Elementary analysis of variance for the behavioural sciences

Ray Meddis

Lecturer in Psychology
Bedford College
University of London

London · New York · St Louis · San Francisco · Düsseldorf
Johannesburg · Kuala Lumpur · Mexico · Montreal · New Delhi
Panama · Paris · São Paulo · Singapore · Sydney · Toronto

Published by

McGRAW-HILL Book Company (UK) Limited

MAIDENHEAD · BERKSHIRE · ENGLAND

---

07 084009 1

Copyright © 1973 McGraw-Hill Book Company (UK) Limited.
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, electronic, mechanical, photocopying,
recording, or otherwise, without the prior permission of
McGraw-Hill Book Company (UK) Limited.

PRINTED AND BOUND IN GREAT BRITAIN

Contents

Preface vii

1 Introduction 1

2 Calculating the mean and variance 4

3 Estimating from samples 13

4 Comparing two variances 23

5 The difference between means 31

6 One factor analysis of variance (quick method) 47

7 Two factor analysis of variance 52

8 Two factor analysis of variance with replications 61

9 Linear trend test 72

10 Multiple trend tests 83

11 Planned comparisons 97

12 Three factor analysis of variance 108

13 Some theoretical considerations 121

Appendix: Instructions for use of F-tables 127

Preface

For the user of the book

This book has been especially written for non-statisticians who wish to master the use of analysis of variance techniques. Because the reader is a non-statistician, an abstract and mathematical treatment has been carefully avoided and replaced by a more practical and intuitive approach to the underlying concepts. The emphasis has been moved from the theory to the practice of these useful techniques. At the same time an intellectually honest attempt is made to explain what analysis of variance is about and why the various calculations are necessary and successful in the business of testing scientific hypotheses.

The various chapters aim to impart both an appreciation of the techniques and a confidence in one's own ability to carry them out successfully. This confidence can only come from practical work following a clear exposition. Each chapter deals with a different procedure. It gives the rationale underlying this procedure as well as an explicit example of how to carry it out. It helps the reader step by step through a structured problem which he himself must complete.

Each chapter is intended as a separate unit and should be attempted at a single sitting. If the session is seriously interrupted, you should begin the chapter again. Under no circumstances should you skip the practical work. This is a practical course and will only be effective if you exercise your new skills as you acquire them.

As you work, sit with large sheets of plain white paper beside the book. When each structured example appears, you should copy out its frame and then go on to fill in the blanks. When you have finished, your sheet should look exactly like the completed version on the next page. If you experience difficulty while working the example, by all means seek a prompt from the completed answer. It is better to 'cheat' at this stage than to compound errors, one on another. Every effort has been made to render the examples arithmetically easy and you should have no trouble in arriving quickly at the correct solutions.

The sequence of the chapters has been carefully compiled. Each chapter

depends upon a knowledge and understanding of previous chapters. Therefore, do not attempt to read chapters out of sequence. This will cause unnecessary confusion. Under no circumstances begin the book in the middle even if you feel that you already understand the more elementary material. The book was conceived as a sequence. Treated as such it works well. Begin at the beginning and work carefully and intelligently through to the end.

Each chapter has been carefully tested with students whose opinions and suggestions have had a considerable influence on the details of this book. The material has worked for them, many of whom had despaired of ever understanding the regular text-book treatments of the topic. If you work through the chapters in sequence and carry out the practical work, the chances are that it will work for you too.

For the teacher

No attempt has been made in this book to cover the mathematics of sampling theory or the partitioning of variance. Some of the more intuitively acceptable points have been discussed, however, and their relationship to analysis of variance procedures explained. Some such approach is necessary to replace the theoretically oriented treatments which usually fail to make much impression upon mixed groups of students. Many social science students remember only casual and unpleasant encounters with mathematics and that often many years ago. Such folk do benefit, however, from a more conversational approach with regular rewarding practical work. To call such a teaching procedure a 'cook-book' is to underestimate both the culinary arts and the understanding to be gained from the regular and successful practical exercise of any skill.

