij} =$ Interaction term between A_i and B_j
$e_{ijk} =$ Experimental error

Each of these terms can be estimated, as illustrated in Box 10.1, and the significance of each term determined by the analysis of variance. These techniques are thoroughly described in most statistics books and are not repeated here.

It is important to remember that there is a priority of testing in factorial designs. You should first ask whether the interactions are significant statistically. If they are, you should stop there and find out what is happening. It is misleading to present and analyze tests of significance for the main effects in a factorial when the interaction term is significant. The important thing is to explain the interaction. Figure 10.6 illustrates why. Testing for the main effects is the same as asking the simple questions (1) and (2) above (page 443).

In order to calculate the complete analysis of variance for a factorial design, each cell (or combination of treatments) must have 2 or more replicates. Treatment replication allows us to calculate the interaction terms and to judge their statistical significance. But this means that the total number of replicates grows very large in a factorial design when there are many factors or many levels within a factor. This alone seems to constrain most ecological experiments to two or three factor designs with a maximum of 4-5 levels of each factor. While there are no theoretical limits to the complexity of factorial designs, in ecology there are very real practical limits that restrain both field and laboratory studies.

Box 10.1 ESTIMATION OF MAIN EFFECTS AND INTERACTIONS IN A TWO FACTOR COMPLETELY RANDOMIZED DESIGN

For two variables of classification we can define:

$$A_i = \text{Main effect of level } i \text{ of factor } A = \bar{x}_i - \bar{\bar{x}}$$

$$B_j = \text{Main effect of factor } B \text{ level } j = \bar{x}_j - \bar{\bar{x}}$$

$$AB_{ij} = \text{Interaction effect for cell } ij = \bar{x}_{ij} - A_i - B_j + \bar{\bar{x}}$$

Example

Mean growth rates of trees (millimeters per year):

		Habitat (= factor *A*)			
		A	*B*	*C*	
Tree species	*X*	50	55	60	(Note that these are
	Y	60	67	60	cell means)
	Z	70	80	90	

Grand mean = $\bar{\bar{x}}$ = 65.777

Habitat mean: $\bar{x}_A = 60.00$ Species means: $\bar{x}_X = 55.00$

$\bar{x}_B = 67.33$ $\bar{x}_Y = 62.33$

$\bar{x}_C = 70.00$ $\bar{x}_Z = 80.00$

Main effects of habitats:

1. Habitat *A* = 60 - 65.777 = -5.78

2. Habitat *B* = 67.33 - 65.7777 = +1.56

3. Habitat *C* = 70 - 65.777 = +4.22

Main effects of tree species:

1. Species *X* = 55 - 65.777 = -10.78

2. Species *Y* = 62.33 - 65.777 = -3.44

3. Species *Z* = 80 - 65.777 = +14.22

Interaction effects:

1. Cell *A*-*X* = 50 - 60 - 55 - 65.777 = +0.78

(Alternatively, cell *A*-*X* deviates from the population mean by -15.78 units,

of which -5.78 can be ascribed to a low average growth in habitat *A* and -10.78 can be ascribed to low average growth of species *X*; the remainder +0.78 is interaction.)

2. Cell *B-X* = 55 - 67.33 - 55 + 65.777 = -1.56

3. Cell *C-X* = 60 - 70 - 55 - 65.777 = +0.78

4. Cell *A-Y* = 60 - 60 - 62.33 + 65.777 = +3.45

5. Cell *B-Y* = 67 - 67.33 - 62.33 + 65.777 = + 3.11

6. Cell *C-Y* = 60 - 70 - 62.33 + 65.777 = -6.56

7. Cell *A-Z* = 70 - 60 - 80 + 65.777 = -4.22

8. Cell *B-Z* = 80 - 67.33 - 80 + 65.777 = -1.56

9. Cell *C-Z* = 90 - 70 - 80 + 65.777 = +5.78

Thus the largest interaction are found in cell *C-Y* and *C-Z*. Note that the sum of the main effects and the interaction effects is zero (within rounding errors).

