

# Analyzing ANOVA Designs

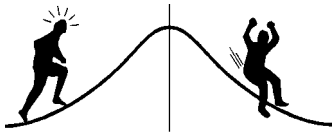
## Biometrics Information Handbook No.5

---

07/1995



Province of British Columbia  
Ministry of Forests Research Program



# Analyzing ANOVA Designs

Biometrics Information Handbook No.5

---

Vera Sit



**Province of British Columbia**  
Ministry of Forests Research Program

Citation:

Sit. V. 1995. Analyzing ANOVA Designs. Biom. Info. Hand. 5. Res. Br., B.C. Min. For., Victoria, B.C. Work. Pap. 07/1995.

Prepared by:

Vera Sit

for:

B.C. Ministry of Forests

Research Branch

31 Bastion Square

Victoria, BC V8W 3E7

© 1995 Province of British Columbia

Copies of this report may be obtained, depending upon supply, from:

B.C. Ministry of Forests

Research Branch

31 Bastion Square

Victoria, BC V8W 3E7

## ACKNOWLEDGEMENTS

---

I would like to thank the regional staff in the B.C. Ministry of Forests for requesting a document on analysis of variance in 1992. Their needs prompted the preparation of this handbook and their encouragement fuelled my drive to complete the handbook. The first draft of this handbook was based on two informal handouts prepared by Wendy Bergerud (1982, 1991). I am grateful to Wendy for allowing the use of her examples (Bergerud 1991) in Chapter 6 of this handbook.

A special thanks goes to Jeff Stone for helping me focus on the needs of my intended audience. I thank Gordon Nigh for reading the manuscript at each and every stage of its development. I am indebted to Michael Stoehr for the invaluable discussions we had about ANOVA. I am grateful for the assistance I received from David IZard in producing the figures in this handbook. Last, but not least, I want to express my appreciation for the valuable criticism and suggestions provided by my reviewers: Wendy Bergerud, Rob Brockley, Phil Comeau, Mike Curran, Nola Daintith, Graeme Hope, Amanda Nemeč, Pasi Puttonen, John Stevenson, David Tanner, and Chris Thompson.

Acknowledgements .....	iii
<b>1 Introduction .....</b>	<b>1</b>
<b>2 The Design of an Experiment .....</b>	<b>1</b>
2.1 Factor .....	2
2.2 Quantitative and Qualitative Factors .....	2
2.3 Random and Fixed Factors .....	2
2.4 Variance .....	3
2.5 Crossed and Nested Factors .....	3
2.6 Experimental Unit .....	5
2.7 Element .....	6
2.8 Replication .....	6
2.9 Randomization .....	7
2.10 Balanced and Unbalanced Designs .....	7
<b>3 Analysis of Variance .....</b>	<b>9</b>
3.1 Example .....	10
3.2 Hypothesis Testing .....	10
3.3 Terminology .....	11
3.3.1 Model .....	11
3.3.2 Sum of squares .....	11
3.3.3 Degrees of freedom .....	12
3.3.4 Mean squares .....	13
3.4 ANOVA Assumptions .....	13
3.5 The Idea behind ANOVA .....	14
3.6 ANOVA <i>F</i> -test .....	17
3.7 Concluding Remarks about ANOVA .....	18
<b>4 Multiple Comparisons .....</b>	<b>19</b>
4.1 What is a Comparison? .....	19
4.2 Drawbacks of Comparisons .....	20
4.3 Planned Contrasts .....	21
4.4 Multiple Comparisons .....	22
4.5 Recommendations .....	23
<b>5 Calculating ANOVA with SAS: an Example .....</b>	<b>24</b>
5.1 Example .....	24
5.2 Experimental Design .....	24
5.3 ANOVA Table .....	26
5.4 SAS Program .....	28
<b>6 Experimental Designs .....</b>	<b>33</b>
6.1 Completely Randomized Designs .....	34
6.1.1 One-way completely randomized design .....	34
6.1.2 Subsampling .....	36
6.1.3 Factorial completely randomized design .....	38

6.2	Randomized Block Designs .....	40
6.2.1	One-way randomized block design .....	41
6.2.2	Factorial randomized block design .....	43
6.3	Split-plot Designs .....	45
6.3.1	Completely randomized split-plot design .....	47
6.3.2	Randomized block split-plot design .....	49
<b>7</b>	<b>Summary .....</b>	<b>51</b>
Appendix 1	How to determine the expected mean squares .....	52
Appendix 2	Hypothetical data for example in Section 5.1 .....	57
References	.....	58

**TABLES**

1	Hypothetical sources of variation and their degrees of freedom ...	12
2	Partial ANOVA table for the fertilizer and root-pruning treatment example .....	27
3	Partial ANOVA table for the fertilizer and root-pruning treatment example .....	27
4	ANOVA table for the fertilizer and root-pruning treatment example .....	28
5	ANOVA table for a one-way completely randomized design .....	34
6	ANOVA table for a one-way completely randomized design with subsamples .....	37
7	ANOVA table for a factorial completely randomized design .....	40
8	ANOVA table for a one-way randomized block design .....	42
9	ANOVA table for a two-way factorial randomized block design ...	44
10	ANOVA table with pooled mean squares for a two-way factorial randomized block design .....	45
11	ANOVA table for a completely randomized split-plot design .....	48
12	ANOVA table for a randomized block split-plot design .....	50
13	Simplified ANOVA table for a randomized block split-plot design .....	50

FIGURES

1	Design structure of an experiment with two factors crossed with one another .....	3
2	Alternative stick diagram for the two-factor crossed experiment ..	4
3	Design structure of a one-factor nested experiment .....	4
4	Forty seedlings arranged in eight vertical rows with five seedlings per row .....	5
5	Stick diagram of a balanced one-factor design .....	8
6	Stick diagram of a one-factor design: unbalanced at the element level .....	8
7	Stick diagram of a one-factor design: unbalanced at both the experimental unit and element levels .....	9
8	Central $F$ -distribution curve for (2,27) degrees of freedom .....	17
9	$F$ -distribution curve with critical $F$ value and decision rule .....	18
10	Design structure for the fertilizer and root-pruning treatment example .....	25
11	SAS output for fertilizer and root-pruning example .....	31
12	Design structure for a one-way completely randomized design ....	34
13	Design structure for a one-way completely randomized design with subsamples .....	36
14	Design structure for a factorial completely randomized design ....	39
15	Design structure for a one-way randomized block design .....	42
16	Design structure for a two-way factorial randomized block design .....	44
17	A completely randomized factorial arrangement .....	46
18	Main-plots of a completely randomized split-plot design .....	46
19	A completely randomized split-plot layout .....	46
20	Design structure for a completely randomized split-plot design ...	48
21	Design structure for a randomized block split-plot design .....	49

Analysis of variance (ANOVA) is a powerful and popular technique for analyzing data. This handbook is an introduction to ANOVA for those who are not familiar with the subject. It is also a suitable reference for scientists who use ANOVA to analyze their experiments.

Most researchers in applied and social sciences have learned ANOVA at college or university and have used ANOVA in their work. Yet, the technique remains a mystery to many. This is likely because of the traditional way ANOVA is taught — loaded with terminology, notation, and equations, but few explanations. Most of us can use the formulae to compute sums of squares and perform simple ANOVAs, but few actually understand the reasoning behind ANOVA and the meaning of the  $F$ -test.

Today, all statistical packages and even some spreadsheet software (e.g., EXCEL) can do ANOVA. It is easy to input a large data set to obtain a great volume of output. But the challenge lies in the correct usage of the programs and interpretation of the results. Understanding the technique is the key to the successful use of ANOVA.

The concept of ANOVA is really quite simple: to compare different sources of variance and make inferences about their relative sizes. The purpose of this handbook is to develop an understanding of ANOVA without becoming too mathematical.

It is crucial that an experiment is designed properly for the data to be useful. Therefore, the elements of experimental design are discussed in Chapter 2. The concept of ANOVA is explained in Chapter 3 using a one-way fixed factor example. The meaning of degrees of freedom, sums of squares, and mean squares are fully explored. The idea behind ANOVA and the basic concept of hypothesis testing are also explained in this chapter. In Chapter 4, the various techniques for comparing several means are discussed briefly; recommendations on how to perform multiple comparisons are given at the end of the chapter. In Chapter 5, a completely randomized factorial design is used to illustrate the procedures for recognizing an experimental design, setting up the ANOVA table, and performing an ANOVA using SAS statistical software. Many designs commonly used in forestry trials are described in Chapter 6. For each design, the ANOVA table and SAS program for carrying out the analysis are provided. A set of rules for determining expected means squares is given in Appendix 1.

This handbook deals mainly with balanced ANOVAs, and examples are analyzed using SAS statistical software. Nonetheless, the information will be useful to all readers, regardless of which statistical package they use.

Experimental design involves planning experiments to obtain the maximum amount of information from available resources. This chapter



defines the essential elements in the design of an experiment from a statistical point of view.

### 2.1 Factor

A **factor** is a variable that may affect the **response** of an experiment. In general, a factor has more than one **level** and the experimenter is interested in testing or comparing the effects of the different levels of a factor. For example, in a seedlot trial investigating seedling growth, height increment could be the response and seedlot could be a factor. A researcher might be interested in comparing the growth potential of seedlings from five different seedlots.

### 2.2 Quantitative and Qualitative Factors

A **quantitative factor** has levels that represent different amounts of a factor, such as kilograms of nitrogen fertilizer per hectare or opening sizes in a stand. When using a quantitative factor, we may be concerned with the relationship of the response to the varying levels of the factor, and may be interested in an equation to relate the two.

A **qualitative factor** contains levels that are different in kind, such as types of fertilizer or species of birds. A qualitative factor is usually used to establish differences among the levels, or to select from the levels.

A factor with a numerical component is not necessarily quantitative. For example, if we are considering the amount of nitrogen in fertilizer, then the amount of nitrogen must be specific (e.g., 225 kg N/ha, 450 kg N/ha, 625 kg N/ha) for the factor to be quantitative. If, however, only two levels of nitrogen are to be considered — none and some amount — and the objective of the study is to compare fertilizers with and without nitrogen, then nitrogen is a qualitative factor (Mize and Schultz 1985; Mead 1988, Section 12.4).

### 2.3 Random and Fixed Factors

A factor can be classified as either fixed or random, depending on how its levels are chosen. A **fixed factor** has levels that are determined by the experimenter. If the experiment is repeated, the same factor levels would be used again. The experimenter is interested in the results of the experiment for those specific levels. Furthermore, the application of the results would only be extended to those levels. The objective is to test whether the means of each of the treatment levels are the same. Suppose an experimenter is interested in comparing the growth potentials of seedlings from seedlots A, B, C, D, and E in a seedlot trial. Seedlot is a fixed factor because the five seedlots are chosen by the experimenter specifically for comparison.

A **random factor** has levels that are chosen randomly from the population of all possible levels. If the experiment is repeated, a new random set of levels would be chosen. The experimenter is interested in generalizing the results of the experiment to a range of possible levels and not just to the levels used. In the seedlot trial, suppose the five seedlots are instead randomly selected from a defined population of seedlots in a particular seed orchard. Seedlot then becomes a random factor; the researcher would not be interested in comparing any particular seedlots, but in testing whether the variation among seedlots is the same.

It is sometimes difficult to determine whether a factor is fixed or random. Remember that it is not the nature of the factor that differentiates the two types, but the objectives of the study and the method used to select factor levels. Comparisons of random and fixed factors have been discussed in several papers (Eisenhart 1947; Searle 1971a, 1971b (Chapter 9); Schwarz 1993; Bennington and Thayne 1994).

2.4 Variance

The sample variance is a measure of spread in the response. Large variance corresponds to a wide spread in the data, while small variance corresponds to data being concentrated near the mean. The dispersion of the data can have a number of sources.

In the fixed factor seedlot example, different seedling heights could be attributed to the different seedlot, to the inherited differences in the seedlings, or to different environmental conditions. Analysis of variance compares the sources of variance to determine if the observed differences are caused by the factor of interest or are simply a part of the nature of things. Many sources of variance cannot be identified. The variance of these sources are combined and referred to as residual or error variance. A complete list of sources of variance in an experiment is a non-mathematical way of describing the “model” used for the experiment (refer to Section 3.3.1 for the definition of model). Section 5.3 presents a method to compile the list of sources in an experiment. Examples are given in chapters 5 and 6.

2.5 Crossed and Nested Factors

An experiment often contains more than one factor. These factors can be combined or arranged with one another in two different ways: **crossed** or **nested**.

Two factors are **crossed** if every level of one factor is combined with every level of the other factor. The individual factors are called **main effects**, while the crossed factors form an **interaction effect**. Suppose we wish to find out how two types of fertilizers (OLD and NEW) would perform on seedlings from five seedlots (A, B, C, D, and E.) We might design an experiment in which 10 seedlings from each seedlot would be treated with the OLD fertilizer and another 10 seedlings would be treated with the NEW fertilizer. Both fertilizer and seedlot are fixed factors because the levels are chosen specifically. This design is illustrated in Figure 1.

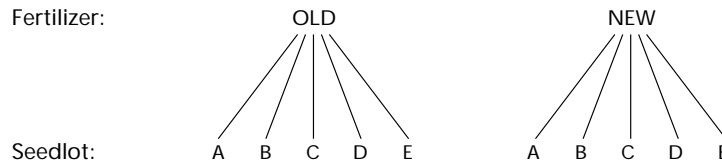


FIGURE 1 *Design structure of an experiment with two factors crossed with one another.*

Figure 1 is called a **stick diagram**. The lines, or sticks, indicate the relationships between the levels of the two factors. Five lines are extended from both the OLD and NEW fertilizers to the seedlots, indicating that the OLD and NEW fertilizers are applied to seedlings from the five seedlots. The same seedlot letter designations appear under the OLD and NEW fertilizers because the seedlings are from the same five seedlots. Hence, the two factors are crossed. The crossed, or interaction factor is denoted as **fertilizer\*seedlot**, or **seedlot\*fertilizer**, as the crossed relationship is symmetrical. We could also draw the stick diagram with seedlot listed first, as in Figure 2.

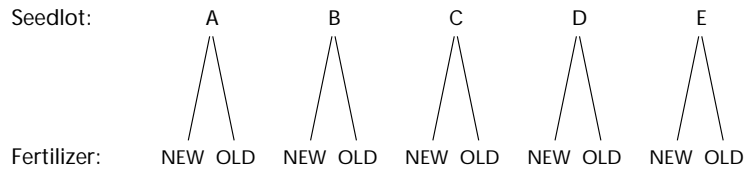


FIGURE 2 *Alternative stick diagram for the two-factor crossed experiment.*

A factor is **nested** when each level of the nested factor occurs with only one or a combination of levels of the other factor or factors. Suppose ten seedlots are randomly selected from all seedlots in an orchard, and seedlings from five of these are treated with the OLD fertilizer, while seedlings from the other five are treated with the NEW fertilizer. The assignment of the fertilizer to the seedlots is completely random. Fertilizer is a fixed factor but seedlot is a random factor. This design is displayed in Figure 3.

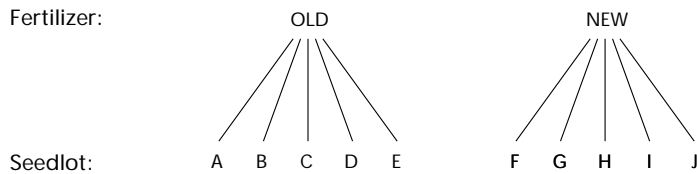


FIGURE 3 *Design structure of a one-factor nested experiment.*

Figure 3 is very similar to Figure 1 except at the seedlot level. The five seedlots under the OLD fertilizer are different from those under the NEW fertilizer because seedlings from different seedlots are treated by the two fertilizers. The non-repeating seedlot letter designations indicate that the seedlot is nested within the fertilizer treatments. The nested factor “seedlot” would never occur by itself in an ANOVA table; rather it would always be denoted in the nested notation **seedlot(fertilizer)**. The nested factor is usually random, but unlike the crossed relationship, the nested

relationship is not symmetrical. In a stick diagram, the nested factor is always drawn below the main factor.

A nested factor never forms an interaction effect with the main factor(s) within which it is nested. It can, however, be crossed with other factors. For example, the factor relationship temperature\*seedlot (fertilizer) is valid. It indicates that each level of the temperature factor combines with each level of the seedlot factor which is nested within the fertilizer factor. If the temperature factor has three levels (e.g., high, medium, and low), then there are thirty ( $3 \times 10 = 30$ ) combinations of temperature, seedlot, and fertilizer. On the other hand, the factor relationship fertilizer\*seedlot(fertilizer) is not valid because the seedlot factor cannot be nested with, and crossed with, the fertilizer factor at the same time.

The stick diagram is a very useful tool to identify the structure of an experiment. Its construction is discussed in detail in Section 5.2.

2.6  
Experimental  
Unit

An **experimental unit (e.u.)**, also referred to as a **treatment unit**, is the smallest collection of the experimental material to which one level of a factor or some combination of factor levels is applied. We need to identify the experimental unit in a study to determine the study's design. To do this, we must know how the factor levels are assigned to the experimental material. Suppose we want to study the effect of a mycorrhizal fungus treatment (e.g., mycelial slurry) on seedling performance. "Fungus" is a fixed and qualitative factor; it has two levels: with fungus treatment and without. The experiment could be performed on 40 seedlings, as arranged in Figure 4.



FIGURE 4 Forty seedlings arranged in eight vertical rows with five seedlings per row.

There are many ways to perform the experiment. Three possible cases are:

1. Randomly assign a treatment level to each seedling so that 20 seedlings are inoculated with the fungus, and 20 are not. Here, a seedling is the experimental unit for the "fungus" factor.
2. Randomly assign a treatment level to a row of seedlings (five seedlings per row) so that exactly four rows of seedlings are inoculated with the fungus and four are not. In this case, a row of seedlings is the experimental unit for the "fungus" factor.

3. Randomly assign a treatment level to a plot of seedlings (four rows of seedlings per plot) so that one plot of seedlings is inoculated with the fungus, and one is not. A plot of seedlings is now the experimental unit for the “fungus” factor.

#### 2.7 Element

An **element** is an object on which the response is measured. In the fungus example, if the response (e.g., height or diameter) is measured on each seedling, then a seedling is the element. Elements can be the same as the experimental units, as in the first case of the previous example, where the fungus treatment is assigned to each seedling. If the two are different, then elements are nested within experimental units.

#### 2.8 Replication

A **replication** of a treatment is an independent observation of the treatment. A factor level is replicated by applying it to two or more experimental units in an experiment. The number of replications of a factor level is the number of experimental units to which it is applied.