Freedom from the algebra of statistical theory has permitted a reasonably flexible sequence of topics. Each chapter is intended to add a small, digestible amount of new ideas to those already learned and leaps of learning have been avoided wherever possible. One-way analysis of variance is first described in terms of deviation scores. This makes the procedures more understandable. Only then are the standard formulae using raw scores given.

Trend tests are given before planned comparisons since they offer an intuitively plausible introduction to lambda coefficients. Multiple trend tests introduce the notion of orthogonality. Both of these are required for understanding multiple planned comparisons when the additional problem of selecting one's own coefficients arises.

Various technical issues, notably choice of 'model', have been avoided, not because they are unimportant but because of the confusion they create in the beginner. The strategies used to avoid such problems are always 'conservative'.

Monte Carlo techniques have been used twice in this book, firstly to justify the use of degrees of freedom in the calculation of variance estimates from

samples and secondly to illustrate the F-values in the significance tables. They do not constitute proofs, of course, but they are enormously convincing demonstrations.

The teacher who uses this book as course material should encourage students to work at a topic *before* the relevant lecture. This removes the necessity to do long worked examples on the board. It also, in my experience, promotes an intelligent discussion in class of the more subtle features of analysis of variance techniques.

Ray Meddis

1

Introduction

If all scientific experiments were repeatable, there would be little need for analysis of variance. If we could guarantee that second attempts would yield the same result as the first, you would not be reading this book. There are, of course, fields of scientific endeavour where experiments do yield the same results, however often they are attempted. These are mainly confined to the 'hard' sciences such as physics and chemistry and their various offshoots.

In these areas there is little need for recourse to statistical methods to help evaluate the results of an experiment. If a chemist adds chemical A to chemical B and observes a given chemical reaction, he can be fairly certain that under the same conditions he will observe the same result in future. Unfortunately, this is not always the case in the behavioural sciences. Not only do different investigators produce different results, but the same investigator repeating the same measure on the same group of persons would be surprised to get the same set of scores twice in succession.

We may say with some evidence that drunken driving causes accidents. It does not always cause accidents, only sometimes. On most days we expect that drunken drivers will cause accidents out of all proportion to the small number of drunks on the road, but the statistics for one day will be different from those for another day. One does not doubt, however, that the proposition is still true despite all this variability. The fact that the statistics vary does not mean that drunkenness causes no accidents. It merely makes the figures more difficult to analyse.

In general, our experiments are not merely attempts to measure phenomena but to develop theories about them. When our results are gathered in, we must decide whether they do or do not provide support for the theories. We must accept one of these two possibilities. We always run the risk of accepting the wrong alternative.

Statistics involves a large collection of techniques to help estimate the magnitude of the risks involved in making inferences from a set of results. Analysis of variance is one of the most general of these tools as it covers a wide range of possible experimental designs.

A simple example will illustrate the kind of problem involved. It is suspected that the average weight of schoolboys aged 13 years is greater now than it

was ten years ago. Consulting the records, we compare the weights of 100 boys aged 13 years with the weights of 100 of their counterparts 10 years ago and find that the average weight is indeed a few ounces higher.

The objection immediately arises that this difference may be true of our small sample but not true of all schoolboys eligible for membership of the two groups. We cannot weigh all 13 years old schoolboys and, of course, the time has passed for weighing all the boys who were aged 13 ten years ago.

We must try to make a decision based on the data at hand. We could use *analysis of variance* to obtain a statement of the likelihood of making a mistake if we were to conclude from the figures that modern schoolboys are heavier. We could use this likelihood statement to help us make a decision as to what we should conclude from the results.

It is common for people to be confused when faced with statements of this kind. After all, we want a firm decision one way or another. Are there reliable differences between the age groups or not? Unfortunately, we know of no way of arriving at such a definite decision. In general, we may only sample reality and then make inferences from our sample results. We never see the whole truth; we merely peek at fractions of it. Experimental statistics is concerned with the risky business of inferring the truth from small samples.