For example,

$$\left\{\begin{matrix}\text{Main effect of habitat } A \\ + \text{ main effect of habitat } B \\ + \text{ main effect of habitat } C\end{matrix}\right\} = 0$$

$$-5.78 + 1.56 + 4.22 \approx 0$$

This follows from the definitions of these terms.

The magnitude of the main effects and the interactions are the interesting parts of an ANOVA to an ecologist. The usual ANOVA table of mean squares and *F* values is singularly uninformative about the ecological effects studied in the experiment and should almost never be presented in a paper.

10.2.3 Randomized Block Design

Randomized block designs always begin with the identification of "blocks". Blocks are relatively homogeneous groups of experimental units. For example, a muskrat litter could be a block of 5 individual animals. It is well-known that litter mates are more similar than individual muskrats in a random sample. A grassland on a SE-facing slope could be a block. In this case the environmental conditions (soil, temperature, rainfall) would be more similar within this block of grassland than between this site and a nearby SW-facing plot. *Blocks* can be constructed around any known or suspected source of variation. Plots of habitat are the most obvious type of blocks in field ecology but a block could also be a room in a greenhouse, a

lab of aquaria tanks, a day of the week, a group of animals of similar weight, the measurements taken on instrument *X*, or the data collected by technician *Y*. Block-to-block variation is a known or suspected source of variation, and because of this it is not usual to ask in an analysis of variance whether the blocks are significantly different from one another - you already *know* that they are. The important feature of block designs is that the variation between blocks is removed from the experimental error term in an ANOVA, thereby increasing the precision of the experiment. Box 10.2 illustrates how this is achieved.

Box 10.2 COMPARISON OF THE RELATIVE EFFICIENCY OF RANDOMIZED BLOCK AND COMPLETELY RANDOMIZED DESIGNS

A plant ecologist measured the growth response of cotton grass (*Eriophorum angustifolium*) to four fertilizer treatments in five tundra locations in northern Alaska. She obtained these data:

	Location					
Fertilizer	B	M	R	S	Q	Means
None	10	6	11	2	5	6.8
N	58	45	55	50	37	49.0
N + P	63	43	68	41	39	50.8
N + P + K	68	47	63	43	40	52.2
Means	49.75	35.25	49.25	34.0	30.25	39.7

The main factor she wishes to analyze is *fertilizer response*, and she considers two analyses. In the first five tundra locations are considered as *replicates* and the design is a one-way completely randomized design:

$$Y_{ij} - \mu = T_i + e_{ij}$$

where:

Y_{ij} = Growth rate of plants receiving fertilizer i in location j
μ = Average growth rate of plants
T_i = Main effect on growth caused by fertilizer i
e_{ij} = Error variability

Using the conventional approach outlined in Sokal and Rohlf (1995, p. 214) or Zar (1996, p. 186), she obtains

Source	d.f.	Sum of squares	Mean square	*F* value

Fertilizer	3	7241.8	2413.9	22.7***
Error	16	1700.4	106.27	
Total	19	6942.2		

Note: *** means $p < .001$.

The second analysis is a randomized complete block design in which the five locations are now considered as *blocks* rather than replicates. The linear additive model is

$$Y_{ij} - \mu = T_i + B_j + e_{ij}$$

where $B_j =$ effect of blocks on growth rate

and the other terms are as defined above.

Using the randomized block calculations outlined in Zar (1996, p. 254) or Sokal and Rohlf (1995, p. 343), she obtains

Source	d.f.	Sum of squares	Mean square	*F* value
Fertilizer	3	7241.8	2413.9	79.3***
Locations (blocks)	4	1335.2	333.8	11.0***
Error	12	365.2	30.4333	
Total	19	8942.2		

Note: *** means $p < .001$.

As you can see clearly from the linear additive models, the error term in the completely randomized design contains the sum of squares due to variation among the blocks in the randomized block design.

