

R Module 5



**Factorial
Analysis of Variance**

Certificate in EnviroStats (Non-Award)

This document is part of an online Certificate in EnviroStats (Non-Award) by the University of Canberra. Course enquiries can be directed to the address below. Expressions of interest in the course can be made online through:

<http://aerg.canberra.edu.au/envirostats>

Copies of this publication are available from:

The Institute for Applied Ecology
University of Canberra ACT 2601
Australia

Email: bernd.gruber@canberra.edu.au, georges@aerg.canberra.edu.au

Copyright © 2011 Arthur Georges [V 6.2], Bernd Gruber [Converted the manuscript from SAS to R]

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, including electronic, mechanical, photographic, or magnetic, without the prior written permission of the author.

R is an open source statistical programme. It is developed by:

R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

SPONSORED BY:**Materials development team:**

Author:	Arthur Georges, 2002, 2006
Instructional designer:	Peter Donnan, 2002
Editor:	Loretta Barnard, 2002
Graphic Design:	Peter Delgado, 2002
Desktop Publishing:	Kristi McDonald, 2004 Sue Bebbington, 2004
FDDU Project Manager:	Deborah Veness, 2002
Dynamic Web Page Design:	TCNI Software Solutions PO Box 47 LATHAM ACT 2615 Australia

First prepared in January, 2002 for Semester 1, 2002.

Reprinted January 2003 for Semester 1, 2003.

Reprinted January 2004 for Semester 1, 2004.

Reprinted November 2004 for Semester 1, 2005.

Revised and reprinted, June 2006

Reprinted February 2007 for Semester 1, 2007

Revised former SAS Module 5 and converted to R, February 2011

Published by Technology & Educational Design Services

(TEDS)

University of Canberra
ACT 2601, AUSTRALIA

Module 5: Factorial ANOVA

Lesson 1: Key Concepts in Factorial ANOVA	5
Overview	5
Rationale	7
Why do we need a new approach?	7
Option 1: Random Allocation of Sexes	8
Option 2: Even Spread of Sexes across Treatments	10
Option 3: Discard One Sex	10
Option 4: Factorial ANOVA	10
Examining the marginal means	12
Single-factor ANOVA on cell means	12
Partitioning the among-cell variation	14
The ANOVA table	16
An intuitive view	17
Where have we come?	20
Lesson 2: The concept of interaction	22
Six simple scenarios	22
Scenario 1: The null case	23
Scenario 2: A main effect of Factor A, no interaction	24
Scenario 3: Main effects of factor A and B, no interaction	25
Scenario 4: Main effect of factor A, with synergistic interaction	26
Scenario 5: Main effects of factors A and B, with interference interaction	27
Scenario 6: Nullifying interaction	29
Where have we come?	30
Lesson 3: Models in Factorial ANOVA	31
Fixed, Random or Mixed Model	31
Constructing an appropriate F ratio	31
Fixed Model	31
Random Model	32
Mixed Model	33
A special case — no replication	33
Where have we come?	34
Lesson 4: Follow-up analysis	35
In the absence of interaction	35
In the presence of interaction	36
Preliminary Examination	36
<i>A priori</i> contrasts	36
Exhaustive <i>post hoc</i> contrasts	37
Restricted <i>post hoc</i> contrasts	37
Splitting the analysis	38
Recommended approach	38
Where have we come?	38
Lesson 5: Application	40
Assumptions of Factorial ANOVA	40
Randomness in sampling	40
Independence	40
Equality of variances	41
Normality	42
Additivity	43
Robustness	43
Summary of the assumptions of ANOVA	44
Unbalanced designs	45
Crossed versus nested factors	45
Summary of the Factorial ANOVA Procedure	46
Where have we come?	48
Lesson 6: Step-through Examples	50

Example 5-1: Bait Acceptability by Feral Pigs	50
The problem.....	50
The data.....	51
Comments on the design	51
The analysis.....	51
<i>Start a R Session</i>	51
<i>Data entry and exploratory examination</i>	51
Analysis of Residuals.....	54
Analysis of Variance	56
Overall test for differences.....	56
Multiple Comparisons	59
Results summary.....	61
Discussion	62
Example 5-2: Soil phosphorus, topography and soil type	63
The problem.....	63
The data.....	63
Comments on the design	64
The analysis.....	64
Start a R Session.....	64
Data entry and exploratory examination	64
Analysis of Residuals.....	66
Analysis of variance	67
Explore the interaction	69
Splitting the analysis.....	70
Multiple comparisons	71
Results summary.....	73
Discussion	74
Where have we come?	76
Lesson 7: Some Challenging Exercises	77
Exercise 5-1: Chase duration in Chats (Mark II)	77
Exercise 5-2: Trace Metals in Botany Bay	79
Exercise 5-3: Frequent burns and ant communities	82
Exercise 5-4: Lake fertilization and fish growth (BACI)	84
Exercise 5-5: Competition from exotics on a rare native plant	86
Exercise 5-6: Artificial substrates in pond aquaculture	88
Exercise 5-7: Phytoplankton in Peel-Harvey Estuary	91
Where have we come?	94

Lesson 1: Key Concepts in Factorial ANOVA

Overview

In single-factor analysis of variance, each value of the response variable was classified according to a single criterion, a single factor. The factor might have been SITE in a river or LABORATORY in Chlorophyll-a testing trials. But what do we do when there is more than one factor? Factorial analysis of variance is one class of analyses for dealing with the effects of two or more factors on a single response variable.

Take for example a study of soil phosphorus undertaken by Anne-Marie Clements in the Sydney region (Clements, 1983). Three factors were thought to influence soil phosphorus – proximity to suburban development, soil type, and topographic location. Each soil phosphorus measurement could be classified as to whether it was close or remote to suburban development, whether it was taken from shale-derived or sandstone-derived soils, and whether it was taken from a hilltop, a north facing slope, a south facing slope, or a valley.

If we take three phosphorus measurements for each combination of the levels of each of the factors, then we have a fully-crossed, balanced design. It is called a factorial design (Figure 5-1).

Figure 5-1.
Diagrammatic
representation
of the data
classified
according to a
three-way
factorial design.

	URBAN			
	NON-URBAN			
SHALE	144 51 100	73 141 74	138 102 153	238 310 210
SANDSTONE	45 85 62	74 102 63	101 63 115	100 134 93
	HILLTOP	NORTH SLOPE	SOUTH SLOPE	VALLEY

There are three factors – topographic location has four levels ($a = 4$), soil type has two levels ($b = 2$) and proximity to suburban development has two levels ($c = 2$). The response variable is soil phosphorus. There are three values of the response variable in each cell ($n = 3$), that is, three values for each combination of the factor levels.

A number of research questions can be asked of these data.

Are there significant differences in soil phosphorus between topographic locations? Are there significant differences in soil phosphorus between shale-derived and sandstone-derived soils? Does proximity to urban development influence soil phosphorus? These questions relate to the **main effects** of the three factors.

There are also questions to do with how the factors interact. If there is a difference between soil types in phosphorus, does the magnitude of the difference depend on which topographic location the soil samples were taken? In other words, does soil type interact with topographic location? This is a question about a **first-order interaction**. Such an interaction would occur if the magnitude of the difference in soil phosphorus between shale-derived and sandstone-derived soils was greater in the valleys than on the hilltops.

Similar questions could be asked about the possible interaction between proximity to urbanisation and topographic location, or between proximity to urbanisation and soil type. In a three-way factorial design, there are three first-order interactions.

There is also a research question related to a **second-order interaction**. Say there was a first-order interaction between soil type and topographic location. If the magnitude of this interaction depended on whether the samples were taken close to or distant from urbanisation, then we would have a second-order interaction.

So in this three-way factorial ANOVA, we can ask:

- Does soil type have an impact on soil phosphorus?
- Does topographic location have an impact on soil phosphorus?
- Does proximity to urbanisation have an impact on soil phosphorus?
- Does the magnitude of the effect of soil type depend on topographic location (or vice versa) – is there an interaction between soil type and topographic location?
- Does the magnitude of the effect of soil type depend on proximity to urbanisation (or vice versa), ie – is there an interaction between soil type and proximity to urbanisation?
- Does the magnitude of the effect of topographic location depend on proximity to urbanisation (or vice versa), ie – is there an interaction between topographic location and proximity to urbanisation?
- Does the magnitude of the interaction between soil type and topographic location depend upon proximity to urbanisation, ie –

is there a three-way interaction between soil type, topographic location and proximity to urbanisation.

This is a remarkable array of hypotheses and research questions that can be addressed by this analysis, compared with what was possible in the single-factor design.

The concept of interaction is the special contribution factorial ANOVA makes to our understanding of variation in the natural world.

Three models of factorial ANOVA are recognised. In the **Fixed Model**, all the factors are fixed. In the **Random Model**, all the factors are random. In the **Mixed Model**, some factors are fixed and some are random. In single-factor ANOVA, choice of model influenced the approach taken in following up a significant result, but did not influence the calculation of the F value in the test of significance. In factorial ANOVA, choice of the model will influence both the calculation of the F statistics and the approach taken to following up a significant result. In some cases, the analysis can proceed without replication, and therefore without a calculable MS_{within} .

A significant result for a fixed factor in the factorial design indicates significant variation among the means over and above that expected to occur by chance alone, but it does not provide information on which factor classes differ from which others. Such a significant result must be followed by a set of comparisons to determine where the differences lie. The appropriate procedures to follow-up a significant result in a fixed model include testing the significance of differences between pairs of means using one of several **multiple comparison procedures**. Most statistical packages provide a wide range of options for undertaking multiple comparisons following a significant result in a factorial ANOVA.

The appropriate follow-up procedure for a significant random factor is to estimate the added variance component due to the effect of the factor. Most statistical packages provide a procedure for estimating the added variance component.

Rationale

Why do we need a new approach?

The advance factorial ANOVA affords over single-factor ANOVA is best illustrated by an example.

Feral pigs are not native to Australia and, when running wild, are significant vertebrate pests. Poisoning with sodium mono-fluoroacetate (1080) is a common method of feral pig control. Field

observations suggest, however, that pigs become bait-shy to 1080. Hone and Kelba (1984) wished to investigate this possibility using penned feral pigs and a rigorous experimental design.

On Day 1, pigs were offered wheat only and their intake was recorded. On Day 2, they were offered one of the following combinations:

- Wheat only
- Wheat and water
- Wheat, water and dye
- Wheat, water and 1080
- Wheat, water, dye and 1080

The dye is necessary for safety reasons. Water is necessary to add the dye and 1080. Intake of the bait was again recorded. Intake on Day 1 was subtracted from intake on Day 2, and the response variable was change in intake in kg (Table 5-1).

The problem Hone and Kleba wished to address was to determine if there was a significant difference in the response of pigs to the different bait mixtures. Let us consider some options for undertaking this experiment and analysing the data so obtained.

Option 1: Random Allocation of Sexes

One design appropriate to this problem is shown in Table 5-1. Here we have a single-factor design, with four pigs per pen, each pen subjected to a different bait treatment.

A significant result for the ANOVA could be followed by suitable contrasts (see Module 4) between Wheat as the control for the Wheat & Water treatment, and between Wheat & Water as the control for the remaining treatments.

*Table 5-1.
Change in bait
intake by feral
pigs in response
to different bait
constituents.*

Wheat	Wheat & water	Wheat, water & dye	Wheat, water & 1080	Wheat, water, dye & 1080
0.188	0.050	0.058	-0.172	-0.610
-0.058	-0.138	-0.082	-1.280	-0.830
-0.280	-0.540	-0.260	-0.894	-0.837
-0.062	-0.336	-0.123	-0.672	-1.202

The single-factor ANOVA approach assumes that all other potentially influential factors are held constant.

What if a second potentially influential factor exists and is not held constant? What is the impact of the sex of the pig? What will happen to our analysis if the pigs chosen for experimentation were a mixture of males and females, and if males and females respond differently to the addition of water, dye and 1080 to their food?

The first issue that arises is **artificial inflation of the error term** of the F ratio in the ANOVA, leading to loss of power. If for example we include both male and female pigs in our analysis at random (Table 5.2) and the sexes differ in their responses, you can see that the within cell variation will contain a contribution from differing responses of males and females and so will be inflated by those sex differences. Recall from Module 4 that the *F* ratio in the single factor ANOVA is calculated as

$$F = \frac{MS_{among}}{MS_{within}}$$

If MS_{within} is artificially inflated, then this F ratio will be artificially deflated. Power of the analysis will be compromised, and possibly severely so. Real differences in the response of the pigs to the different bait treatments will be potentially obscured by inflation of the error term MS_{within} . If the difference between the sexes is large, we will lose any chance of detecting even a substantial treatment effect.

Table 5-2. Change in bait intake by feral pigs in response to different bait constituents. Note the unbalanced distribution of male pigs (bold) and female pigs across the factor classes.

Wheat	Wheat & water	Wheat, water & dye	Wheat, water & 1080	Wheat, water, dye & 1080
0.188	0.050	(0.058)	-0.172	(-0.210)
(-0.058)	(-0.138)	-0.082	(-1.280)	-0.230
-0.280	(-0.540)	-0.260	(-0.894)	-0.837
(-0.062)	-0.336	(-0.123)	-0.672	(-1.202)

A more serious issue is that of **confounding**. If we for some reason preferentially pick out the male pigs and allocate pigs to pens from left to right, then the problem for the analysis is fairly obvious (Table 5-3). We will not be able to disentangle differences in response owing to bait additives from differences in response between males and females. Any differential response of the pigs to Wheat, Water, Dye & 1080 that we detect could be because male pigs dominate that treatment. Our design will have been confounded.

Table 5-3. Change in bait intake by feral pigs in response to different bait constituents. Note the unbalanced distribution of male pigs (bold) and female pigs across the factor classes.

Wheat	Wheat & water	Wheat, water & dye	Wheat, water & 1080	Wheat, water, dye & 1080
0.188	0.050	(0.058)	(-0.172)	(-0.210)
-0.058	(-0.138)	-0.082	(-1.280)	(-0.230)
-0.280	-0.540	-0.260	(-0.894)	(-0.837)
-0.062	-0.336	(-0.123)	-0.672	(-1.202)

Confounding in a single factor ANOVA occurs when a second uncontrolled factor, *correlated in some way with the first*, has an influence on the response variable. In a confounded experiment, such as shown in Table 5-3, we would not be able to unequivocally assign a

significant result to the variable we have chosen to control — bait composition. The significant result may have arisen instead because of the confounding effect of the uncontrolled variable, sex.

Option 2: Even Spread of Sexes across Treatments

To overcome the confounding, we could systematically allocate male and female pigs across the pens (Table 5-4), so that each pen had two males and two females. Any difference between the sexes would then cancel out.

Table 5-4. Change in bait intake by feral pigs in response to different bait constituents. Note the balanced distribution of male pigs (bold) and female pigs across the factor classes.

Wheat	Wheat & water	Wheat, water & dye	Wheat, water & 1080	Wheat, water, dye & 1080
(0.188)	(0.050)	(0.058)	(-0.172)	(-0.210)
(-0.058)	(-0.138)	(-0.082)	(-1.280)	(-0.230)
-0.280	-0.540	-0.260	-0.894	-0.837
-0.062	-0.336	-0.123	-0.672	-1.202

The downside of this approach is that, if there is a difference between the sexes, we are still presented with the problem of artificial inflation of the within sample variance, with attendant loss of power, as outlined above.

Option 3: Discard One Sex

What if instead we select only pigs of one sex (Table 5-5)? This will solve the problems of confounding and of potential inflation of the error term, but at the expense of **loss of generality**. Our conclusions will only apply to male pigs. We could regain the generality by repeating the experiment for female pigs, but at double the cost.

Table 5-5. Change in bait intake by male feral pigs in response to different bait constituents.

Wheat	Wheat & water	Wheat, water & dye	Wheat, water & 1080	Wheat, water, dye & 1080
(0.188)	(0.050)	(0.058)	(-0.172)	(-0.210)
(-0.058)	(-0.138)	(-0.082)	(-1.280)	(-0.230)
(-0.280)	(-0.540)	(-0.260)	(-0.894)	(-0.837)
(-0.062)	(-0.336)	(-0.123)	(-0.672)	(-1.202)

Clearly, approaching this problem from the perspective of single-factor ANOVA is fraught with difficulty.

Option 4: Factorial ANOVA

Factorial analysis of variance overcomes these difficulties. Hone and Kelba allocated their pigs to experimental treatments in a two-way

crossed design. There were ten pens in all, with two pigs per pen. Five pens contained only female pigs, and five only males (Table 5-6).

Note that we have data for each combination of the factor levels, where there are five levels for the BAIT factor and two levels for the SEX factor. It is a **crossed design**. The equal sample sizes in each cell indicate that it is also a **balanced design**.

Table 5-6. Change in bait intake by feral pigs in response to different bait constituents. Pigs are classified according to two criteria – bait treatment and sex.

	Wheat	Wheat & water	Wheat, water & dye	Wheat, water & 1080	Wheat, water, dye & 1080
Male	0.188 -0.058	0.050 -0.138	0.058 -0.082	-0.712 -1.280	-0.610 -0.830
Female	-0.280 -0.062	-0.540 -0.336	-0.260 -0.123	-0.894 -0.672	-0.837 -1.202

In this design, a two-way factorial design, we consider the variability among mean responses for the different bait types, after having averaged out the effect of sex. This averaging out is a simple process if the sample sizes are equal. We simply consider the column means of Table 5-6.

We can test the variability among column means (ie bait types) against what we would expect given observed variation within columns, after taking out the effect of sex. We take out the effect of sex from the within column variation by averaging the within cell variance for each column.

Similarly, we can test the variability among row means (ie sexes) against what we would expect given observed variation with rows, after taking out the effect of bait type.

These are tests of the **main effects**.

Finally, we can test to see if the two factors BAIT and SEX are acting in concert, one influencing the effect of the other, to influence change in bait intake of the pigs. This is a test of **interaction**.

We have overcome the problem of possible confounding by controlling both influential factors in a fully crossed design. We overcome the potential inflation of the mean square error (in this case MS_{within}) by taking out the effect of sex before we consider the significance of variation among the means for bait type, and vice versa. We maintain the generality of our conclusions by including both sexes in a combined analysis. This is what factorial ANOVA delivers.

Let us look at this approach in more detail.

Examining the marginal means

The experiment of Hone and Kleba has two factors. Factor A is bait type (BAIT) with $a = 5$ levels, and Factor B is sex (SEX) with $b = 2$ levels. There are two replicates in each cell of the 5×2 data matrix ($n = 2$). The response variable is change in grain intake (kg).

A useful exploratory tool is to present the data together with the row and column **marginal means**, and the grand mean (Table 5-7). The column marginal means are for BAIT, and are based on $bn = 4$ measurements. The row marginal means are based on $an = 10$ measurements.

Table 5-7
Cell and marginal means for a two-way classification of response of male and female pigs to different bait constituents.

	Wheat	Wheat & water	Wheat, water & dye	Wheat, water & 1080	Wheat, water, dye & 1080	
Male	0.07	-0.04	-0.01	-1.00	-0.72	-0.34
Female	-0.17	-0.44	-0.19	-0.78	-1.02	-0.52
	-0.05	-0.24	-0.10	-0.89	-0.87	-0.43

The terminology used to describe the cell means, marginal column and row means and the grand mean is shown in Table 5-8.

Table 5-8
Terminology for cell and marginal means in the two-way factorial ANOVA. Factor A (BAIT) has five factor levels ($a=5$). Factor B (SEX) has two factor levels ($b=2$).

	Wheat	Wheat & water	Wheat, water & dye	Wheat, water & 1080	Wheat, water, dye & 1080	
Male	\bar{Y}_{11}	\bar{Y}_{21}	\bar{Y}_{31}	\bar{Y}_{41}	\bar{Y}_{51}	\bar{B}_1
Female	\bar{Y}_{12}	\bar{Y}_{22}	\bar{Y}_{32}	\bar{Y}_{42}	\bar{Y}_{52}	\bar{B}_2
	\bar{A}_1	\bar{A}_2	\bar{A}_3	\bar{A}_4	\bar{A}_5	\bar{Y}

There is a difference in the response variable, change in grain intake, between males (340 g reduction) and females (520 g reduction), but is this difference statistically significant? Similarly, grain intake changes by a greater amount in the treatments with dye and 1080 added, but is this statistically significant? Are the trends across bait types the same for males and females, or does the differential response to bait additives vary between males and females?

Single-factor ANOVA on cell means

While a single factor ANOVA is not appropriate to compare between bait treatments, or between sexes, it is appropriate to look for significant differences among the 10 cells means of the two-way table of Table 5-7. This can be achieved with a simple single-factor ANOVA comparing the means of the $a.b = 10$ cells each with $n = 2$ values.

Clearly from Table 5-9, there is a significant difference among cells (in this case, pig pens) ($F = 8.98$; $df = 9,10$; $p < 0.001$).

Table 5-9.
Analysis of variance
table for a comparison
among pens in the
mean response of
male and female pigs
to different bait
constituents.