Analysis of variance in particular concerns itself with those experiments where the measures come from two or more samples and where we wish to know whether the groups from which the samples are taken are characterized by different average scores. Do tall children have a higher I.Q. than small children? Do students under the influence of drugs perform worse at examinations than those who are drug free? Do old people have a poorer ability to memorize lists of nonsense syllables than young people? There are many such experimental situations which involve taking samples of people, measuring the characteristics or performance of each member of the sample and then asking whether the differences between the samples represent differences between the groups.

The difficulty is that you will find some difference between the average score of your two samples whether or not there are real differences between the groups from which the samples are drawn. It is unlikely that the average score for both samples will be exactly the same. If you infer a true difference from a small sample difference then you run the risk of drawing the wrong conclusion. Analysis of variance will help assess just how big the risk is.

The basic concept underlying the analysis of variance procedure is that of consistency of scores within the sample. If the sample scores show marked fluctuations themselves, then this reduces the significance of small differences between the average scores for the two samples. A single illustration at this stage should suffice to indicate the direction we are moving in.

Two sets of data for an experiment which compares the effect of a pep talk on examination performance are given below. Two groups of boys (10 in each group) sit an examination; one group is given a pep talk, the other

(the control group) is not. The imaginary examination results are given in table 1.1. In both cases, the difference between the averages for the two groups is 3 marks. In example A, however, the fluctuations of marks within both groups are so violent that the 3 marks difference in average score is puny by comparison. In this case, we would be tempted to infer that pep talks do not produce a reliable effect.

In example B, the picture is quite different. Here, the variability of scores within the samples is so small that the 3 marks difference commands respect. Here, we might be tempted to infer that this is a reliable effect and that pep talks are a good idea when it comes to improving examination performance.

The procedures to be described in this book rely heavily upon the concept of variance or variability. Special emphasis is placed on the comparison of variance within samples with variance between samples. Accordingly, we begin our studies in chapter 2 with ways of measuring the variability of groups of scores.

Example A (high within groups variance)

GROUPS:	NO PEP TALK	PEP TALK
	68	48
	46	84
	77	78
	40	68
	64	80
	51	49
	75	57
	45	36
	81	67
	33	43
TOTALS:	580	610
AVERAGES:	58·0	61·0

Example B (low within groups variance)

GROUPS:	NO PEP TALK	PEP TALK
	53	56
	54	64
	63	60
	57	65
	54	60
	62	61
	57	57
	61	64
	61	66
	58	57
TOTALS:	580	610
AVERAGES:	58·0	61·0

TABLE 1.1 Two fictitious sets of data. Examination results for 20 college students under two experimental conditions.

2

Calculating the mean and variance

The mean and variance are the two vital measures in analysis of variance. Usually, we are dealing with two or more groups of scores which have different *means*. We wish to test the idea that these means are significantly different and not trivially different. To do this, we need to take into account the inherent variability or *variance* of the individual scores. This chapter will briefly illustrate some simple techniques for calculating the mean and variance of a collection of scores.

First of all we shall need some scores to work on. These we derive from a simple experiment with 20 boys to compare teacher instruction with teaching machine instruction. Ten boys were instructed for 1 hour by their teacher on the causes of the Wars of the Roses. The other 10 boys spent 1 hour with a teaching machine which had been programmed to instruct on exactly the same topic. Both groups were then tested on their knowledge with a very searching set of questions. Final scores were between 0 and 10 where 10 indicates complete mastery of the subject matter. The final scores are given in table 2.1. We had predicted that neither method of instruction would be superior but that machine instruction would be more variable in its effects.

TEACHER GROUP	TEACHING MACHINE GROUP
2	0
2	0
2	0
3	2
4	2
4	3
4	3
6	5
6	7
7	8

TABLE 2.1 Examination marks for 20 schoolboys learning about the Wars of the Roses using two different techniques of instruction.

Calculating the mean

The mean is exactly the same statistic as the familiar average. Unfortunately, the term average has taken on a number of meanings in statistics. In the interests of precision, we shall always use the term 'mean' in this book.

Calculating the mean of a group is straightforward. Add up all of the scores to find the group total. Divide the total by the number of scores in the group. Table 2.2 illustrates this for the teacher group.