The gain in efficiency that is obtained from using a randomized block design over a completely randomized design can be calculated from equations 6.1 and 6.2 in Damon and Harvey (1987):

$$\left\{\begin{matrix}\text{Relative efficiency of} \\ \text{randomized block}\end{matrix}\right\} = 100 \left(\frac{MS_{CR}}{MS_{RB}}\right) \frac{(n_1 + 1)(n_2 + 3)}{(n_1 + 3)(n_2 + 1)}$$

where MS_{CR} = Error mean square obtained in a completely randomized design
MS_{RB} = Error mean square obtained in a randomized complete block design
n_1 = Error degrees of freedom for randomized block design
n_2 = Error degrees of freedom for completely randomized design

For this particular example:

$$\text{Relative efficiency} = 100\left(\frac{106.27}{30.433}\right)\frac{(13)(19)}{(15)(17)}$$
$$= 338\%$$

This means that to achieve the same precision you would need to take about 3.4 times as many samples in a completely randomized design as you need in a randomized block design. Clearly, it pays to use a block design in this situation.

It is possible for the relative efficiency of a block design to be less than 100%, which means a completely randomized design is better. If blocks have a small effect, you will get a relative efficiency less than 100% because you lose degrees of freedom to blocks. I have never seen an ecological case where this was true.

There are many types of randomized block designs. The most common one is the randomized complete block design in which each treatment appears once in each block, and each block thus contains *t* experimental units (where *t* = number of treatments). There are many other incomplete block designs in which each block does not contain all the treatments. Steel and Torrie (1980 Chap. 12) and Damon and Harvey (1987, Chap. 7) discuss some incomplete block designs that are commonly used in biological experiments.

10.2.4 Split-Unit Designs

When the treatment structure of an experiment involves two or more factors, the best treatment design is a factorial one. But there are a class of designs that have been called *split-plot designs* because they originated in agriculture but are better labeled *split-unit designs* because not all treatment structures involve physical plots: Mead (1988, Chapter 14) has an excellent discussion of split-unit designs with many examples.

The key usage of split-unit designs occurs when two (or more) treatment factors are being studied and the size of the experimental unit which is appropriate to the first factor is much larger than that required for the second factor. Figure 10.7 gives a simple experiment from growth chambers for plants. The two treatment factors in this simple experiment are CO_2 level which is held in large greenhouses at two levels (ambient or enhanced CO_2) and soil temperature which is heated to 25°C or 30°C. The key element here is that the experimental units for the two treatments differ in size. In an ideal world with infinite funding it would be possible for the

experimenter to construct a series of large size experimental units that could be treated replicates, but this is very inefficient, and fewer treatment replicates would be available.