Source	DF	SS	MS	F value	Pr > F
Among cells	9	3.1386	0.3487	8.98	0.0010
Within	10	0.3885	0.03885		
Total	19	3.5271			

With only little adjustment of the formulae given in Module 4, the formulae for the mean squares are:

$$MS_{within} = \frac{\sum_a \sum_b S^2}{ab} = \frac{\sum_a \sum_b \sum_n (Y - \bar{Y})^2}{ab(n-1)} \sim \sigma^2$$

$$MS_{among\ cells} = \frac{\sum_a \sum_b \sum_n (\bar{Y} - \bar{\bar{Y}})^2}{ab-1} \approx \sigma^2 + n\sigma_{treatment\ effect}^2$$

A test for an overall effect of the treatments is given by

$$F = \frac{MS_{among\ cells}}{MS_{within}} \approx \frac{\sigma^2 + n\sigma_{treatment\ effect}^2}{\sigma^2}$$

The partition of the sums of squares and partition of the degrees of freedom are

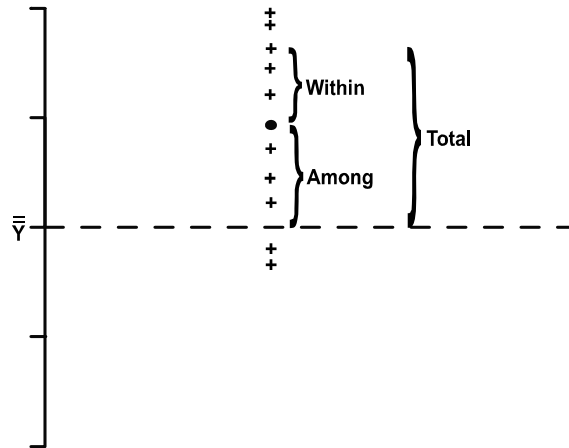
$$SS_{total} = SS_{among\ cells} + SS_{within}$$

$$\sum_a \sum_b \sum_n (Y - \bar{\bar{Y}})^2 = \sum_a \sum_b \sum_n (\bar{Y} - \bar{\bar{Y}})^2 + \sum_a \sum_b \sum_n (Y - \bar{Y})^2$$

$$abn - 1 = ab - 1 + ab(n - 1)$$

in agreement with the standard partition of the sums of squares for the single-factor ANOVA described in Module 4 (Figure 5-2). So far, nothing really new.

Figure 5-2.
Diagrammatic
representation of
the partition of
the sums of
squares in
single-factor
ANOVA. In this
case, we are
applying the
theory to the
variation among
and within cells
of the two-way
classification of
Table 5-6.



Partitioning the among-cell variation

We might now ask, how much of the variation among cell means can be attributed to a differential response of the pigs to bait type, or to differences between the sexes?

We can define a Mean Square representing the variation among means for the bait types (columns of Table 5-7) as

$$MS_{\text{among } A} = \frac{\sum^a \sum^b \sum^n (\bar{A} - \bar{Y})^2}{a-1} \approx \sigma^2 + bn\sigma_A^2$$

and test for an added variance component due to differences among bait types in the mean change in bait intake using the $MS_{\text{within cells}}$ as the error term. That is, we can test for an effect of Factor A (bait type) using

$$F = \frac{MS_{\text{among } A}}{MS_{\text{within}}} \approx \frac{\sigma^2 + bn\sigma_A^2}{\sigma^2}$$

This is a test for the significance of differences in the response of the pigs to the differing bait constituents.

Similarly,

$$MS_{\text{among } B} = \frac{\sum^a \sum^b \sum^n (\bar{B} - \bar{Y})^2}{b-1} \approx \sigma^2 + an\sigma_B^2$$

$$F = \frac{MS_{\text{among } B}}{MS_{\text{within}}} \approx \frac{\sigma^2 + an\sigma_B^2}{\sigma^2}$$

This is a test for the significance of differences in the response of male and female pigs.

A component of the sums of squares among cell means can be explained ***in part*** by a component attributable to variation among bait types (among marginal column means) and a component attributable to variation between the sexes (among the marginal row means).

It turns out that if you subtract $SS_{\text{among A}}$ and $SS_{\text{among B}}$ from SS_{total} there is something left over. This residual quantity is called the sums of squares interaction, for reasons that will become evident later. So we have

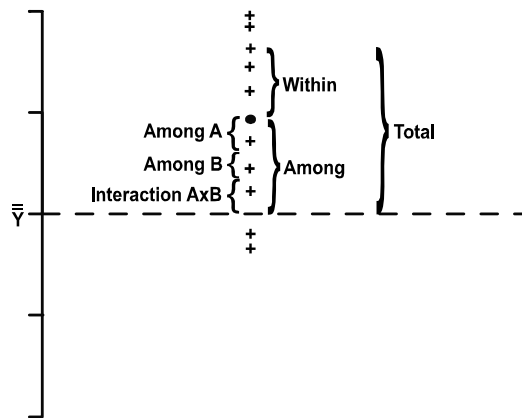
$$SS_{\text{among cells}} = SS_A + SS_B + SS_{A \times B}$$

$$\sum \sum \sum (\bar{Y} - \bar{Y})^2 = \sum \sum \sum (\bar{A} - \bar{Y})^2 + \sum \sum \sum (\bar{B} - \bar{Y})^2 + \sum \sum \sum (\bar{A} - \bar{Y})(\bar{B} - \bar{Y})$$

$$ab - 1 = (a - 1) + (b - 1) + (a - 1)(b - 1)$$

So now we have a complete partition of the total sums of squares (Figure 5-3). The total sums of squares can be split into a component attributable to variation among the cell means and a component attributable to variation within cells. The among cells sums of squares can be further partitioned into a component attributable to the effect of Factor A, a component attributable to the effect of Factor B, and an additional component attributable to the interaction between Factor A and Factor B.

Figure 5-3.
Diagrammatic
representation of
the partition of
the sums of
squares for the
factorial ANOVA.



The interaction mean square is given by

$$MS_{AxB} = \frac{\sum^a \sum^b \sum^n (\bar{A} - \bar{Y})(\bar{B} - \bar{Y})}{(a-1)(b-1)} \approx \sigma^2 + n\sigma_{AxB}^2$$

This can be tested by placing the mean square interaction over the mean square within as the error term.

$$F = \frac{MS_{AxB}}{MS_{within}} \approx \frac{\sigma^2 + n\sigma_{AxB}^2}{\sigma^2}$$

The interpretation placed on a significant interaction will be discussed later.

The ANOVA table

The partition of the sums of squares shown in Figure 5-3 is translated into an ANOVA table (Table 5-10).

Table 5-10. An ANOVA table for a two-way factorial design with a=5 levels of Factor A and b=2 levels of Factor B (both fixed). There are 2 replicates in each cell.

Source	DF	SS	MS	F	Pr > F
Among cells	9	3.1387	0.3487	8.98	0.0010
BAIT	4	2.7604	0.6901	17.77	0.0002
SEX	1	0.1606	0.1606	4.13	0.0694
SEX.BAIT	4	0.2177	0.05441	1.40	0.3023
WITHIN	10	0.3885	0.03885		
TOTAL	19	3.5271			

Confirm for yourself that the degrees of freedom and the sums of squares for BAIT, SEX and SEX.BAIT add up to those of AMONG CELLS. Confirm also that the sums of squares and the degrees of freedom for AMONG CELLS and WITHIN add up to the TOTAL sums of squares and degrees of freedom.

The tests of the overall effects of the treatments, the main effects of BAIT and SEX, and of the interaction SEX.BAIT, all have MS_{within} as the mean square error. This is true only of the fixed design, where all factors are fixed.

$$F_{overall} = \frac{MS_{among\ cells}}{MS_{within}} \approx \frac{\sigma^2 + n\sigma_{treatment\ effect}^2}{\sigma^2} \quad ab - 1$$

$$F_A = \frac{MS_{among\ A}}{MS_{within}} \approx \frac{\sigma^2 + bn\sigma_A^2}{\sigma^2} \quad a - 1$$

$$F_B = \frac{MS_{\text{among } B}}{MS_{\text{within}}} \approx \frac{\sigma^2 + an\sigma_B^2}{\sigma^2} \quad b - 1$$

$$F_{A \times B} = \frac{MS_{A \times B}}{MS_{\text{within}}} \approx \frac{\sigma^2 + n\sigma_{A \times B}^2}{\sigma^2} \quad (a - 1)(b - 1)$$

$$MS_{\text{within}} \approx \sigma^2 \quad ab(n - 1)$$

The following interpretation can be placed on the ANOVA table.

There were significant differences among the cell means ($F = 8.98$; $df = 9,10$; $p < 0.001$), of which 87.9% could be explained by an effect of differing bait constituents ($F = 17.77$; $df = 4,10$; $p < 0.0002$). No significant difference could be demonstrated between the sexes ($F = 4.13$; $df = 1,10$; $p = 0.07$) and there was no interaction between bait type and sex ($F = 1.40$; $df = 4,10$; $p = 0.30$).

An intuitive view

The foundation of ANOVA, as we saw in Module 4, is based on a relationship between the variance of sample means expected under the null hypothesis and the variance of measurements within samples.

$$\sigma_{\bar{Y}}^2 = \frac{\sigma^2}{n} \quad \text{and} \quad S_{\bar{Y}}^2 \cong \frac{S^2}{n}$$

We can use this formula to calculate our expectation.

In plain English, this says that, if the null hypothesis is true, the variation we observe among the means should roughly equal what we expect using the above relationship, that is, on consideration of the variation within samples.

Our expectation for the variation among means generally, if the null hypothesis is true, is

$$\text{Expected} = \frac{S^2}{n}$$

If the null hypothesis is not true, then we might expect the observed variation among the means to be significantly greater than expectation, and we can test this with an F test.

We put observed variation among means over expected:

$$F = \frac{\text{Observed}}{\text{Expected}} = \frac{S_Y^2}{S^2/n} = \frac{nS_Y^2}{S^2} = \frac{MS_{\text{among}}}{MS_{\text{within}}}$$

The F ratio in ANOVA can be interpreted intuitively as a test of the observed variation among means over expected.

Refer to Module 4 if you do not follow the argument, where it is presented in more detail.

Consider again the table of marginal means for the feral pig study (Table 5-11). The marginal column means represent the mean response for each bait type, averaged across the two sexes. Note that the sexes contribute equally to each bait treatment. If sex and bait are additive in their effects, that is, if there are no complicating interactions between the two, the pattern we see across the marginal means for BAIT is attributable to the effect of the baits alone, independent of any effect of sex.

In other words, by averaging across the sexes in a fully-crossed, balanced design, we have effectively cancelled out the effect of sex when computing the mean response for each bait type. **The variation among the marginal means for BAIT represents among BAIT variation, after having corrected for or controlled the effect of SEX.**

Table 5-11
Cell and
marginal means
for a two-way
classification of
response of
male and female
pigs to different
bait constituents.

	Wheat	Wheat & water	Wheat, water & dye	Wheat, water & 1080	Wheat, water, dye & 1080	
Male	0.188 -0.058	0.050 -0.138	0.058 -0.082	-0.172 -1.280	-0.610 -0.830	-0.34
Female	-0.280 -0.062	-0.540 -0.336	-0.260 -0.123	-0.894 -0.672	-0.837 -1.202	-0.52
	-0.05	-0.24	-0.10	-0.89	-0.87	-0.43

You can see by perusing the marginal means that there is considerable observed variation among the bait treatments (Table 5-11). But is it statistically significant?

What is a reasonable expectation for variation among bait treatments under the null hypothesis?

We have seen that we cannot use the within column variation as a basis for calculating our expectation, because it includes a component attributable to differences between the sexes. To do so would inflate our expectation artificially, and our observed variation would appear less significant (falsely so).

What we can do instead is take out the component of variation within columns attributable to sex by averaging the within cell variance for

each column. This gives us a good estimate of the within column variance that would occur if sex were not at all influential.

We calculate

$$\frac{\sum^b S^2}{b}$$

for each column in Table 5-7. We then average this across columns to get our estimate of the within-column variance

$$S_{A|B}^2 = \frac{\sum^a \left(\frac{\sum^b S^2}{b} \right)}{a} = \frac{\sum^a \sum^b S^2}{ab} = MS_{within}$$

We call this $S_{A|B}^2$ to indicate that it is the within-column variance after we have taken out the effect of sex (Factor B). By simple rearrangement, this is equivalent to the average within-cell variance, MS_{within} .

Each column mean is based on bn values, so using this estimate of the within column variance (after taking out the sex component), we would expect the variance of the column marginal means to be

$$\frac{S_{A|B}^2}{bn}$$

We now have

$$F_{A|B} = \frac{\text{Observed}}{\text{Expected}} = \frac{S_A^2}{S_{A|B}^2 / bn} = \frac{bnS_A^2}{S_{A|B}^2} = \frac{MS_{Among A}}{MS_{within}}$$

So this provides a basis for a more intuitive statement about what we are testing using the F ratios in factorial ANOVA.

We are comparing how variable the marginal means are, say for factor BAIT, against how variable we would expect them to be under the null hypothesis.

In doing so, we test the effect of the main effect of BAIT **after we have taken out the effect of SEX**. Similarly, the test of the main effect of SEX is undertaken **after we have taken out the effect of BAIT**. This is a critical point.

We could adjust the way in which we present sources of variation in the ANOVA table to reflect this new insight (Table 5-12), but it is not common practice.

Table 5-12. An ANOVA table for a two-way factorial design with $a=5$ levels of Factor A and $b=2$ levels of Factor B (both fixed). There are 2 replicates in each cell.

Source	DF	SS	MS	F	Pr > F
Among cells	9	3.1387	0.3487	8.98	0.0010
BAIT SEX, μ	4	2.7604	0.6901	17.77	0.0002
SEX BAIT, μ	1	0.1606	0.1606	4.13	0.0694
SEX.BAIT μ	4	0.2177	0.05441	1.40	0.3023
WITHIN	10	0.3885	0.03885		
TOTAL	19	3.5271			

We can use this new insight to explain why the sums of squares for the main effects do not fully account for the sums of squares among cells. If SEX and BAIT have a component of influence on the response variable that is in common, then we have taken it out on both occasions when we test the main effects.

This component in common is the **interaction** between the two factors. It occurs when the value taken by one Factor influences the effect of the other Factor.

Where have we come?

In this introduction to Factorial ANOVA, you should have come to appreciate that

- Single Factor ANOVA is often a limited approach because typically the response variable is not solely influenced by a single influential factor. We acknowledge that many factors may influence our response variable, but through the process of randomisation, simply add to the background noise in the data (the error variance or MS_{within}).
- When the effect of secondary factors is substantial, we need to accommodate them in the analysis, and one way of doing this is to control them experimentally, by including them as additional factors in the analysis. An ANOVA with more than one factor, in a fully crossed design, is called Factorial ANOVA.
- Factorial ANOVA allows us to investigate the simultaneous effects of more than one factor and the interactions among those factors in determining the value of the response variable.

In addition, you should now understand the additional partitioning of the total sums of squares that underpins factorial ANOVA. In particular, you should appreciate that the factorial ANOVA begins with a single factor ANOVA on the cells in the crossed design, and that the among cells variation can be further partitioned into

components attributable to the main effects of the factors and interactions. This is typically reported in the form of a standard ANOVA Table.

With this theory under your belt, it is now time to look more seriously at the concept of interaction.

Lesson 2: The concept of interaction

Six simple scenarios

Let us consider a series of simple hypothetical examples to tease out what is meant by the significance of the main effects and interaction.

Reaction time in humans can be measured by placing the hands 10 cm apart, holding a metre rule between them, dropping it and measuring how many cm pass through the hands before the ruler is caught. This is our response variable, in cm.

We know that alcohol taken for recreational purposes will adversely influence our reaction time. We know also that some medicines have the same effect, such as antihistamines. Consider a design where we assign five subjects to each of four categories at random. They are subject differentially to the following four treatments

- Orange juice and a placebo tablet
- Orange juice and an antihistamine
- Orange juice laced with vodka and a placebo tablet
- Orange juice laced with vodka and an antihistamine

So we have a fully crossed design, with some subjects receiving each combination of the levels of two factors, ALCOHOL and DRUG. The orange juice is provided only to disguise the presence of the alcohol.

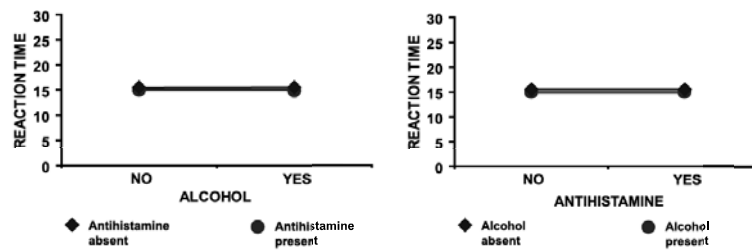
Scenario 1: The null case

Under Scenario 1 (Figure 5-4), the null case, there is no effect of either alcohol or antihistamine on reaction time. It stays steady at the average of 15 cm in all cells. The marginal means are constant at 15 cm, and the plots of the cell means show two coincident horizontal lines. None of the effects in the ANOVA table are significant.

Figure 5-4.
Reaction time in
20-year-old
males under the
influence of
various
combinations of
alcohol and
antihistamine.

	NO	YES	
NO	15	15	15
YES	15	15	15
	15	15	

There is no
effect of any
treatment on
reaction times.



Source	DF	Significance
Among cells	3	NO
A B	1	NO
B A	1	NO
AxB	1	NO
WITHIN	16	
TOTAL	19	

A = ALCOHOL B = ANTIHISTAMINE

Scenario 2: A main effect of Factor A, no interaction

Under Scenario 2 (Figure 5-5), alcohol increases reaction time by 5 units and antihistamine has no effect. The influence of alcohol is the same whether or not anti-histamine is present, so there is no interaction between the two factors.

In the absence of an interaction, the true influence of the factors can be seen as differences among the marginal means and the significance of these differences is accurately reflected in the F tests of the ANOVA table.

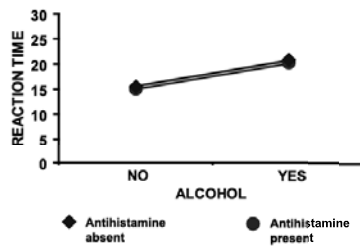
Factor A – Main Effect

Factor B – No Main Effect

No Interaction

Figure 5-5.
Reaction time in
20-year-old males
under the
influence of
various
combinations of
alcohol and
antihistamine. The
main effect of
alcohol is
significant, adding
5 units to reaction
time.

	NO	YES	
NO	15	20	17.5
YES	15	20	17.5
	15	20	



Source	DF	Significance
Among cells	3	YES
A B	1	YES
B A	1	NO
AxB	1	NO
WITHIN	16	
TOTAL	19	

A = ALCOHOL B = ANTIHISTAMINE

The graphs of the response variable against each of the factors provide a useful aid to interpretation. It is clear from both graphs that alcohol is adding 5 units to reaction time, and that antihistamine

is not having any influence. The fact that the lines are parallel (coincident in one case) indicates that there is no interaction.

Scenario 3: Main effects of factor A and B, no interaction

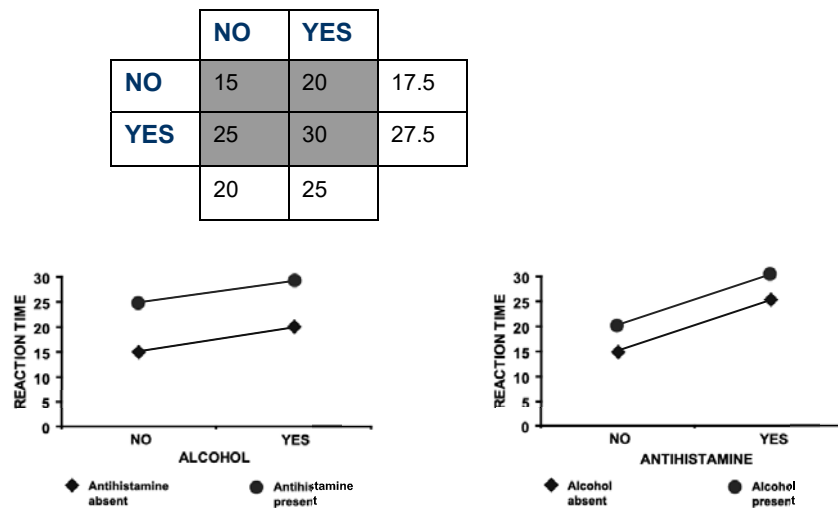
Under Scenario 3 (Figure 5-6), the main effect of alcohol adds 5 units to reaction time as before; antihistamine adds 10 units. When both alcohol and antihistamine are consumed, 15 units are added to reaction time, so the effect of the two factors is **additive**. There is no interaction.

Factor A – Main Effect

Factor B – Main Effect

No Interaction

Figure 5-6. Reaction time in 20-year-old males under the influence of various combinations of alcohol and antihistamine. The main effect of alcohol adds 5 units to reaction time as before; antihistamine adds 10 units. When both alcohol and antihistamine are consumed, 15 units are added to reaction time, so the effect of the two factors is additive.



Source	DF	Significance
Among cells	3	YES
A B	1	YES
B A	1	YES
AxB	1	NO
WITHIN	16	
TOTAL	19	

A = ALCOHOL B = ANTIHISTAMINE

Again, in the absence of an interaction, the true influence of the factors can be seen as differences among the marginal means and the significance of these differences is accurately reflected in the F tests of the ANOVA table.