TEACHER GROUP	
2	
2	
2	
3	
4	
4	$\text{Mean} = \frac{40}{10} = 4{\cdot}0$
4	
6	
6	
7	
TOTAL = 40	

TABLE 2.2 Calculating the mean.

It is a simple matter to explain how the mean is calculated. It will not be easy to describe the more complex calculations that lie ahead in simple English. A little shorthand will become essential. The following signs are intended merely as shorthand NOT as algebra. No algebra will be attempted in this book.

Some of the signs to be used as as follows:

X: the letter 'X' refers to the scores in any particular group.

Σ: the Greek letter 'S' (sigma) is an instruction to 'add up all the scores', so ΣX means 'add up all the scores in the group.'
Referring to table 2.2 the total is 40

$$\text{So } \Sigma X = 40$$

n: the small letter 'n' refers to the number of scores in the group. In table 2.2, there are 10 scores

$$\text{So } n = 10$$

M: the letter 'M' refers to the mean of the group.

We calculate the mean when we divide the total sum of scores by the number of scores in the group. In shorthand,

$$M = \frac{\Sigma X}{n}$$

Substituting the values from table 2.2,

$$M = \frac{40}{10} = 4{\cdot}0$$

We can now rewrite table 2.2 using the shorthand notation. Notice that we write in table 2.3 the number of scores in the group. This is a good practice and reduces the possibility of error.

TEACHER GROUP	
2	
2	
2	$n=10$
3	
4	$M=\frac{\Sigma X}{n}=\frac{40}{10}=4{\cdot}0$
4	
4	
6	
6	
7	
$\Sigma X=40$	

TABLE 2.3 Calculation of the mean incorporating shorthand notation.

Table 2.4 gives the layout for calculating the mean for the teaching machine group. Copy it out and fill in the blanks after the manner of table 2.3. It will be standard procedure throughout this book to offer practice examples at all stages. These examples will take the form of a layout with the scores marked but the workings absent. Usually, on the next page the same example will be found, this time completed. You should check your own working against the correct solution and make sure that you understand the reason for any mistakes before proceeding. Calculation of the mean is a trivial example but the full procedure shall be followed, none the less, by way of illustration.

TEACHING MACHINE GROUP	
0	
0	
0	
2	$n=$
2	
3	$M=\frac{\Sigma X}{n}=\frac{\quad}{\quad}=$
3	
5	
7	
8	
$\Sigma X=$	

TABLE 2.4 Computation layout for calculating the mean. Fill out this table and compare your answer with table 2.5.

TEACHING MACHINE GROUP	
0	
0	
0	
2	$n=10$
2	
3	$M=\frac{\Sigma X}{n}=\frac{30}{10}=3{\cdot}0$
3	
5	
7	
8	
$\Sigma X=30$	

TABLE 2.5 Completed version of table 2.4.

Calculating the variance

We have already established that the two groups have different mean scores. Careful examination of the two sets of scores will reveal another important difference; the spread of scores is not the same. Scores in the teacher group all lie compactly within the range 2–7. Scores in the teaching machine group are spread out more widely across the greater range 0–8. This could be an important distinction and we need to be able to measure it. It is possible that the teaching machines are suited to some students but not to others. The teacher, on the other hand, is flexible and produces more even results across the full range of ability.

The first step in measuring spread or variance is to treat each score as a deviation from the mean. Obviously, the bigger the spread the larger the deviations will be. Those scores below the mean will be negative deviations, those above are positive deviations. We calculate the deviation by subtracting the group mean from each individual score. The deviate scores for the teacher group are given in the second column in table 2.6. The raw scores are called 'X' scores, but the deviate scores are called 'D' scores. The D scores in any group must add up to zero ($\Sigma D = 0$).

The second step in this two-step procedure for measuring variance is to square each of the D scores (column 3) and then find the average of the squares. The variance is the average squared deviate.

$$\text{Variance} = \frac{\Sigma D^2}{n}$$

The bigger the spread, the larger the value of the variance.