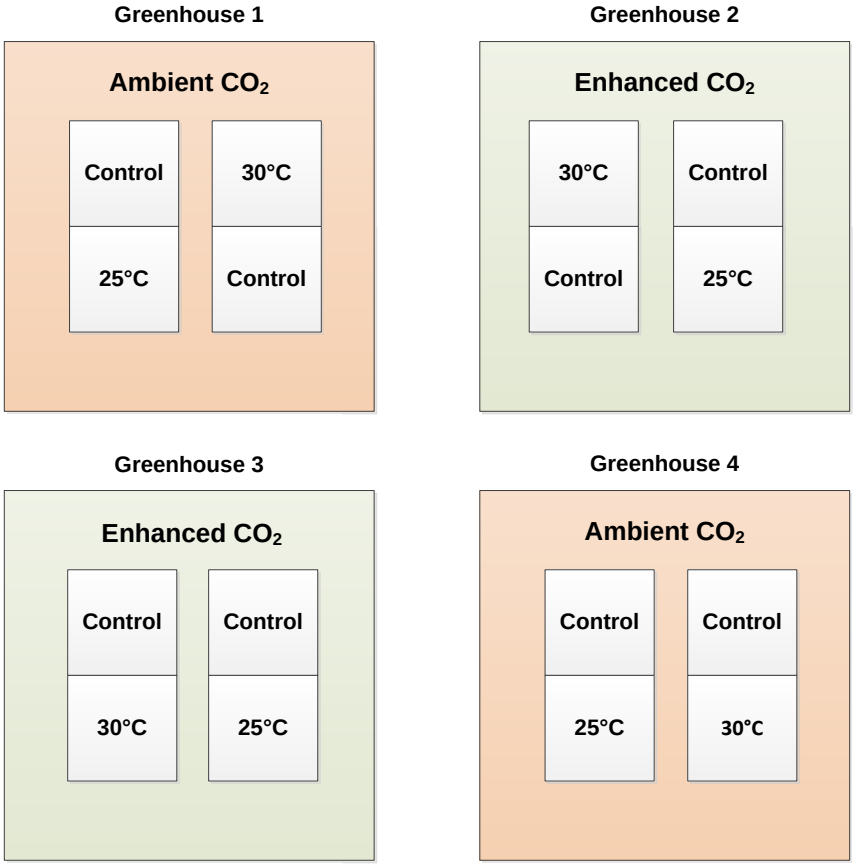


Figure 10.7 Schematic illustration of a split-unit design for a greenhouse experiment on plant growth. Greenhouses are held at either ambient or enhanced CO_2. Within each greenhouse smaller units have heated soils. This is a simple two-factor design to investigate CO_2 treatment effects and soil temperature effects.

Split-unit designs have several drawbacks (Mead 1988). In particular split-unit designs are difficult to interpret when there is interaction between the two or more factors involved in the treatment structure. These designs must assume independence of the sub-units of the design structure so that what happens in one sub-unit will not affect the processes in the other sub-units.

Hurlbert (2012) has pointed out that many analyses of ecological data commit the error of *pseudofactorialism* in the analysis of their data. This is a form or invalid statistical analysis that results from confusing the response structure (actual

variables measured) of a study with the treatment factor. Hurlbert (2012) found this error in 9% of papers published in 2009 in *Ecology* and the *Journal of Animal Ecology*, and he suggests it is a common statistical mistake in many papers. The key is to recognize what the *experimental unit* is in any particular study and what the treatment structure is. In any two-or three-way ANOVA response variables (like species) can be mistakenly identified as a treatment factor when they are part of the measurements taken as defined by the response structure. To avoid this error, rigorously set out in your study the treatment structure of your experiment, the design structure, and the response structure and if there is only one treatment factor, you should use only one-way ANOVAs in your analysis.

The statistical analysis of split-unit designs is complex. If you cannot avoid this design for practical reasons, consult a professional statistician about the appropriate analysis for your data.

10.2.5 Latin Square Designs

Randomized block designs are useful when there is one source of variation known before the experiments are carried out. In a few cases ecologists can recognize two sources of variation in the experimental units and wish to correct for *both* sources in doing an experiment. One way to achieve this is to use the latin square design. A latin square design is a simple extension of the randomized complete block design involving blocking in two directions.

Latin square designs are highly constrained designs because the number of levels of each factor must be equal. For example, assume you wish to compare the population increase rate of 3 species of aphids at different seasons of the year and you have available several plots in different field locations. In this case *plots* is one blocking factor (a known source of variation) and *season* is a second blocking factor (a second known source of variation). *Species* is the third factor of classification and you wish to know whether there is significant variation in the response variable, species' growth rates. Since there are 3 species, you are immediately constrained to a 3x3 latin square in which:

Number of plots = number of seasons = number of species = 3

If you cannot have this symmetry, you cannot use a latin square and must use a factorial design.

The linear additive model for a latin square is:

$$Y_{ijk} - \mu = A_i + B_j + C_k + e_{ijk}$$

where:

Y_{ijk} = Observed value of level i of A, level j of B, and level k of C
μ = Grand mean of all observations
A_i = Effect of level i of blocking factor A
B_j = Effect of level j of blocking factor B
C_k = Main effect of treatment factor C
e_{ijk} = Error deviation
$i = j = k$ = Number of levels of each factor

Table 10.3 gives some examples of latin square designs. The smallest possible latin square is 3x3 and in practice few are larger than 5x5. Note the symmetry of the latin square: each of the treatments appear once in each row and once in each column. Thus each row is a complete block and each column is also a complete block.