It is clear from both graphs (Figure 5-6) that alcohol is adding 5 units to reaction time in the absence of antihistamine, and that

antihistamine adding 10 units to reaction time in the absence of alcohol. The fact that the lines are parallel again indicates that there is no interaction, that the effects of the two factors are additive.

Scenario 4: Main effect of factor A, with synergistic interaction

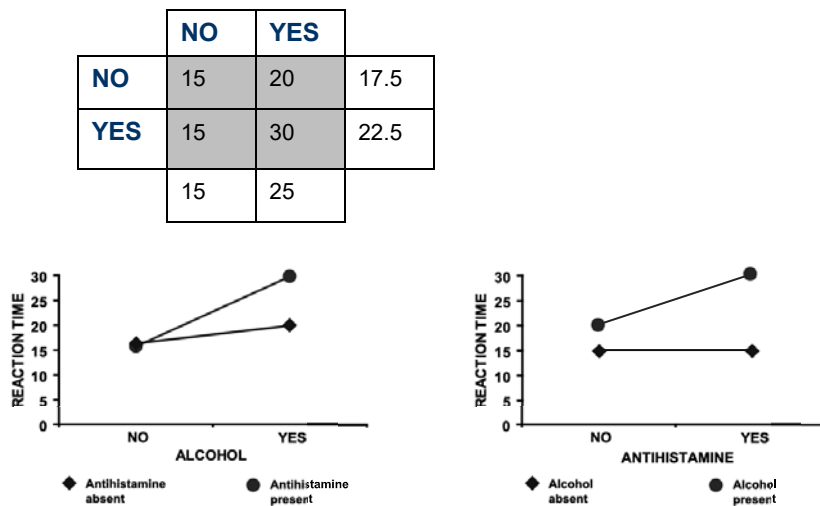
In Scenario 4 (Figure 5-7), alcohol adds 5 units to reaction time as before. Antihistamine has no effect in the absence of alcohol, but magnifies the effect of alcohol when both are present. The effect of alcohol depends on whether or not antihistamine is consumed – the two factors are interacting. The nature of the interaction is **synergistic**, because the presence of antihistamine magnifies the effect of alcohol.

Factor A – Main Effect

Factor B – No Main Effect

Synergistic Interaction

Figure 5-7. Reaction time in 20-year-old males under the influence of various combinations of alcohol and antihistamine. The main effect of alcohol adds 5 units to reaction time in the absence of the drug as before; antihistamine has no effect in the absence of alcohol, but magnifies the effect of alcohol when both are present.



Source	DF	Significance
Among cells	3	YES
A B	1	YES
B A	1	YES
AxB	1	YES
WITHIN	16	
TOTAL	19	

A = ALCOHOL B = ANTIHISTAMINE

So how is this scenario reflected in the outcomes of an analysis? First of all, the graphs show lines that are no longer parallel, indicating a potential interaction between the two factors. Close inspection reveals that the two factors are synergising not interfering. The

significance of this interaction is demonstrated by the appropriate entry in the ANOVA table.

The main effect of alcohol is evident in the marginal means and in the ANOVA table, but curiously, so too is the main effect of antihistamine. The true situation is that antihistamine exerts its influence only through its interaction with alcohol, so why does the ANOVA table indicate that it is also exerting a main effect?

It is because Factor B (Antihistamine) is exerting an influence on the means for each level of Factor A (alcohol), through their interaction. In the presence of an interaction, the variation among marginal means for each factor will be inflated (synergism) or deflated (interference), with obvious consequences for the F tests of significance of the main effects.

The lesson is that, in the presence of an interaction, interpretation of the significance of the main effects is very difficult. In this case, it has led to a spurious significant result for the main effect of antihistamine, when in fact we know that it exerts its effect only through its influence on the action of alcohol.

Scenario 5: Main effects of factors A and B, with interference interaction

In Scenario 5 (Figure 5-8), alcohol adds 5 units to reaction time and antihistamine adds 10 units as before. However, when both are present, they interfere with each other so that reaction time is not the sum of the two effects (yielding 30 units), but something less (together adding only 3 units to yield 18 units)

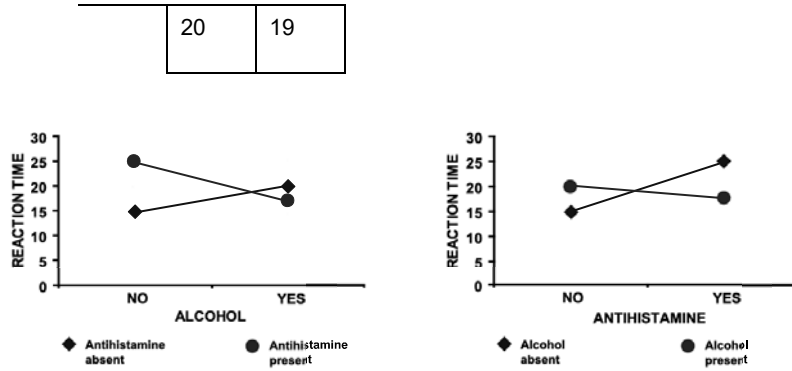
The magnitude of the effect of alcohol depends on whether or not antihistamine is consumed, and the magnitude of the effect of antihistamine depends on whether or not alcohol is consumed – the two factors are interacting. The nature of the interaction is **interference**, because the presence of antihistamine depresses the effect of alcohol and vice versa.

Factor A – Main Effect
Factor B – Main Effect
Interference Interaction

Figure 5-8.
Reaction time in
20-year-old males
under the
influence of
various
combinations of
alcohol and
antihistamine. The

	NO	YES	
NO	15	20	17.5
YES	25	18	21.5

main effect of alcohol adds 5 units to reaction time in the absence of antihistamine as before; antihistamine adds 10 units in the absence of alcohol. Together, though, they only add 3 units. The two compounds are interfering.



Source	DF	Significance
Among cells	3	YES
A B	1	YES
B A	1	YES
AxB	1	YES
WITHIN	16	
TOTAL	19	

A = ALCOHOL B = ANTIHISTAMINE

Examination of the graphs reveals the nature of the interaction. The ANOVA table provides the test of its significance. The relationship of the marginal means to each other is affected by the interaction, and so the tests of the main effects in the ANOVA table cannot be easily interpreted.

Scenario 6: Nullifying interaction

The final scenario (figure 5-9) shows a case where there is a clear influence of both alcohol and antihistamine, but because of a strong interaction, the evidence of their effects is obscured.

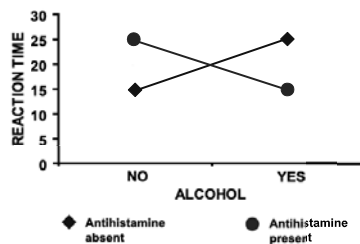
Alcohol has a strong influence on reaction time. In the absence of antihistamine, it increases reaction time by 10 units. However, in the presence of antihistamine, consumption of alcohol decreases reaction time by 10 units. The two effects cancel out in determining the marginal means. This interpretation is clear from the graphs.

Note that in the ANOVA table, tests of the significance of the main effects will fail to demonstrate a main effect of alcohol or of antihistamine. It would be wrong however to conclude that alcohol does not influence reaction time. Clearly it does, but the magnitude and direction of its effect is greatly influenced by the consumption of antihistamine.

Factor A – Main Effect
 Factor B –Main Effect
 Nullifying Interaction

Figure 5-9. Reaction time in 20-year-old males under the influence of various combinations of alcohol and antihistamine. The main effect of alcohol adds 10 units to reaction time in the absence of antihistamine, but subtracts 10 units in the presence of antihistamine. This strong interaction leads to a cancellation of the main effects on the marginal means, and so to lack of significance of the main effects in the ANOVA table.

	NO	YES	
NO	15	25	20
YES	25	15	20
	20	20	



Source	DF	Significance
Among cells	3	YES
A B	1	NO
B A	1	NO
AxB	1	YES
WITHIN	16	
TOTAL	19	

A = ALCOHOL B = ANTIHISTAMINE

Hence, in the presence of an interaction, we cannot even place a reliable interpretation on lack of significance of the main effects. In this case, the interaction has led to lack of significance of the main effects of alcohol and antihistamine, when in fact we know that they exert a substantial effect. It is just that the effect of each factor is in different directions depending on the value of the other.

Where have we come?

In this lesson, you explored the concept of interaction. If this is a new concept for you, it will have been a steep learning curve. You need to appreciate that:

- An interaction occurs in the two-way classification when the effect of one factor is moderated by the value taken by the second factor.
- The presence of an interaction indicates that the main effects of the two factors do not predict perfectly the individual cell means – that some variability among the cell means is not attributable to the sum of the effects of the two factors. The effects of the two factors are not additive.
- In the presence of interaction, we cannot give an unequivocal statement about the effect of one factor, without qualifying the statement in terms of the value taken by the other factor.
- An interaction may be synergistic, whereby the effect of the two factors in combination is greater than would be expected from the sum of their two independent effects.
- An interaction may be interference, whereby the effect of the two factors in combination is less than would be expected from the sum of their two independent effects.
- Interaction may lead to either inflation of the variance among marginal means for each factor and spurious significant results, or deflation in the variance among marginal means and spurious non-significant results.

Hence, in the presence of a strong interaction, it is not possible to unequivocally interpret the significance or lack of significance of the main effects in factorial ANOVA.

We now need to consider an additional important dimension of Factorial ANOVA, that of different models of ANOVA.

Lesson 3: Models in Factorial ANOVA

Fixed, Random or Mixed Model

Two models were recognised in single-factor ANOVA — the fixed model and the random model. If you are unclear on the distinction, you should revise the material presented in Module 4. In the fixed model, the criterion upon which the factor levels are chosen are fixed and repeatable. In the random model, the factor levels are chosen at random from a substantial pool of possible choices. The two models differed little in the computation leading to the final F statistic in the single factor ANOVA. Where they differed was in the direction taken during follow-up analysis.

In factorial ANOVA, three models are recognised. In the **Fixed Model**, all factors are fixed. In the **Random Model**, all factors are random. In the **Mixed Model**, some factors are fixed and some random. Unlike single-factor ANOVA, choice of model in the factorial ANOVA affects both the F tests of significance and the direction taken in follow-up analysis.

In particular, choice of the model determines the ratio of mean squares that is appropriate for isolating a particular effect. We need to know the model we are working with in order to construct the appropriate tests of the main effects, and unfortunately, most software packages do not do this automatically.

Constructing an appropriate F ratio

Fixed Model

In the fixed model factorial ANOVA, the error term in the denominator of the F ratios used to test the main effects and the interaction is in all cases MS_{within} . The basis for this decision lies in the mean squares and what they estimate. For the fixed model, they were

$$MS_{Among A} \approx \sigma^2 + nb\sigma_A^2$$

$$MS_{Among B} \approx \sigma^2 + na\sigma_B^2$$

$$MS_{AxB} \approx \sigma^2 + n\sigma_{AxB}^2$$

$$MS_{within} \approx \sigma^2$$

Placing the mean square for each Factor and for the interaction over MS_{within} isolates the treatment effects of Factor A, Factor B and the interaction respectively.

$$F_A = \frac{MS_{among A}}{MS_{within}} \approx \frac{\sigma^2 + bn\sigma_A^2}{\sigma^2}$$

$$F_B = \frac{MS_{among B}}{MS_{within}} \approx \frac{\sigma^2 + an\sigma_B^2}{\sigma^2}$$

$$F_{AxB} = \frac{MS_{AxB}}{MS_{within}} \approx \frac{\sigma^2 + n\sigma_{AxB}^2}{\sigma^2}$$

Under the corresponding null hypotheses, each of these F ratios should be approximately equal to 1. If these F ratios differ significantly from 1, then it must be because of a bias due to the effect of the treatment or interaction respectively. Each F ratio isolates the effect we are interested in testing.

Random Model

For the Random Model, with both factors random, the expected mean squares are

$$MS_{Among A} \approx \sigma^2 + n\sigma_{AxB}^2 + nb\sigma_A^2$$

$$MS_{Among B} \approx \sigma^2 + n\sigma_{AxB}^2 + na\sigma_B^2$$

$$MS_{AxB} \approx \sigma^2 + n\sigma_{AxB}^2$$

$$MS_{within} \approx \sigma^2$$

To isolate the main effects using an F ratio, we need to place the mean square for the main factors over the mean square interaction as the error term.

$$F_A = \frac{MS_{among A}}{MS_{AxB}} \approx \frac{\sigma^2 + n\sigma_{AxB}^2 + bn\sigma_A^2}{\sigma^2 + n\sigma_{AxB}^2}$$

$$F_B = \frac{MS_{among B}}{MS_{AxB}} \approx \frac{\sigma^2 + n\sigma_{AxB}^2 + an\sigma_B^2}{\sigma^2 + n\sigma_{AxB}^2}$$

The interaction is tested by placing the mean square interaction over MS_{within} as before.

Mixed Model

For the Mixed Model, with Factor A fixed and Factor B random, the expected mean squares are

$$MS_{\text{Among A}} \approx \sigma^2 + n\sigma_{\text{AxB}}^2 + nb\sigma_A^2$$

$$MS_{\text{Among B}} \approx \sigma^2 + na\sigma_B^2$$

$$MS_{\text{AxB}} \approx \sigma^2 + n\sigma_{\text{AxB}}^2$$

$$MS_{\text{within}} \approx \sigma^2$$

Hence, the mean square error for the fixed factor is the mean square interaction, and the mean square error for the random factor is

$$MS_{\text{within}}.$$

$$F_A = \frac{MS_{\text{among A}}}{MS_{\text{AxB}}} \approx \frac{\sigma^2 + n\sigma_{\text{AxB}}^2 + bn\sigma_A^2}{\sigma^2 + n\sigma_{\text{AxB}}^2}$$

$$F_B = \frac{MS_{\text{among B}}}{MS_{\text{within}}} \approx \frac{\sigma^2 + an\sigma_B^2}{\sigma^2}$$

The interaction is tested by placing the mean square interaction over MS_{within} as before.

Clearly, choice of model in factorial ANOVA has a profound effect on the F ratios for testing the main effects. Many computer packages require you to explicitly state the error term for each test before constructing the appropriate F ratio.

A special case — no replication

When there is no replication, it is not possible to calculate MS_{within} . If the ANOVA were a Fixed Model, then no testing would be possible. However, the Mixed Model two-way ANOVA requires that the mean square for the fixed factor is tested against the interaction mean square. This test can therefore be conducted in the absence of MS_{within} and so in the absence of replication.

This is an important special case of the two-way factorial ANOVA that will be introduced as a worked example later in this Module.

When the fixed factor has only two factor levels, this analysis is equivalent to, and replaces, the Paired T-Test.

Where have we come?

In this lesson, we revisited the concept of Models in ANOVA. You should now appreciate that:

- A factor can be considered as fixed or random. This means that a Factorial ANOVA can be fixed (all factors fixed), random (all factors random) or mixed (some factors fixed and some random).
- In single factor ANOVA, whether the factor was fixed or random did not alter the calculations, but had a profound effect on how we followed up a significant result with further analysis.
- In factorial ANOVA, the choice of model profoundly affects both the calculation of the F ratios in the ANOVA and how we follow up a significant result. There is also the possibility of mixtures of fixed and random factors in the design.

You should also be able to use a table of mean squares and what they estimate to determine the appropriate F ratios for isolating an effect of interest.

We move now on to how to follow up a significant result in factorial ANOVA.

Lesson 4: Follow-up analysis

Recall from Module 4 that the issue of following up a significant result in ANOVA requires that the compounding of error must be avoided (the probability of any Type I error at all in the overall experiment must be controlled to 0.05) and requires that we use all available data in estimating the error variance in our hypothesis tests. These same considerations apply in factorial ANOVA.

Choice of follow-up analysis in factorial ANOVA is influenced by:

- the type of factors involved.
- whether or not there is a substantive interaction between the factors in the design.
- whether the comparisons to be undertaken are restricted *a priori* or *post hoc* and exhaustive.

In the absence of interaction

When the factors do not interact, or when they interact only weakly, the multiple comparison procedures discussed in Module 4 for single-factor ANOVA apply to factorial ANOVA with only minor modifications.

- Where there are only two factor levels for a fixed factor, no correction is necessary as only one comparison is possible and its significance has already been determined by the F statistic in the ANOVA table.
- If there are more than two factor levels for a fixed factor, and all or a large number of comparisons are to be made between levels, then the Tukey-Kramer procedure is appropriate. In the context of factorial ANOVA, $MS_{\text{within cells}}$ is used in the Tukey-Kramer calculations to compare the marginal means, with improvement in power.
- When only a few pairwise comparisons are to be made, and are specified in advance of the analysis, the Bonferroni procedure is appropriate. Again, $MS_{\text{within cells}}$ is used in the calculations to compare the marginal means.
- If there is a single control and a number of treatment classes among the factor levels, then Dunnett's procedure is appropriate. Again, $MS_{\text{within cells}}$ is used in the calculations.
- For random factors, the added variance component due to the overall influence of the factor can be estimated. Multiple comparisons among the factor levels are not appropriate.

In the presence of interaction

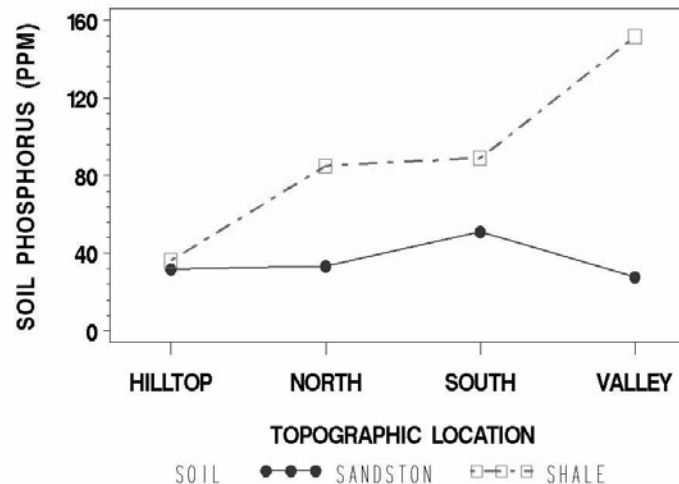
When a strong interaction is present, and cannot be removed by transformation, we are faced with the problem that the main effects of one factor cannot be described in general – the effect of one factor will depend in magnitude or direction on the level of the other. An unqualified statement about the main effects of each of the two factors cannot be made.

Statisticians vary in their approaches to following up a significant result in a factorial ANOVA where there is a substantial interaction.

Preliminary Examination

The very first step in interpreting a factorial ANOVA with interaction is to plot the treatment means. These are called **profile plots**. The response variable is assigned to the Y-axis, the factor with the most factor levels to the X-axis, and the levels of the other factor are distinguished in the profile plot with different colours or symbols (Figure 5-10).

Figure 5-10. Profile plot of treatment means for a factorial ANOVA of soil phosphorus against topographic location and soil type. Clearly topographic location has a greater effect on soil phosphorus in shale-derived soils than in sandstone-derived soils – an interaction.



The nature of the interaction should be readily apparent from the profile plot because the trend lines will not be parallel in the presence of interaction. Refer to Figures 5-4 to 5-9 for other examples.

In the presence of such a strong and no doubt significant interaction, the F tests of the main effects are unreliable. So too would be any follow-up tests on the marginal means.

A priori contrasts

One option is to partition the analysis into a series of contrasts specified explicitly to address hypotheses of interest. Some statisticians would require that these contrasts be decided in advance

of the factorial analysis (that is, that they be *a priori*). The difficulty with this approach is that it takes no account of the compounding of the Type I Error rate when we undertake a chain of related tests. Nevertheless, it is an approach that is widely used.

If the sample sizes are equal, a profile plot of the treatment means can be prepared. The least significant difference (LSD) among the cell means can be added to the plot to assist interpretation.

Exhaustive *post hoc* contrasts

If there are no prior expectations to guide the selection of contrasts, a second approach is to undertake exhaustive multiple comparisons among the $a \times b$ cell means, using the Tukey-Kramer Procedure. The comparisons of interest can then be selected from among them.

While this approach covers all bases in terms of potential pairwise comparisons, it is not recommended because it is unnecessarily conservative. The pool of comparisons of interest usually involves only comparisons within rows or within columns, not comparisons between all cells. For example, a 5×4 table has 20 cells yielding 190 potential comparisons, and they will form the basis of the Tukey-Kramer correction. We will be correcting the error rate for 190 comparisons.

If we restrict our attention to comparisons within rows or within columns, we need to adjust only for 70 comparisons. Hence, applying the Tukey-Kramer Procedure to the 20 cell means is far too conservative.

Restricted *post hoc* contrasts

A third approach is to decide what comparisons among factor levels are of interest at the planning stages of the study (that is, *a priori*), and work out from that what is the pool of potential comparisons. Then use the Bonferroni to correct for multiple comparisons.

For example, in the study of the response of feral pigs to various bait additives, Hone and Kleba designed the experiment as a two-way, fixed model, factorial ANOVA. Bait type was Factor A and pig sex was Factor B. Bait type had two experimental controls, Wheat and Wheat with Water. Sex is included in the design in order to control for that potential source of variation. It was not the primary focus of the study.

Had an interaction sex and bait type been present, we would only be interested in comparisons between the bait types and the control treatments for each sex – 14 comparisons in all. We would apply the Bonferroni correction on the basis of a pool of $k = 14$ comparisons, not the $k = 45$ possible comparisons between all pairs of cell means.