In the fungus treatment example of Section 2.6, all three experimental cases result in 20 seedlings per treatment level. In the first case, where a seedling is the experimental unit, each level of fungus treatment is replicated 20 times. In the second case, where a row of five seedlings is the experimental unit, each level is replicated four times. In the last case, where a plot of seedlings is the experimental unit, the fungus treatments are not replicated.

Too often, elements or repeated measurements are misinterpreted as replications. Some researchers may claim that in the last example of the fungus treatment, the experiment is replicated 20 times as there are 20 seedlings in a plot. This is incorrect because the 20 seedlings in a plot are considered as a group in the random assignment of the treatments — all 20 seedlings receive the same treatment regardless of the treatment type. This situation is called **pseudo-replication**.

Replication allows the variance among the experimental units to be estimated. If one seedling is inoculated with mycorrhiza and another seedling is used as a control, and a difference in seedling height is then observed, the only possible conclusion is that one seedling is taller than the other. It is not possible to determine if the observed height difference is caused by the fungus treatment, or by the natural variation among seedlings, or by some other unknown factor. In other words, treatment effect is **confounded** with natural seedling variability.

Replication increases the **power** of an experiment. It increases the probability of detecting differences among means if these differences exist. In the fungus example, an experiment performed according to the first case with treatments replicated twenty times would be more powerful than the same experiment with treatments replicated four times.

Replication can be accomplished in many ways and will depend on the type of comparison desired. But regardless of how it is done, it is essential. Without replication we can not estimate experimental error or decide objectively whether the differences are due to the treatments or to other hidden sources. See Bergerud (1988a), Mead (1988, Chapter 6), and Thornton (1988) for more discussion of the importance of replication.

Milliken and Johnson (1989) also offer some advice about performing unreplicated experiments.

## 2.9 Randomization

In many experimental situations, there are two main sources of experimental error: unit error and technical error. **Unit errors** occur when different experimental units are treated alike but fail to respond identically. These errors are caused by the inherent variation among the experimental units. **Technical errors** occur when an applied treatment can not be reproduced exactly. These errors are caused by the limitation of the experimental technique. The technical errors can be reduced with more careful work and better technologies. Unit errors can be controlled, in a statistical sense, by introducing an element of randomness into the experimental design (Wilk 1955). This should provide a valid basis for the conclusions drawn from an experiment. In any experiment, the following randomization steps should occur:

1. experimental units are selected randomly from a well-defined population, and
2. treatments are randomly assigned to the experimental units.

If a seedling is the experimental unit, as in the fungus study example, we would first select 40 seedlings at random from all possible seedlings in the population of interest (e.g., Douglas-fir seedlings from a particular seed orchard). Then one of the two fungus treatments would be assigned at random to each seedling.

These randomization steps ensure that the chosen experimental units are representative of the population and that all units have an equal chance of receiving a treatment. If experimental units are selected and assigned treatments in this manner, they should not show different response values in the presence or absence of a treatment. Thus, any significant differences that are found should be related only to the treatments and not to differences among experimental units.

Randomization of treatments can be imposed in an experimental design in many ways. Two common designs are the completely randomized design and the randomized block design.

- In a **completely randomized design**, each treatment is applied randomly to some experimental units and each unit is equally likely to be assigned to any treatment. One method of random assignment is used for all experimental units.
- In a **randomized block design**, homogeneous experimental units are first grouped into blocks. Each treatment is assigned randomly to one experimental unit in each block such that each unit within a block is equally likely to receive a treatment. A separate method of random assignment should be used in each block.

These designs are detailed in Chapter 6.

## 2.10 Balanced and Unbalanced Designs

An experiment is **balanced** if all treatment level combinations have the same number of experimental units and elements. Otherwise, the experiment is **unbalanced**. To illustrate this idea, let us consider the following example.

A study is designed to compare the effectiveness of two different fertilizers (NEW and OLD) on seedling growth. The experiment is performed on six styroblocks of seedlings. Each styroblock of seedlings is randomly assigned a fertilizer treatment. Three styroblocks are treated with each fertilizer. After three weeks, 10 seedlings from each styroblock are randomly selected for outplanting. Heights are measured three months after outplanting.

Fertilizer is a fixed and qualitative factor with two levels, the styroblock is the experimental unit for the fertilizer, and the individual seedling is the element. This study has a balanced design because each level of the fertilizer factor has the same number of experimental units and each of these has the same number of elements. The corresponding stick diagram is shown in Figure 5.

Unbalanced data can occur at the experimental unit level, at the element level, or at both. For instance, the design given in Figure 6 is unbalanced at the element level since styroblock 6 has only five elements. In contrast, the design shown in Figure 7 is unbalanced at both the

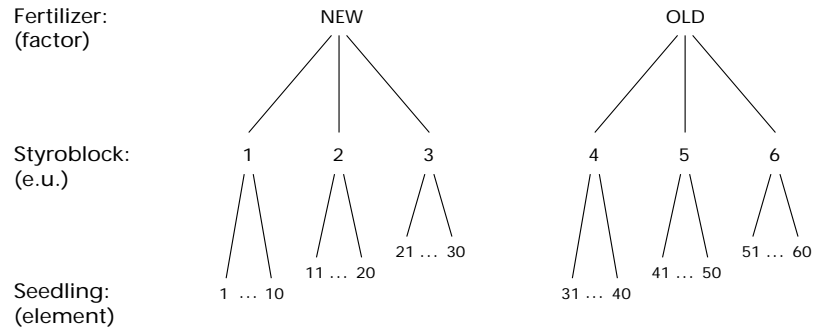


FIGURE 5 Stick diagram of a balanced one-factor design.

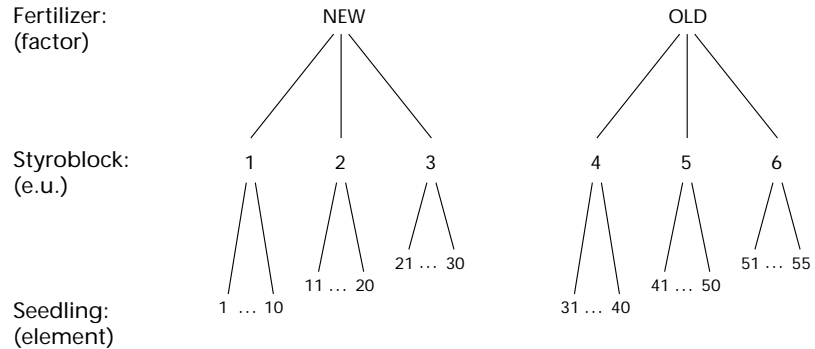


FIGURE 6 Stick diagram of a one-factor design: unbalanced at the element level.

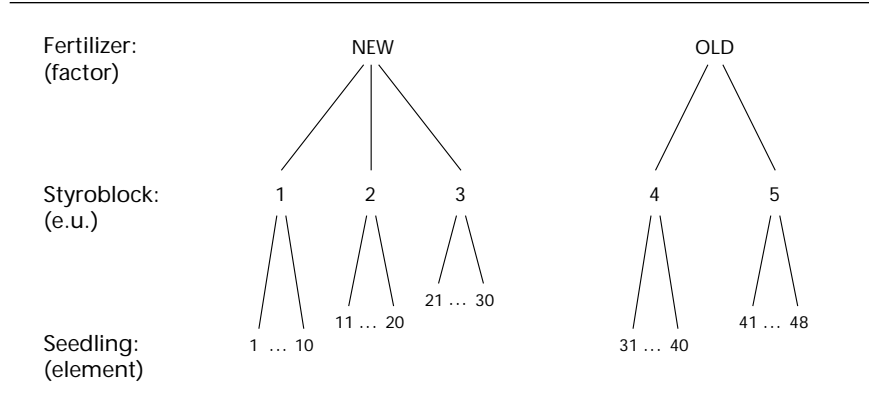


FIGURE 7 *Stick diagram for a one-factor design: unbalanced at both the experimental unit and element levels.*

experimental unit and element levels. Many experiments are designed to be balanced but produce unbalanced data. In a planting experiment, for example, some of the seedlings may not survive the duration of the experiment, resulting in unequal sample sizes.

We should note that balanced data is simpler to analyze than unbalanced data. For unbalanced data, analysis is easier if the design is balanced at the experimental unit level. An example of unbalanced data analysis is given in Section 6.1.2.

### 3 ANALYSIS OF VARIANCE

---

We must consider the method of analysis when designing a study. The method of analysis depends on the nature of the data and the purpose of the study. Analysis of variance, ANOVA, is a statistical procedure for analyzing continuous data<sup>1</sup> sampled from two or more populations, or from experiments in which two or more treatments are used. It extends the two-sample  $t$ -test to compare the means from more than two groups.

This chapter discusses the concept of ANOVA using a one-factor example. The basic idea of hypothesis testing is briefly discussed in Section 3.2. Terms commonly used in ANOVA are defined in Section 3.3, while the basic assumptions of ANOVA are discussed in Section 3.4. Section 3.5 presents the idea behind ANOVA. The meanings of the sum of squares and mean square terms in the ANOVA model are explained in detail. The testing mechanism and the ANOVA  $F$ -test are described in Section 3.6. Some concluding remarks about ANOVA are given in

<sup>1</sup> Continuous data can take on any value within a range. For example, tree heights, tree diameters, and percent cover of vegetation are all continuous data.



Section 3.7. We will begin by introducing an example which will be used throughout this chapter.

### 3.1 Example

A study is designed to investigate the effect of two different types of fertilizer (types A and B) on the height growth of Douglas-fir seedlings. Thirty Douglas-fir seedlings are randomly selected from a seedlot; each seedling is randomly assigned one of the three treatments: fertilizer A, fertilizer B, or no fertilizer, so that ten seedlings receive each type of fertilizer treatment. Seedling heights are measured before the treatments are applied and again five weeks later. The purpose of the study is to determine whether fertilizer enhances seedling growth over this time period and whether the two fertilizers are equally effective in enhancing seedling growth for this particular seedlot.

This is a **one-factor completely randomized design**.<sup>2</sup> It can also be called a **one-way fixed factor completely randomized design**, with fertilizer as the fixed factor. An individual seedling is both an experimental unit and an element. The fertilizer has three levels (A, B, or none) with each level replicated 10 times.

The objective of the study is to compare the true mean height increments (final height – initial height) of three populations of seedlings:

- all seedlings treated with fertilizer A,
- all seedlings treated with fertilizer B, and
- all seedlings not fertilized.

Since the true mean height increments of the three populations of seedlings are unknown, the comparison is based on the sample means of the three groups of seedlings in the study.

### 3.2 Hypothesis Testing

Hypothesis testing is a common technique that is used when comparing several population means. It is an integral part of many common analysis procedures such as ANOVA and contingency table analysis. The objective is to reject one of two opposing mutually exclusive hypotheses based on observed data. The null hypothesis states that the population means are equal; the alternative hypothesis states that not all the populations means are the same.

In hypothesis testing, the null hypothesis is first assumed to be true. Under this assumption, the data are then combined into a single value called the “test statistic.” The choice of the test statistic depends on the analysis method (e.g.,  $F$  for ANOVA and  $\chi^2$  for chi-square contingency table analysis). The distribution of the test statistic or the probability of observing different values of the test statistic, is often known from statistical theory. The probability of obtaining a test statistic at least as extreme as the observed value is computed from the distribution. This probability is called the  $p$ -value. A **large**  $p$ -value (e.g.,  $p = 0.9$ ) implies that the observed data represent a likely event under the null hypothesis; hence there is no evidence to reject the null hypothesis as false. On the

<sup>2</sup> This and other experimental designs are described in detail in Chapter 6.

other hand, a **small**  $p$ -value (e.g.,  $p = 0.001$ ) suggests that the observed data represent a highly improbable event if the null hypothesis is true. Therefore, the observed data contradict the initial assumption and the null hypothesis should be rejected.

The cut-off point for the  $p$ -value is called the level of significance, or alpha ( $\alpha$ ) level. It depends on how often we can afford to be wrong when we conclude that some treatment means are different. By setting  $\alpha = 0.05$ , for example, we declare that events with probabilities under 5% (i.e.,  $p < \alpha$ ) are considered rare. These events are not likely to occur by chance and indicate that the null hypothesis is false; that is, the population means are not equal. When we reject the null hypothesis, there is, at most, a 5% chance that this decision will be wrong. We would retain the null hypothesis when  $p > \alpha$ , as there is not enough evidence in the data to reject it. More specific discussion of testing hypotheses in ANOVA is given in Section 3.6.

### 3.3 Terminology

Before proceeding with the discussion of analysis of variance, several frequently used terms must be defined.

**3.3.1 Model** An ANOVA **model** is a mathematical equation that relates the measured response of the elements to the sources of variation using a number of assumptions. In the fertilizer example, height increment is the response. The assumptions associated with an ANOVA model will be discussed in Section 3.4.

The ANOVA model of the fertilizer study can be expressed in words as:

observed	average	height increment			
height	= height	+ due to treatment	+ experimental error		
increment	increment	applied			

This model assumes that the observed height increment of a seedling can be divided into three parts. One part relates to the average performance of all seedlings; another part to the fertilizer treatment applied. The third part, called **experimental error**, represents all types of extraneous variations. These variations can be caused by inherent variability in the experimental units (i.e., seedlings) or lack of uniformity in conducting the study. The concept of experimental error is crucial in the analysis of variance and it plays an important part in the discussions in subsequent sections.

**3.3.2 Sum of squares** The term **sum of squares** refers to a sum of squared numbers. For example, in the calculation of variance ( $s^2$ ) of a sample of  $n$  independent observations ( $Y_1, Y_2, \dots, Y_n$ ),

$$s^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n - 1}, \quad (1)$$

the numerator of  $s^2$  is a sum of squares: the squares of the differences between the observed values ( $Y_i$ ) and the sample mean ( $\bar{Y}$ ). In ANOVA,

the sum of squares of a source of variation is a measure of the variability due to that source. Sum of squares is usually denoted as “SS” with a subscript identifying the corresponding source.

**3.3.3 Degrees of freedom** The degrees of freedom refers to the number of independent observations that are calculated in the sum of squares (Keppel 1973). It is often denoted as “df.” For example, suppose  $Y_1, Y_2, \dots, Y_n$  are  $n$  observations in a random sample. The quantity  $\sum_{i=1}^n Y_i^2$  has  $n$  degrees of freedom because the  $n$   $Y_i$  values are independent. Independent means that the value of one observation does not affect the values of the other observations. On the other hand, the quantity  $n\bar{Y}^2$  has only one degree of freedom because for any specific value of  $n$ , its value depends only on  $\bar{Y}$ .

Li (1964: 35–36) postulated that the degrees of freedom for the difference between (or sum of) two quantities is equal to the difference between (or sum of) the two corresponding number of degrees of freedom. Thus, the quantity

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n Y_i^2 - n\bar{Y}^2 \quad (2)$$

has  $(n - 1)$  degrees of freedom.

The degrees of freedom of a source of variation can also be determined from its structure, regardless of whether it contains a single factor or several factors combined in a nested or crossed manner. The following are rules for determining degrees of freedom of a source in balanced designs.

- A source containing a single factor has degrees of freedom one less than its number of levels.
- A source containing nested factors has degrees of freedom equal to the product of the number of levels of each factor inside the parentheses, and the number of levels minus one of each factor outside the parentheses.
- A source containing crossed factors has degrees of freedom equal to the product of the number of levels minus one of each factor in the source.
- The total degrees of freedom in a model is one less than the total number of observations.

TABLE 1 *Hypothetical sources of variation and their degrees of freedom*

Source	df	Comments
A	$a - 1$	single factor
B(A)	$(b - 1)(a)$	B nested within A
A*B	$(a - 1)(b - 1)$	A crossed with B
D*B(A)	$(d - 1)(b - 1)(a)$	D crossed with B nested within A
B(AD)	$(b - 1)(a)(d)$	B nested within A and D

As an illustration, assume factor A has  $a$  levels, factor B has  $b$  levels, and factor D has  $d$  levels. Table 1 shows some hypothetical sources of variations and the formulae to compute their degrees of freedom.

We will not discuss the rules for determining degrees of freedom in unbalanced designs, as they are quite involved. Interested readers can consult Sokal and Rohlf (1981, Sections 9.2 and 10.3).

**3.3.4 Mean squares** The **mean square** of a source of variation is its sum of squares divided by its associated degrees of freedom. It is denoted as “MS,” often with a subscript to identify the source.

$$MS = \frac{SS}{df} \quad (3)$$

Mean square is an “average” of the sum of squares. It is an estimate of the variance of that source. For instance, as shown in the last section, the sum of squares

$$SS = \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (4)$$

has  $(n - 1)$  degrees of freedom. Dividing this sum of squares by its degrees of freedom yields

$$MS = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n - 1}, \quad (5)$$

which is the sample variance ( $s^2$ ) used to estimate the population variance based on a set of  $n$  observations ( $Y_1, Y_2, \dots, Y_n$ ).

#### 3.4 ANOVA Assumptions

The objective of the fertilizer study outlined in Section 3.1 is to compare the true mean height increments of the three groups of seedlings, and determine whether they are statistically different. This objective cannot be achieved unless some assumptions are made about the seedlings (Cochran 1947; Eisenhart 1947).

Analysis of variance assumes the following:

1. Additivity. Height measurement is viewed as a sum of effects, which includes (1) the average performance in the seedling population, (2) the fertilizer treatment applied, and (3) the experimental error (inherent variation in the seedlings and variation introduced in conducting the experiment).
2. Equality of variance (homoscedasticity). The experimental errors should have a common variance ( $\sigma^2$ ).
3. Normality. The experimental error should be normally distributed with zero mean and  $\sigma^2$  variance.

4. Independence. The experimental errors are independent for all seedlings. This means that the experimental error for one seedling is not related to the experimental error for any other seedlings.

The last three assumptions restrict the study to subjects that have similar initial characteristics and have known statistical properties. The assumption of independence can be met by using a randomized design.