The most critical assumption of the latin square design is that there are *no interactions* between any of the three factors A, B, and C. For this reason the linear additive model is very simple and the analysis itself is simple. If you do not know enough to make this assumption of no interactions, you should use a full factorial design with replication so that interactions can be estimated.

The latin square design is not commonly found in field ecological research but when its restrictive assumptions can be met it is a very efficient experimental design to apply particularly in laboratory or common garden experiments.

10.2.6 Repeated Measure Designs

Ecological experiments often involve repeated measurements over time on individual organisms or individual plots. For example, growth studies may involve measuring the height of individual plants at several times over months or years. The response structure in repeated measure designs are the multiple measurements made on each individual subject, often over time. Repeated measurements on the

same evaluation unit or subject are not independent, and treating them as independent violates the basic assumption of ANOVA. Clearly two measurements of the height of the same plant (a single evaluation unit) cannot be independent, especially if time intervals are close. The key assumption for all statistical analysis is that experimental units are independent.

The analysis of data from repeated measure designs is difficult because a series of statistical decisions must be made before tests of significance can be calculated (Barcikowski and Robey 1984, Zar 1996 p. 259). Rowell and Walters (1976) discuss the validity of the assumptions made in most repeated measures experiments and provide an alternative approach when the assumptions are not satisfied. Gurevitch and Chester (1986) discuss another approach to repeated measures data that attempts to analyze the time trends in ecological variables. Most of the common statistical packages (SYSTAT, SAS, SPSS, JMP, NCSS) provide computations for the repeated measures design.

Hurlbert (2013) suggests that many repeated measure designs that involve time can be most simply analyzed as a series of one-way ANOVAs, one ANOVA for each time period.

10.3 WHERE SHOULD I GO NEXT?

The analysis of variance and the design of manipulative experiments are complex and rapidly growing subdisciplines of statistics, and you could spend a lifetime becoming an expert in one of these areas. This chapter has covered 0.1% of the material available and at this stage you need to do two things:

1. Take a course in experimental design or read a good elementary text like Cox (1958), Mead (1988), or Winer *et al.* (1991).
2. Think out your experimental design and take it to a professional statistician *before* you begin work. If you have a simple design, you may find the texts listed above adequate.

You can also get some guidance on design by looking at previous studies in your area of ecology. But read Hurlbert (1984, 2009) and Peters (1991) before you accept what is already published too easily. Alternatively, you may get some advice from a statistical computer packages such as SigmaStat by Jandel Software.

Unfortunately most statistical packages assume you already have a fully implemented experimental design and thus give little guidance, and I do not recommend using computer packages until you have thought out and planned your design.

Finally, remember that good and interesting ecology comes from exciting ideas *and* good experimental design. Being a first class statistician will not make you a good ecologist, but many exciting ideas have floundered on poor experimental design.

10.4 SUMMARY

The general principles of experimental design are often overlooked in the rush to set up ecological experiments or ignored in the rush to utilize computer packages of statistical programs. The first step is in designing a good experiment is to define carefully the treatment structure (what treatments are you applying?), the design structure (what rules are you using to assign treatments?) and the response structure (what measurements are you taking?). The *experimental units* must be unambiguously defined, and these units must be sampled *randomly* to satisfy the assumption that all observations are independent and to reduce bias. *Replication* of experimental units is needed to estimate experimental "error", the yardstick of statistical significance. Treatments should be *interspersed* in space and in time to minimize the possibility that chance events will affect the results of the experiment. If interspersion is not used, replicates may not be independent and the error of *pseudoreplication* may occur.

Many experimental designs are available to ecologists to assist in increasing the precision of manipulative field experiments. All of them can be represented by a linear additive model. The elementary strategy for increasing precision is to group together experimental units into *blocks* that are homogeneous. Any known or suspected source of variation can be blocked out of the experiment by this simple grouping.