Splitting the analysis

A fourth option is to split the analysis. For example, we might conduct a series of single-factor ANOVAs on Factor A separately for each level of Factor B, with significance in the ANOVAs followed up with an appropriate multiple comparison procedure. However, this controls the "experiment-wide error rate" for each single factor ANOVA and not for the whole experiment, and the mean square error for each single factor ANOVA is based on only a portion of the available data. Nevertheless, this approach is sometimes recommended because of its simplicity. It is conservative, so if significant results emerge from the analysis and meet the objectives of the study, then they can be believed. The loss of power will only be a problem when non-significant results are obtained for important questions.

Recommended approach

The approach taken in this Module is to split the analysis into a series of comparisons when a factor involved in the interaction has more than two levels. This requires specifying sub-analyses where the effect of one factor is assessed separately for each level of the other, but still using the overall $MS_{\text{within_cells}}$ as the error term to maintain power. There is no correction at this stage of the analysis for multiple comparisons.

Significant results arising within each of the split analyses are followed-up with appropriate pairwise comparisons (refer to Module 4). Corrections for compounding error are made only for a restricted pool of potential comparisons decided as part of the overall study design, or in the absence of any specific restrictions, for the restricted pool arising from comparisons within rows and within columns only.

This approach enables the comparison among means of one factor for each level of another in the presence of interaction. It splits the analysis, but maintains power through the use of the overall within cell variance in comparisons. The impact of compounding error is controlled to an extent considered to be acceptable, but not to the fullest extent possible.

This approach will be illustrated in the second step-though example of this Module.

Where have we come?

In this lesson, we have covered the application of multiple comparison tests in the context of factorial ANOVA and seen that when the assumption of additivity is met (no substantive interactions), that the procedures we covered in *Module 4: Single Factor ANOVA* apply with little modification.

In addition, we have explored the complexity that occurs when there is a substantial interaction. The key take home message is that, in the presence of a substantial interaction, it is not possible to report the significance of the main effects without qualification. The analysis needs to be split into sub-analyses, with a range of attendant challenges for the analyst – how do you maintain power, how do you maintain the experiment-wide error to 0.05.

Various options have been presented, from simply graphing the data to explore the nature and strength of the interaction before deciding on how to report the main effects qualitatively, through to splitting the analysis into sub-analyses so that statistical significance can be assigned properly in reporting the main effects.

These concepts are best conveyed by way of example, and further attention to them will be paid during the worked examples.

Lesson 5: Application

Assumptions of Factorial ANOVA

Up to this point, factorial ANOVA has been presented without much attention paid to the assumptions of the technique. This is the approach adopted by Sokal and Rohlf (1981), in the belief that non-mathematical audiences learn better if they come to understand the structure and purpose of the analysis, without being distracted from the central theme by whether or not the data are strictly amenable to such an analysis. However, it is essential for the practitioner to verify that the assumptions are reasonable. If they are not, steps should be taken to ensure that the assumptions are met.

In this section, I describe the assumptions of factorial ANOVA, how to check if they are reasonable, and how to proceed in the face of perceived violations. Many of the assumptions of factorial ANOVA are held in common with single-factor ANOVA, so the following treatment will be brief. You are referred to Module 4 for revision of the assumptions of single-factor ANOVA.

Factorial ANOVA has five assumptions, namely, randomness and independence in sampling, equality of variances across samples, normality and additivity.

Randomness in sampling

Factorial ANOVA assumes that the items, individuals or entities allocated to each combination of the factor levels (that is, to each cell in the $a \times b$ classification) are done so at random. It is important that the only systematic difference between the cells, if any, is attributable to the differential effects of the factors. Non-randomness may manifest itself as lack of independence of the entities, or in unequal variances or in non-normality. Violation of the assumption of randomness in sampling cannot be overcome easily, and typically the data must be discarded, the sampling protocols redesigned and the data recollected. Adequate attention must be paid at the time of designing an experiment, or when sampling from natural populations, to ensure random sampling.

Independence

Independence requires that knowledge of the value of one measurement in a particular cell provides no information on the value of any other measurement in the same cell, with respect to its deviation from the cell mean. Violation of this assumption may lead to deflation of the within-cell variance, and so artificially inflate the F-value in tests of significance that use MS_{within} as the denominator. This in turn will lead to the discovery of spurious significant

differences or trends, so it is fatal. This form of lack of independence is sometimes called **pseudo-replication of the sampling units**.

Independence also requires that knowledge of the value of one measurement in a particular cell provides no information on the value of any other measurement in any other cell, with respect to its deviation from the overall mean. Failure of this assumption may lead to deflation of the among column or among row means, and so artificially deflate the F-value in tests of significance using $MS_{\text{Among } A}$ or $MS_{\text{Among } B}$ as the numerator. The power of the analysis will be compromised. This form of lack of independence is sometimes called **pseudo-replication of the treatment levels**.

The assumption of independence can be met through appropriate experimental design. Essentially, as with randomness in sampling, violations of the assumption of independence of errors cannot be overcome easily, and typically the data must be discarded, the sampling protocols redesigned and the data recollected. Adequate attention must be paid at the time of designing an experiment, or when sampling from natural populations, to ensure independence.

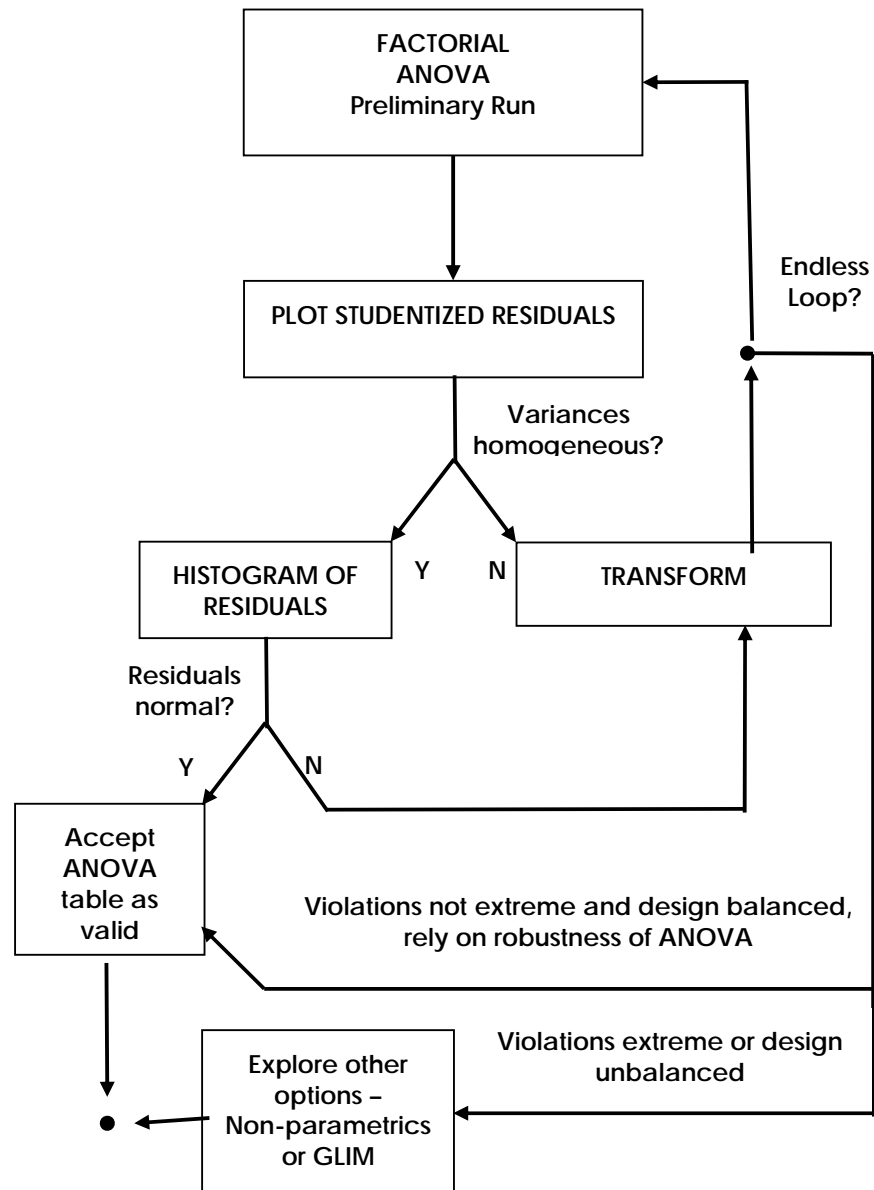
Equality of variances

An assumption of factorial ANOVA, in common with single-factor ANOVA, is that the individual sample variances within each cell estimate a common population variance, that is, that the population variances are equal.

This assumption is evaluated using a residual analysis (Figure 5-11). We visually examine the scatter of sample values about their predicted values, the cell means. Ideally, this scatter should be random across the cells. There should be no systematic trend or difference in the scatter of values about their respective means. The visual examination is achieved by constructing a plot of residuals, that is, by plotting the observed deviation of each sample value from its cell mean against the predicted value for that cell, namely the cell mean. Often, the residuals are scaled by dividing by their standard errors, called Studentized Residuals, but in this Module we will use unstandardised residuals. The method of plotting residuals is demonstrated in the worked examples that follow later in this Module.

If the variances appear heterogenous, then a transformation may be applied to bring the variance of the residuals closer to equality, or the original data for a single suspect outlier may be scrutinised in case an error has been made. An analysis of residuals should be undertaken routinely as part of the factorial ANOVA.

Figure 5-11.
A Decision Tree
for Residual
Analysis.



Normality

A factorial ANOVA assumes that the individual measurements in each cell are normally distributed about the true cell mean. What we must do first is centre the data on the cell means, pool the resulting residuals, examine the distribution of the residuals, and apply tests of normality.

This approach is conditional on the assumption of equality of variances. It is assumed that we have rectified any heterogeneity of variances in advance of the test of normality.

Additivity

In two-way or higher-order ANOVA it is necessary to assume no interaction if tests of the main effects are to be efficient and reliable. Tests of the main effects in the ANOVA table are based on the assumption that the factors involved in the ANOVA are acting independently, that is, that their effects are additive.

Tests of the main effects are not always the focus of attention. Sometimes it is the test of the interaction itself that is the focus of the study. If the main effects are unimportant, then no action is necessary in the presence of an interaction.

In considering the impact of interaction on our ability to interpret the main effects, it is important to make the clear distinction between significance and strength of result. With sufficient data, an interaction of no real consequence biologically or for its impact on the main effects, can be highly significant. It is not sufficient that an interaction is significant to obscure interpretation of the main effects – it must be substantial. The strength of the interaction, and an indication of its importance in governing the direction of the subsequent analysis, can be assessed from the plots of treatment means (Figures 5-4 to 5-9).

When the main effects are the focus of study, and a substantial interaction is present, it may be possible to remove the interaction by applying an appropriate transformation. A classic case is when the factors are multiplicative rather than additive. Applying a log transformation to the response variable will convert the multiplicative interaction between the factors to an additive one. The significance of the main effects can then be determined from the F-tests in the ANOVA table.

Robustness

The approach to checking assumptions recommended in this Module, that is, through qualitative examination of residuals, is not particularly rigorous. It relies in part on a general belief that analysis of variance is robust to moderate violations of the assumptions of normality and equality of variances. All but moderate violations would be evident on examination of the residuals in the manner described.

The foundation for this belief lies in Monte Carlo simulations undertaken in the middle of this century and reported by Lindquist (1953:78) and Keppel (1973). These studies show that moderate violations of normality and equality of variances do not constitute a serious problem provided the samples sizes are reasonable (greater than 10, say) and equal or nearly so.

If you are to rely heavily upon the robustness of ANOVA to violations of the assumption of equality of variances, in designing experiments, it is important to balance the design, that is, to ensure that the size of samples in each factor class are the same.

Balancing the design carries the additional benefit of relying upon mathematical argument for which there is greater consensus. The issues of how to analyse unbalanced designs, particularly those where some cells contain no data, remain controversial.

Of course, the assumptions of independence and randomness in sampling continue to be important regardless of sample size, and additivity remains important for the interpretation of the main effects.

Summary of the assumptions of ANOVA

The assumptions of randomness and independence in sampling must be ensured by paying adequate attention to the random selection and allocation of items to the experimental classes or, if the design is constrained by the logistics of working with natural populations, by paying adequate attention to the random selection of items from within the experimental classes. If the assumptions of randomness or independence are violated, the results of the analysis can be profoundly affected, and the only recourse is to discard the data, redesign and repeat the experiment.

Departures from the assumption of equality of variances can be detected in a qualitative way by examining a plot of residuals, and a suitable transformation might be suggested by the pattern of scatter of those residuals. The effectiveness of the transformation may be evaluated by examination of the residuals following the transformation.

Having convinced yourself that the assumption of equality of variances is tenable, the assumption of normality may be tested by examining a histogram of the pooled residuals, and by applying one or more of the tests introduced in Module 2 (Shapiro-Wilks test, probability plots etc) to the pooled residuals. Transformation may improve the normality of the residuals.

Factorial ANOVA is robust to modest violations of the assumptions of normality and homogeneity of variances provided the sample sizes are reasonably large (say, > 10) and the design is balanced.

The presence of a significant and substantial interaction is an indication of a violation of the assumption of additivity. If the main effects are the primary focus of the study, then the interaction may be removed by applying an appropriate transformation.

Multiplicative effects are commonplace and can be rendered additive with a log transformation of the response variable.

If no suitable transformation presents itself, the analysis can be split into separate analyses for each level of one of the factors involved in the interaction.

Unbalanced designs

The theory presented on factorial ANOVA involving a comparison of the variation among the marginal means with that expected on consideration of variation within cells rests upon the design being balanced. By balanced, I mean equal sample sizes across all cells. When the sample sizes are unequal, or worse, when some cells lack data altogether, adjustments have to be made to the computations. Statisticians have been working on this problem since a seminal paper by Yates was published in the earliest days of ANOVA in 1933.

In the case of unequal sample sizes (but not empty cells), the solution is to use weighting to adjust the contribution of the values in a cell to their marginal mean on the basis of the sample size. Using Type III Sums of Squares you need to use the function `Anova()` for within the R-package `car`.

In the extreme case of cells with data missing altogether, the problem is less tractable. One approach is to use linear modelling to estimate the values of the means in the missing cells from information in the other cells. Using Type IV Sums of Squares in the `Anova()` function makes these adjustments, but the usefulness of this approach is very much debated among statisticians.

The take home message here is that when the design is unbalanced, simply considering the marginal means in a summary table, or considering a profile plot of the means such as one constructed to examine interactions, may give very misleading results. Instead, one should use or plot the means adjusted to cater for the unequal cell sample sizes or for missing cells.

There are clearly major advantages in striving for a balanced design. First, the analysis will be more robust in the presence of modest violations of the assumptions, and second, the need to adjust the computations to cater for unequal samples sizes or missing cells is often an avoidable complication.

Crossed versus nested factors

Factorial ANOVAs apply to fully crossed designs, where data has been collected for each combination of the factor levels for two or more factors.

Consider an example where we wish to evaluate feral donkey density in three areas that differ in their degree of aridity. In each area we establish three 50 km transects by placing prominent white cones on the ground to mark the beginning and end of the transects. Each transect is flown five times using a fixed wing aircraft and standard protocols for counting donkey numbers (Table 5-13). Is this a crossed design?

Table 5-13.
Counts of feral donkeys along 50 km transects in coastal, semi-arid and intermediate regions of the Victoria River catchment, Northern Territory.

	Transect	LOCATION					
		SEMI-ARID		INTERMEDIATE		COASTAL	
	1	10	15	10	19	11	20
		9	17	13	12	14	13
		12	11	11	17	12	18
	2	9	14	12	17	10	12
		15	10	9	21	9	19
		9	15	11	11	17	12
	3	15	15	12	19	19	27
		10	17	15	18	13	21
		12	13	19	16	20	15

At first glance, it might appear to be a crossed design, but it is not. It is a single-factor ANOVA with location as the fixed factor and nested replication. This becomes apparent when you realise that you could readily swap the contents of Transect 1 and 2 in the semi-arid column without swapping the entire rows. Transect 1 in the semi-arid region is not particularly linked to Transect 1 of the Intermediate region or Transect 1 of the coastal region. If it were a crossed design, this would not be so.

This example could be made into a crossed design if there were some attribute that linked Transect 1 across the levels of the factor Location, and similarly for Transects 2 and 3. For example, all the transects labeled 1 might be from the uplands, all of the transects labeled 3 might be from the lowlands, and all of the transects labeled 2 might be from intermediate topographic positions. Then the analysis would be crossed. No longer could you sensibly swap transect 1 (upland) and 2 (intermediate topography) in the semi-arid column without corrupting the analysis.

It is important before contemplating a factorial ANOVA that you check that all factors in the design are indeed crossed.

Summary of the Factorial ANOVA Procedure

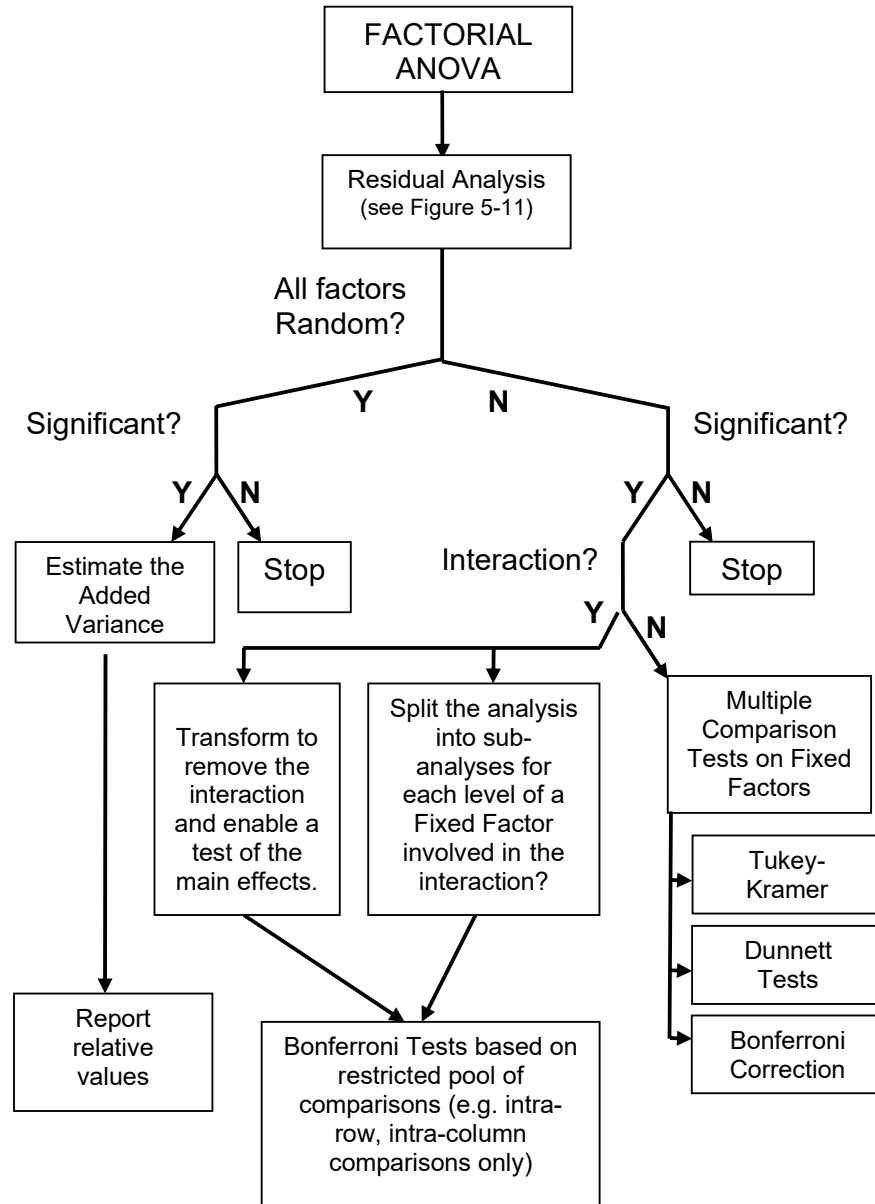
Factorial analysis of variance is an important analysis. It carries us beyond the realm of single-factor ANOVA, which was seen merely as an extension of the t-test to cater for comparisons among more than two means. It allows us to investigate simultaneously the independent effects of two factors as well as the interaction between

them. The notion of interaction is the special contribution factorial ANOVA makes to our understanding of variation in natural systems.

The general procedure for undertaking a study involving factorial ANOVA is summarised as follows:

- Decide the questions you wish to address, involving the joint action of two or more factors.
- Carefully select your factors and factor classes so that differences among the factor classes or interaction between the factors will unambiguously address the research questions.
- Ensure that the design is fully crossed. Where possible, replicate the measurements for each combination of the factor classes. A balanced design (sample sizes equal) is desirable.
- Design your experiment and sampling protocols to ensure that the entities to be measured are either randomly allocated to each combination of factor classes, or if class membership is beyond your control, to ensure that the entities are selected at random from the populations represented by each cell in the factorial classification. Ensure independence of the entities selected within each cell.
- If the model is fixed, or important hypotheses relate to a fixed factor in a mixed design, plan your comparisons in advance of beginning the experiment if at all possible.
- Collect the data.
- Undertake an exploratory analysis, based on graphical techniques, preliminary runs of the ANOVA and examination of residuals to verify that the assumptions of ANOVA are tenable. Include plots of the treatment means. Transform the data where necessary.
- Perform the final ANOVA, and follow by multiple comparison tests if the model is fixed and no strong interaction is present. If there is a strong interaction, and the main effects are important to the research questions, split the analysis before undertaking multiple comparison procedures. Estimate the added variance components for random factors.
- Interpret significant results in the context of the initial question that you wished to address.
-

Figure 5–12.
A Decision Tree
for factorial
ANOVA.