3.5 The Idea  
behind ANOVA

If fertilizer treatments in the study were applied to thousands and thousands of seedlings, we could then infer that the height increment ( $Y_{ij}$ ) of any seedling that received a particular treatment, say fertilizer A, is simply the average height ( $\mu_i$ ) increment of all seedlings treated with fertilizer A, plus the experimental error ( $e_{ij}$ ). This relationship can be expressed in symbolic form as:

$$Y_{ij} = \mu_i + e_{ij}. \quad (6)$$

The index  $i$  indicates the fertilizer treatment that a seedling received; the index  $j$  identifies the seedling that received each treatment. The group mean,  $\mu_i$ , can be further split into two components: the grand mean ( $\mu$ ), which is common to all seedlings regardless of the fertilizer treatment applied, and the group mean ( $\alpha_i$ ), which represents the effect of the treatment applied. That is,

$$\mu_i = \mu + \alpha_i. \quad (7)$$

Combining equations (6) and (7) we get:

$$Y_{ij} = \mu_i + \alpha_i + e_{ij}. \quad (8)$$

This equation is equivalent to the ANOVA model introduced in Section 3.3.1. Note that:

$$\alpha_i = \mu_i - \mu \quad (\text{group mean} - \text{grand mean})$$

$$\text{and } e_{ij} = Y_{ij} - \mu_i \quad (\text{individual observation} - \text{group mean}).$$

We can express equation (8) solely in terms of  $Y_{ij}$ ,  $\mu$ , and  $\mu_i$  as

$$Y_{ij} = \mu + (\mu_i - \mu) + (Y_{ij} - \mu_i)$$

$$\text{or } Y_{ij} - \mu = (\mu_i - \mu) + (Y_{ij} - \mu_i). \quad (9)$$

Now we can use the corresponding sample values:

$$\bar{Y} \approx \mu \quad (\text{sample grand mean for true grand mean}), \text{ and}$$

$$\bar{Y}_i \approx \mu_i \quad (\text{sample group mean for true group mean})$$

to express the model as:

$$Y_{ij} - \bar{Y} = (\bar{Y}_i - \bar{Y}) + (Y_{ij} - \bar{Y}_i). \quad (10)$$

Equation (10) provides an important basis for computations in ANOVA.

The objective of the fertilizer experiment is to compare the true mean height increment of seedlings given fertilizer A, fertilizer B, or no fertilizer. Suppose that the average height increment of the 10 seedlings given fertilizer A was 10 cm, of those given fertilizer B was 12 cm, and of those given no fertilizer was 9 cm. Before we conclude that fertilizer B best enhances height growth in seedlings, we need to establish that the difference in their means is statistically significant; that is, the difference is caused by the fertilizer treatment and not by chance. Natural variation among seedlings can be measured by the variance among seedlings within groups. Variation attributed to the treatments can be measured by the variance among the group means. If the estimated natural variation is large compared to the estimated treatment variation, then we would conclude that the observed differences among the group means is caused largely by natural variation among the seedlings and not by the different treatments. On the other hand, if the estimated natural variation is small compared to the estimated treatment variation, then we would conclude that the observed differences among group means is caused mainly by the treatments. Hence, comparison of treatment means is equivalent to comparison of variances.

To extend our discussion to a general one-factor fixed design, suppose that the fixed factor has  $k$  levels (compared to three levels in the fertilizer example). Level 1 is applied to  $n_1$  experimental units, level 2 is applied to  $n_2$  experimental units, and so on up to level  $k$ , which is applied to  $n_k$  experimental units. There are in total  $N = n_1 + n_2 + \dots + n_k$  experimental units (compared to  $10 + 10 + 10 = 30$  seedlings in the fertilizer example).

If we square each side of the model expressed in equation (10) and sum over all experimental units in all levels (compared to 10 seedlings receiving each of the 3 levels of fertilizer treatment), we obtain:

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2 = \sum_{i=1}^k n_i (\bar{Y}_{ij} - \bar{Y})^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2, \quad \text{or}$$

$$SS_{\text{total}} = SS_{\text{treat}} + SS_e. \quad (11)$$

This is the sum of squares identity for the one-factor fixed ANOVA model. We will now examine each sum of squares term in the equation more closely.

The term on the left is called ‘‘Total Sum of Squares’’ ( $SS_{\text{total}}$ ). The squares are performed on the differences between individual observations and the grand mean. It is a measure of the total variation in the data. The sum is performed on all  $N$  observations; hence it has  $N - 1$  degrees of freedom (compared to  $30 - 1 = 29$   $df$  in the fertilizer example).

The first term on the right-hand side is called ‘‘Treatment Sum of Squares’’ ( $SS_{\text{treat}}$ ). The squares are performed on the differences between treatment group means and the grand mean. It is a measure of the variation among the group means. The sum is performed on the  $k$  group means weighted by the number of experimental units in each group;

hence it has  $(k - 1)$  degrees of freedom (compared to  $3 - 1 = 2$  *df* in the fertilizer example).

The last term on the right is called “Error Sum of Squares” ( $SS_e$ ). It measures the variation within each group of experimental units. This sum of squares is a sum of  $k$  sums of squares. The squares are performed on the differences between individual observations and their corresponding group means. The degrees of freedom of the  $k$  sums of squares are  $(n_1 - 1), (n_2 - 1), \dots, (n_k - 1)$ . By the rule stated in Section 3.3.3, the degrees of freedom for the error sum of squares is the sum of the individual degrees of freedom,  $\Sigma (n_i - 1) = N - k$  (compared to  $30 - 3 = 27$  *df* in the fertilizer example).<sup>3</sup>

We can divide the sums of squares by their corresponding degrees of freedom to obtain estimates of variance:

$$MS_{\text{total}} = SS_{\text{total}} / (N - 1) \quad (12)$$

$$MS_{\text{treat}} = SS_{\text{treat}} / (k - 1) \quad (13)$$

$$MS_e = SS_e / (N - k) \quad (14)$$

The meaning of these equations can be found when we examine the structure of the sums of squares terms.

In equation (12),  $MS_{\text{total}}$  represents differences between individual data and the grand mean; therefore, it is an estimate of the total variance in the data. In equation (13),  $MS_{\text{treat}}$  represents differences between group means and the grand mean; it is an estimate of the variance among the  $k$  groups of experimental units. This variance has two components that are attributed to:

1. the different treatments applied, and
2. the natural differences among experimental units (experimental error).

In equation (14),  $MS_e$  represents  $k$  sum of squares. It is the pooled variance; that is, a weighted average of the variances within each group. We can pool the variances because of the equality of variances assumption in ANOVA. The error mean squares is an unbiased estimator for the true variance of experimental error,  $\sigma^2$ .

ANOVA uses  $MS_{\text{treat}}$  and  $MS_e$  to test for treatment effects. A ratio of  $MS_{\text{treat}}$  to  $MS_e$  is computed as:

$$F = \frac{MS_{\text{treat}}}{MS_e} \quad (15)$$

If treatments have no effect (all group means are equal), then  $MS_{\text{treat}}$  will contain only the variance that is due to experimental error, and the *F*-ratio will be approximately one. If the treatments have an effect on the

<sup>3</sup> The sum of the individual degrees of freedom would be computed as:  
 $(n_1 - 1) + (n_2 - 1) + \dots + (n_k - 1) = n_1 + n_2 + \dots + n_k - k = N - k$

response measure (at least two group means are different), then  $MS_{\text{treat}}$  will tend to be bigger than  $MS_e$ , and the  $F$ -ratio will be larger than one.

### 3.6 ANOVA $F$ -test

The  $F$ -ratio is the test statistic for the null hypothesis that treatment means are equal. According to probability theory, the ratio of two independent variance estimates of a normal population will have an  $F$ -distribution. Since means squares in an ANOVA model are estimated variances of normally distributed errors,  $F$  has an  $F$ -distribution. The shape of the distribution depends on the degrees of freedom corresponding to the numerator and denominator mean squares in the  $F$ -ratio, and the non-centrality parameter  $\lambda$ . We will not discuss  $\lambda$  at length here. Interested readers can refer to Nemeč (1991). We can simplify the meaning of  $\lambda$  by saying that it is a measure of the differences among the treatment means. In ANOVA, the null hypothesis is assumed to be true unless the data contain sufficient evidence to indicate otherwise. Under the null hypothesis of equal treatment means,  $\lambda = 0$  and  $F$  has a central  $F$ -distribution; otherwise,  $\lambda > 0$ , and  $F$  has a non-central  $F$ -distribution.

In the fertilizer example, the fertilizer factor has three levels, and each level is applied to 10 experimental units. Therefore the numerator mean squares has  $3 - 1 = 2$  degrees of freedom, and the denominator mean squares has  $30 - 3 = 27$  degrees of freedom. The central  $F$ -distribution curve at  $(2,27)$  degrees of freedom is shown in Figure 8. This is the curve on which hypothesis testing will be based.

The  $X$ -axis in Figure 8 represents the  $F$ -values; the  $Y$ -axis represents the probability of obtaining an  $F$ -value. The total area under the  $F$ -distribution curve is one.

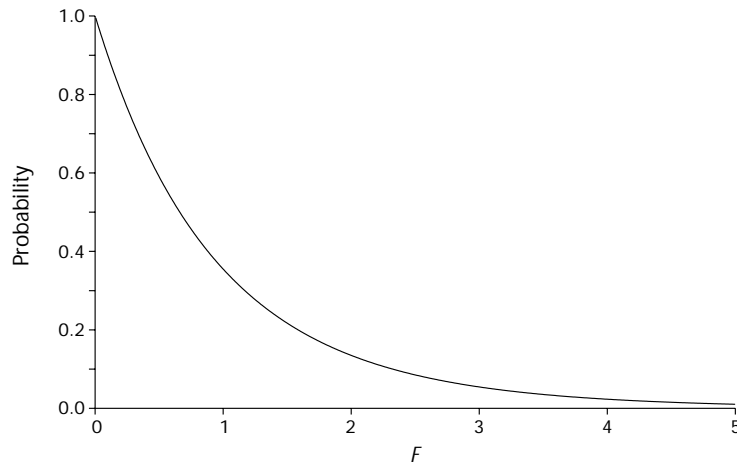


FIGURE 8 Central  $F$ -distribution curve for  $(2,27)$  degrees of freedom.

Let  $F_c$  be the cut-off point such that the null hypothesis is rejected when the observed  $F$ -ratio,  $F_o$ , is greater than  $F_c$ . The tail area to the right



of  $F_c$  is the probability of observing an  $F$ -value at least as large as  $F_c$ . Hence, it represents the probability of rejecting the null hypothesis. On the central  $F$ -curve (which assumes the null hypothesis is true), this tail area also represents the probability of rejecting the null hypothesis by mistake. Therefore, we can set the maximum amount of error (called the significance level  $\alpha$ ) that we will tolerate when rejecting the null hypothesis, and find the corresponding critical  $F_c$  value on the central  $F$ -curve. Figure 9 shows the critical  $F$ -value and decision rule for the fertilizer example. The shaded area on the right corresponds to  $\alpha = 0.05$ , with  $F_c = 3.35$ .

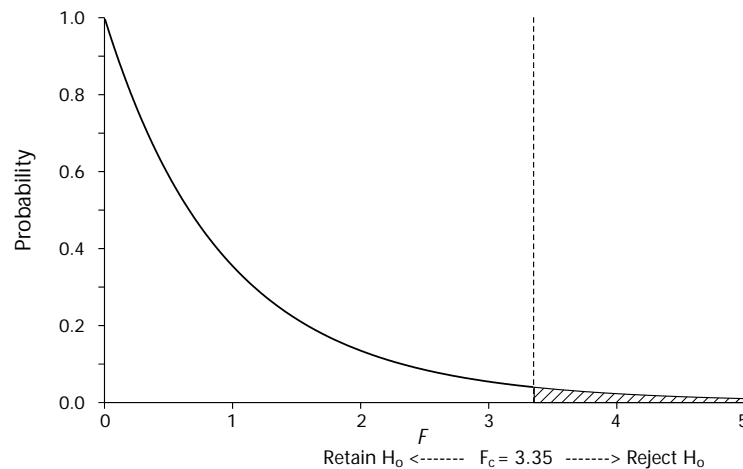


FIGURE 9  $F$ -distribution curve with critical  $F$  value and decision rule.

Alternatively, we can find the right tail area that corresponds to  $F_o$ . This area, called the  $p$ -value, represents the probability that an  $F$ -value at least as large as  $F_o$  is observed. If this probability is large, then the observed event is not unusual under the null hypothesis. If the probability is small, then it implies that the event is unlikely under the null hypothesis and that the assumed null hypothesis is not true. In this case,  $\alpha$  is the minimum probability level below which the event's occurrence is deemed improbable under the null hypothesis. Therefore, when  $p < \alpha$ , the decision is to reject the null hypothesis of no treatment effect. For this example, suppose that  $F_o$  is 3.8. At (2,27) degrees of freedom, the corresponding  $p$ -value is 0.035. At  $\alpha = 0.05$ , we would reject the null hypotheses and conclude that not all the fertilizer treatments are equally effective in promoting growth. However, at  $\alpha = 0.01$  (more conservative), we would not reject the null hypothesis of equal population means because of lack of evidence.

3.7 Concluding  
Remarks about ANOVA

The  $F$ -test is used to draw inferences about differences among the treatment means. When this is not significant, we may conclude that there is no evidence that the treatment means are different. Before finishing the

analysis, however, we should determine the **power** of the  $F$ -test to ensure that it is sensitive enough to detect the presence of real treatment differences as often as possible. Interested readers can consult Cohen (1977), Peterman (1990), and Nemeč (1991).

When the  $F$ -test is significant, we may conclude that treatment effects are present. Because the alternative hypothesis is inexact (i.e., “not **all** treatment means are the same”), so is the conclusion based on the ANOVA. To investigate which treatment means are different, more detailed comparisons need to be made. This is the topic of the next chapter.

ANOVA assumes that the additive model is appropriate, the experimental errors are independent, normally distributed, and have equal variance. These assumptions must be checked before accepting the results. ANOVA is a robust procedure in that it is insensitive to slight deviations from normality or equal variance. However, the procedure is invalid if the experimental errors are dependent. In practice, it is unlikely that all ANOVA assumptions are completely satisfied. Therefore, we must remember that the ANOVA procedure is at best an approximation. For more discussion about methods for checking the ANOVA assumptions and the consequences when they are not satisfied, see Bartlett (1947), Cochran (1947), and Hahn and Meeker (1993).

#### 4 MULTIPLE COMPARISONS

---

There are two ways to compare treatment means: planned comparisons and post-hoc, or unplanned, comparisons. Planned comparisons are determined before the data are collected. They are therefore relevant to the experiment objectives and represent concerns of the experimenter. They can be performed regardless of the outcome of the basic  $F$ -test of equal means. Unplanned comparisons occur after the experimenter has seen the data and are performed only if the basic  $F$ -test of equal means is significant. They are exploratory and are used to search for interesting results but with no particular hypothesis in mind.

In this chapter, we will first examine comparisons in general. This is followed by separate discussions on planned and unplanned comparisons. Finally, some recommendations on performing planned and unplanned comparisons are provided.

##### 4.1 What is a Comparison?

All comparisons can be expressed as an equation. For example, suppose we have three treatment groups (as in the fertilizer example in Section 3.1):

- Group 1: control group to which no treatment is applied
- Group 2: treatment group to which fertilizer A is applied
- Group 3: treatment group to which fertilizer B is applied

Let  $\mu_1$ ,  $\mu_2$ , and  $\mu_3$  denote the true means of the response variable, (height increments) of the three groups. The null hypothesis that the average height increment of seedlings treated with fertilizer A is the same as the

average height increment of seedlings treated with fertilizer B can be expressed as:

$$\mu_2 = \mu_3 \tag{16}$$

or 
$$\mu_2 - \mu_3 = 0 \tag{17}$$

or 
$$\mu_3 - \mu_2 = 0. \tag{18}$$

We can include  $\mu_1$  in equation (17) explicitly by writing the coefficients associated with the means:

$$0\mu_1 + 1\mu_2 + (-1)\mu_3 = 0. \tag{19}$$

The left side of equation (19) is a weighted sum of the means  $\mu_1$ ,  $\mu_2$ , and  $\mu_3$ ; the weights (0, 1, -1), which are the coefficients of the means, sum to zero.

Equation (19) is also a comparison. It is usually referred to as a **contrast**; that is, a weighted sum of means with weights that sum to zero. The weights are called **contrast coefficients**.

We can also develop a contrast from equation (18), with contrast coefficients (0, -1, 1). This set of weights has the opposite sign to those in equation (17). In fact, any comparison, or contrast, can be expressed by two sets of weights that have opposite signs.

The comparison  $\mu_2 = \mu_3$  can also be expressed as  $2\mu_2 = 2\mu_3$ , or  $5\mu_2 = 5\mu_3$ , or  $11.2\mu_2 = 11.2\mu_3$ . The corresponding weights are (0, 1, -1), (0, 2, -2), (0, 5, -5), and (0, 11.2, -11.2), respectively. All four sets of weights represent the same comparison of  $\mu_2$  versus  $\mu_3$ . Indeed, there are an infinite number of possible weights for each contrast; the sets of weights differ only by a multiplicative factor. For convenience, it is customary to express a contrast with the lowest set of whole numbers; in this example, they are (0, 1, -1) or (0, -1, 1).

All comparisons, planned or unplanned, can be expressed as contrasts. Nevertheless, it is conventional to call planned comparisons “contrasts,” and unplanned comparisons “multiple comparisons.” This nomenclature relates to the way these two types of comparisons are usually carried out. In planned comparisons, experimenters design the contrast beforehand; that is, they set the weights or contrast coefficients. In unplanned comparisons, all possible comparisons are made between pairs of means. In this handbook, the term **planned contrasts** refers to planned comparisons and the term **multiple comparisons** refers to unplanned comparisons.

4.2 Drawbacks of Comparisons

Additional comparisons, planned or unplanned, can inflate the overall error rates in an experiment.

All statistical tests are probabilistic. Each time a conclusion is made, there is a probability or chance that the conclusion is incorrect. If we

conclude that the null hypothesis of equal means is false when it is actually true, we are making a type I error. The probability of making this error is  $\alpha$ . On the other hand, if we conclude that the null hypothesis is true when it is actually false, we are making a type II error. The probability of making this error is  $\beta$ . These two types of errors are related to each other like the two ends of a seesaw: as one type of error is kept low, the other type of error is pushed up.

If only one comparison is performed in an experiment, the type I and type II error rates of this test are  $\alpha$  and  $\beta$ , respectively. If two comparisons are performed, and the error rates for each comparison are still  $\alpha$  and  $\beta$ , the overall error rates would increase to approximately  $2\alpha$  and  $2\beta$  as there are more chances to make mistakes. In general, the more comparisons we make, the more inflated are the overall error rates. If we conduct  $n$  independent comparisons, each with type I error rate of  $\alpha$ , then the overall type I error rate,  $\alpha_E$  is

$$\alpha_E = 1 - (1 - \alpha)^n, \quad (20)$$

or approximately

$$\alpha_E = n\alpha \quad (21)$$

when  $\alpha$  is small (Keppel 1973). For example, suppose we want to consider five pairwise comparisons simultaneously. To maintain an overall type I error of  $\alpha_E = 0.05$ , we should use  $\alpha = 0.01$  for each comparison. Many procedures are available to control the overall error rates. These procedures are briefly discussed in Section 4.4.