X (TEACHER GROUP)	D (DEVIATE SCORES)	D^2 (SQUARED DEVIATES)	
2	−2	4	
2	−2	4	
2	−2	4	
3	−1	1	
4	0	0	n = 10
4	0	0	
4	0	0	
6	+2	4	
6	+2	4	
7	+3	9	
$\Sigma X = 40$	$\Sigma D = 0$	$\Sigma D^2 = 30$	

$M = 4{\cdot}0$

$$\text{Variance} = \frac{\Sigma D^2}{n} = \frac{30}{10} = 3{\cdot}0$$

Comment
The scores in the teacher group have a mean of 4·0 and a variance of 3·0.

TABLE 2.6 Calculating the mean and variance using deviate scores.

Now calculate the mean and variance of the scores in the teaching machine group using the layout in table 2.7. Remember that we expect the variance score for this group to be greater than that of the teacher group. Follow this sequence of steps exactly:

1. Find the sum of the scores in the group (ΣX).
2. Find the mean (M).
3. Convert each score (X) to a deviate score (D) by subtracting the value of the mean from each individual score.
4. CHECK: the deviate scores must sum to zero ($\Sigma D = 0$).
5. Square each deviate score and note the value (D^2) in column 3.
6. Find the sum of the squared deviates (ΣD^2).
7. Calculate the variance by dividing ΣD^2 by the number of scores in the group.

X (TEACHING MACHINE GROUP)	D (DEVIATE SCORES)	D^2 (SQUARED DEVIATES)
0		
0		
0		
2		n=
2		
3		
3		
5		
7		
8		
$\Sigma X=$	$\Sigma D=$	$\Sigma D^2=$
$M=$		$\text{Variance}=\frac{\Sigma D^2}{n}=\frac{\quad}{\quad}=$

Comment

TABLE 2.7 Layout for calculating the mean and variance of a set of scores. Check your results against table 2.8.

X (TEACHING MACHINE GROUP)	D (DEVIATE SCORE)	D^2 (SQUARED DEVIATE)	
0	−3	9	
0	−3	9	
0	−3	9	
2	−1	1	
2	−1	1	n=10
3	0	0	
3	0	0	
5	2	4	
7	4	16	
8	5	25	
$\Sigma X=30$	$\Sigma D=0$	$\Sigma D^2=74$	
$M=3{\cdot}0$		$\text{Variance}=\frac{\Sigma D^2}{n}=\frac{74}{10}=7{\cdot}4$	

Comment
The scores in the teaching machine group have a mean of 3·0 and a variance of 7·4.

TABLE 2.8 Completed version of table 2.7.

A second formula for variance

Hitherto, we have calculated the variance in two stages. The first stage is to convert the raw scores into deviate scores. The second stage was to find the average of the squared deviates. This procedure was chosen because it more clearly illustrates the nature of variance. The two-stage procedure is, however, cumbersome and computationally tedious when large numbers of scores are used. There is a single stage procedure which does not involve converting the raw scores to deviate scores. Instead we find the sum of squares of the raw scores and make a simple correction. The procedure is described by this formula:

$$\frac{\Sigma X^2-\frac{(\Sigma X)^2}{n}}{n}$$

and is illustrated by the example in table 2.11.

X (TEACHER GROUP)	X^2	
2	4	n=10
2	4	
2	4	$\text{Variance}=\frac{\Sigma X^2-\frac{(\Sigma X)^2}{n}}{n}$
3	9	
4	16	
4	16	$=\frac{190-\frac{40\times 40}{10}}{10}$
4	16	
6	36	
6	36	$=\frac{190-160}{10}=3{\cdot}0$
7	49	
$\Sigma X=40$	$\Sigma X^2=190$	

Comment
The variance of scores in the teacher group is 3·0.

TABLE 2.9 Calculating the variance using the single stage procedure.