Where have we come?

Key concepts with which you need to be broadly familiar include:

- The partition of the sums of squares and degrees of freedom for the factorial ANOVA.
- The difference between fixed, random and mixed models in factorial ANOVA, and the practical consequences of these differences for the computation of appropriate F ratios and for follow-up analysis.

- The issues central to choosing an appropriate multiple comparison procedure for fixed main effects, and a sensible position on a workable set of procedures to cover common circumstances.
- The concept of interaction and its interpretation in a biological context.
- The assumptions of factorial ANOVA, how to detect violations and how to overcome them, with emphasis on displaying and interpreting residuals.

It is now appropriate to put this knowledge to use and to reinforce your understanding of the key concepts via worked examples and exercises.

Lesson 6: Step-through Examples

Example 5-1: Bait Acceptability by Feral Pigs

This is a fixed model, two-way factorial analysis of variance with no interaction.

The problem

Feral pigs are not native to Australia. They were introduced as domestic pigs at the time of European settlement. Unfortunately many of these pigs were allowed to roam free. During the 19th Century, many domestic pigs were transported to settlements across Australia and owing either to insufficient enclosures or to deliberate releases, it was not long before feral populations of pigs became established.

Poisoning with sodium monofluoroacetate (1080) is a common method of feral pig control. Field observations suggest, however, that pigs become bait-shy to 1080. Hone and Kelba (1984) wished to investigate this possibility using penned pigs and a rigorous experimental design.

Ten pens were constructed. Feral pigs were captured using baited pig traps, and from them 10 male and 10 female pigs were selected to be of approximately the same age and weight. Two male pigs were allocated randomly to each of five pens and two female pigs were allocated randomly to each of the remaining five pens.

On Day 1, pigs were offered wheat only and their intake was recorded. On Day 2, they were offered one of the following combinations:

- Wheat only
- Wheat and water
- Wheat, water and dye
- Wheat, water and 1080
- Wheat, water, dye and 1080

The dye is necessary for safety reasons. Water is necessary to add the dye and 1080.

Intake of the bait was again recorded. Intake on Day 1 was subtracted from intake on Day 2, and the response variable was change in intake in kg (Table 5-14).

The central problem Hone and Kleba wished to address was to determine if there was a significant difference in the response of pigs to the different bait mixtures.

The data

Table 5-14. Change in bait intake by feral pigs in response to different bait constituents. Pigs are classified according to two criteria – bait treatment and sex.

	Wheat	Wheat & water	Wheat, water & dye	Wheat, water & 1080	Wheat, water, dye & 1080
Male	0.188 -0.058	0.050 -0.138	0.058 -0.082	-0.712 -1.280	-0.610 -0.830
Female	-0.280 -0.062	-0.540 -0.336	-0.260 -0.123	-0.894 -0.672	-0.837 -1.202

Comments on the design

The response variable is change in grain intake in kg, which will be potentially influenced by the composition of the bait and by the sex of the pig. Each measurement of the response variable is classified according to two factors – bait and sex. The design is fully crossed and balanced.

A factorial analysis will be able to address the following questions:

- Is there a significant difference in change in grain intake among the different bait treatments?
- Is there a significant difference in grain intake between males and females?
- If there is a difference in response of feral pigs to bait composition, do the trends vary in magnitude or direction between males and females?

The first question is the one of greatest interest.

The analysis

Start a R Session



Double click on the Tinn-R icon and launch R from within Tinn-R (Click in the Menu on R->Initiate/Close Rgui->Initiate preferred Rgui)

Data entry and exploratory examination

R expects the data for a two-way classification in the form of three variables. Two variables contain the factor class data, and the remaining variable contains the values of the response variable.

```

bait    sex    intake
W       MALE   0.188
W       MALE   -0.058
W       FEMALE -0.280
W       FEMALE -0.062
WW      MALE   0.050
WW      MALE   -0.138
WW      FEMALE -0.540
WW      FEMALE -0.336
WWD     MALE   0.058
WWD     MALE   -0.082
WWD     FEMALE -0.260
WWD     FEMALE -0.123
WW1080 MALE   -0.712
WW1080 MALE   -1.280
WW1080 FEMALE -0.894
WW1080 FEMALE -0.672
WWD1080 MALE   -0.610
WWD1080 MALE   -0.830
WWD1080 FEMALE -0.837
WWD1080 FEMALE -1.202

```

Read the data into R, creating a new object `pigs`. We quickly peruse it and compute some basic statistics.

```

> setwd("H:\\Biometry\\data\\")
> pigs <- read.table("PIGBAIT.DAT", header=TRUE)
> summary(pigs)

```

```

      bait      sex      intake
W       :4  FEMALE:10  Min.    :-1.2800
WW      :4   MALE :10  1st Qu.: -0.7415
WW1080 :4                      Median :-0.3080
WWD     :4                      Mean   :-0.4310
WWD1080:4                      3rd Qu.: -0.0770
                          Max.    : 0.1880

```

```

> dim(pigs)

```

```

[1] 20 3

```

```

> names(pigs)

```

```

[1] "bait" "sex" "intake"

```



Submit the above program for execution.

The `pigs` data.frame should contain 3 variables – `bait`, `sex` and `intake` – and 20 values (4 values for each level of `bait`). You can further check the data at this point to see if it has been read as intended.

Next we should construct a table of cell and marginal means. The easiest way to achieve that is to use the `tapply()` function. The

`tapply` function has three arguments, the first is the vector (or column of the data frame) on which a function should be performed. The second is the factor (groups) which subsets the first vector and the third argument is the function (here we use the predefined function `mean`) that should be performed on each group of the data.

```
> tapply(pigs$intake, pigs$bait, mean)

      W      WW  WW1080      WWD  WWD1080
-0.05300 -0.24100 -0.88950 -0.10175 -0.86975
```

This calculates all means for all levels of `bait`. So if we want to check how many cases are there in each level of `bait` we could have typed:

```
> tapply(pigs$intake, pigs$bait, length)

      W      WW  WW1080      WWD  WWD1080
      4      4      4      4      4
```

Which tells us that there are four pigs and its intake measured in each level.

The mean for each `sex` is straight forward as well. First we check how many cases there are in each level.

```
> tapply(pigs$intake, pigs$sex, length)
```

```
FEMALE  MALE
      10    10
```

```
> tapply(pigs$intake, pigs$sex, mean)
```

```
FEMALE  MALE
-0.5206 -0.3414
```

And finally we want to calculate the mean intake for both factors at the same time. Here we use the `tapply()` function, specifying both levels inside a list function in the second argument of `tapply()`

```
> tapply(pigs$intake, list(pigs$sex, pigs$bait), mean)
```

```
      W      WW  WW1080      WWD  WWD1080
FEMALE -0.171 -0.438 -0.783 -0.1915 -1.0195
MALE    0.065 -0.044 -0.996 -0.0120 -0.7200
```



Extra task

Can you calculate how many cases there are within both levels, i.e. how many real repetition there are?

We need the overall mean intake for all pigs regardless of bait and sex. This is easy:

```
> mean (pigs$intake)

[1] -0.431
```

Finally we copy past all means into a nice, shiny table (Table 5-15).

*Table 5-15.
Cell and
marginal means
for a two-way
classification of
response of
male and female
pigs to different
bait constituents.*

	Wheat	Wheat & water	Wheat, water & 1080	Wheat, water & dye	Wheat, water, dye & 1080	
FEMALE	-0.171	-0.438	-0.783	-0.1915	-1.0195	-0.5206
MALE	0.065	-0.044	-0.996	-0.012	-0.72	-0.3424
	-0.053	-0.241	-0.8895	-0.10175	-0.86975	-0.431

There is some suggestion that males and females differ in their overall response to the experiment, with males reducing their intake by less than females, but is this result significant? Similarly, the data suggest an adverse response to baits containing 1080, but is this a significant result?

Analysis of Residuals

Before we launch into the factorial ANOVA, it is wise to pay attention to the assumptions. A residual analysis will allow us to assess the assumptions of homogeneity of variances and normality.

In principle we can run our ANOVA using the `aov()` function.

```
> anova.pigs <- aov(intake ~ bait * sex, data=pigs)
```

Note that there are now two factors specified, and that the “model” specified in the `aov()` function specifies that variation in the response variable `intake` is to be explained by two main effects and the interaction. In all other respects the approach is similar to that used in single-factor ANOVA. You should read the specification in the `aov()` function as follow:

We want to explain the variation in `intake` by the factors `bait` and `sex`. To specify that we want to have the interaction between `bait` and `sex` taken into account aswell (as we have no reason why we should not) we specify our ANOVA (model) by the “*” sign between the factors. If we want to specify an ANOVA where no interaction

between `bait` and `sex` should be used, we would specify our model by:

```
intake ~ bait + sex
```

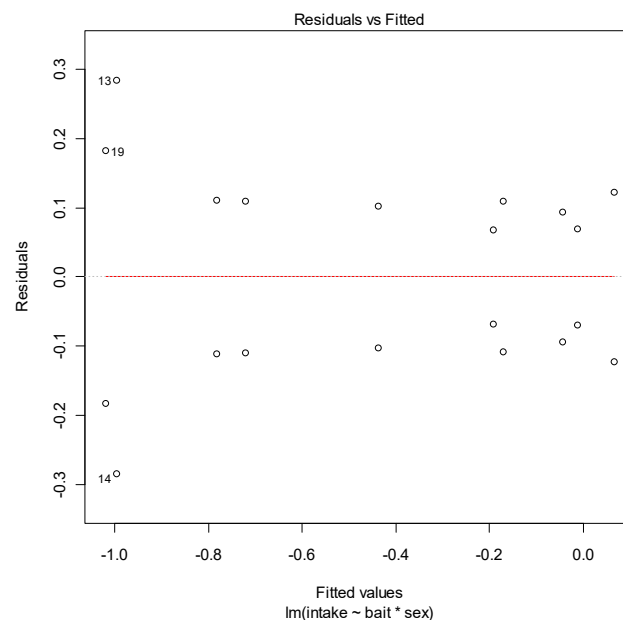
Before we have a look at our ANOVA table let us check the residuals as we have done with a single factor ANOVA.

```
> hist(resid(anova.pigs), col="darkolivegreen")
> plot(anova.pigs) #do not forget to select History ->
  recording to record your plots
```

Submit the above two programs for execution.

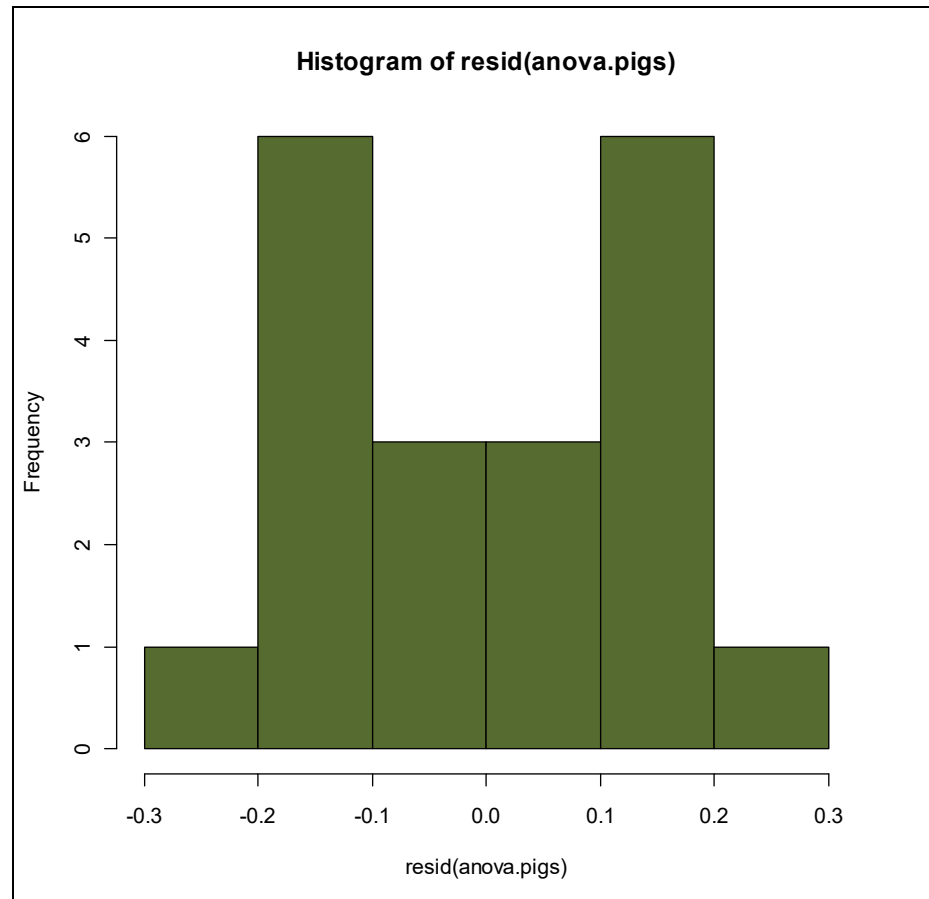
The distribution of residuals is acceptable, despite wider than average variances for the two samples with the largest response (Figure 5-13).

*Figure 5-13
Residual plot for
change in grain
intake by feral
pigs versus bait
type and sex in a
factorial design.*



A histogram plot of the residuals shows no substantive deviation from normality (though admittedly it does not look great).

Box 5-2.
 Histogramm for
 the residuals
 from a factorial
 ANOVA of
 change in bait
 intake versus
 bait type and sex
 in feral pigs.



Submit the above program for execution.

Analysis of Variance

There is no need for transformation, as the assumptions of homogeneity of variances and normality have been upheld.

Overall test for differences

Before we have a look at the ANOVA table we would like to know if there is an overall difference between the means (disregarding the different factors). If this would not be the case then the ANOVA is quite pointless as at least some means between any combination of factors should be different.

For this we use a new function called `lm()` [for linear model]. This function will be our new working horse in the future and therefore it is a good idea to get used to it. In principle you can forget the `aov()` function as the `lm()` function does provide the same and some additional features. Fortunately the code is the same, just exchange `aov` by `lm`:


```

> lm.pigs <- lm(intake ~ bait * sex, data=pigs)
> summary(lm.pigs)

Call:
lm(formula = intake ~ bait * sex, data = pigs)

Residuals:
    Min       1Q   Median       3Q      Max
-2.840e-01 -1.092e-01 -2.082e-17  1.092e-01  2.840e-01

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.1710     0.1394  -1.227  0.24793
baitWW         -0.2670     0.1971  -1.355  0.20533
baitWW1080    -0.6120     0.1971  -3.105  0.01115 *
baitWWD       -0.0205     0.1971  -0.104  0.91922
baitWWD1080  -0.8485     0.1971  -4.305  0.00155 **
sexMALE        0.2360     0.1971   1.197  0.25876
baitWW:sexMALE  0.1580     0.2787   0.567  0.58331
baitWW1080:sexMALE -0.4490     0.2787  -1.611  0.13828
baitWWD:sexMALE -0.0565     0.2787  -0.203  0.84343
baitWWD1080:sexMALE  0.0635     0.2787   0.228  0.82438
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1971 on 10 degrees of freedom
Multiple R-squared:  0.8899,    Adjusted R-squared:  0.7907
F-statistic: 8.978 on 9 and 10 DF,  p-value: 0.00099

```

At this stage we are not interested in the estimates of the parameters of the model, but at the end of this output we can see the overall fit of the ANOVA model. The output contains an overall test for differences among cell means ($F = 8.98$; $df = 9, 10$; $p < 0.001$). Overall the differences are highly significant, therefore finally, let us have a look at our ANOVA table.

```
> summary(anova.pigs)
```

To confuse and demonstrate you that the `lm()` function is equivalent to the `aov()` function, just type in:

```
> anova(lm.pigs)
```

This should show you the same ANOVA table as the summary command on the `aov()` object.



Submit the above program for execution.

The results of the analysis are shown in Box 5-3.

*Box 5-3.
Results of a
factorial ANOVA
testing the
effects of bait
type and sex on
the intake of bait
by feral pigs.*

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
bait	4	2.76045	0.69011	17.7658	0.0001538 ***
sex	1	0.16056	0.16056	4.1334	0.0694460 .
bait:sex	4	0.21766	0.05441	1.4008	0.3022677
Residuals	10	0.38845	0.03885		

We can now construct our standard ANOVA Table (Table 5-16) using the data from the R output. Of the significant variation among the cell means, a significant component can be attributed to differences in the response of the feral pigs to bait type ($F = 17.77$; $df = 4,10$; $p < 0.0002$).

Differences between the sexes were not significant ($F = 4.13$; $df = 1,10$; $p = 0.069$). There was no significant interaction between sex and bait type ($F = 1.40$; $df = 4,10$; $p = 0.30$), that is, the differential response of the feral pigs to bait type was similar for males and females.

Table 5-16. Factorial ANOVA to determine the effects of bait constituents and sex on intake of bait by feral pigs. (see text).

Source	DF	SS	MS	F	Pr > F
Among cells	9	3.1387	0.3487	8.98	0.0010
A B	4	2.7604	0.6901	17.77	0.0002
B A	1	0.1606	0.1606	4.13	0.0694
AxB	4	0.2177	0.05441	1.40	0.3023
WITHIN	10	0.3885	0.03885		
TOTAL	19	3.5271			

FACTOR A: BAIT FACTOR B: SEX

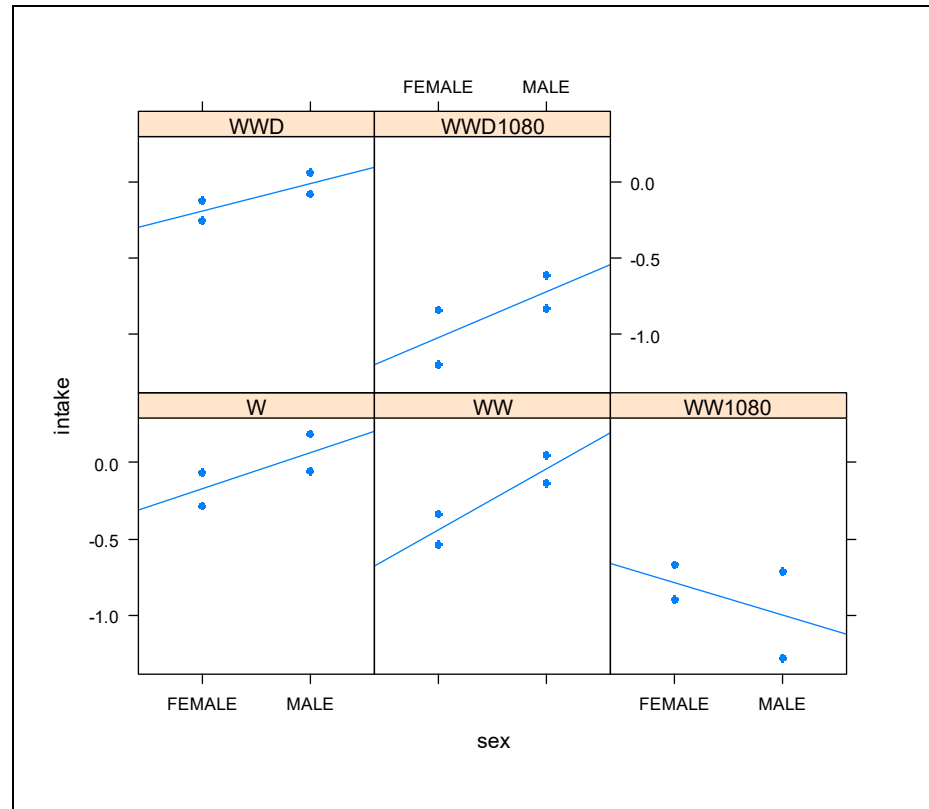
As you know R has great graphic capabilities and it would be great to have a plot on the whole dataset, but a separated plot for each combination of factors. This requires an additional package named: lattice (see the moodle site how to install a package if you are not working with you own computer). So we type:

```
> library(lattice)
> install.packages("lattice", lib="H:\\Biometry")
#at home just type install.packages("lattice")
> xyplot(intake ~ sex | bait, data=pigs, ,
type=c("p", "r"), pch=16)
```

Box 5-4 shows the result of these commands. Here you can see the intake for each combination of factors. First we notice that there are two points for each combination, which are our repetitions. We have five levels in the factor bait and two in the factor sex. Graphically we see an increase in intake from females to males and also differences between bait treatments. Treatments that have 1080 showing a reduced intake. A slight trend of an interaction can be seen, so males do eat always more than females except for the WW1080 treatment.