#### 4.3 Planned Contrasts

Planned contrasts are constructed before the data are collected to test for specific hypotheses. They are useful when analyzing experiments with many types and levels of treatments (Mize and Schultz 1985).

Planned contrasts can be used to compare treatment means when qualitative treatments are grouped by similarities. For example, consider a planting experiment in which seedlings are outplanted at six different dates, three during the summer and three during the spring. If the objective is to compare the effect of summer and spring planting, then a contrast can be used to test the difference between the averages of the three summer and spring means.

Contrasts can also be used in experiments that involve quantitative treatments. Consider an experiment that examines the effect of various amounts of nitrogen fertilizer on seedling performance. Suppose four levels are used: 200 kg N/ha, 150 kg N/ha, 100 kg N/ha, and a control with no fertilizer. We could construct a contrast to compare the control to the three other levels to test whether the fertilizer applications improve seedling performance. We could also construct a contrast to test whether a linear trend existed among the means; that is, if seedling performance got better or worse with increased amounts of nitrogen fertilizer.

We will not discuss how to obtain contrast coefficients for various types of contrasts. Interested readers should consult Keppel (1973), Rosenthal and Rosnow (1985), or Bergerud (1988b, 1989c).

Some authors (Hays 1963; Kirk 1968) suggest that planned contrasts must be statistically independent or **orthogonal** to one another. Winer (1962) argued that, in practice, contrasts are constructed for meaningful comparisons; whether they are independent or not makes little or no difference. Mize and Schultz (1985) supported this view and stated that contrasts do not need to be orthogonal if they are meaningful and have a sound basis for consideration. I agree with Winer (1962) and Mize and Schultz (1985), but emphasize that the number of contrasts tested must be kept to a minimum to reduce the overall error rates.

#### 4.4 Multiple Comparisons

Unplanned multiple comparisons are often “data sifting” in nature. They are performed to squeeze as much information as possible from the data. A common approach is to conduct multiple comparison tests on all possible pairs of means. There are many procedures available for multiple comparisons. In these tests, a critical difference is computed for each pair of means. Two means are declared significantly different if their difference is larger than the critical difference. The critical difference is computed differently for each of the tests. Some tests use the same critical difference for all comparisons while others compute a new value for each comparison. We will briefly describe the characteristics of several popular procedures in this section. Readers should refer to standard textbooks such as Keppel (1973), Steel and Torrie (1980), Sokal and Rohlf (1981), and Milliken and Johnson (1992) for the formulae and exact test procedures.

**Least Significant Difference (LSD)** is one of the simplest multiple comparison procedures available for comparing pairs of means. Each comparison is a paired  $t$ -test. No protection is made against inflating the  $\alpha$ -level, so this procedure tends to give significant results. It is called **unrestricted LSD** if it is used without regard to the outcome of the basic  $F$ -test of equal means. Fisher (1935) recommended that LSD tests should be performed only if the basic  $F$ -test is significant, in which case it is called **Fisher’s LSD** or **restricted LSD**. In an experiment with several treatments, if two treatment means are different but the remaining means are equal, it is possible that the basic  $F$ -test is non-significant. Therefore, while Fisher’s LSD improves the overall  $\alpha$ -level, it may not detect existing differences between some pairs of treatment means (Milliken and Johnson 1992, Chapter 3).

A quick way to adjust the overall  $\alpha$ -level is to apply the **Bonferroni** correction. Suppose an experimenter wants to make  $k$  comparisons. The  $k$  tests will give an overall error of less than or equal to  $\alpha$  if the error rate of each test is set at  $\alpha/k$ . The Bonferroni procedure is valid for data with either equal or unequal sample sizes.

**Scheffé’s** test is very general, allowing an infinite number of comparisons to be made while maintaining a reasonable overall type I error rate. Scheffé’s test is more conservative than the LSD as it is less likely to declare differences between means. It has very low power and will not identify any significant differences unless the basic  $F$ -test of equal means is significant. Scheffé’s test does not require equal numbers of observations for each mean.

**Tukey's Honestly Significant Difference** (Tukey's HSD) requires that the means be based on the same number of observations. Nonetheless, an approximation procedure can be used if the observations are not too unequal. Tukey's test gives an overall error rate of less than or equal to  $\alpha$ . When only pairs of means are compared, Tukey's HSD will always use a smaller critical difference than Scheffé's test.

**Newman-Keul's test** does not make all possible pairwise comparisons. Instead, it compares the means in a systematic way. A  $t$ -test is first used to compare the largest and smallest means. If the means are not significantly different, the procedure will stop. Otherwise, the means will be grouped into two sets, one containing all of the means except the largest, the other containing all of the means except the smallest. A  $t$ -test will be performed on the largest and the smallest means in each set. The procedure continues to examine smaller subsets of means until all  $t$ -tests are non-significant. Since the number of comparisons performed varies according to the means, the exact  $\alpha$  error rate is unknown. Newman-Keul's test is more liberal than Tukey's HSD, but is more conservative than LSD.

**Duncan's Multiple Range test** is one of the most popular multiple comparison procedures. This procedure also requires equal numbers of observations for each mean. It is similar to Newman-Keul's test except that it uses a variable  $\alpha$  level depending on the number of means involved at each stage. As a result, Duncan's test is less conservative than the Newman-Keul's test.

#### 4.5 Recommendations

There are many debates over the use of multiple comparison procedures. The following are suggestions made by different authors.

- Saville (1990) stated that all multiple comparison procedures, except unrestricted LSD, are inconsistent. A given procedure can return a verdict of "not significant" for a given difference in one experiment, but return a verdict of "1% significant" for the same difference in a second experiment, with no change in the standard error of the difference or the number of error degrees of freedom. He felt that the biggest misuse of multiple comparison procedures was the attempt to formulate and test hypotheses simultaneously. Saville suggested that unrestricted LSD should be used to generate hypotheses of differences. These differences must be confirmed in subsequent studies.
- Mize and Schultz (1985) stated that test selection should be based on how conservative or liberal the researcher wants to be in declaring means to be significantly different. They quoted Chew's (1976) ranking of multiple comparison procedures by increasing the likelihood of declaring differences significant: Scheffé's test, Tukey's HSD, Newman-Keul's test, Duncan's multiple range test, and Fisher's LSD.
- Mead (1988, Section 12.2) opposed the use of multiple comparison methods. He argued that these are routinely used in inappropriate situations and thereby divert many experimenters from proper analysis of their data.
- Milliken and Johnson (1992, Section 8.5) recommended that if the basic  $F$ -test of equal means is significant, LSD should be used for any planned comparison and Scheffé's test for unplanned comparisons. If

the basic  $F$ -test of equal means is not significant, then only planned contrasts with the Bonferroni procedure should be performed.

I agree with some of the above suggestions and recommend the following:

1. Formulate planned contrasts according to the objectives of the study. These contrasts should be performed regardless of the result of the basic  $F$ -test of equal means.
2. If the basic  $F$ -test of equal means is **significant**, then a multiple comparison method of the experimenter's choice can be used for data exploration. The method chosen will depend on how conservative or liberal the experimenter wishes to be in declaring means to be significantly different. The experimenter must ensure that the selected method is appropriate. For example, use Duncan's test only if there are equal numbers of observations for all treatment means.
3. Use the Bonferroni correction with the LSD method.
4. Remember that significant results from unplanned comparisons are only indications of possible differences. Any differences of interest must be confirmed in later studies.
5. If the basic  $F$ -test of equal means is not significant, then unplanned comparisons should **not** be carried out.
6. Always plot the means on a graph for visual comparison.
7. Always examine the data from the biological point of view as well as from the statistical point of view. For example, a difference of 3 cm in seedling growth may be statistically insignificant but biologically significant.

## 5 CALCULATING ANOVA WITH SAS: AN EXAMPLE

---

In this chapter, we demonstrate how to:

- recognize the design of an experiment,
- set up the ANOVA table, and
- write an SAS program to perform the analysis.

All discussions are based on the following example.

### 5.1 Example

Suppose we want to test the effect of three types of fertilizer and two root-pruning methods on tree growth. We have twelve rows of trees, with four trees in each row. We randomly assign a fertilizer and a root-pruning treatment to a row of trees so that each fertilizer and root treatment combination is replicated twice. We measure the height of each tree at the end of five weeks.

### 5.2 Experimental Design

To determine the design of an experiment, we must ask the following questions:

- What are the factors?
- Are the factors fixed or random, nested or crossed?
- What are the experimental units and elements?

We can draw a "stick" diagram to help answer some of these questions.

A stick diagram is drawn in a hierarchical manner to illustrate the design structure of an experiment. In the diagram, the experiment's factors are positioned so that the one with the largest experimental unit is listed first, followed by its experimental unit. The factor with the next largest experimental unit and its experimental unit come next, and so on to the smallest elements. If several factors have the same experimental unit, then all of these would be listed before the experimental unit. The order is not usually important except for nested factors, where the nested factor must follow the main factor. The levels of each factor are numbered, with distinct levels being denoted by different numbers; each experimental unit has a different number as each unit is unique. The relationships between levels of different factors and experimental units are indicated by straight lines or sticks. For the example, we observe that:

- there are two factors, fertilizer, F, and root-pruning treatment, T;
- both F and T are fixed factors; they are crossed with each other; and
- the experimental unit for both F and T is a row of trees; a tree is an element.

The stick diagram for this example is shown in Figure 10.

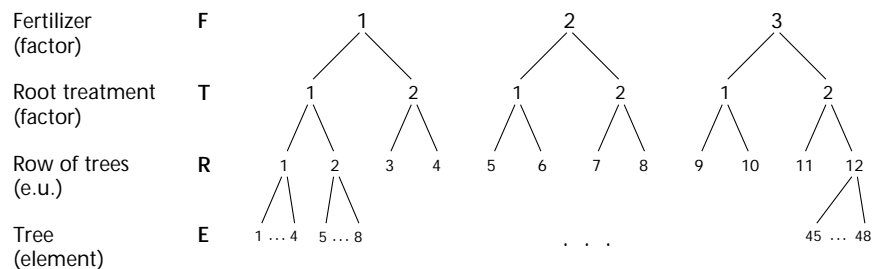


FIGURE 10 Design structure for the fertilizer and root-pruning treatment example.

Note that we could also begin the diagram with T since both F and T have the same experimental unit (see Figures 1 and 2, for example). Two sticks come under each level of F, one for each of the two root-pruning methods. Since each level of F is combined with the same two levels of T, the numbers for T repeat across F. This implies that F and T are crossed.

The next level in the stick diagram corresponds to the row of trees, R, the experimental units for F and T. Two sticks come out from each F\*T combination as each is replicated twice. A different number is used for each row of trees because each row is distinct. The non-repeating numbers imply that the row of trees unit is nested within the fertilizer and root-pruning treatment factors; that is, R(FT). Finally, four trees in each row are sampled. As the trees are different, a unique number is assigned to each. The element tree is nested within the row of trees unit, fertilizer, and root-pruning treatment factors; that is, E(RFT).

This is a completely randomized factorial design. It is a completely randomized design because the levels of the treatment factors are



randomly assigned to the experimental units with no restriction. It is factorial because more than one factor is involved, and the factors have the same experimental unit and are crossed with each other. A complete description of this and other common designs is provided in Chapter 6.

### 5.3 ANOVA Table

The ANOVA table contains the stick diagram information and the necessary values to perform the  $F$ -test. Because it reveals the experimental design, it is useful even when ANOVA is not the method of analysis. The ANOVA table generated by statistical software usually has six columns: Source,  $df$ , SS, MS,  $F$ , and  $p$ ; the last four columns (SS, MS,  $F$ , and  $p$ ) are excluded in the planning stage of a study when no data are available. Another column called "Error" is often added to the ANOVA table to indicate the error term (that is, the denominator in the  $F$ -ratio) used to test the significance of a source of variation. To simplify the discussion in this chapter, factors, experimental units, and elements are referred to collectively as **variates** (in this example, F, T, R, and E are variates).

To complete an ANOVA table, we must know all the sources of variation in the study, the degrees of freedom ( $df$ ) for each source of variation, and the nature (fixed or random) of each source of variation.

To compile the list of sources:

- List all variates. This should include all the entries in the stick diagram.
- List all possible interaction terms by combining all the variates in every possible way.
- Delete all meaningless interaction terms; that is, terms in which the same variate appears both inside and outside of any parentheses.

Steps 2 and 3 can be tedious for a complicated model. To save time, we could examine the stick diagram and note how the numbers vary at each level; repeated numbers imply a crossed relationship, while unique numbers imply a nested relationship. For example, in Figure 10, the numbers for T repeat for each level of F because T is crossed with F. The numbers for the nested factors, R and E, do not repeat. Note that for crossed relations, both of the main factors (F and T in our example) and their interaction term (F\*T) are sources of variation. However, for nested relations, only the nested terms are sources. For our example, R(FT) and E(RFT) are sources of variation, but R and E alone are not. A source can be fixed or random depending on its composition; it is fixed if every variate in that source is fixed, otherwise it is random. For our example, F, T, and F\*T are fixed sources while R(FT) and E(RFT) are random.

When the list is complete, we can find the degrees of freedom for each source of variation by using the rules stated in Section 3.3.3. The sum of all the degrees of freedom should be the total number of elements minus one. Examining the sum provides a good check to see if all the sources are included. Table 2 below shows a partial ANOVA table for our example with only the sources and their degrees of freedom.

The sum of squares (SS) and mean square (MS) values are required to calculate the  $F$ -ratio. Formulae for SS exist but are quite involved for complex designs. Statistical software such as SAS, SYSTAT, and S-Plus are often used to do the calculations. Therefore, we will use the software and not concern ourselves with the SS computation equations.

Recall from Section 3.5 that the MS of a source is a variance estimate at that source level and that the variance can have more than one component. For the fertilizer example of Chapter 3, the MS for fertilizer has a component linked to the fertilizer treatment applied and a component linked to the natural differences among the seedlings. The makeup of an MS is often revealed in the expected means square (EMS) equation. This equation shows a particular source's variance components. It is used to determine the error term for the source's  $F$ -ratio. The rules for determining the EMS for any source are given in Appendix 1. Table 3 below is a partial ANOVA table for our factorial example. It is essentially Table 2 with an added column for EMS.

TABLE 2 *Partial ANOVA table for the fertilizer and root-pruning treatment example*

Source	$df$
Fertilizer, F	$3 - 1 = 2$
Root treatment, T	$2 - 1 = 1$
F * T	$(3 - 1)(2 - 1) = 2$
Row, R (FT)	$(2 - 1)(3)(2) = 6$
Tree, E (RFT)	$(4 - 1)(2)(3)(2) = 36$
Total	$(3)(2)(2)(4) - 1 = 47$

TABLE 3 *Partial ANOVA table for the fertilizer and root-pruning treatment example*

Source	$df$	EMS
Fertilizer, F	2	$\sigma^2_{E(RFT)} + 4\sigma^2_{R(FT)} + 16\phi_F$
Root treatment, T	1	$\sigma^2_{E(RFT)} + 4\sigma^2_{R(FT)} + 24\phi_T$
F * T	2	$\sigma^2_{E(RFT)} + 4\sigma^2_{R(FT)} + 8\phi_{F*T}$
Row, R (FT)	6	$\sigma^2_{E(RFT)} + 4\sigma^2_{R(FT)}$
Tree, E (RFT)	36	$\sigma^2_{E(RFT)}$
Total	47	

Notice in the EMS equations that a variance component linked to a random source is denoted by  $\sigma^2$  and one linked to a fixed source is denoted by  $\phi$ .

The  $F$ -ratio of a source of variation is the ratio of its MS to another MS. The denominator MS is called the "error term." The correct error term for a source of interest can be found by comparing its EMS to the EMS of the other sources. To determine the error term of a source of variation, *find another source that has all the same terms in its EMS as the source of interest, except for the term directly related to the source of interest.*

Applying this rule, we see from Table 3 that the error term for testing a fertilizer effect is the mean square of the row of trees,  $MS_{R(FT)}$ . This is because the expected mean square for R(FT) is the same as the expected mean square for F, except for the term  $16\phi_F$ , which depends only on the fertilizer treatment means. If a fertilizer effect does not exist or is

minimal, then the term  $16\phi_F$  would be close to zero and the ratio of  $MS_F$  to  $MS_{R(FT)}$  would be approximately one. Conversely, if a fertilizer effect is significant, the variance component attributed to  $16\phi_F$  would be large and the  $F$ -ratio would be greater than one. The complete ANOVA table for our fertilizer and root-pruning example is given in Table 4.

In this example,  $R(FT)$  is the experimental error and  $E(RFT)$  is the sampling error. In most textbooks and papers, the last source —  $E(RFT)$  in this example — in the ANOVA table is denoted as “Error” without identifying its composition. I recommend specifying the composition of each source because this helps us to recognize the nature and origin of the variation associated with the source and determine its degrees of freedom, EMS equation, and error term.

TABLE 4 ANOVA table for the fertilizer and root-pruning treatment example

Source	<i>df</i>	EMS	Error
Fertilizer, F	2	$\sigma_{E(RFT)}^2 + 4\sigma_{R(FT)}^2 + 16\phi_F$	$MS_{R(FT)}$
Root treatment, T	1	$\sigma_{E(RFT)}^2 + 4\sigma_{R(FT)}^2 + 24\phi_T$	$MS_{R(FT)}$
F*T	2	$\sigma_{E(RFT)}^2 + 4\sigma_{R(FT)}^2 + 8\phi_{F*T}$	$MS_{R(FT)}$
Row, R(FT)	6	$\sigma_{E(RFT)}^2 + 4\sigma_{R(FT)}^2$	$MS_{E(RFT)}$
Tree, E(RFT)	36	$\sigma_{E(RFT)}^2 + 4\sigma_{R(FT)}^2$	—
Total	47		

#### 5.4 SAS Program

The SAS procedures `PROC GLM` and `PROC ANOVA` can be used to perform an analysis of variance. The ANOVA procedure, `PROC ANOVA`, is suitable for one-factor ANOVAs, balanced or not, and for any other models that are balanced. The general linear models procedure, `PROC GLM`, is appropriate for any balanced or unbalanced design. This procedure can save residuals and perform contrasts, but is slower and requires more memory than `PROC ANOVA`. We will use `PROC GLM` in all of our examples.