A step by step analysis of the procedure runs as follows:

1. Calculate ΣX.
2. Square each of the scores and note them in column 2.
3. Calculate the sum of the squares of X; (ΣX^2).
4. Substitute (ΣX) and (ΣX^2) in the formula;

$$\text{Variance} = \frac{\Sigma X^2 - \dfrac{(\Sigma X)^2}{n}}{n}$$

The result using this procedure must be exactly the same as the result using the two-stage procedure. In this case, the resulting variance is 3·0, the same result as was found in table 2.6. The two formulae are exactly equivalent.

$$\text{Variance} = \frac{\Sigma D^2}{n} = \frac{\Sigma X^2 - \dfrac{(\Sigma X)^2}{n}}{n}$$

From this, we see that

$$\Sigma D^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{n}$$

We call $\dfrac{(\Sigma X)^2}{n}$ the CORRECTION TERM because it 'corrects' the value of ΣX^2 to equal the value of ΣD^2 which is needed for the calculation of variance.

Calculate the variance of the scores in the teaching machine group using the one-stage formula. Use the layout in table 2.12. We know that the variance is 7·4. Hopefully, the one-stage procedure will give the same result. Follow the procedure steps given above carefully. Check your results against table 2.11.

X (TEACHING MACHINE GROUP)	X^2
0	
0	
0	
2	
2	
3	
3	
5	
7	
8	
$\Sigma X =$	$\Sigma X^2 =$

$n =$

$$\text{Variance} = \frac{\Sigma X^2 - \dfrac{(\Sigma X)^2}{n}}{n}$$

$$= \frac{\qquad}{\qquad}$$

$$= \frac{\qquad}{\qquad} =$$

Comment

TABLE 2.10 Layout for variance calculation using the one-stage procedure.

X (TEACHING MACHINE GROUP)	X^2
0	0
0	0
0	0
2	4
2	4
3	9
3	9
5	25
7	49
8	64
$\Sigma X = 30$	$\Sigma X^2 = 164$

$$n = 10$$

$$\text{Variance} = \Sigma X^2 - \frac{(\Sigma X)^2}{n}$$

$$= \frac{164 - \frac{30 \times 30}{10}}{10}$$

$$= \frac{164 - 90}{10} = 7{\cdot}4$$

Comment
The variance of scores in the teaching machine group is 7·4.

TABLE 2.11 Completed version of table 2.10.

3

Estimating from samples

In chapter 2, we treated mean and variance as purely descriptive statistics. We described the examination results of the 'teacher group' as having a mean of 4·0 and a variance of 3·0. The 'teaching machine group' yielded a set of examination marks having a mean of 3·0 and a variance of 7·4. These values merely *describe* the results which we observed. Usually, we want to go a step beyond simple description; we want to predict.

Our experiment involved only 20 schoolboys but we hope that our results will be typical of *all* schoolboys. We were not concerned merely with those particular individuals; we were interested in how schoolboys in general would react to the change in teaching method. Our teacher group provided better examination results on average. Would we get the same result if we repeated the experiment but this time included every schoolboy? Our teaching machine group yielded examination results with a greater variance than the teacher group. This is an interesting finding, but can we generalize it to all schoolboys?

The general problem can be expressed simply. We know the mean and variance of a small sample. What are the mean and variance of the general population? This chapter will deal with this particular problem, carry out an experiment and draw conclusions as to the best strategy for estimating.

A sampling experiment

In order to make the discussion quite straightforward, we will create an artificial example and focus on its properties. This example has been made crudely rigid so that the calculations will be easy (for me) and easy to follow (for you). It concerns a hypothetical survey of the travelling habits of university undergraduates in a single small department of 100 students. Each student was asked to specify how many different foreign countries he had visited since the age of ten. The highest reply was 9 and the lowest 0. The results were very uniform for reasons already given and are shown in table 3.1.

There are only 100 students in this department and no more; this is a total survey. Consequently, we can calculate exactly the mean and variance. I

n = 100

0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

TABLE 3.1 Results of a survey of 100 undergraduates concerning the number of different foreign countries they had visited since the age of ten.

have already done this; *the mean is* 4·50 *and the variance* 8·25. You are invited to check these calculations if you wish.

We have learned these results at the expense of a time consuming total survey. Supposing that we attempt to estimate the mean and variance from only a small sample of say 5 answers, could we generalize from our sample to the whole population of 100 students? We could try the experiment. I selected 5 students, completely at random, from the department and put the question to them. The sample answers are given in table 3.2 along with the sample mean and sample variance.