From the ANOVA table we see that this effect is not significant. Once we have an idea of the overall effects of the factors, we can now have a look, if all bait levels are significantly different or not. So we need to do multiple comparisons, equivalent to single factor ANOVAs (R Module 4).

Box 5-4.
A plot of intake
for each
combination of
factors (bait &
sex)



Multiple Comparisons

We now need to determine which bait combinations are significantly different from which others.

We have different options:

- Tukey-Kramer Procedure (Honestly Significant Difference)

```
> TukeyHSD(anova.pigs)
```

- Dunnett's Procedure (single control)

```
> require(multcomp)
> dunnett <- glht(anova.pigs, linfct=
mcp(bait=contrMat(c(4,4,4,4,4), type="Dunnett", base=2) ,
interaction_average=T)
> summary(dunnett)
```

■ Bonferroni Correction

```
> t.test (and use bonferroni correction for p value)
```

Two approaches suggest themselves. The first, given the complex relationship of the factor levels, with at least three levels that can be regarded as experimental controls for other levels, is to conduct exhaustive post hoc comparisons using the Tukey-Kramer Procedure (refer to Module 4). Here we need to be careful that we do not want to test for all interactions, therefore we



Submit the above statements for execution.

Box 5-4. Results of Tukey-Kramer comparisons among mean response of feral pigs to various bait types produced by an LSMEANS statement in PROC GLM.

```
Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = intake ~ bait * sex, data = pigs)

$bait
      diff          lwr          upr      p adj
WW-W      -0.18800 -0.6466610  0.2706610  0.6699170
WW1080-W  -0.83650 -1.2951610 -0.3778390  0.0009644
WWD-W      -0.04875 -0.5074110  0.4099110  0.9962384
WWD1080-W -0.81675 -1.2754110 -0.3580890  0.0011619
WW1080-WW  -0.64850 -1.1071610 -0.1898390  0.0062714
WWD-WW      0.13925 -0.3194110  0.5979110  0.8500012
WWD1080-WW -0.62875 -1.0874110 -0.1700890  0.0077296
WWD-WW1080  0.78775  0.3290890  1.2464110  0.0015342
WWD1080-WW1080 0.01975 -0.4389110  0.4784110  0.9998918
WWD1080-WWD -0.76800 -1.2266610 -0.3093390  0.0018596

$sex
      diff          lwr          upr      p adj
MALE-FEMALE 0.1792 -0.01719271  0.3755927  0.069446
```

We can peruse the output matrix (Box 5-4) for non-significant and significant pairwise comparisons to yield non-significant subsets. Clearly, bait combinations including 1080 are showing significantly greater reduction in intake by feral pigs. Adding water or dye or both had no significant impact on intake.

Had we wished to use the potentially more powerful approach based on specifying in advance of the study the comparisons we wished to make (a priori comparisons), then Dunnett's Procedure would have been appropriate. Say we decided in advance that we wished to compare mean response for each bait type against the control of Wheat and Water (WW), we would use the following statement to produce Box 5-5.

```
> require(multcomp) #should work now, if you have already
installed the package multcomp, otherwise refer to module 4
> dunnnett <- glht(anova.pigs, linfct=
mcp(bait=contrMat(c(4,4,4,4,4), type="Dunnnett", base=2 ) ,
interaction_average=T))
> summary(dunnnett)
```



Submit the above statements for execution.

The results are essentially the same as for the Tukey-Kramer Procedure (Box 5-5). Adding water or dye to the wheat has no impact on bait intake; both treatments including 1080 differ significantly from the control.

Box 5-5. Results of Dunnett's comparisons among mean response of feral pigs to various bait types produced by an LSMEANS statement in PROC GLM. The experimental control is wheat and water (WW).

```
Simultaneous Tests for General Linear Hypotheses
Multiple Comparisons of Means: User-defined Contrasts

Fit: aov(formula = intake ~ bait * sex, data = pigs)

Linear Hypotheses:
      Estimate Std. Error t value Pr(>|t|)
1 - 2 == 0    0.1880    0.1394   1.349  0.49977
3 - 2 == 0   -0.6485    0.1394  -4.653  0.00297 **
4 - 2 == 0    0.1393    0.1394   0.999  0.71927
5 - 2 == 0   -0.6287    0.1394  -4.512  0.00386 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

Results summary

The following is an example of an appropriate results summary, as would appear in your report or publication.

Mean response of feral pigs to the various combinations of bait components and sex in a two-way factorial ANOVA differed significantly ($F = 8.98$; $df = 9,10$; $p < 0.001$). The percentage variation that could be attributed to the effect of the two factors BAIT and SEX, individually and in combination, was high ($\text{adj. } R^2 = 0.79$).

Of the significant variation among the cell means, a significant component could be attributed to differences in the response of the feral pigs to bait type ($F = 17.77$; $df = 4,10$; $p < 0.0002$). The mean response varied from a reduction in intake of only 53 g for unadulterated wheat to a reduction of 870-890 g for treatments involving 1080.

Differences between the sexes were not significant ($F = 4.13$; $df = 1,10$; $p = 0.069$). There was no significant interaction between sex and bait type ($F = 1.40$; $df = 4,10$; $p = 0.30$), that is, the differential response of the feral pigs to bait type was similar for males and females.

Tukey-Kramer pairwise comparisons among the treatment means for bait type (Table 5-14) revealed no significant difference between the treatments of Wheat, Wheat & Water, and Wheat, Water & Dye ($p = 0.67-1.00$), but pigs showed significantly lower response to these treatments than to those involving the addition of 1080 ($p < 0.01$). Addition of dye with 1080 did not alter the high response.

Note that the results as described include a statement of the overall significance of the ANOVA as well as a statement of the significance or lack of significance for each of the main effects and the interaction. The strength of the result is indicated by the overall R^2 and by the magnitude and direction of the difference between specific means of interest. Both significance and strength of result are covered. The ANOVA table is rarely published, but the results of the comparisons between means for each bait type (Table 5-17) can be optionally included.

Discussion

It appears that feral pigs can distinguish grain that is laced with 1080, and that neither the mandatory addition of water to the grain nor the optional addition of dye is influencing their response.

Nearly all the pigs used in this study were from a property where 1080 had been used for feral pig and rabbit control, and they may have learned to recognise 1080 on the basis of bad experiences with its ingestion in the past.

This behaviour is likely to provide a partial explanation for reduced efficacy of 1080 for feral pig control, and the response of feral pigs to alternative poisons such as Warfarin or brodifacoum should be explored in future studies.



Tidy up the program by ensuring there are no elements remaining of the program that did not work. Save the program to disk within Tinn-R for future reference. You may also want to save your workspace and Rhistory, if you want to return to this example at some stage.

Exit from R and Tinn-R by choosing File->Exit from the Menu Bar.

Example 5-2: Soil phosphorus, topography and soil type

This is a fixed model, two-way factorial analysis of variance with interaction.

The problem

Soil phosphorus is important for the invasion of native vegetation by exotic weeds. Anne-Marie Clements undertook a study of soil phosphorus in the Sydney region to investigate how soil phosphorus varies with topographic location and soil type (Clements, 1983).

Bushland sites were chosen in Brisbane Waters National Park, Kuring-gai Chase National Park and Royal National Park, which are relatively unaffected by suburban development. The sites were free from immediate roadside or track effects and had not been burned for at least two years.

Shale-derived and sandstone-derived soils in four topographic locations were examined, with three 250 m² quadrats in each of the eight combinations of soil type and topography.

Cores of soil of 75 mm depth and 25 m diameter, free from surface litter, were collected from each of five randomly selected points in each quadrat. The five soil samples were pooled and the total soil phosphorus (ppm) was determined for each pooled sample.

Anne-Marie was interested to learn:

- if there were significant differences in soil phosphorus among the four topographic locations – valleys, north-facing slopes, south-facing slopes and hilltops;
- if there were significant differences in soil phosphorus between shale-derived and sandstone-derived soils; and
- if any difference between topographic locations in soil phosphorus depended in magnitude on whether the soils were shale-derived or sandstone-derived.

The data

Table 5-18. Soil phosphorus (ppm) collected from the vicinity of Sydney according to a two-way factorial design. The factors are topography and soil type.

	VALLEY	NORTH	SOUTH	HILLTOP
SHALE	98	78	117	83
	172	77	54	12
	185	100	96	14
SANDSTONE	19	27	28	55
	39	49	53	21
	25	24	72	19

Comments on the design

The response variable is soil phosphorus (`phos`), which will be potentially influenced by topographic location (`topo`) and soil type (`soil`).

The factor `topo` has four factor levels – valleys (VALLEY), north-facing slopes (NORTH), south facing slopes (SOUTH) and hilltops (HILLTOP). The factor `soil` has two factor levels – shale-derived soils (SHALE) and sandstone-derived soils (SANDSTONE). The design is fully crossed and balanced ($n = 3$).

All hypotheses relating to the main effects and interaction are of interest.

The analysis

Start a R Session



Double click on the Tinn-R icon and launch R from within Tinn-R (Click in the Menu on R->Initiate/Close Rgui->Initiate preferred Rgui)

Data entry and exploratory examination

R expects the data for a two-way classification in the form of three variables. Two variables contain the factor class data, and the remaining variable contains the values of the response variable.

We start by reading in the data from SYDNEY.DAT and store it as `data.frame` in an R-object called `phosphor`.

```
> setwd("H:\\Biometry\\data")
> sydney <- read.table("SYDNEY.DAT", header=T)
```



Submit the above commands for execution.

We check if the data were read in correctly.

```
> dim(sydney)
[1] 24 3
> summary(sydney)
      soil      topo      phosphor
SANDSTONE:12 HILLTOP:6  Min.    : 12.00
SHALE       :12      NORTH  :6    1st Qu.: 24.75
```



```

SOUTH :6 Median : 53.50
VALLEY :6 Mean : 63.21
          3rd Qu.: 86.25
          Max. :185.00

```

Next we should construct a table of cell and marginal means. As in the first example we need to do that by “hand”, as there is no function available.

```

> mean(sydney$phosphor)

[1] 63.20833

> tapply(sydney$phosphor, sydney$topo, mean)

HILLTOP NORTH SOUTH VALLEY
34.00000 59.16667 70.00000 89.66667

> tapply(sydney$phosphor, sydney$soil, mean)

SANDSTONE SHALE
35.91667 90.50000

> tapply(sydney$phosphor, list(sydney$soil, sydney$topo),
mean)

          HILLTOP NORTH SOUTH VALLEY
SANDSTONE 31.66667 33.33333 51 27.66667
SHALE      36.33333 85.00000 89 151.66667

```



Submit the above program for execution.

The collected output is shown, rearranged to conform to a more conventional form, in Table 5-19. There look to be major differences in soil phosphorus with both topographic location and soil type. Shale-derived soils appear to have more phosphorus than sandstone-derived soils, but is this result significant? Similarly, the data suggest differences in soil phosphorus with topographic location, but is this a significant result? Furthermore, the trends across topographic location look to be different for the two soil types, so we might anticipate a significant interaction.

Table 5-19.
Cell and marginal means of soil phosphorus for a two-way classification against topographic location and soil type.

	HILLTOP	NORTH	SOUTH	VALLEY	
SANDSTONE	31.67	33.33	51	27.67	35.92
SHALE	36.33	85	89	151.67	90.5
	34.0	59.17	70.0	89.67	63.21

Analysis of Residuals

Before we launch into the factorial ANOVA, we need to attend to the assumptions. A residual analysis will allow us to assess the assumptions of homogeneity of variances and normality. This time we use the `lm()` function all the time.

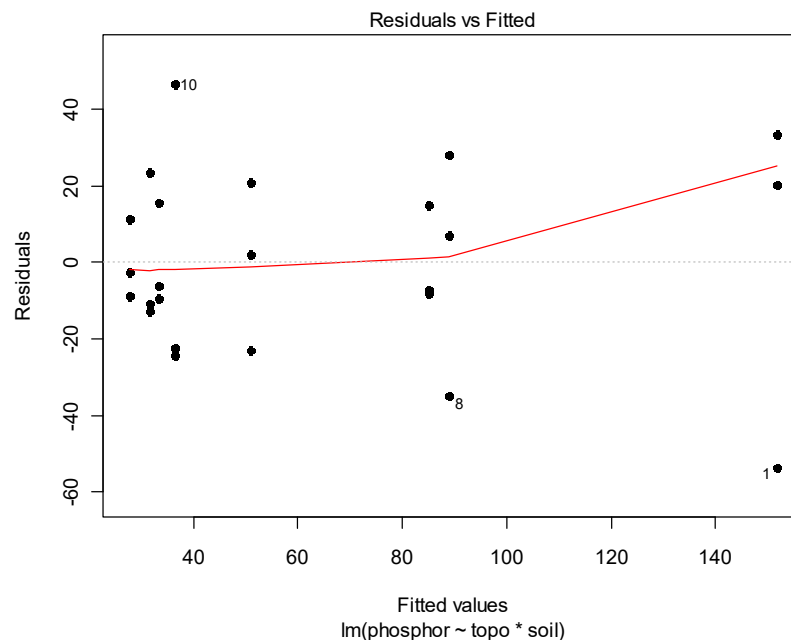
```
> lm.sydney <- lm(phosphor ~ topo * soil, data=sydney)
> plot(lm.sydney, pch=16)
> hist(resid(lm.sydney), col="gray58")
```



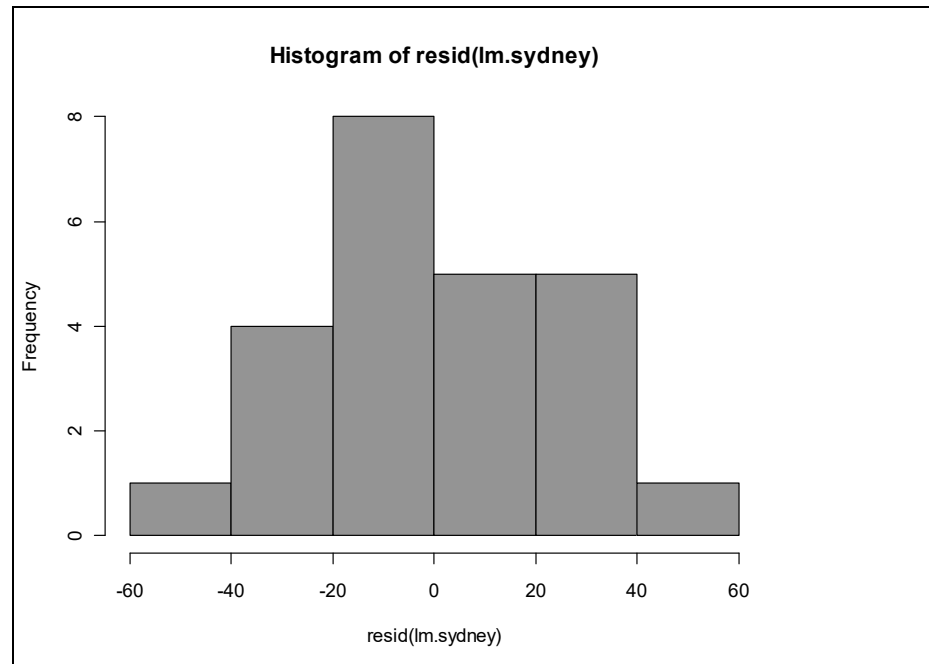
Submit the above two programs for execution.

The distribution of residuals is acceptable (Figure 5-14).

Figure 5-14.
Residual plot for
soil phosphorus
versus soil type
and topographic
location in a
factorial design.



Box 5-7.
 Histogram of residuals for soil phosphorus versus soil type and topographic location in a factorial design..



Submit the above program for execution.

Analysis of variance

There is no need for transformation, as the assumptions of homogeneity of variances and normality have been upheld. So we finally can have a look at our ANOVA table.

```
> summary(lm.sydney)
> anova(lm.sydney)
```



Submit the above program for execution.

The results of the analysis are shown in Box 5-8.

Box 5-8.
Results of a
factorial ANOVA
testing the
effects of soil
type and
topographic
location on soil
phosphorus.

```
> summary(lm.sydney)

Call:
lm(formula = phosphor ~ topo * soil, data = sydney)

Residuals:
    Min       1Q   Median       3Q      Max
-53.67 -11.17  -4.50   16.83   46.67

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    31.667     16.102   1.967  0.06682 .
topoNORTH       1.667     22.772   0.073  0.94256
topoSOUTH      19.333     22.772   0.849  0.40840
topoVALLEY     -4.000     22.772  -0.176  0.86277
soilSHALE       4.667     22.772   0.205  0.84021
topoNORTH:soilSHALE  47.000     32.204   1.459  0.16380
topoSOUTH:soilSHALE  33.333     32.204   1.035  0.31603
topoVALLEY:soilSHALE 119.333     32.204   3.706  0.00192 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 27.89 on 16 degrees of freedom
Multiple R-squared:  0.7579,    Adjusted R-squared:  0.652
F-statistic: 7.156 on 7 and 16 DF,  p-value: 0.0005729

Analysis of Variance Table

Response: phosphor
      Df Sum Sq Mean Sq F value    Pr(>F)
topo   3  9693.8   3231.3   4.1542 0.0235128 *
soil   1 17876.0  17876.0  22.9818 0.0001988 ***
topo:soil  3 11390.8   3796.9   4.8814 0.0134826 *
Residuals 16 12445.3    777.8
```

The differences among cell means are significant ($F = 7.16$; $df = 7,16$; $p < 0.001$) and the percentage variation that can be attributed to the effect of the two factors TOPO and SOIL, individually and in combination is high ($R^2 = 0.76$).

Table 5-20. Factorial ANOVA to determine the effects of topographic location and soil type on soil phosphorus. Note that in the presence of a significant interaction, the significant tests for the main effects are unreliable.

Source	DF	SS	MS	F	Pr > F
Among cells	7	38960.6	5565.8	7.16	0.0006
A B	3	9693.8	3231.3	4.15	0.0235
B A	1	17876.0	17876.0	22.98	0.0002
AxB	3	11390.8	3796.9	4.88	0.0135
WITHIN	16	12445.3	777.8		
TOTAL	23	51406.0			

FACTOR A: TOPO FACTOR B: SOIL

Of the significant variation among the cell means, a significant component can be attributed to differences in the topographic location ($F = 4.15$; $df = 3,16$; $p < 0.05$; Table 5-20). **However, in the presence of a significant interaction ($F = 4.88$; $df = 3,16$; $p < 0.02$), this result is not reliable.** For example, the differences in soil phosphorus among topographic locations reflected in this result may occur only for one soil type. Similarly, we cannot

believe the significant result for the other main effect, soil type, in the presence of the interaction.

Explore the interaction

We must first explore the nature of the interaction. The best way to do this is to plot the treatment means. Again we need to load the package lattice (you should have already installed on your computer, so a simple library command should do the trick, if not refer to the moodle site, how to install a package).

```
> library(lattice)
> xyplot(phosphor ~ soil | topo, data=sydney,
type=c("p","r"), pch=16)
```

This is clearly the sort of program you file away for future reference. The first two procedures average the values for each cell in the two-way classification (we need to sort the data first) and store it in a new work file WORK.SUMMARY.

The GPLOT procedure plots the cell means in a plot with the response variable PHOS on the Y-axis and one of the factors on the X-axis. It is conventional to choose the factor with the most factor levels to assign to the X-axis. The levels of the second factor are distinguished with different symbols, lines or colours.

This is achieved with the statement

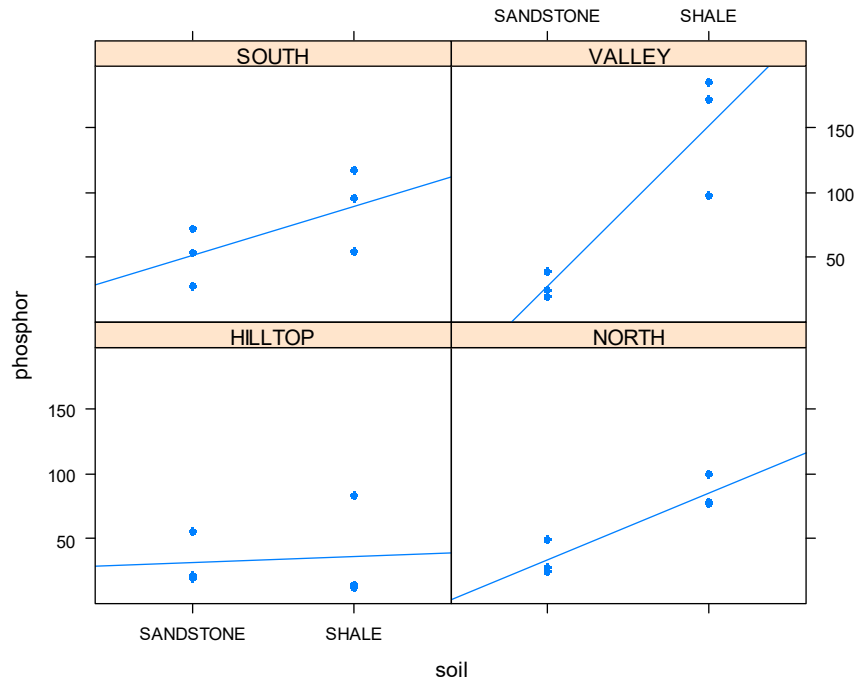
```
      PLOT MNPPOS*TOPO=SOIL
```

which asks for a plot of mean soil phosphorus against topographic location, with separate graphs for each value of soil type (Figure 5-15).



Submit the above program for execution.

Figure 5-15.
Plot of treatment means for a factorial ANOVA of soil phosphorus against topographic location and soil type. Clearly topographic location has a greater effect on soil phosphorus in shale-derived soils than in sandstone-derived soils – an interaction.



The nature of the interaction is now clearly evident. Let us view it first from the perspective of topographic location.

Topographic location appears to have a marked affect on soil phosphorus in shale-derived soils, but its effect in sandstone-derived soils is not strong if present at all. We have an interaction because the magnitude of the effect of topography depends on which soil type you consider. We cannot make a general statement about the effect of topographic location without qualifying the statement for soil type.

The interaction can also be viewed from the perspective of soil type. There is clearly no difference between shale-derived and sandstone-derived soils when samples are taken from hilltops. In the valleys however, shale-derived soils have much higher levels of soil phosphorus than sandstone-derived soils. We have an interaction because the magnitude of the effect of soil type depends upon topographic location. We cannot make a general statement about the effect of soil type without qualifying the statement for topographic location.

Splitting the analysis

The significance of the main effects for topographic location is based on averaging soil phosphorus across soil types, and so is compromised by the interaction. So too is the significance of soil type.

So how then do we assess the significance of the main effects?

We first conduct a separate sub-analysis of the effect of topographic location for each soil type. We split the data set using the subset command.