The SAS `PROC GLM` program used to run an ANOVA is similar for most designs. The following SAS statements are essential:

- `CLASS` defines the class variables or factors. These variables have only a fixed number of values. Examples are fertilizer types, root-pruning methods, or the row of trees;
- `MODEL` defines the model to be fitted. The `MODEL` statement has the form:

```
MODEL response = sources;
```

where: `response` is the response variable(s) (e.g., tree heights) and `sources` is the list of sources, excluding the last source in the ANOVA table.

For the above example, if height (HT) is the response variable, then an appropriate `MODEL` statement is:

```
MODEL HT = F T F*T R(F T);
```

The sources F, T, and F\*T can be specified with the shorthand crossed variable symbol F|T. Therefore, the last MODEL statement can be shortened to:

```
MODEL HT = F|T R(F T);
```

As shown in Table 4, the tree term, E(RFT), is the last source listed on the ANOVA table (the last entry, Total, does not count). It is not included in the MODEL statement and is used by SAS as the default error term for all *F*-tests. This means that all the *F*-ratios are computed using  $MS_{E(RFT)}$  as the denominator. Since R(FT) is the correct error term for F, T, and F\*T, the ANOVA table generated by SAS is wrong for these three sources.

To perform the correct *F*-tests, we can use the TEST statement to specify the correct error term for each test. The TEST statement has the form:

```
TEST H=sources E=error;
```

More than one source can be listed in the option "H=" if all have the same error term. For example, to test the main effects F and T, and the interaction effect F\*T, we could use the TEST statement:

```
TEST H=F T F*T E=R(F T);
```

We can also use the shorthand crossed variable symbol in the TEST statement:

```
Test H=F|T E=R(F T);
```

Besides testing for differences among treatment levels of a source of variation, we can also test contrasts or make multiple comparisons. Contrasts can be specified with the CONTRAST statement which has the form:

```
CONTRAST 'label' source coefficients / E = error;
```

where: *label* labels the contrast (it is optional),  
*source* lists the source on which contrast tests will be performed,  
*coefficients* specifies the contrast to be tested, and  
*error* specifies the error term to be used in the contrast test.

For example, to perform a contrast on the first fertilizer versus the average of the second and third fertilizers, we would use the SAS statement:

```
CONTRAST 'First vs Second & Third F' F -2 1 1 / E=R(F T);
```

Sometimes SAS will declare that a contrast is “non-estimable.” In this case, we have to compute the contrast  $F$ -test by hand. See Bergerud (1993) for a demonstration of calculating contrasts by hand.

Multiple comparisons can be requested as an option in the MEANS statement in PROC GLM. The MEANS statement has the form:

```
MEANS source / method option;
```

where: `source` is the source whose means will be computed,

`method` is the name of the multiple comparison procedure to be used on all main effect means in the MEANS statement. Some possible tests are:

DUNCAN for Duncan’s test,  
 LSD or T for pairwise  $t$ -tests or Least Significance Difference test,  
 SCHEFFÉ for Scheffé’s test, and  
 TUKEY for Tukey’s test.

`option` are options to specify details for the multiple comparison procedure. For example:

CLDIFF requests confidence intervals for all pairwise differences between means, and  
 E=effect specifies the error term to use in the multiple comparisons.

For example, suppose we want to compute the means of the factor F and perform LSD tests with confidence intervals. We would issue the statement:

```
MEANS F / LSD E=R(F T) CLDIFF;
```

For more information on these and other SAS statements in PROC GLM, consult the SAS/STAT User’s Guide (1989).

The following is a SAS program to perform an ANOVA for the two-way fixed factor design described in this chapter. In addition to the basic ANOVA, a contrast comparing the first and third fertilizers, and LSD tests on the factor F with confidence intervals are also requested.

```
PROC GLM DATA=EXAMPLE;
  CLASS F T R;
  MODEL HT = F|T R(F T);
  TEST H = F|T E=R(F T);
  CONTRAST `First vs Third in F' F -1 0 1 / E = R(F T);
  MEANS F / LSD E = R(F T) CLDIFF;
RUN;
```

By putting the CONTRAST statement before the MEANS statement, the contrast results come out on the same page as the ANOVA table. The SAS output, based on the hypothetical data set given in Appendix 2, is displayed in Figure 11.

Page 1 of the SAS output is generated by the CLASS statement. It is a summary of variables listed in the CLASS statement. Page 2 of the output is generated by the MODEL, TEST, and CONTRAST statements. The first ANOVA table only has two sources: Model and Error. Model corresponds to the sum of all the sources in the MODEL statement. Error represents the sources not listed in the MODEL statement, which in this case is the source of variation attributed to tree, E(RFT). The source Model is then partitioned into its component sources as specified in the MODEL statement. The table showing the Type I SS should be discarded because this type of SS depends on the order of the sources listed in the MODEL statement. When interpreting results, we should look at the Type III SS ANOVA table. All of the  $F$ -values in the ANOVA table are calculated by using the Error mean square in the first ANOVA table as the denominator. Since only the source R(F\*T) should be tested by E(RFT), all the other  $F$ -tests are wrong. The next ANOVA table comes from the TEST statement. As stated in the heading, R(F\*T) is used as the error term; hence the results are correct. Since the interaction effect F\*T is not significant at  $\alpha = 0.05$ , we might conclude that the effect of fertilizer treatments is consistent across the root-pruning treatments (at least, there is no evidence to the contrary), and vice versa. Therefore, tests on the separate main effects are meaningful. The tests on F and T suggest that they are highly significant ( $p = 0.001$ ). The last ANOVA test on page 2 gives the contrast  $F$ -test. The contrast is also highly significant ( $p = 0.001$ ), which suggests that trees treated with fertilizer 1 reach a different height than those treated with fertilizer 3.

Finally, the last page of the SAS output gives the multiple comparisons requested by the MEANS statement. See the NOTE that the experimentwise, or overall type I error rate is not controlled for the LSD test. For more information on interpreting SAS outputs, see examples given in the SAS/STAT User's Guide (1989, Chapter 24).

---

The SAS System	1
General Linear Models Procedure Class Level Information	
Class	Levels      Values
F	3            1 2 3
T	2            1 2
R	12           1 2 3 4 5 6 7 8 9 10 11 12
Number of observations in data set = 48	

---

FIGURE 11 SAS output for fertilizer and root-pruning example.

## General Linear Models Procedure

Dependent Variable: HT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	13933.94309	1266.72210	38.79	0.0001
Error	36	1175.65898	32.65719		
Corrected Total	47	15109.60207			

R-Square	C.V.	Root MSE	HT Mean
0.922191	16.72028	5.714647	34.17795

Source	DF	Type I SS	Mean Square	F Value	Pr > F
F	2	12601.15203	6300.57601	192.93	0.0001
T	1	1280.15974	1280.15974	39.20	0.0001
F*T	2	14.58631	7.29316	0.22	0.8010
R(F*T)	6	38.04501	6.34083	0.19	0.9764

Source	DF	Type III SS	Mean Square	F Value	Pr > F
F	2	12601.15203	6300.57601	192.93	0.0001
T	1	1280.15974	1280.15974	39.20	0.0001
F*T	2	14.58631	7.29316	0.22	0.8010
R(F*T)	6	38.04501	6.34083	0.19	0.9764

Tests of Hypotheses using the Type III MS for R(F\*T) as an error term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
F	2	12601.15203	6300.57601	993.65	0.0001
T	1	1280.15974	1280.15974	201.89	0.0001
F*T	2	14.58631	7.29316	1.15	0.3777

Tests of Hypotheses using the Type III MS for R(F\*T) as an error term

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
First vs Third in F	1	12582.82905	12582.82905	1984.41	0.0001

FIGURE 11 *Continued*

## General Linear Models Procedure

T tests (LSD) for variable: HT

NOTE: This test controls the type I comparisonwise error rate  
not the experimentwise error rate.

Alpha= 0.05 Confidence= 0.95 df= 6 MSE= 6.340835

Critical Value of T= 2.44691

Least Significant Difference= 2.1784

Comparisons significant at the 0.05 level are indicated by `\*\*\*`.

F Comparison		Lower Confidence Limit	Difference Between Means	Upper Confidence Limit	
3	- 2	16.3405	18.5190	20.6974	***
3	- 1	37.4808	39.6592	41.8377	***
2	- 3	-20.6974	-18.5190	-16.3405	***
2	- 1	18.9618	21.1402	23.3187	***
1	- 3	-41.8377	-39.6592	-37.4808	***
1	- 2	-23.3187	-21.1402	-18.9618	***

FIGURE 11 *Continued*

## 6 EXPERIMENTAL DESIGNS

In this chapter, we look at three types of experimental designs: completely randomized designs, randomized block designs, and split-plot designs. Several examples<sup>4</sup> are presented for each design. Stick diagrams, ANOVA tables, and sample SAS programs are provided for the examples.

All of the examples assume that:

- the responses of all experimental units and elements are independent of one another (this can be ensured by proper randomization);
- the experimental errors can be modelled reasonably by a normal distribution with a constant variance; and
- the ANOVA is balanced; that is, each treatment is applied to the same number of experimental units and an equal number of elements is measured.

4 All the examples used in this chapter are adapted from Bergerud (1991).



The sample sizes used in the examples are too small for practical purposes (the power of the various tests would be low), but are easier to handle for demonstration purposes. This chapter does not include all the theoretical details of the designs. For more in-depth discussions, consult Keppel (1973), Anderson and McLean (1974a), Steel and Torrie (1980), Sokal and Rohlf (1981), Mead (1988), Milton and Arnold (1990), and Milliken and Johnson (1992).

In a **completely randomized design**, all experimental units are assumed to be homogeneous and the treatments are assigned to the experimental units completely at random. Generally, the treatments are assigned to an equal number of experimental units, although this is not required (Milliken and Johnson 1992, Section 4.2.1).

**6.1.1 One-way completely randomized design** This is the simplest kind of ANOVA in which only one factor is involved. The main objective is to compare the mean responses attributed to the different levels of the factor.

**Example:** Suppose we have six fertilizer treatments and each treatment is applied to three trees, each in its own pot. The eighteen trees in the experiment are selected at random from a well-defined population, and each tree is assigned randomly to one of the six treatments. We would like to test whether the six fertilizer treatments are equally effective in promoting tree growth.

- Factor: Fertilizer, F with  $f = 6$  levels (with an unfertilized control).
- F is fixed.
- Experimental unit for F is a pot containing one tree, R.
- There are  $r = 3$  experimental units per level of F; that is, each level of F is replicated three times.
- R is random, nested in F.
- A tree is an element (same as experimental unit).

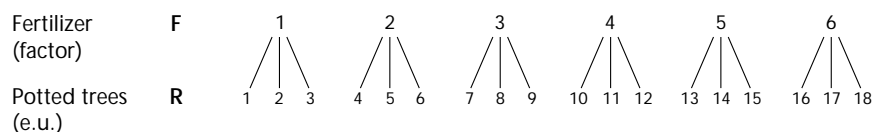


FIGURE 12 Design structure of a one-way completely randomized design.

TABLE 5 ANOVA table for a one-way completely randomized design

Source of variation	$df$	EMS	Error
Fertilizer, F	$f - 1 = 5$	$\sigma_{R(F)}^2 + 3\phi_F$	R(F)
Potted trees, R(F)	$(r - 1)f = 12$	$\sigma_{R(F)}^2$	—
Total	$fr - 1 = 17$		

Suppose that the six treatments are varying amounts of fertilizer: 0, 5, 10, 15, 20, and 25 (kg N/ha), where 0 represents the unfertilized control. Two contrasts of interest are:

1. to test the overall effect of the fertilizer against the control, and
2. to test whether the fertilizer effect is linear with the application rate.

For the treatment levels:	0	5	10	15	20	25
the contrast coefficients are:						
Contrast (1):	-5	1	1	1	1	1
Contrast (2):	-5	-3	-1	1	3	5

See Bergerud (1988b) for a discussion on how to choose the coefficients for the linear contrast.

**SAS Program:**

```
PROC GLM;
  CLASS F;
  MODEL Y = F;
  /** Fertilizer Levels:      0  5 10 15 20 25  ***/
  CONTRAST 'LINEAR IN F'      F -5 -3 -1  1  3  5 ;
  CONTRAST 'FIRST VS LAST 5' F -5  1  1  1  1  1 ;
  MEANS F / SCHEFFE;
RUN;
```

**Note:**

- Scheffé’s tests are requested to perform multiple comparisons on the means of F.
- The source R(F) is not included in the MODEL statement and therefore is the default error term for all tests.
- CONTRAST statements assume a certain ordering to the treatment levels. This order can be checked by looking at the first page produced by PROC GLM. For example, the above SAS program would produce the following first page, which gives the assumed order of the six levels of fertilizer application:

```
SAS

General Linear Models Procedure
Class Level Information

Class  Levels      Values
F      6           0 5 10 15 20 25
```

Number of observations in data set = 18

- Be aware that SAS sorts character values alphabetically. For example, if the fertilizer levels were HIGH, MEDIUM, and LOW, then SAS would order the levels as HIGH, LOW, and MEDIUM, and the proper

contrast coefficients for testing HIGH versus LOW are (1 -1 0). To avoid confusion, a useful trick is to number your factors in order. For example, using the variable names 1LO, 2MED, and 3HI would ensure SAS sorting them in the “logical” order.

- In all the SAS programs in this chapter,  $Y$  in the MODEL statement is the response variable. For this example,  $Y$  could be height increment or diameter.

**6.1.2 Subsampling** In the previous example, a tree is both an experimental unit and an element. In many experimental situations, however, experimental units may contain a number of smaller units that are actually measured. For instance, an experimental unit could be a row of 10, 20, or 50 trees, and all or some of the trees may be selected from each row for measurements. These trees are the elements, sometimes referred to as subsamples. Differences among elements within an experimental unit are observational rather than experimental unit differences (Steel and Torrie 1980, Section 7.9). The mean square of the variation attributed to the elements is generally referred to as sampling error. It can be used to test for differences among experimental units.

**Example:** Suppose each fertilizer treatment is applied to three rows of four trees. How does this change the structure of the design, the ANOVA table, and the SAS program?

- Factor: Fertilizer, F with  $f = 6$  levels (with an unfertilized control).
- F is fixed.
- Experimental unit for F is a row of four trees, R.
- There are  $r = 3$  experimental units per level of F, that is, each level of F is replicated three times.
- R is random, nested in F.
- A tree, E, is an element.
- There are  $e = 4$  trees per row.
- E is random, nested in F and R.

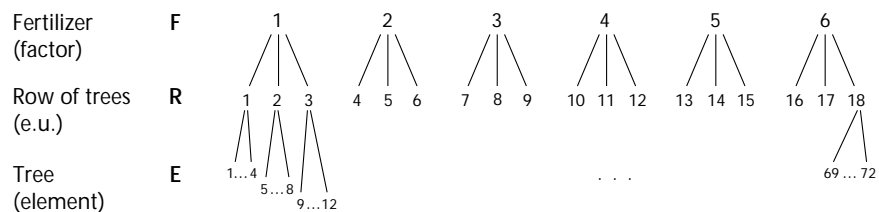


FIGURE 13 Design structure for a one-way completely randomized design with subsamples.

TABLE 6 ANOVA table for a one-way completely randomized design with subsamples

Source of variation	df	EMS	Error
Fertilizer, F	$f - 1 = 5$	$\sigma_{E(FR)}^2 + 4\sigma_{R(F)}^2 + 12\phi_F$	R(F)
Row of trees, R(F)	$(r - 1)f = 12$	$\sigma_{E(FR)}^2 + 4\sigma_{R(F)}^2$	E(FR)
Tree, E(FR)	$(e - 1)fr = 54$	$\sigma_{E(FR)}^2$	—
Total	$fre - 1 = 71$		

**SAS Program:** There are several ways to calculate the ANOVA for this design.

- 1. Analysis of individual data:** Use the raw data in the analysis. Notice in the following program that the error term used to test F is stated explicitly in the TEST and CONTRAST statements to override the use of the term — in this case, E(FR).

```
PROC GLM;
  CLASS F R;
  MODEL Y = F R(F);
  /** Fertilizer Levels:      0  5 10 15 20 25  ***/
  CONTRAST 'LINEAR IN F'      F -5 -3 -1  1  3  5 / E=R(F);
  CONTRAST 'FIRST VS LAST 5' F -5  1  1  1  1  1 / E=R(F);
  TEST H = F      E = R(F);
  MEANS F / SCHEFFE E = R(F);
RUN;
```

- 2. Analysis of experimental unit means:** The subsamples within each experimental unit are averaged and the analysis is performed on the means as if there was only one observation for each unit. In our example, the tree responses are averaged for each row (using PROC MEANS) and the row means are used in the ANOVA (i.e., 18 data points instead of the 72 used in method 1).

```
PROC MEANS NWAY NOPRINT MEAN;
  CLASS F R;
  VAR Y;
  OUTPUT OUT=YMEANS MEAN=YMEAN;
PROC GLM DATA=YMEANS;
  CLASS F;
  MODEL YMEAN = F;
  /** Fertilizer Levels:      0  5 10 15 20 25  ***/
  CONTRAST 'LINEAR IN F'      F -5 -3 -1  1  3  5 ;
  CONTRAST 'FIRST VS LAST 5' F -5  1  1  1  1  1 ;
  MEANS F / SCHEFFE;
RUN;
```

If the design is balanced, then methods (1) and (2) would result in the same  $F$ -test for the main effect. Method (2) is quite popular because it

simplifies the design of an experiment — by using the experimental unit means, the element level is removed from the design. Also, the means are more likely (according to Central Limit Theorem) to be normally distributed. The main disadvantage of analyzing the experimental unit means is a loss of potentially interesting information about the variation among the elements within the experimental units (e.g., information to determine the optimum number of rows and trees for future experiments).

When the design is unbalanced, method (2) is undesirable as it ignores the unbalanced nature of the experimental units. To account for this lack of balance, Rawlings (1988, Section 16.4) recommends a weighted analysis of the experimental unit means.

**3. Weighted analysis of the experimental unit means:** In this case, sample sizes of the experimental units are used as weights in the analysis. This results in best linear unbiased estimates — the most desirable kind of estimates for the model parameters. The following program shows how to do a weighted analysis using SAS. Row means and sample sizes are computed with PROC MEANS and saved in a SAS data set called YMEANS. Then, a weighted analysis of the means is performed using PROC GLM in which the weights are specified with the WEIGHT statement.

```
PROC MEANS NWAY NOPRINT MEAN;
  CLASS F R;
  VAR    Y;
  OUTPUT OUT=YMEANS MEAN=YMEAN N=NUM;
PROC GLM DATA=YMEANS;
  CLASS F;
  MODEL YMEAN = F;
  WEIGHT NUM;
  /*** Fertilizer Levels:      0  5 10 15 20 25  ***/
  CONTRAST 'LINEAR IN F'      F -5 -3 -1  1  3  5  ;
  CONTRAST 'FIRST VS LAST 5' F -5  1  1  1  1  1  ;
  MEANS    F / SCHEFFE;
RUN;
```

For balanced designs, all three methods generate the same results. For unbalanced designs, methods (1) and (3) give similar  $F$ -tests and sum of squares. When a design involves subsampling, the following steps are recommended:

- If possible, use the full data set in the analysis.
- If experimental unit means must be used (to simplify the model, for instance), perform a weighted analysis on the means with the sample sizes as weights.