Other experimental designs are less commonly used. *Split-unit* designs can be used when two treatments are applied and one occurs in much larger sampling units

than the other. *Repeated measure* designs can be used when a time series of data are obtained on a set of experimental units.

There is an enormous complex literature on experimental design for manipulative experiments. Refer to a comprehensive text like Mead (1988) and consult your local statistician for details before you launch into any project with a complex design.

SELECTED REFERENCES

Cox, D.R. 1958. *Planning of Experiments*. John Wiley and Sons, New York.

Eberhardt, L.L. 2003. What should we do about hypothesis testing? Journal of Wildlife Management **67**(2): 241-247.

Heffner, R.A., Butler, M.J., & Reilly, C.K. 1996. Pseudoreplication revisited. *Ecology* **77**: 2558-2562.

Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* **54**: 187-211.

Hurlbert, S.H. 2013. Pseudofactorialism, response structures and collective responsibility. Austral Ecology **38**.

Mead, R. 1988. *The Design of Experiments: Statistical Principles for Practical Application*. Cambridge University Press, Cambridge.

Rice, W.R. &. Gaines, S.D. 1994. 'Heads I win, tails you lose': testing directional alternative hypotheses in ecological and evolutionary research. *Trends in Ecology and Evolution* **9**: 235-237.

Shaw, R.G. & Mitchell-Olds, T. 1993. ANOVA for unbalanced data: an overview. *Ecology* **74**: 1638-1645.

Yates, F. 1964. Sir Ronald Fisher and the Design of Experiments. *Biometrics* 20: 307-321.

QUESTIONS AND PROBLEMS

10.1. Examine a sample of 40 papers in an ecological journal of your choice and classify the frequency of errors in the use of the analysis of variance, following the discussion of pseudofactorialism of Hurlbert (2013). Compare your results with those of Hurlbert (2013) for *Ecology* and the *Journal of Animal Ecology* in 2009.

10.2 Mead (1988, p. 7) states:

"I shall emphasize the need to think separately about the properties of the experimental units available for the experiment and about the choice of

experimental treatments. Far too often, the two aspects are confused at an early stage, and an inefficient and useless experiment results".

Read any experimental paper in a recent ecology journal and determine if this problem exists.

10.3 Hurlbert (1984 p. 200) gives the following problem: A beetle population in a large field has a mean density of 51 beetles/m^2. We wish to test if a herbicide has an effect on beetle density. Two possible designs are:
(a) The field is divided into two parts, and a preapplication sampling of 20 quadrats in each part shows no significant difference between them. The herbicide is applied to one part of the field and the other part is kept as a control. After 48 hours each part is sampled with 20 quadrats.
(b) The field is partitioned into a grid of 1000 4x4 m plots. Twenty of these are selected at random and used as controls and another 20 plots are selected randomly as herbicide plots. A preapplication sampling is done on each plot and no significant difference is found. The herbicide is applied and 48 hours latter each plot is sampled again.

Are designs (a) and (b) equivalent? Why or why not?

10.4. Pseudoreplication was found in 50% of the field experiments done on small rodents and only 10% of the field experiments done on freshwater plankton (Hurlbert 1984, Table 3). Discuss why this might occur.

10.5. Hayne (1978) states in discussing experimental design:

"If you insist on a replicate area being an exact duplication of your first (study area), then you've restricted your scientific interest to a very narrow subset of the whole universe. It is a misconception that replicates have to be identical. There's no such thing... Any decent naturalist-type investigator will find differences between any two areas you choose."

Do you agree? What are the implications for field experiments in ecology?

10.6. List the sources of variability in measurements of abundance for the design shown in Figure 10.1 (page 000). Which sources of variability can you reduce or eliminate? Describe exactly how these sources of variability change if you increase the number of control areas from one to four.

10.7. Are there important ecological issues that cannot be investigated with an experimental design?