```
> s.shale <- subset(sydney, soil=="SHALE")
> s.sandstone <- subset(sydney, soil=="SANDSTONE")
```

Then we perform an ANOVA on the subsets.

```
> lm.shale <- lm(phosphor ~ topo, data=s.shale)
> anova(lm.shale)
```

Analysis of Variance Table

```
Response: phosphor
      Df Sum Sq Mean Sq F value Pr(>F)
topo   3  20124   6707.9   5.3294 0.02605 *
Residuals 8  10069   1258.7
```

```
> lm.sandstone <- lm(phosphor ~ topo, data=s.sandstone)
> anova(lm.sandstone)
```

Analysis of Variance Table

```
Response: phosphor
      Df Sum Sq Mean Sq F value Pr(>F)
topo   3   960.92   320.31   1.0785 0.4115
Residuals 8 2376.00   297.00
```



Submit the above program for execution.

On sandstone-derived soils, there is no significant effect of topographic position on soil phosphorus ($F = 1.07$; $df = 3,8$; $p = 0.41$), whereas a effect was evident for shale-derived soils ($F = 5.33$; $df = 3,8$; $p < 0.05$) (Figure 5-15).

Multiple comparisons

We are now in a position to undertake pairwise comparisons, with corrections for compounding error, across the levels of topographic position in shale-derived soils and across soil types in each topographic position.

The pool of potential comparisons in a 4x2 classification is 16, if we restrict our attention to comparisons within rows and columns. This

comprises 6 potential comparisons in each of the 2 rows and 1 comparison per each of the 4 columns.

Pairwise comparisons will be significant under the Bonferroni Method (see Module 4).

$$p \leq \frac{0.05}{16} = 0.003$$

Thus, if we apply a significance level of $\alpha = 0.003$ to individual comparisons, we will control the probability of getting a spurious significant result in any of the comparisons to 0.05.

To do a test for only a subsample we could select all the 16 necessary pairs of test data and then do a simple t.test. If the p-value is below 0.003 then we have a significant difference. A much easier way is to do an exhaustive comparison between all levels (as in a Tukey test), but look only at the selected 16 tests, if they are below 0.003.

Unfortunately the `TukeyHSD()` function does not allow to set for an uncorrected p-value, therefore we need to use the function, `pairwise.t.test()`. Before we can use the function we need to combine both factors into a single one, as this function allows only one factor as an argument.

```
> onefactor <- as.factor(paste(sydney$topo,
                               sydney$soil, sep=""))
> onefactor

 [1] VALLEYSHALE      VALLEYSHALE      VALLEYSHALE
NORTHSHALE        NORTHSHALE        NORTHSHALE
SOUTHSHALE        SOUTHSHALE
 [9] SOUTHSHALE        HILLTOPSHALE     HILLTOPSHALE
HILLTOPSHALE      VALLEYSANDSTONE VALLEYSANDSTONE
VALLEYSANDSTONE  NORTHSANDSTONE
[17] NORTHSANDSTONE   NORTHSANDSTONE   SOUTHSANDSTONE
SOUTHSANDSTONE   SOUTHSANDSTONE   HILLTOPSANDSTONE
HILLTOPSANDSTONE HILLTOPSANDSTONE
Levels: HILLTOPSANDSTONE HILLTOPSHALE NORTHSANDSTONE
NORTHSHALE SOUTHSANDSTONE SOUTHSHALE VALLEYSANDSTONE
VALLEYSHALE
```

So `onefactor` is a single factor combined of the levels of `topo` and `soil`. Now we can use the `pairwise.t.test()` function, specifying that we do not want to have an adjustment by the argument `p.adj="none"`.