In subsequent sections, only the SAS programs for method (1) are provided.

**6.1.3 Factorial completely randomized design** A factorial experiment consists of more than one treatment factor, and all the treatment factors

are crossed with one another. In addition, all the factors must have the same experimental unit. An experiment is a “complete factorial” if all combinations of levels are represented. A major advantage of the factorial design is that it allows the interactions of factors and individual factors to be examined. Tests of main factors are tests of one factor averaged over the others. If there is no interaction between two factors (i.e., the effect of one factor is consistent across all levels of the other factor), then the tests on the main factors are logical. If interactions among factors exist, we must interpret the results from the tests of main effects carefully. Factorial designs also provide “hidden replications” because each main effect is examined over a range of conditions (i.e., other factors).

Factorial experiments may incorporate completely randomized or randomized block designs (see Section 6.2.2). In a completely randomized factorial design, each experimental unit is randomly assigned one of the treatment combinations.

**Example:** Suppose that the six treatments in the previous examples are combinations of two different fertilizers each applied at a low, moderate, and high rate. How does this change the structure of the design, the ANOVA table, and the SAS program?

- Factors: Fertilizer, F with  $f = 2$  levels,  
Amount, A with  $a = 3$  levels.
- F and A are fixed.
- F and A are crossed with each other.
- Experimental unit for F and A is a row of four trees, R.
- There are  $r = 3$  experimental units for each F\*A combination; that is, each F\*A combination is replicated three times.
- R is random, nested in F and A.
- A tree, E, is an element.
- There are  $e = 4$  trees per row of trees.
- E is random, nested in F, A, and R.

The degrees of freedom for F, A, and F\*A sum to 5. This is the same as the one-way ANOVA case when F had six levels. The sums of squares for the F source in Section 6.1.1 has now been **partitioned** into three sources: F, A, and F\*A.

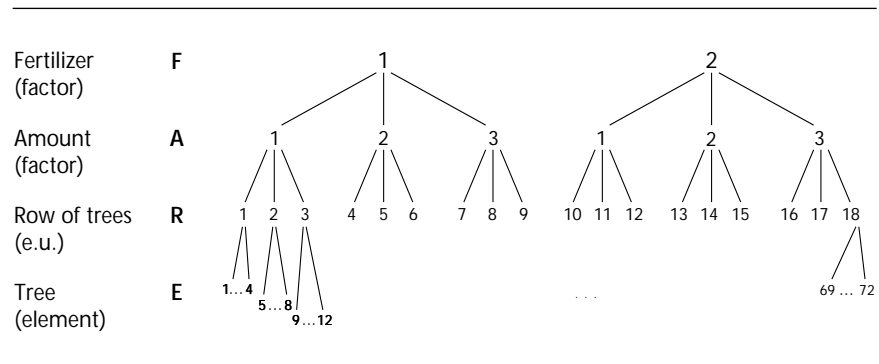


FIGURE 14 Design structure for a factorial completely randomized design.

TABLE 7 ANOVA table for a factorial completely randomized design

Source of variation	df	EMS	Error
Fertilizer, F	$f - 1 = 1$	$\sigma_{E(FAR)}^2 + 4\sigma_{R(FA)}^2 + 36\phi_F$	R (FA)
Amount, A	$a - 1 = 2$	$\sigma_{E(FAR)}^2 + 4\sigma_{R(FA)}^2 + 24\phi_A$	R (FA)
F*A	$(f - 1)(a - 1) = 2$	$\sigma_{E(FAR)}^2 + 4\sigma_{R(FA)}^2 + 12\phi_{F*A}$	R (FA)
Row of trees, R(FA)	$(r - 1)fa = 12$	$\sigma_{E(FAR)}^2 + 4\sigma_{R(FA)}^2$	E (FAR)
Tree, E (FAR)	$(e - 1)far = 54$	$\sigma_{E(FAR)}^2$	—
Total	$fare - 1 = 71$		

**SAS Program: Analysis of individual data**

```

PROC GLM;
  CLASS F A R;
  MODEL Y = F|A R(F A);

  /*** A Levels:                1L 2M 3H *** /

  CONTRAST 'A : LINEAR'      A 1 0 -1 / E=R(F A);
  TEST H = F|A E = R(F A);
  MEANS F|A / SCHEFFE CLDIFF E = R(F A);
RUN;

```

This SAS program requests a linear contrast to be performed on A. Means for F, A, and F\*A will be calculated. Scheffé’s test will be performed on the main effects F and A only; the test results will be presented as confidence intervals for all possible pairwise differences between means. Tests for F, A, and F\*A are requested with the proper error term.

6.2 Randomized Block Designs

The completely randomized design requires that all experimental units be homogeneous. In practice, the experimenter often does not have enough homogeneous experimental units available to construct a completely randomized design with adequate replication. Instead, smaller collections of homogeneous experimental units, called **blocks**, are identified. The complete set of treatment combinations appears in each block as in the completely randomized design. This set-up is called a **randomized block design**, and is really a group of completely randomized designs. Usually, each block has only one experimental unit per treatment combination, though replication within blocks is possible.

Many things can act as a block. For example, locations, ecosystems, batches of seedlings, or units of time may act as blocks in an experiment. In the last case, there may be only enough resources to run one set of treatment combinations each year. One could use time as the block design criteria and repeat the experiment over many years; the number of years that the experiment is carried out would be equivalent to the number of blocks.

One advantage of block design is that it removes variations attributed to the block from the experimental error. In many cases, this makes the test on the main effect more powerful than if a non-block design were used. This design is most effective if the blocks are very different from

one another, but the experimental units within a block are fairly uniform. On the other hand, if the experimental units within blocks are more heterogeneous than that among blocks, then a completely randomized design without blocks would be more appropriate (Lentner et al. 1989). Block design also helps to expand the inference space when treatment combinations will be examined over a range of conditions (via block).

The block factor is usually considered random. Its levels are often selected randomly from all possible blocks and the study's conclusions can be generalized beyond the blocks used in the study.

The block design criteria, such as location, moisture level, slope, or aspect, should be independent of the treatment of interest. That is, the effects of the treatment on the measured response should be consistent across the blocks. Otherwise, testing of treatment effects is not possible.

Finally, within each block treatments should be applied randomly to the experimental units. A separate randomization scheme should be used in each block.

**6.2.1 One-way randomized block design** Only one factor is involved and the levels of the factor are repeated from block to block.

**Example:** Suppose that the fertilizer trial will be conducted in three different orchards. In each orchard, six rows of four trees each will be set aside for the trial. Since the orchard locations could be widely separated, there is great potential for tree response at each orchard to vary because of weather, soils, or management practices. Therefore, a complete set of treatments should occur at each orchard. Each treatment would be randomly assigned to one row in each orchard. Orchard is now functioning as block.

- Factor: Fertilizer, F with  $f = 6$  levels (with an unfertilized control).
- F is fixed.
- Block, B is random.
- There are  $b = 3$  blocks (i.e., orchards).
- B and F are crossed.
- Experimental unit for F is a row of four trees, R.
- There is  $r = 1$  experimental unit per treatment level per block; that is, no replication within blocks.
- R is random, nested in F and B.
- A tree, E, is an element.
- There are  $e = 4$  trees per row of trees.
- E is random, nested in F and B.

According to the expected mean squares, B and B\*F are tested by rows of trees, R(BF). However, since there is only one row of trees per treatment level in each block, R(BF) has zero degrees of freedom, making the tests for B, B\*F and R(BF) impossible. This inability to test for block and block interactions can also be reasoned from another angle. The following discussion is adapted from a similar one in Sokal and Rohlf (1981: 350–352).



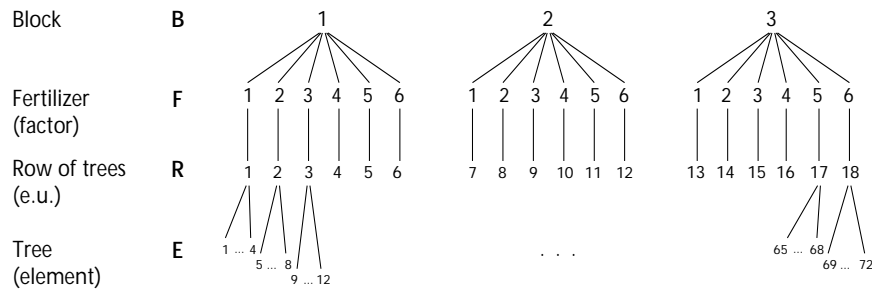


FIGURE 15 Design structure for a one-way randomized block design.

TABLE 8 ANOVA table for a one-way randomized block design

Source of variation	df	EMS	Error
Block, B	$b - 1 = 2$	$\sigma_{E(BFR)}^2 + 4\sigma_{R(BF)}^2 + 24\sigma_B^2$	—
Fertilizer, F	$f - 1 = 5$	$\sigma_{E(BFR)}^2 + 4\sigma_{R(BF)}^2 + 4\sigma_{B^*F}^2 + 12\phi_F$	B*F
B*F	$(b - 1)(f - 1) = 10$	$\sigma_{E(BFR)}^2 + 4\sigma_{R(BF)}^2 + 4\sigma_{B^*F}^2$	—
Row of trees, R (BF)	$(r - 1)bf = 0$	$\sigma_{E(BFR)}^2 + 4\sigma_{R(BF)}^2$	—
Tree, E (BFR)	$(e - 1)bfr = 54$	$\sigma_{E(BFR)}^2$	—
Total	$bfr - 1 = 71$		

In our orchard example, each orchard is unique in many aspects, such as location, environment, or management. These characteristics can not be duplicated even for the same orchard selected at a different time. Thus, there is a random error deviation, called “restriction error,”  $\sigma_r^2$  (Anderson and McLean 1974b), attached to each of the orchards because of its unique set-up. Therefore, the block EMS should be  $\sigma_{E(BFR)}^2 + 4\sigma_{R(BF)}^2 + 24\sigma_B^2 + \sigma_r^2$ . Unless  $\sigma_r^2$  is negligible, the block cannot be tested even if there were more than one row of trees per level of F in each block. The restriction error is usually not identified in the EMS. We must be aware of its existence and the impact it has on ANOVA tests.

Testing the block is usually not of interest, as it is expected to be different. However, researchers often attempt to test the block to see if a block design was necessary in their studies. The effectiveness of the block design can be assessed by an estimate of the relative efficiency of the randomized block design compared to the completely randomized design. The theoretical development of relative efficiency is beyond the scope of this handbook but can be found in Kempthorne (1952) and Lentner and Bishop (1986). Lentner et al. (1989) showed that the estimated relative efficiency of the randomized block design compared to the completely randomized design is proportional to the ratio:

$$H = \frac{MS(\text{BLOCK})}{MS(\text{BLOCK} * \text{TREATMENT})}$$

In particular,  $H > 1$  implies a randomized block design is more effective than a completely randomized design;  
 $H = 1$  implies no gain attributed to the block design; and  
 $H < 1$  implies a completely randomized design is more effective.

Both numerator and denominator are mean squares from the ANOVA table.

The source R(BF) has zero degrees of freedom because the fertilizer treatments are not replicated within blocks. This source of variation is usually excluded in the ANOVA table. We include it to show the full design of the experiment. It is not needed in the MODEL statement in the following SAS program.

### SAS Program: Analysis of individual data

```
PROC GLM;
  CLASS B F;
  MODEL Y = B|F ;
  /*** Fertilizer Levels:      0  5 10 15 20 25  ***/
  CONTRAST 'LINEAR IN F'      F -5 -3 -1  1  3  5 / E=B*F ;
  CONTRAST 'FIRST VS LAST 5' F -5  1  1  1  1  1 / E=B*F ;
  TEST H = F      E = B*F ;
  MEANS F / TUKEY E = B*F ;
RUN;
```

**6.2.2 Factorial randomized block design** This design is similar to that in the last section except that each block has a factorial design similar to that in Section 6.1.3.

**Example:** If the six treatments are again split up into two factors of two different fertilizers and three different amounts, then a two-way factorial randomized block design is obtained.

- Factors: Fertilizer, F with  $f = 2$  levels,  
Amount, A with  $a = 3$  levels.
- F and A are fixed.
- Block, B is random with  $b = 3$  levels.
- B, F, and A are crossed with one another.
- Experimental unit for F and A is a row of four trees, R.
- There is  $r = 1$  experimental unit per F\*A combination per block; that is, no replication within block.
- R is random, nested in B, F, and A.
- A tree, E, is an element.
- There are  $e = 4$  trees per row of trees.
- E is random, nested in B, F, A, and R.

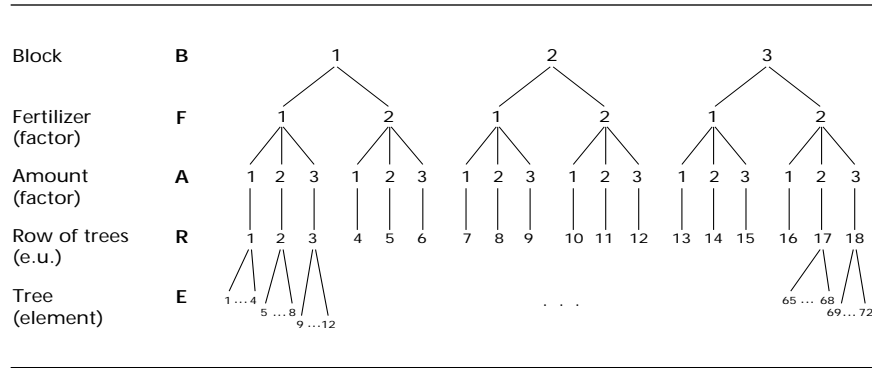


FIGURE 16 Design structure for a two-way factorial randomized block design.

TABLE 9 ANOVA table for a two-way factorial randomized block design

Source of variation	$df$	EMS	Error
Block, B	$b - 1 = 2$	$\sigma_{E(BFAR)}^2 + 4\sigma_{R(BFA)}^2 + 24\sigma_B^2$	—
Fertilizer, F	$f - 1 = 1$	$\sigma_{E(BFAR)}^2 + 4\sigma_{R(BFA)}^2 + 12\sigma_{B^*F}^2 + 36\phi_F$	B*F
Amount, A	$a - 1 = 2$	$\sigma_{E(BFAR)}^2 + 4\sigma_{R(BFA)}^2 + 8\sigma_{B^*A}^2 + 24\phi_A$	B*A
A*F	$(a - 1)(f - 1) = 2$	$\sigma_{E(BFAR)}^2 + 4\sigma_{R(BFA)}^2 + 4\sigma_{B^*F^*A}^2 + 12\phi_{A^*F}$	B*F*A
B*F	$(b - 1)(f - 1) = 2$	$\sigma_{E(BFAR)}^2 + 4\sigma_{R(BFA)}^2 + 12\sigma_{B^*F}^2$	—
B*A	$(b - 1)(a - 1) = 4$	$\sigma_{E(BFAR)}^2 + 4\sigma_{R(BFA)}^2 + 8\sigma_{B^*A}^2$	—
B*F*A	$(b - 1)(f - 1)(a - 1) = 4$	$\sigma_{E(BFAR)}^2 + 4\sigma_{R(BFA)}^2 + 4\sigma_{B^*F^*A}^2$	—
Row of trees, R (BFA)	$(r - 1)bfa = 0$	$\sigma_{E(BFAR)}^2 + 4\sigma_{R(BFA)}^2$	—
Tree, E (BFAR)	$(e - 1)bfar = 54$	$\sigma_{E(BFAR)}^2$	—
Total	$bfare - 1 = 71$		

The denominator degrees of freedom for testing F, A, and A\*F are small (2, 4, and 4 respectively), and this results in tests that are not very powerful (i.e., insensitive to significant difference). If there is reason to believe that the B\*F, B\*A, and B\*F\*A interaction effects are negligible (based on strong prior knowledge or conservative  $F$ -tests), then the expected mean squares of B\*F, B\*A, B\*F\*A, and R(BFA) simply estimate  $\sigma_{E(BFAR)}^2 + 4\sigma_{R(BFA)}^2$ . Therefore, we could pool these mean squares together according to:

$$MS_{R(BFA)}^* = \frac{SS_{R(BFA)} + SS_{B^*F^*A} + SS_{B^*F} + SS_{B^*A}}{\text{pooled degrees of freedom}}$$

The pooled degrees of freedom for  $MS_{R(BFA)}^*$  is the sum of the degrees of freedom of R(BFA), B\*F\*A, B\*F, and B\*A. The pooled mean squares for R(BFA) can be used to test for F, A, and F\*A. Since the new error term has larger degrees of freedom ( $df = 10$ ), the pooled tests are more powerful. The pooled ANOVA table is given in Table 10.

The danger of pooling is that an error can be made when deciding that B\*F, B\*A, or B\*F\*A interaction effects are negligible when in fact they

are not. Thus, the pooled mean square could be actually larger than expected, making significant effects more difficult to detect. Montgomery (1991, Section 8.3) suggested that terms to be pooled only if they have fewer than six degrees of freedom and if the  $F$ -test for each term is not significant at a large  $\alpha$  value, say  $\alpha = 0.25$ .

TABLE 10 ANOVA table with pooled mean squares for a two-way factorial randomized block design

Source of variation	$df$	EMS	Error
Block, B	2	$\sigma_{E(BFAR)}^2 + 4\sigma_{R(BFA)}^2 + 24\sigma_B^2$	—
Fertilizer, F	1	$\sigma_{E(BFAR)}^2 + 4\sigma_{R(BFA)}^2 + 36\phi_F$	R(BFA)*
Amount, A	2	$\sigma_{E(BFAR)}^2 + 4\sigma_{R(BFA)}^2 + 24\phi_A$	R(BFA)*
A*F	2	$\sigma_{E(BFAR)}^2 + 4\sigma_{R(BFA)}^2 + 12\phi_{A*F}$	R(BFA)*
Row of trees, R(BFA)*	10	$\sigma_{E(BFAR)}^2 + 4\sigma_{R(BFA)}^2$	—
Tree, E(BFAR)	54	$\sigma_{E(BFAR)}^2$	—
Total	71		

### SAS Program: Analysis based on pooled ANOVA

```

PROC GLM;
  CLASS B F A;
  MODEL Y = B F | A R(BFA);
  / *** A Levels:                1L 2M 3H *** /
  CONTRAST 'A : LINEAR'      A 1 0 -1 / E=R(BFA) ;
  TEST H = F | A E=R(BFA) ;
  MEANS F | A / LSD E=R(BFA) ;
RUN;

```

**Note:** Any source of variation not listed in the MODEL statement is pooled into the next level. Therefore, B\*F, B\*A, and B\*F\*A are pooled into R(BFA), which is used to test F, A, and F\*A.