```
> pairwise.t.test(sydney$phosphor, sf, p.adj="none")
```



Submit the above statements for execution.

Box 5-10.
Uncorrected
pairwise
comparisons
between all
treatment cells in
the factorial
ANOVA to
assess the
effects of
topographic
location and soil
type on soil
phosphorus.
Bonferroni
correction to the
level of
significance
yields $\alpha = 0.003$
for each
comparison. The
experimentwise
error rate
remains at
 $\alpha \leq 0.05$.

Pairwise comparisons using t tests with pooled SD

data: sydney\$phosphor and sf

	HILLTOPSANDSTONE	HILLTOPSHALE	NORTHSANDSTONE
HILLTOPSHALE	0.84021	-	-
NORTHSANDSTONE	0.94256	0.89683	-
NORTHSHALE	0.03244	0.04837	0.03746
SOUTHSANDSTONE	0.40840	0.52866	0.44918
SOUTHSHALE	0.02284	0.03437	0.02646
VALLEYSANDSTONE	0.86277	0.70851	0.80665
VALLEYSHALE	7.6e-05	0.00011	8.8e-05

	NORTHSHALE	SOUTHSANDSTONE	SOUTHSHALE	VALLEYSANDSTONE
HILLTOPSHALE	-	-	-	-
NORTHSANDSTONE	-	-	-	-
NORTHSHALE	-	-	-	-
SOUTHSANDSTONE	0.15487	-	-	-
SOUTHSHALE	0.86277	0.11462	-	-
VALLEYSANDSTONE	0.02284	0.32077	0.01599	-
VALLEYSHALE	0.00986	0.00043	0.01418	5.4e-05

Perusal of the results of the pairwise comparisons (only bold faced comparison are the correct ones) reveals the following as significant at the $\alpha = 0.003$ level of significance:

SHALE [Valley vs Hilltop] $p = 0.00011 < 0.003$

VALLEY [Shale vs Sandstone] $p < 5.4e-05 < 0.003$

Results summary

The following is an example of an appropriate results summary, as would appear in your report or publication.

Mean soil phosphorus varied significantly across the various combinations of topographic location and soil type in a two-way factorial ANOVA ($F = 7.16$; $df = 7,16$; $p < 0.001$). The percentage variation that could be attributed to the effect of the two factors, individually and in combination, was high ($R^2 = 0.76$).

There was a significant interaction between soil type and topographic location in their influence on soil phosphorus ($F = 4.88$; $df = 3,16$; $p < 0.02$). Topographic location had no effect on soil phosphorus in sandstone-derived soils ($F = 1.07$; $df = 3,8$; $p = 0.41$), whereas a pronounced effect was evident for shale-derived soils ($F = 5.33$; $df = 3,8$; $p < 0.05$) (Figure 5-15). In particular, soil phosphorus in shale-derived soils of valleys (151.67 ppm) was significantly greater than for

similar soils on the hilltops (36.33 ppm) (Bonferroni correction; $p < 0.003$).

The interaction was also evident on consideration of soil types. On hilltops, soil phosphorus was very similar for both shale and sandstone-derived soils (36.33 and 31.67 ppm respectively), whereas in valleys the difference between the two soil types was substantial (Figure 5-15). Soil phosphorus in shale-derived soils in valleys (151.67 ppm) was significantly greater than in sandstone-derived soils in valleys (27.67 ppm) (Bonferroni correction; $p < 0.003$).

Note that the analysis for factorial ANOVA with interaction is much more complicated than where interaction is absent. It required us to consider the effects of each factor separately for each level of the other factor.

The results as described include a statement of the overall significance of the ANOVA. The strength of the overall result is indicated by the R^2 and by the magnitude and direction of the difference between specific means of interest (means are given where appropriate). Both significance and strength of result need to be covered.

The interaction is described in full, and the results of the split analysis are presented.

The ANOVA table is rarely published, but the graph of treatment means showing the nature of the interaction is very informative and was published by Anne-Marie Clements.

Discussion

Clearly topographic location and soil type have an influence on the levels of phosphorus found in the soils and this will influence the susceptibility to weed invasion.

In response to the questions posed at the beginning of the study:

- Yes, there were significant differences in soil phosphorus among the four topographic locations – valleys, north-facing slopes, south-facing slopes and hilltops – but the differences were only evident in shale-derived soils;
- Yes, there were significant differences in soil phosphorus between shale-derived and sandstone-derived soils, but this difference was significant only in the valleys; and
- Yes, the difference between topographic locations in soil phosphorus depended in magnitude on whether the soils were shale-derived or sandstone-derived.

The relationships are clearly complex. It appears that shale-derived soils are better at retaining soil phosphorus than are sandstone-derived soils, and that this difference in capacity is most prominent in valleys.



Tidy up the program by ensuring there are no elements remaining of the program that did not work. Save the program to disk within Tinn-R for future reference. You may also want to save your workspace and Rhistory, if you want to return to this example at some stage.

Exit from R and Tinn-R by choosing File->Exit from the Menu Bar.

Where have we come?

With some sound theory behind us from Lessons 1 to 4, it was time to get our hands dirty with some blow-by-blow sample analyses.

In this lesson, you were introduced to a few common examples of factorial ANOVA, namely:

- Fixed design, with replication, no interaction (the Pig Trials)
- Fixed design, with replication, with interaction (the Soil Phosphorus Study)
- Mixed design, without replication (the Marine Turtle Study) as an example of an analysis that extends the Paired T-test to include triplicates, quadruplicates, etc.

The Random design, where the analysis is continued with estimation of the added variance components, is outside the scope of this course. Many other designs, such as three-way ANOVAs and mixed model ANOVAs with replication, are natural extensions of what you have learned and are not beyond your capability.

Skills imparted in Lesson 5 include

- How to compare means of samples classified by two or more criteria in a single factorial design, combining the power of the ANOVA with appropriate follow-up analysis.
- How to use R graphics to display the data in both tabular and graphical form in the exploratory phase of the analysis, and later during the analysis of residuals.
- How to make important decisions that govern the direction of the analysis, such as whether the model is fixed, random or mixed, whether the overall result of the ANOVA was significant, whether or not there is a substantial interaction, what follow-up options are available, and whether there are a priori constraints that need to be accommodated in the follow-up analysis.
- How to report the results of your analysis of variance. There is a very formal and agreed format for this.

And of course, working through these examples should have reinforced a number of skills required to use R for statistical analyses.

Time to try some challenging exercises. This is where the real learning occurs. So there was both practice and theory introduced in these worked examples. Now it is time for you to apply what you have learned in some challenging exercises.

Lesson 7: Some Challenging Exercises

Exercise 5-1: Chase duration in Chats (Mark II)

Three species of Australian Chat (*Epthianura*) can be found in micro-sympatry in mesic coastal, semi-arid and xeric arid regions of Western Australia. *Epthianura aurifrons* is the most physiologically competent to handle aridity, *Epthianura albifrons* is the least physiologically competent and *Epthianura tricolor* is intermediate in competence.

Territorial behaviour is expensive in terms of maintaining water balance, so data was collected for each species in each of the three habitats (mesic coastal, semi-arid, xeric arid) to see if physiological competence and habitat have a bearing on the duration of the territorial chase. A bird is said to engage in a territorial chase when it sees another bird off its territory.

The data comprise variables giving the species, the habitat the duration of chase (in seconds).

Table 5-23.
Duration of territorial chase for three species of Australian Chat in each of three different habitats.

Albifrons				
Mesic coastal	58	60	72	76
Semi-arid	56	58	68	74
Xeric arid	48	24	32	39
Tricolor				
Mesic coastal	72	75	78	79
Semi-arid	60	70	74	77
Xeric arid	66	66	54	51
Aurifrons				
Mesic coastal	84	86	88	88
Semi-arid	76	80	80	84
Xeric arid	72	74	76	70

We need to analyse the data using an appropriate ANOVA model to address hypotheses on differences in duration of chase.

Describe as completely as possible an appropriate analysis and give reasons for your choice. Be sure to specify the nature of the Factor (s) involved, to identify the response variable and to state clearly the null hypotheses to be addressed.

Enter the data in a form suitable for the nominated analysis, and conduct an exploratory analysis based on graphical presentations with box diagrams. Include the box diagrams below. What would you anticipate the results of an appropriate ANOVA to be?

Before preparing an ANOVA table, examine a plot of the residuals to determine whether the assumptions of the analysis are tenable. If not, try some potential remedies, and repeat the ANOVA. Please include any graphs from your residual analysis below.

Perform the Analysis of Variance and summarise the results in the form of an ANOVA table.

If the above analyses demonstrate a significant difference among the mean chase durations, perform an appropriate graphic analysis to facilitate interpretation of the results. Present your results below.

Write a summary of the results of the entire analysis, as might be included in the results section of a report or manuscript. Refer in your summary to an ANOVA table and a figure showing the variation among species and habitats. Include in your results, a statement of any clear and statistically significant trends in chase duration, but do not at this stage attempt to explain them.

Discuss the analysis in the context of the reasons for conducting the study. What might be the causes of the observed variation in chase duration among species and habitats, taking into account differences in their physiological tolerance to aridity?

Exercise 5-2: Trace Metals in Botany Bay

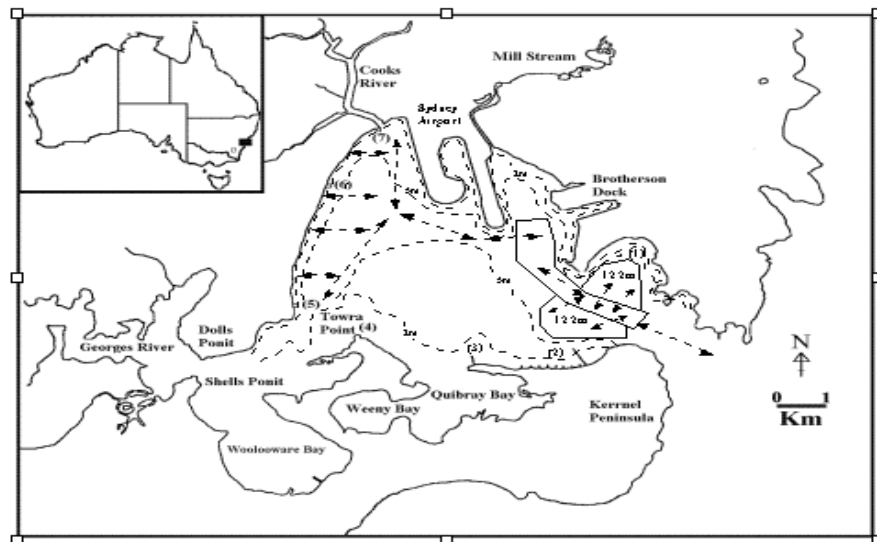
Botany Bay is the first place of European settlement in Australia, and now has the burgeoning metropolis of Sydney on its shores and adjacent hinterland. This urban and industrial development brings with it many pressures on the environment, pressures that may interfere with its ability to deliver for human health, recreation and fisheries.

Daniel Spooner undertook a study of the spatial and temporal distribution of trace metals in the sediments of the bay to identify the zones and extent of contamination by a variety of trace metals. He sampled sediments at locations for three southern sites, three western sites and one eastern site (Figure 5-16). The southern and western sites were selected systematically according to their relative proximity to inflows from the river catchments of metropolitan Sydney. The eastern site was close to the open end of the bay.

Daniel visited the sites on five occasions between January 1998 to May 1999 to capture seasonal variance. Sediment samples were randomly taken from a 10 m x 10 m quadrat for each site at each sampling occasion. Daniel assayed a suite of metals that are of concern (arsenic, zinc, copper, lead, and cadmium) but this exercise will concern just the arsenic data ($\mu\text{g/g}$).

The central issue of concern is whether there are significant differences among sites in the concentration of arsenic in the sediments, and if there is, where do these differences lie and how do they relate to the distribution of inflows to the bay.

Figure 5-16: Botany Bay location map, showing sampling sites, dredging activities, and proposed flow patterns for sediment deposition and erosion. Source: Kinhill, 1990. Arrows show tidal & river flow depositional & erosion path.



The data reside in a file called `botbay.dat` on the data disk. The contents are in the following form, without the headers. Not all the data are shown.

*Table 5-24.
Concentrations of
eight trace elements
in the sediments of
Botany Bay,
Australia.*

TIME	SITE	Al	Mn	Zn	Cu	As	Se	Cd	Pb
Jan-98	SITE-1	1172.52	49.36	21.80	10.95	1.60	0.4374	0.0324	16.61
Jan-98	SITE-1	1170.41	31.39	17.34	3.69	1.54	0.5043	0.0285	14.18
Jan-98	SITE-1	1169.26	31.36	17.32	3.69	1.54	0.5038	0.0285	14.16
Jan-98	SITE-2	1667.29	25.64	20.75	3.65	1.64	0.7296	0.0469	10.56
Jan-98	SITE-2	1092.27	16.44	9.86	1.50	0.99	0.2503	0.0258	7.29
Jan-98	SITE-2	1928.75	29.26	28.77	4.40	1.65	0.9063	0.0812	12.25
Jan-98	SITE-3	1047.30	33.45	8.32	1.16	0.89	0.2574	1.9892	7.43
May-98	SITE-1	2893.57	47.88	50.22	13.20	3.47	0.7173	0.0970	33.61
May-98	SITE-1	2450.45	44.64	40.13	9.07	2.82	0.7737	0.7782	29.01
May-98	SITE-1	3245.13	54.35	60.10	18.13	4.36	1.6431	0.1590	39.77
May-98	SITE-2	1624.02	52.65	27.42	1.98	2.21	0.9481	0.0583	11.80

You are asked to analyse the data using an appropriate ANOVA model to address hypotheses on differences in arsenic concentration across the different sites. Optionally conduct similar analyses for the other trace metals.

- Describe as completely as possible an appropriate analysis and give reasons for your choice. Be sure to specify the nature of the Factor (s) involved, to identify the response variable and to state clearly the null hypotheses to be addressed.
- Access the data using an appropriate DATA step, and peruse it to ensure it has been read as intended. Conduct an exploratory analysis based on a table of cell and marginal means. Include the table below. What would you anticipate the results of an appropriate ANOVA to be?
- Before preparing an ANOVA table, examine a plot of the residuals to determine whether the assumptions of the analysis are tenable. If not, try some potential remedies, and repeat the ANOVA. Are there any outliers, and if so, what is your advice? Please include any graphs from your residual analysis below.
- Perform the Analysis of Variance and summarise the results in the form of a standard ANOVA table.
- If the above analyses demonstrate a significant interaction between site and time, perform an appropriate graphic analysis to facilitate interpretation of the interaction. Present your graph and interpretation below.
- What is an appropriate follow-up analysis to facilitate interpretation of significant differences among sites (Hint: Is the interaction strong or weak?). Perform the analysis, and present the outcome below.

- (g) Write a summary of the results of the entire analysis, as might be included in the results section of a report or manuscript. Refer in your summary to an ANOVA table and a figure showing the variation among sites and times. Include in your results, a statement of any clear and statistically significant trends in arsenic concentration, including any interactions, but do not at this stage attempt to explain them.
- (h) Discuss the analysis in the context of the reasons for conducting the study. Where are the major concentrations of arsenic and how does its distribution relate to the inflows and flow patterns for Botany Bay? What advice would you give to the organisation funding this study? Feel free to use poetic licence for the sake of this exercise.

Exercise 5-3: Frequent burns and ant communities

Periodic low-intensity fires (hazard-reduction burning) are a common management strategy in virtually all of Australia's dry forest communities, primarily to reduce fuel levels and thereby minimise the extent and severity of wildfires. However, surprisingly little is known about how repeated burns affect invertebrate communities of natural ecosystems over long periods of time. Alan York studied the long-term effects of frequent low-intensity fire on forest ant communities by comparing frequently burned sites with long-unburned sites.

In 1970, six pairs of plots were selected along Sandy Hollow Road in Bulls Ground State Forest on the mid north coast of NSW. Membership of each pair was decided on the basis of similarity in their soil, aspect and overstory characteristics. The pairs were treated as blocks in the design.

Two fire treatments, burning and no burning, were randomly allocated to the plots with the constraint that there was one site in each block scheduled for burning and one site in each block scheduled to be free of burning. This is a randomized block design.

The regime applied to the burnt plots conformed with the Zone 1 standards laid down by the Hazard Reduction Zoning System for managed forests in NSW. The burnt plots were subjected to low intensity burns in autumn whenever fuel buildup permitted, usually every three years (1970, 1973, 1977, 1980, 1983, 1986, 1989 and 1992). The unburnt plots were not burned during this period.

In February of 1994, two years after the final fire, four subplots were selected within each plot to capture within plot variation. Four 1.0 m² randomly selected samples of leaf litter were collected from each subplot. The litter collections for each subplot were pooled and invertebrates were extracted over a four day period using modified Tullgren funnels. Samples were returned to the lab and sorted under a microscope to quantify and identify the insects.

Sampling 2 years postfire largely avoids the known short-term responses of invertebrates to a single fire event. The samples represent a snapshot of the effect of frequent low intensity fires on a time scale appropriate to the dynamics of insect communities (20 years).

The data comprise total ant abundance and ant species richness, respectively, classified according to whether they were in a burnt or unburnt plot in each of the six blocks. The data reside in the file ants.dat.

The problem is to determine if there is an effect of low intensity fires, over 20 years, on ant species abundance and richness, after correcting for the natural variation that would be expected to occur among blocks that differ in overstory, soil and aspect.

- (a) Describe as completely as possible an appropriate analysis and give reasons for your choice. Be sure to specify the nature of the Factor (s) involved, to identify the response variable and to state clearly the null hypotheses (plural) to be addressed.
- (b) Read the dataset into a R workfile suitable for analysis. Construct a table of cell and marginal means and peruse the marginal means as a preliminary assessment of likely outcomes of the analysis. Include the table of marginal and cell means below. What would you anticipate the results of an appropriate ANOVA to be?
- (c) What is the hypothesis of greatest interest? Why?
- (d) Before preparing an ANOVA table, examine a plot of the residuals to determine whether the assumptions of the analysis are tenable. If not, try some potential remedies, and repeat the ANOVA. Please include any graphs from your residual analysis below.
- (e) Perform the Analysis of Variance and summarise the results in the form of an ANOVA table.
- (f) What do you conclude from the analysis? Provide a concise summary of the results, such as might appear in the results section of a manuscript or report. Remember to distinguish between the magnitude of the result and its statistical significance.
- (g) Discuss the analysis in the context of the reasons for conducting the study. This analysis is referred to as a randomised blocks design. Explain what the benefits were of blocking the data into pairs with similar characteristics. How did your conclusions benefit from the blocking of treatment and control?

Exercise 5-4: Lake fertilization and fish growth (BACI)

Lake fertilization is a fishery management tool that brings about increased productivity at different trophic levels of a lacustrine food web. Lake productivity is strongly influenced by phosphorus inputs yet a relationship between total phosphorus (TP) and fish biomass or fish yield varies from year to year. A debate continues as to whether the biomass patterns are more reliant on resource inputs (“bottom-up” controls) or predation (“top-down” controls). The relative importance of bottom-up and top-down controls depend partially on the efficiency of prey consumption by predators. A sophisticated and ambitious field test of the competing models of trophic structure is through whole-lake experiments.

Based on outcomes of previous studies, additions of fertilizer to lake ecosystems are predicted to alter feeding, growth, survival, fecundity, abundance, biomass, and production of the aquatic flora or fauna. Johnston et al. (1999) documented the responses of rainbow trout (*Oncorhynchus mykiss*) and their prey to 5 years of increased nutrient loading. Their aims were to test the predictive relationships between productivity (nutrient concentration) and trophic-level biomass patterns and to identify specific changes in the trout populations’ parameters. The detection of environmental change in response to a perturbation is addressed by the class of ANOVA models known as Before-After Control-Impact studies or BACI designs.

The study site is the Twin Lakes in British Columbia, Canada. The two lakes are small (4.5 and 7.7 ha surface area) and oligotrophic. Physical and chemical characteristics of the lakes were similar, although West Twin Lake was marginally deeper and had less littoral area. East Twin Lake was the treatment lake (it received fertilizer) and West Twin Lake was used as a control (it received none). Lakes were without fish prior to 1982. The study design includes a pre-fertilization period (1982-1989) and a fertilization period (1990-1994). Data for a post-fertilization period (1995-1996) are not included.

The data reside in a datafile called twinlakes.dat, and comprise two categorical variables LAKE and YEAR and total phosphorus.

You are asked to analyse the data to determine if phosphorus level is affected by fertilization.

- (a) Describe as completely as possible an appropriate analysis and give reasons for your choice. Be sure to specify the nature of the Factor (s) involved, to identify the response variable and to state clearly the null hypotheses (plural) to be addressed.

- (b) Read the data into a R in a form suitable for the nominated analysis, and conduct an exploratory analysis based on graphical presentations of the treatment means. Include the graphs below. What would you anticipate the results of an appropriate ANOVA to be?
- (c) What is the hypothesis of greatest interest? Why?
- (d) Before preparing an ANOVA table, examine a plot of the residuals to determine whether the assumptions of the analysis are tenable. If not, try some potential remedies, and repeat the ANOVA. Please include any graphs from your residual analysis below.
- (e) Perform the Analysis of Variance and summarise the results in the form of an ANOVA table.
- (f) Write a summary of the results of the entire analysis, as might be included in the results section of a report or manuscript. Refer in your summary to an ANOVA table and a figure showing the variation across time for each lake. Include in your results, a statement of any clear and statistically significant trends in total phosphorus over time, but do not at this stage attempt to explain them.
- (g) Discuss the analysis in the context of the reasons for conducting the study. This analysis is referred to as a BACI design. Explain how your conclusion is strengthened by having a spatial as well as a temporal control.

Exercise 5-5: Competition from exotics on a rare native plant

Introduced plant species threaten many terrestrial and aquatic ecosystems by altering geomorphological processes, nutrient cycles and fire regimes, and by displacing native plant species. Rare plants in some ecosystems may be particularly vulnerable because of small population size, poor competitiveness or small geographic ranges. Although many endangered or threatened plants are thought to be declining because of introduced plants, few studies have quantified the impact of introduced species on the establishment, survival, growth, and reproduction of rare plants.

Solidago shortii is an endangered plant endemic to a 12.2 km² area in north-eastern Kentucky (USA). Jeffery Walck and his colleagues studied the impacts of introduced plants on the viability of this relict population. They gathered data designed to determine the effect of removing exotic plants on the reproductive performance of *Solidago shortii*.

The field study established twenty replicate plots with each plot containing a clump of ramets (stems) of *S. shortii*. Soil nutrients and light are relatively high at the site and thus do not appear to limit growth of *S. shortii*. Ten plots were selected randomly for treatment (removal of non-native vegetation) and the other ten plots served as control plots (non-native vegetation not removed). The experimental plots were maintained free of non-native vegetation from 1993-1996. The number of flowering and nonflowering ramets of *S. shortii* were counted each September, in each year from 1992 to 1996 on both the treatment and control sites.

The data are held in the file `solidago.dat` and comprise five columns – plot number, removal treatment (CONTROL, TRT), year, number of flowering ramets, number of non-flowering ramets.

You are asked to analyse the data on total number of ramets per plot to determine if there was a significant effect of removal of non-native species on the reproductive capacity of this rare plant.

- (a) Describe as completely as possible an appropriate analysis and give reasons for your choice. Be sure to specify the nature of the Factor (s) involved, to identify the response variable and to state clearly the null hypotheses to be addressed. Which hypothesis is likely to be of greatest interest?
- (b) Access the data using an appropriate DATA step, and peruse it to ensure it has been read as intended. Conduct an exploratory analysis based on a table of cell and marginal means. Include the table below. What would you anticipate the results of an appropriate ANOVA to be?

- (c) Before preparing an ANOVA table, examine a plot of the residuals to determine whether the assumptions of the analysis are tenable. If not, try some potential remedies, and repeat the ANOVA. Are there any outliers, and if so, what is your advice? Please include any graphs from your residual analysis below.
- (d) Perform the Analysis of Variance and summarise the results in the form of a standard ANOVA table.
- (e) If the above analyses demonstrate a significant interaction between treatment and year, perform an appropriate graphic analysis to facilitate interpretation of the interaction. Present your graph and interpretation below.
- (f) What is an appropriate follow-up analysis to facilitate interpretation of significant differences among treatments and across years. Perform the analysis, and present the outcome below.
- (g) Write a summary of the results of the entire analysis, as might be included in the results section of a report or manuscript. Refer in your summary to an ANOVA table and a figure showing the variation among sites and times. Include in your results, a statement of any clear and statistically significant trends, including any interactions, but do not at this stage attempt to explain them.
- (h) Discuss the analysis in the context of the reasons for conducting the study. What advice might you give to the management agencies responsible for the conservation of this endangered species?

Exercise 5-6: Artificial substrates in pond aquaculture

Asia accounts for 90% of the world's aquaculture production, the bulk of which occurs in ponds or rice fields. Pond production systems in southern Asian countries are increasingly reliant on external resources (feed, fertilizers) to supplement or stimulate autochthonous food production for pond fish. Ways that enhance the conversion of these nutrients into harvestable products, through adoption of periphyton-based production into existing pond systems, are worth exploring.

Periphyton is defined as the entire complex of all sessile biota attached to the substratum, plus associated detritus and microorganisms. A less technical though accurate name for periphyton is 'pond scum'. Culturing fish that graze on the periphyton is an old and proven approach. The approach derives from traditional fishing practices, such as the 'acadjas' of the Ivory Coast, the 'samarahs' of Cambodia and the 'katha' fisheries of Bangladesh where tree branches are placed in shallow open waters to attract fish and enhance productivity. Some studies suggest that fish yields from periphyton grown on these artificial substrates could be up to 8 tons per hectare a year, which is potentially 8 times higher than in control areas without artificial substrates.

However, the results from experiments in aquaculture ponds can be quite variable within and between substrate types, and the design of many trials allows no conclusion about the causal factors responsible for differences.

This need for a systematic research program on the potential of periphyton-based aquaculture stems was a stimulus for experiments conducted by Mohammed E. Azim and his colleagues to determine:

- if there were significant differences in quantity and quality of periphyton grown on artificial substrates of three locally available plant materials, in the absence of fish;
- If there significant differences in periphyton occurred with depth, which would potentially indicate some need to control water depth for optimal growth.

Field trials were carried out in 9 earthen ponds of 1.2 m depth. Three common substrate types (bamboo, kanchi, and hizol) were evaluated in triplicate using a complete randomised design. Ponds were drained and prepared similarly. Ponds were filled from ground water and topped up weekly to replace any evaporative losses. A traditional schedule of fortnightly fertilization was maintained after pond filling and through the experimental period of 6 weeks. The substrates were made available by putting clusters of bamboo poles, or kanchi poles,

or hizol branches in a 1 m² arrangement that yielded a total submerged substrate area about the same as the pond surface area.

Each week the periphyton biomass growing on a substrate was sampled to determine the dry matter (DM) and pigment concentration (Chlorophyll a). Areas were scraped from the pole substrate and replaced to the pond, but marked so that they were excluded from other subsequent sampling. Data on dry mass of the periphyton (DM) are given for the final sample (Week 6), broken down by substrate (control or B, K, H), and depth (0,30,60,90 cm).

The data reside in the file PERIPHYTON.DAT.

Table 5-25. Dry weight of periphyton grown on different substrates at different depths in a two-way factorial design.

Substrate	Depth	DM
bamboo	0	1.67
bamboo	30	1.67
bamboo	60	0.83
bamboo	90	1.67
hizol	0	0.83
hizol	30	4.17
hizol	60	7.5
hizol	90	4.17
kanchi	0	0.83
etc.	etc.	etc

You are asked to analyze the data to determine if there are significant differences periphyton growth among the different substrates, if there is a significant difference in periphyton with depth, and if there is a difference in periphyton growth on the substrates, does the difference depend in magnitude or direction on the water depth?

The first question is the one of greatest interest.

- (a) Describe as completely as possible an appropriate analysis and give reasons for your choice. Be sure to specify the nature of the Factor (s) involved, to identify the response variable and to state clearly the null hypotheses to be addressed.
- (b) Access the data using an appropriate DATA step, and peruse it to ensure it has been read as intended. Conduct an exploratory analysis based on a table of cell and marginal means. Include the table below. What would you anticipate the results of an appropriate ANOVA to be?
- (c) Before preparing an ANOVA table, examine a plot of the residuals to determine whether the assumptions of the analysis

are tenable. If not, try some potential remedies, and repeat the ANOVA. Are there any outliers, and if so, what is your advice?

- (d) Perform the Analysis of Variance and summarise the results in the form of a standard ANOVA table.
- (e) If the above analyses demonstrate a significant interaction between treatment and year, perform an appropriate graphic analysis to facilitate interpretation of the interaction. Present your graph and interpretation below.
- (f) What is an appropriate follow-up analysis to facilitate interpretation of significant differences among treatments and across years. Perform the analysis, and present the outcome below.
- (g) Write a summary of the results of the entire analysis, as might be included in the results section of a report or manuscript. Refer in your summary to an ANOVA table and a figure showing the variation among sites and times. Include in your results, a statement of any clear and statistically significant trends, including any interactions, but do not at this stage attempt to explain them.
- (h) Discuss the analysis in the context of the reasons for conducting the study. What advice might you give to the aquaculturalists?

Exercise 5-7: Phytoplankton in Peel-Harvey Estuary

Peel-Harvey estuary is a shallow estuary located approximately 80 km south of Perth in the south-west of Western Australia. Clearing and agriculture in the catchments of the estuary have led to a dramatic increase in the levels of nutrients entering the estuary. For more than 30 years prior to 1994, the estuary was plagued by algal blooms, including blooms of the toxic blue-green algae *Nodularia*. Build-up of macroalgae was also a problem and the shorelines were regularly cleared by front end loaders. The only opening to the ocean was by a small channel running north from the Peel Inlet and flushing of the system was poor (the tidal range in the estuary was only 10% of that in the ocean – Hale and Paling, 1999).

A three part strategy was developed to reduce eutrophication in the Peel-Harvey Estuary (Peel Inlet Management Authority 1994):

- reduce the nutrient run-off from the catchment;
- continue harvesting macroalgae as necessary; and
- increase flushing to the ocean.

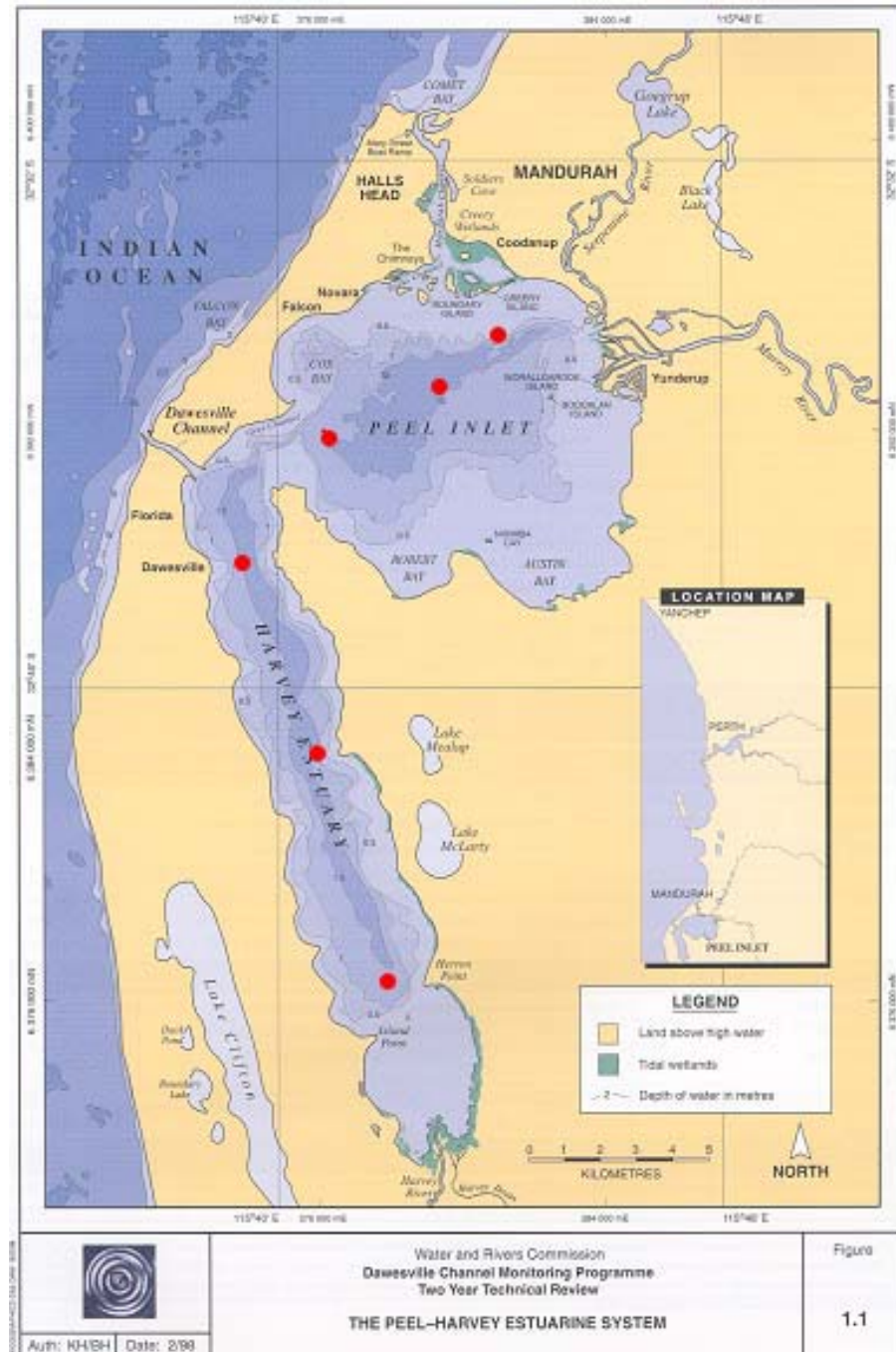
The third point was addressed by the construction of the Dawesville Channel (see Figure 5-17.) It is a \$65 million engineering approach to improving water quality. The Channel is approximately 130 to 200m wide, 3.5 km long and 4.5 to 6.5 metres deep and was opened in 1994.

This study aims to use the extensive data set collected before and after the opening of the Channel to see if the construction of the Channel has had the desired effect of reducing algal blooms throughout the whole year. Chlorophyll-a concentrations will be used as an indicator of algal blooms.

You are asked to analyze the data to determine if there are significant differences in chlorophyll-a pre-channel and post-channel, if there are significant differences in chlorophyll-a with season, and if there is a difference in chlorophyll-a pre and post channel, does the difference depend in magnitude or direction on the season?

- (a) Describe as completely as possible an appropriate analysis and give reasons for your choice. Be sure to specify the nature of the Factor (s) involved, to identify the response variable and to state clearly the null hypotheses to be addressed.
- (b) Access the data using an appropriate DATA step, and peruse it to ensure it has been read as intended. Conduct an exploratory analysis based on a table of cell and marginal means. Include the table below. What would you anticipate the results of an appropriate ANOVA to be?

Figure 5-17:
Peel-Harvey
Estuary in
Western
Australia. Red
dots show
sampling sites



- (c) Before preparing an ANOVA table, examine a plot of the residuals to determine whether the assumptions of the analysis are tenable. If not, try some potential remedies, and repeat the ANOVA. Are there any outliers, and if so, what is your advice?
- (d) Perform the Analysis of Variance and summarise the results in the form of a standard ANOVA table.
- (e) If the above analyses demonstrate a significant interaction between treatment and year, perform an appropriate graphic

analysis to facilitate interpretation of the interaction. Present your graph and interpretation below.

- (f) What is an appropriate follow-up analysis to facilitate interpretation of significant differences among treatments and across seasons. Perform the analysis, and present the outcome below.
- (g) Write a summary of the results of the entire analysis, as might be included in the results section of a report or manuscript. Refer in your summary to an ANOVA table and a figure showing the variation among sites and times. Include in your results, a statement of any clear and statistically significant trends, including any interactions, but do not at this stage attempt to explain them.
- (h) Discuss the analysis in the context of the reasons for conducting the study. What advice might you give to the Peel Inlet Management Authority?

Where have we come?

This lesson is where the real learning occurs. In earlier lessons, you have read and understood written material and been led through worked examples. In this lesson, you were required to recall and integrate the information to complete some challenging real-world exercises. Recall in the context of problem solving is one of the best ways of achieving lasting learning. It is hard yakka.

In completing this module successfully, you will have achieved a number of core competencies, namely,

- Knowledge of the analysis options available to you under the broad heading of Factorial ANOVA.
- Understanding the distinction between the various classes of analysis (fixed, random, mixed), and the questions that can be addressed by them.
- A working knowledge of the `lm()` procedure, and how to specify a model with more than one factor including an interaction term.
- The ability and confidence to interpret the results of the analyses in a biological context based on demonstrated understanding of the analyses. In particular, you should have a sound understanding of the biological meaning of interaction.
- The ability to present findings in a style appropriate to the scientific literature.
- Appropriate attitudes and efficient strategies for extending your abilities to conduct analyses and solve problems beyond the scope of this module, by using resource materials such as statistical texts, software manuals, and your colleagues.

In terms of practical skills, on completing this Module, you should be able to competently undertake the following analyses:

- Two-way and three-way Fixed Model Factorial ANOVA with replication, with or without significant interactions.
- Two-way Mixed Model Factorial ANOVA with or without replication.