### 6.3 Split-plot Designs

Split-plot designs are common in forestry. Their distinguishing feature is that levels of one factor are randomly assigned to experimental units called **main-plots**; each main-plot is further divided into smaller units called **split-plots** to which levels of another factor(s) are randomly assigned. Such designs may incorporate completely randomized and randomized block designs (see Sections 6.3.1 and 6.3.2).

Split-plot designs are often misinterpreted as factorials. Both designs result in factors crossed with one another; however, split-plot designs restrict the random assignment of treatments. To compare the two designs, let's consider a study in which two factors are of interest: factor A with four levels and factor B with two levels; both factors are fixed. In a completely randomized factorial design, a total of sixteen experimental units are needed if we want to replicate the treatment combinations twice. A possible arrangement is shown in Figure 17. Each square is an experimental unit for factors A and B. The levels of each factor are assigned randomly to the experimental units without restriction.

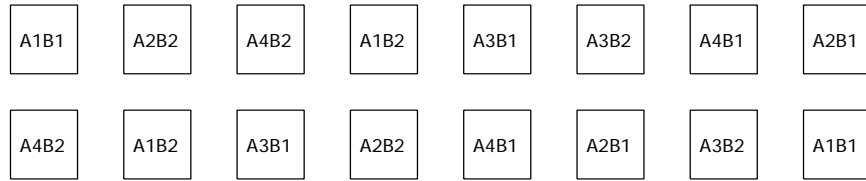


FIGURE 17 A completely randomized factorial arrangement.

In a completely randomized split-plot design, the same amount of experimental material would be used differently. For example, we could pair the 16 squares and randomly assign a level of factor A to each pair, as shown in Figure 18. A pair of squares, called a main-plot, is the experimental unit for A.

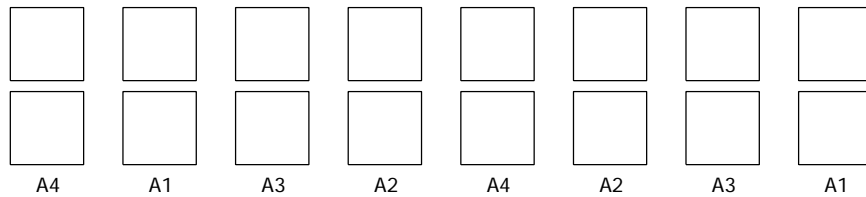


FIGURE 18 Main-plots of a completely randomized split-plot design. Each pair of squares is an experimental unit for factor A.

Then we could divide each main-plot into two split-plots (in this case, two squares) and randomly assign the two levels of B to the split-plots in each main-plot. A split-plot is the experimental unit for factor B. Notice that the random assignment of the levels of B is restricted: both levels must appear in each main-plot. A possible layout is displayed in Figure 19 in which pairs of squares are main-plots whereas squares within a pair are split-plots.

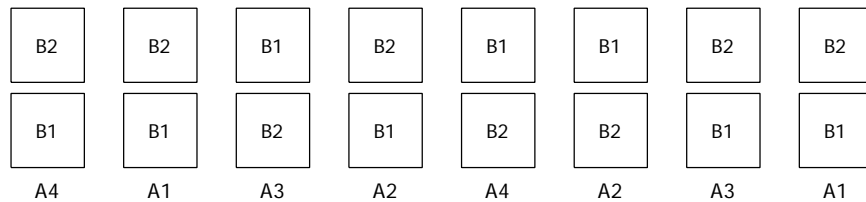


FIGURE 19 A completely randomized split-plot layout.

Split-plot designs may be used when some treatments associated with the levels of one or more factors require larger amounts of experimental

material in an experimental unit than do treatments for other factors (Steel and Torrie 1980). For example, in a fertilizer and root-pruning study, we could use the split-plot design and apply fertilizer treatments to plots (main-plot) and root-pruning treatments to rows of trees within a plot (split-plot). As each main-plot is fairly uniform and all levels of the split-plot factor appear within it, variation among split-plots is expected to be less than among main-plots. This design yields more precise information on the split-plot factor while at the expense of losing information on the main-plot factor. In summary, the following factors are assigned to a split-plot:

- those that require smaller amounts of experimental material,
- those that are of major importance,
- those that are expected to exhibit smaller differences, or
- those that require greater precision for any reasons (Steel and Torrie 1980, Chapter 16).

Consult Snedecor and Cochran (1967, Section 12.12) or Steel and Torrie (1980, Chapter 16) for more discussion of the split-plot design.

### 6.3.1 Completely randomized split-plot design

**Example:** Suppose that in the example of Section 6.1.2 we want to test the effectiveness of extra boron. We will split each row into two pairs of trees: one pair will receive extra boron, while the other pair will not.

- Factors: Fertilizer, F with  $f = 6$  levels (main-plot factor), Boron, N with  $n = 2$  levels (split-plot factor).
- F and N are fixed factors, crossed with each other.
- Experimental unit for F is a row of four trees, R.
- There are  $r = 3$  experimental units per level of F; that is, each level of F is replicated 3 times.
- R is random, nested in F.
- Experimental unit for N is a pair of trees, P.
- There is  $p = 1$  experimental unit per level of N.
- P is random, nested in F, R, and N.
- A tree, E, is an element.
- There are  $e = 2$  trees per pair of trees.
- E is random, nested in F, R, N, and P.

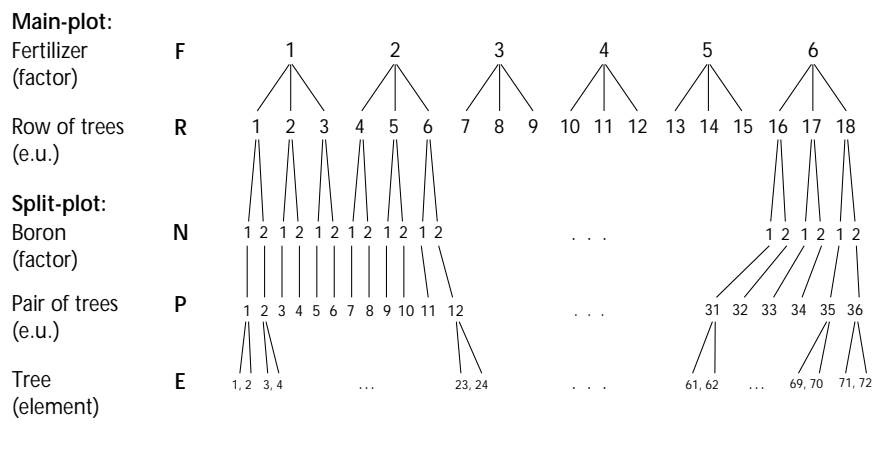


FIGURE 20 Design structure for a completely randomized split-plot design.

TABLE 11 ANOVA table for a completely randomized split-plot design

Source of variation	df	EMS	Error
<b>Main-plot</b>			
Fertilizer, F	$f - 1 = 5$	$\sigma_{E(FRNP)}^2 + 2\sigma_{P(FRN)}^2 + 4\sigma_{R(F)}^2 + 12\phi_F$	R(F)
Row of trees, R(F)	$(r - 1)f = 12$	$\sigma_{E(FRNP)}^2 + 2\sigma_{P(FRN)}^2 + 4\sigma_{R(F)}^2$	—
<b>Split-plot</b>			
Boron, N	$n - 1 = 1$	$\sigma_{E(FRNP)}^2 + 2\sigma_{P(FRN)}^2 + 2\sigma_{N^*R(F)}^2 + 36\phi_N$	N*R(F)
F*N	$(f - 1)(n - 1) = 5$	$\sigma_{E(FRNP)}^2 + 2\sigma_{P(FRN)}^2 + 2\sigma_{N^*R(F)}^2 + 6\phi_{F*N}$	N*R(F)
N*R(F)	$(n - 1)(r - 1)f = 12$	$\sigma_{E(FRNP)}^2 + 2\sigma_{P(FRN)}^2 + 2\sigma_{N^*R(F)}^2$	—
Pair of trees,			
P(FRN)	$(p - 1)frn = 0$	$\sigma_{E(FRNP)}^2 + 2\sigma_{P(FRN)}^2$	—
Tree, E(FRNP)	$(e - 1)frnp = 36$	$\sigma_{E(FRNP)}^2$	—
Total	$frnpe - 1 = 71$		

Since the source P(FRN) has zero degrees of freedom, the tests for R(F), NR(F), and P(FRN) are not possible. The source P(FRN) is excluded in the MODEL statement in the following SAS program.

### SAS Program: Analysis of individual data

```

PROC GLM;
  CLASS F N R;
  MODEL Y = F|N R(F) N*R(F);
  TEST H = F          E = R(F);
  TEST H = N  F*N     E = N*R(F);

/*** Fertilizer Levels:      0  5 10 15 20 25      ***/

```

```

CONTRAST `LINEAR IN F'      F  -5  -3  -1   1   3   5 / E=R(F);
CONTRAST `FIRST VS LAST 5' F  -5   1   1   1   1   1 / E=R(F);
MEANS      F / DUNCAN  E = R(F);
MEANS      N  F*N / LSD  CLDIFF  E = N*R(F);
RUN;

```

Note that means are computed for F, N, and F\*N. Duncan's Multiple Range test and the Least Significance Difference test are performed on the main effects F and N, respectively.

### 6.3.2 Randomized block split-plot design

**Example:** Suppose that the one-way randomized block factorial design in Section 6.2.2 is modified with boron added to pairs of trees within rows. This creates another slightly different split-plot design. The discussion of restriction error in Section 6.2 applies here.

- Factors: Fertilizer, F with  $f = 6$  levels (main-plot factor),  
Boron, N with  $n = 2$  levels (split-plot factor).
- F and N are fixed.
- Block, B is random.
- There are  $b = 3$  blocks (i.e., orchards).
- B, F, and N are crossed with one another.
- Experimental unit for F is a row of four trees, R.
- There is  $r = 1$  experimental unit per level of F.
- R is random, nested in B and F.
- Experimental unit for N is a pair of trees, P.
- There is  $p = 1$  experimental unit per level of N.
- P is random, nested in B, F, R, and N.
- A tree, E, is an element.
- There are  $e = 2$  trees per pair of trees.
- E is random, nested in B, F, R, N, and P.

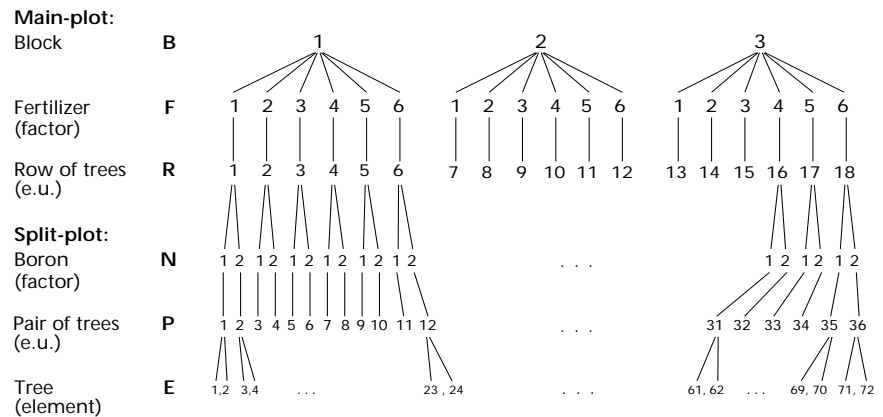


FIGURE 21 Design structure for a randomized block split-plot design.



TABLE 12 ANOVA table for a randomized block split-plot design

Source of variation	df	EMS	Error
<b>Main-plot</b>			
Block, B	$b - 1 = 2$	$\sigma^2_{E(BFRNP)} + 2\sigma^2_{P(BFRN)} + 4\sigma^2_{R(BF)} + 24\sigma^2_B$	—
Fertilizer, F	$f - 1 = 5$	$\sigma^2_{E(BFRNP)} + 2\sigma^2_{P(BFRN)} + 4\sigma^2_{R(BF)} + 4\sigma^2_{B^*F} + 12\phi_F$	B * F
B * F	$(b - 1)(f - 1) = 10$	$\sigma^2_{E(BFRNP)} + 2\sigma^2_{P(BFRN)} + 4\sigma^2_{R(BF)} + 4\sigma^2_{B^*F}$	—
Row of trees, R(BF)	$(r - 1)bf = 0$	$\sigma^2_{E(BFRNP)} + 2\sigma^2_{P(BFRN)} + 4\sigma^2_{R(BF)}$	—
<b>Split-plot</b>			
Boron, N	$n - 1 = 1$	$\sigma^2_{E(BFRNP)} + 2\sigma^2_{P(BFRN)} + 2\sigma^2_{N^*R(BF)} + 12\sigma^2_{B^*N} + 36\phi_N$	B * N
F * N	$(f - 1)(n - 1) = 5$	$\sigma^2_{E(BFRNP)} + 2\sigma^2_{P(BFRN)} + 2\sigma^2_{N^*R(BF)} + 2\sigma^2_{B^*F^*N} + 6\phi_{F^*N}$	B * F * N
B * N	$(b - 1)(n - 1) = 2$	$\sigma^2_{E(BFRNP)} + 2\sigma^2_{P(BFRN)} + 2\sigma^2_{N^*R(BF)} + 12\sigma^2_{B^*N}$	—
B * F * N	$(b - 1)(f - 1)(n - 1) = 10$	$\sigma^2_{E(BFRNP)} + 2\sigma^2_{P(BFRN)} + 2\sigma^2_{N^*R(BF)} + 2\sigma^2_{B^*F^*N}$	—
N * R (BF)	$(n - 1)(r - 1)bf = 0$	$\sigma^2_{E(BFRNP)} + 2\sigma^2_{P(BFRN)} + 2\sigma^2_{N^*R(BF)}$	—
Pair of trees, P(BFRN)	$(p - 1)frn = 0$	$\sigma^2_{E(BFRNP)} + 2\sigma^2_{P(BFRN)}$	—
Tree, E(BFRNP)	$(e - 1)frnp = 36$	$\sigma^2_{E(BFRNP)}$	—
Total	$frnpe - 1 = 71$		

Sources with zero degrees of freedom are usually not listed in the ANOVA table. Most textbooks and published manuscripts would use the following simplified table. (Table 13).

TABLE 13 Simplified ANOVA table for a randomized block split-plot design

Source of variation	df	EMS	Error
<b>Main-plot</b>			
Block, B	$b - 1 = 2$	$\sigma^2_{E(BFRNP)} + 24\sigma^2_B$	—
Fertilizer, F	$f - 1 = 5$	$\sigma^2_{E(BFRNP)} + 4\sigma^2_{B^*F} + 12\phi_F$	B * F
B * F	$(b - 1)(f - 1) = 10$	$\sigma^2_{E(BFRNP)} + 4\sigma^2_{B^*F}$	—
<b>Split-plot</b>			
Boron, N	$n - 1 = 1$	$\sigma^2_{E(BFRNP)} + 12\sigma^2_{B^*N} + 36\phi_N$	B * N
F * N	$(f - 1)(n - 1) = 5$	$\sigma^2_{E(BFRNP)} + 2\sigma^2_{B^*F^*N} + 6\phi_{F^*N}$	B * F * N
B * N	$(b - 1)(n - 1) = 2$	$\sigma^2_{E(BFRNP)} + 12\sigma^2_{B^*N}$	—
B * F * N	$(b - 1)(f - 1)(n - 1) = 10$	$\sigma^2_{E(BFRNP)} + 2\sigma^2_{B^*F^*N}$	—
Tree, E(BFRNP)	$(e - 1)frnp = 36$	$\sigma^2_{E(BFRNP)}$	—
Total	$frnpe - 1 = 71$		

**SAS Program: Analysis of individual data**

```

PROC GLM;
  CLASS B F N;
  MODEL Y = B|F|N;
  TEST H = F      E = B*F;
  TEST H = F*N    E = B*F*N;
  TEST H = N      E = B*N;

  /*** Fertilizer Levels:      0  5 10 15 20 25      *** /

```

```

CONTRAST `LINEAR IN F'      F  -5  -3  -1   1   3   5  / E = B*F;
CONTRAST `FIRST VS LAST 5' F  -5   1   1   1   1   1  / E = B*F;
MEANS  F / SCHEFFE          E = B*F;
MEANS  F*N / SCHEFFE        E = B*F*N;
MEANS  N / SCHEFFE          E = B*N;
RUN;

```

The sources  $B*N$  and  $B*F*N$  are often pooled into the source  $E(BFRNP)$ , and the pooled term is then used to test for  $N$  and  $F*N$ , as in the example in Section 6.2.2. Be aware that pooling of sources is efficient only if the variances  $\sigma^2_{B*N}$  and  $\sigma^2_{B*F*N}$  are negligible. As a general rule, pool sources only if they have fewer than six degrees of freedom and the  $F$ -test for each is not significant at a large  $\alpha$  value, say  $\alpha = 0.25$  (Montgomery 1991, Section 8.3).

---

## 7 SUMMARY

A successful experiment begins with a sound experimental design. A well-designed experiment should have adequate replication and be properly randomized. Equally important, the design should have the capacity to explore the study objectives. Where possible, a simple design should be employed.

When choosing the method of analysis, we must keep in mind the questions we would like answered. The method of analysis should be compatible with the study objectives and the response variables. Analysis of variance is a popular procedure, but it is only appropriate for continuous data and to compare several population means. Routine use of ANOVA without discretion could lead to misleading results.

ANOVA  $F$ -tests and contrasts are powerful tools that can test preconceived hypotheses, whereas multiple comparisons can be used most effectively to generate hypotheses. All results obtained from multiple comparisons should be verified in well-designed studies. Moreover, statistical procedures such as ANOVA can only establish statistical significance. We must also look at the data and evaluate significant difference from the biological point of view.

Finally, the widespread use of computers has made statistical analysis quick and easy. But the automation has also brought misuse and misinterpretation of statistical techniques. We must remember that the results from statistical packages are only as reliable as the input data and our understanding of the software and statistics. The key to successful data analysis lies in good knowledge of the subject matter.

The expected mean squares, EMS, are required to determine the proper error terms for the sources of variation in an ANOVA model. This appendix describes the steps to find both the EMS and the error terms. We will use the factorial completely randomized design example in Section 6.1.3 to demonstrate each step in the process. Recall that in the example, we have:

- Factors: Fertilizer, F with  $f = 2$  levels,  
Amount, A with  $a = 3$  levels.
- F and A are fixed.
- F and A are crossed with each other.
- Experimental unit for F and A is a row of four trees, R.
- There are  $r = 3$  experimental units for each F\*A combination; that is, each F\*A combination is replicated three times.
- R is random, nested in F and A.
- A tree, E, is an element.
- There are  $e = 4$  trees per row of trees.
- E is random, nested in F, A, and R.

In this appendix, the term **variates** refers to the factors, experimental units, and elements in an experiment. The term **sources** refers to all the sources of variation in the data; see Section 5.3 for a description on how to compile the list of sources in an experiment.

**Steps:**

1. Create a table with the variates listed across the top and the sources listed down the left side. Above each variate designate the number of levels it has, as well as whether it is fixed ( $f$ ) or random ( $r$ ).

Sources	Variates			
	$f$ 2 F	$f$ 3 A	$r$ 3 R	$r$ 4 E
F				
A				
F*A				
R (FA)				
E (FAR)				

2. Make another column titled "Variance component" for the sources. Use  $\phi$  if a source is fixed,  $\sigma^2$  otherwise. A source is considered fixed only if all variates comprising the source are fixed.

Sources	Variates				Variance component
	<i>f</i> 2 F	<i>f</i> 3 A	<i>r</i> 3 R	<i>r</i> 4 E	
F					$\phi_F$
A					$\phi_A$
F * A					$\phi_{F * A}$
R (FA)					$\sigma^2_{R (FA)}$
E (FAR)					$\sigma^2_{E (FAR)}$

The entries in the centre of the table will be filled in the next three steps.

3. Begin with the variate in the first column on the left and the top entry in that column. If the source of variation that corresponds to that entry contains the column variate, then leave the entry blank; otherwise, enter the number of levels of the column variate. Repeat for the next entry below until the end of the column is reached, then continue with the next column variate to the right. For example, the variate F appears in all the sources except source A; hence the number 2 (F has 2 levels) is placed in the second entry in the first column. The factor R is not contained in the source F, A, or F \* A; therefore, the number 3 (R has 3 levels) is placed in the first three entries of the R column, corresponding to the sources F, A, and F \* A.

Sources	Variates				Variance component
	<i>f</i> 2 F	<i>f</i> 3 A	<i>r</i> 3 R	<i>r</i> 4 E	
F					$\phi_F$
A	2				$\phi_A$
F * A		3			$\phi_{F * A}$
R (FA)			3		$\sigma^2_{R (FA)}$
E (FAR)				4	$\sigma^2_{E (FAR)}$

4. Identify all the nested sources — that is, all sources with parentheses. For each of these sources, fill the entries across the row as follows: put down “1” if the variate that corresponds to the entry appears inside the parentheses; otherwise, leave the entry blank. For example, R(FA) is a nested source. The variates F and A are inside the parentheses so the number 1 is placed in the first two entries across the R(FA) row, corresponding to the variates F and A.

Sources	Variates				Variance component
	<i>f</i> 2 F	<i>f</i> 3 A	<i>r</i> 3 R	<i>r</i> 4 E	
F		3	3	4	$\phi_F$
A	2		3	4	$\phi_A$
F * A			3	4	$\phi_{F * A}$
R (FA)	1	1		4	$\sigma^2_{R(FA)}$
E (FAR)	1	1	1		$\sigma^2_{E(FAR)}$

5. Work columnwise again. If the variate is fixed, put “0” in each blank entry down the column; put “1” if the variate is random.

Sources	Variates				Variance component
	<i>f</i> 2 F	<i>f</i> 3 A	<i>r</i> 3 R	<i>r</i> 4 E	
F	0	3	3	4	$\phi_F$
A	2	0	3	4	$\phi_A$
F * A	0	0	3	4	$\phi_{F * A}$
R (FA)	1	1	1	4	$\sigma^2_{R(FA)}$
E (FAR)	1	1	1	1	$\sigma^2_{E(FAR)}$

All the entries in the table should be filled at the end of step 5.

6. The expected mean squares of a source are presented as an equation that shows the variance components of that source. The weight (or coefficient) corresponding to each variance component is the product of the entries in the last table. The structure of the source dictates which columns are to be used in the computation of the coefficients. Here are the steps for finding the EMS of a source:
- For a nested source, do not use columns for variates that are outside the parentheses of the source; otherwise, do not use all the columns that correspond to the variates that make up the source.
  - Multiply the remaining entries across each row, including the variance component column.
  - Add the row products.
  - Drop a variance component term if it does not contain all the variates in the source of interest.

For example, to find the EMS of the source F \* A, we would ignore the F and A columns, multiply the entries across each row, and add the variance components. At the end of step (iii), we should get for the source F \* A:

$$\text{EMS} = (3 \times 4) \phi_F + (3 \times 4) \phi_A + (3 \times 4) \phi_{F^*A} + (1 \times 4) \sigma^2_{R(FA)} + (1 \times 1) \sigma^2_{E(FAR)}$$

$$\text{EMS} = 12\phi_F + 12\phi_A + 12\phi_{F^*A} + 4\sigma^2_{R(FA)} + \sigma^2_{E(FAR)}$$

Finally, according to step (iv), we would drop the variance components  $12\phi_A$  and  $12\phi_F$  because they do not contain the F and the A variates.

The EMS of the sources in this factorial completely randomized design example are:

$$\text{F: } \sigma^2_{E(FAR)} + 4\sigma^2_{R(FA)} + 36\phi_F$$

$$\text{A: } \sigma^2_{E(FAR)} + 4\sigma^2_{R(FA)} + 24\phi_A$$

$$\text{F * A: } \sigma^2_{E(FAR)} + 4\sigma^2_{R(FA)} + 12\phi_{F^*A}$$

$$\text{R(FA): } \sigma^2_{E(FAR)} + 4\sigma^2_{R(FA)}$$

$$\text{E(FAR): } \sigma^2_{E(FAR)}$$

It is customary to list the variance components in the EMS in the order shown because it makes the identification of the proper error terms in the next step easier.

7. The correct error term for a source is another source that has the same EMS, except for the variance component attributed to the source of interest. Follow these easy steps:
  - i. Look at the EMS of the source of interest (say F) and identify the variance component term attributed to this source ( $36\phi_F$ ).
  - ii. Write down the EMS with this variance component term removed:  $\sigma^2_{E(FAR)} + 4\sigma^2_{R(FA)}$ .
  - iii. Find the source which has the identical EMS to the one you have just written down in step (ii). In this case, R(FA) is the error term for F.

The following table shows the error terms used to test the sources in the factorial completely randomized design example:

Source	Error
F	R (FA)
A	R (FA)
F * A	R (FA)
R (FA)	E (FAR)
E (FAR)	—

Similar EMS rules can be found in Schultz (1955), Sokal and Rohlf (1981, Section 12.3), and Montgomery (1991, Chapter 8). This set of rules for fixed models (models that involve only fixed factors) and

random models (models that involve only random factors) is recognized by all statisticians. For mixed models (models that involve both fixed and random factors), there is some debate on how the fixed\*random interaction term should be handled. Two different approaches exist using different EMS equations and tests. A full discussion of this controversy is beyond the scope of this appendix. Interested readers can refer to Sit (1992) and Schwarz (1993).

For complicated designs, exact  $F$ -tests may not exist for some of the sources of variation. One possible solution is to pool the mean square of effects that can be considered negligible (see Section 6.2.2 for an example). Another approach is to perform pseudo  $F$ -tests using Satterthwaite's method (1946). More discussion on pseudo  $F$ -tests can be found in Satterthwaite (1946), Gaylor and Hopper (1969), Bergerud (1989a), and Milliken and Johnson (1992: 249–255).

APPENDIX 2 Hypothetical data for example in Section 5.1

Fertilizer Treatment	Root Treatment	Row of Trees	Tree Number	Height (cm)
1	1	1	1	7.3
1	1	1	2	10.3
1	1	1	3	14.8
1	1	1	4	2.9
1	1	2	5	14.9
1	1	2	6	1.8
1	1	2	7	5.1
1	1	2	8	17.7
1	2	3	9	14.1
1	2	3	10	13.0
1	2	3	11	19.3
1	2	3	12	20.7
1	2	4	13	24.9
1	2	4	14	16.0
1	2	4	15	14.9
1	2	4	16	24.9
2	1	5	17	24.3
2	1	5	18	28.6
2	1	5	19	31.8
2	1	5	20	32.7
2	1	6	21	29.1
2	1	6	22	24.7
2	1	6	23	30.1
2	1	6	24	31.9
2	2	7	25	42.2
2	2	7	26	45.6
2	2	7	27	40.7
2	2	7	28	34.5
2	2	8	29	38.4
2	2	8	30	46.4
2	2	8	31	38.0
2	2	8	32	41.7
3	1	9	33	50.7
3	1	9	34	53.2
3	1	9	35	50.1
3	1	9	36	41.4
3	1	10	37	53.0
3	1	10	38	37.9
3	1	10	39	50.9
3	1	10	40	51.0
3	2	11	41	69.5
3	2	11	42	53.8
3	2	11	43	53.8
3	2	11	44	62.1
3	2	12	45	54.3
3	2	12	46	63.0
3	2	12	47	47.0
3	2	12	48	65.4



## REFERENCES

---

- Anderson, V.L. and R.A. McLean. 1974a. Design of experiments: a realistic approach. Marcel Dekker, Inc., New York, N.Y.
- . 1974b. Restriction errors: another dimension in teaching experimental statistics. *Amer. Stat.* 28: 145–152.
- Ayre, M.P. and D.L. Thoma. 1990. Alternative formulation of the mixed-model ANOVA applied to quantitative genetics. *Evolution* 44(1): 221–226.
- Bennington, C.C. and W.V. Thayne. 1994. Use and misuse of mixed model analysis of variance in ecological studies. *Ecology* 75(3): 717–722.
- Bartlett, M.S. 1947. The use of transformations. *Biometrics* 3: 39–52.
- Bergerud, W. 1982. Experimental design: a practical approach. Unpubl. notes.
- . 1988a. Understanding replication and pseudo-replication. B.C. Min. For. Res. Br., Victoria, B.C., *Biom. Info. Pamp.* No. 5.
- . 1988b. Determining polynomial contrast coefficients. B.C. Min. For. Res. Br., Victoria, B.C., *Biom. Info. Pamp.* No. 12.
- . 1989a. ANOVA: pseudo  $F$ -tests. B.C. Min. For. Res. Br., Victoria, B.C., *Biom. Info. Pamp.* No. 19.
- . 1989b. What are degrees of freedom? B.C. Min. For. Res. Br., Victoria, B.C., *Biom. Info. Pamp.* No. 21.
- . 1989c. ANOVA: contrasts viewed as correlation coefficients. B.C. Min. For. Res. Br., Victoria, B.C., *Biom. Info. Pamp.* No. 23.
- . 1991. Calculating ANOVAs with SAS. Paper presented at the 1991 Pacific NW SAS User's Group Meeting in Victoria, B.C., Unpubl. notes.
- . 1993. Calculating contrast  $F$ -test when SAS will not. B.C. Min. For. Res. Br., Victoria, B.C., *Biom. Info. Pamp.* No. 45.
- Cochran, W.G. 1947. Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics* 3: 22–38.
- Cohen, J. 1977. *Statistical power analysis for the behavioral sciences.* Academic Press, New York, N.Y.
- Eisenhart, C. 1947. The assumptions underlying the analysis of variance. *Biometrics* 3: 1–21.
- Fernandez, G.C.J. 1992. Residual analysis and data transformations: important tools in statistical analysis. *Hortscience*, Vol. 27(4).
- Fisher, R.A. 1935. *The design of experiments.* Oliver and Boyd, Ltd., Edinburgh and London.
- Gaylor, D.W. and F.N. Hopper. 1969. Estimating the degrees of freedom for linear combinations of mean squares by Satterthwaite's formula. *Technometrics* 11: 691–705.

- Hahn, G.J. and W.Q. Meeker. 1993. Assumptions for statistical inference. *Amer. Stat.* 47(1): 1–11.
- Hartley, H. and S. Searle. 1969. A discontinuity in mixed model analyses, *Biometrics* 25: 573–576.
- Hays, W.L. 1963. *Statistics for psychologists*. Holt, Rinehart, and Winston, New York, N.Y.
- Hocking, R.R. 1973. A discussion of the two-way mixed model. *Amer. Stat.* 27(4): 148–152.
- Kempthorne, O. 1952. *Design and analysis of experiments*. Wiley, New York, N.Y.
- Keppel, G. 1973. *Design and analysis: a researcher's handbook*. Prentice-Hall, Englewood Cliffs, N.J.
- Kirk, R.E. 1968. *Experimental design: procedures for the behavioral sciences*. Brooks/Cole, Belmont, Calif.
- Lentner, M. and T. Bishop. 1986. *Experimental design and analysis*. Valley Book Company, Blacksburg, Va.
- Lentner, M., C.A. Jesse, and K. Hinkelmann. 1989. The efficiency of blocking: how to use MS(Blocks)/MS(Error) correctly. *Amer. Stat.* 43(2): 106–108.
- Li, C. 1964. *Introduction to experimental statistics*. McGraw-Hill, Inc., New York, N.Y.
- Lindman, H.R. 1974. *Analysis of variance in complex experimental designs*. W.H. Freeman and Co., San Francisco, Calif.
- Maxwell, S.E. and H.D. Delaney. 1990. *Designing experiments and analyzing data: a model comparison perspective*. Wadsworth Publishing Co., Belmont, Calif.
- Mead, R. 1988. *The design of experiments, statistical principles for practical applications*. Cambridge Univ. Press, Cambridge, Eng.
- . 1990. The non-orthogonal design of experiments. *J. R. Stat. Soc. A.* 153(2): 151–201.
- . 1991. Designing experiments for agroforestry research. *In* *Biophysical research for Asian agroforestry*. M.E. Avery, M.G. Cannel, and C.K. Ong (editors). Winrock International and South Asia Books USA, Bangkok, Thailand.
- McLean, R., W. Sanders, and W. Stroup 1991. A unified approach to mixed linear models. *Amer. Stat.* 45(1): 54–64.
- Milliken, G. and D.E. Johnson. 1989. *Analysis of messy data. Vol. 2: Nonreplicated experiments*. Wadsworth Inc., Belmont, Calif.
- . 1992. *Analysis of messy data. Vol 1: Designed experiment*. Wadsworth Inc., Belmont, Calif.
- Milton, J.S. and J.C. Arnold 1990. *Introduction to probability and statistics*. 2nd ed. McGraw-Hill, New York, N.Y.
- Mize, C.W. and R.C. Schultz. 1985. Comparing treatment means correctly and appropriately. *Can. J. For. Res.* 15: 1142–1148.

- Montgomery, D.C. 1991. Design and analysis of experiments. 3rd ed. John Wiley and Sons, Toronto, Ont.
- Nemec, A. 1991. Power analysis handbook for the design and analysis of forestry trials. B.C. Min. For. Res. Br., Victoria, B.C., Biom. Info. Handb. No. 2.
- . 1992. Guidelines for the statistical analysis of forest vegetation management data. B.C. Min. For. Res. Br., Victoria, B.C., Biom. Info. Handb. No. 3.
- Peterman, R.M. 1990. The importance of reporting statistical power: the forest decline and acidic deposition example. *Ecology* 71: 2024–2027.
- Rawlings, J. 1988. Applied regression analysis. Brooks/Cole, Belmont, Calif.
- Rosenthal, R. and R. Rosnow. 1985. Contrast analysis: focused comparisons in the analysis of variance. Cambridge Univ. Press, New York, N.Y.
- SAS Institute Inc. 1989. SAS/STAT user's guide. Version 6. Vol. 2. 4th ed., SAS Institute Inc., Cary, N.C.
- Satterthwaite, F.E. 1946. An approximate distribution of estimates of variance components. *Biom. Bull.* 2: 110–114
- Saville, D.J. 1990. Multiple comparison procedures: the practical solution. *Amer. Stat.* 44(2): 174–182.
- Scheffé, H. 1959. The analysis of variance. Wiley, New York, N.Y.
- Schultz, E.F. Jr. 1955. Rules of thumb for determining expectations of mean squares in analysis of variance. *Biometrics* 11: 123–135.
- Schwarz, C.J. 1993. The mixed-model ANOVA: the truth, the computer packages, the books. Part I: Balance data. *Amer. Stat.* 47(1): 48–59.
- Searle, S. 1971a. Topics in variance component estimation. *Biometrics* 27: 1–76.
- . 1971b. Linear models. John Wiley and Sons, New York, N.Y.
- Shaw, R.G. and T. Mitchell-Olds. 1993. ANOVA for unbalanced data: an overview. *Ecology* 74(6): 1638–1645.
- Sit, V. 1991. Interpretation of probability p-values. B.C. Min. For. Res. Br., Victoria, B.C., Biom. Info. Pamp. No. 30.
- . 1992. Finding the expected mean squares and the proper error terms with SAS. B.C. Min. For. Res. Br., Victoria, B.C., Biom. Info. Pamp. No. 40.
- Snedecor, G.W. and W.G. Cochran. 1967. Statistical method. 6th ed. Iowa State Univ. Press, Ames, Iowa.
- Sokal, R.R. and F.J. Rohlf. 1981. Biometry: the principles and practice of statistics in biological research. 2nd ed. W.H. Freeman and Co., San Francisco, Calif.
- Steel, R.G. and J.H. Torrie. 1980. Principles and procedures of statistics: a biometrical approach. 2nd ed. McGraw-Hill, New York, N.Y.

- Thornton, J. 1988. The importance of replication in analysis of variance. B.C. Min. For. Res. Br., Victoria, B.C., Biom. Info. Pamp. No. 2.
- Warren, W.G. 1986. On the presentation of statistical analysis: reason or ritual. *Can. J. For. Res.* 16: 1185–1191.
- Wester, D.B. 1992 Viewpoint: replication, randomization, and statistics in range research. *J. Range Manage.* 45: 285–290.
- Wetherill, G.B. 1981. *Intermediate statistical methods*. Chapman and Hall, New York, N.Y.
- Wilk, M.B. 1955. The randomization analysis of a generalized randomized block design. *Biometrika* 42: 70–79.
- Winer, B.J. 1962. *Statistical principles in experimental design*. McGraw-Hill, New York, N.Y.