Sample No. 1

n = 5

Answers:	4	7	3	7	2	Sample mean = 4·60 Sample variance = 4·24

TABLE 3.2 Results for random sample (number 1) drawn from the population described in table 3.1.

The sample mean of 4·60 is a good approximation of the department population mean of 4·50. Unfortunately, the sample variance of 4·24 is a gross underestimate of the correct value of 8·25. In order to get a feeling for sampling we shall repeat this process five more times. Each time we select the 5 students completely at random from the department. The results are given in table 3.3.

SAMPLE NUMBER	SAMPLE ANSWERS					SAMPLE MEAN	SAMPLE VARIANCE
2	4	3	5	3	7	4·40	2·24
3	5	8	2	6	8	5·80	4·96
4	8	0	2	3	7	4·00	9·20
5	0	4	6	5	5	4·00	4·40
6	9	5	2	5	5	5·20	4·96

TABLE 3.3 Five more random samples from the population of scores described in table 3.1.

Estimating the population mean

The general impression from these results is that the sample mean is a good estimator. The sample means are fairly closely clustered about the true mean of 4·50. It seems as if the estimate is equally likely to be an overestimate as an underestimate. It looks as if our guesses are *unbiased*. Our hunch is borne out by the results in table 3.6. Here, we have repeated our small sampling experiment one hundred times. Fifty-five estimates of the mean were in excess of 4·5 and 45 estimates were below the true average.

We know from other experiments of this nature and from purely theoretical considerations that the sample mean is an unbiased estimate of the population mean. This can be expressed more simply. In the absence of any other information, the sample mean is your BEST ESTIMATE of the population mean.

Estimating the population variance

The same does not appear to be true of the sample variance. The results in tables 3.2 and 3.3 show a strong tendency to underestimate the population variance of 8·25. The results in table 3.6 (column 3) confirm our opinion. Here, the average estimate from 100 samples is 6·59, a clear underestimate of the true value. Sixty-nine per cent of our samples yield an underestimate and only 31% an overestimate. The sample variance is obviously only a biased estimate of the population variance. Theoretical considerations too support the expectation that the sample variance will, in general, be an underestimate.

The reason for this systematic error is not difficult to spot. The variance estimate is critically dependent upon our estimate of the mean. If our calculation of the mean is in error then the variance estimate will be wrong too. Our sampling technique estimates the true mean only roughly and some error is almost guaranteed. Remember that we are trying to estimate the variance about the population mean, but we are reduced to calculating the variance about the sample mean because we do not normally know what the true population mean is.

In this artificial example we do know what the true mean is; it is 4·50. We can therefore illustrate the principle under discussion by estimating the variance in two different ways, (a) using the sample mean and (b) using the true population mean. The calculations are given in tables 3.4 and 3.5. Table 3.4 uses the sample mean of 5·80. The deviate scores are calculated by subtracting the sample mean from each individual score. Table 3.5 uses the population mean of 4·50. The deviate scores are now calculated by subtracting 4·50 from each individual score. This second procedure produces less of an underestimate of the true population variance of 8·25.

This general impression is borne out by our hundred sample experiment in table 3.6. For each sample, we estimated the population variance using the

X (RANDOM SAMPLE)	D (USING SAMPLE MEAN)	D^2	
3	−2·8	7·87	
8	2·2	4·84	
7	1·2	1·44	n = 5
8	2·2	4·84	
3	−2·8	7·84	
$\Sigma X = 29$	$\Sigma D = 0{\cdot}0$	$\Sigma D^2 = 26{\cdot}80$	

$$M = \frac{29}{5} = 5{\cdot}80 \qquad \text{Variance estimate} = \frac{26{\cdot}80}{5} = 5{\cdot}36$$

Comment
Using the sample mean (5·80) we estimate the population variance to be 5·36. (It is in fact 8·25.)

TABLE 3.4 Estimating the population variance using the sample mean (BIASED VARIANCE).

X (RANDOM SAMPLE)	D (USING 4·50 SA POPULATION MEAN)	D^2	
3	−1·50	2·25	
8	3·50	12·25	
7	2·50	6·25	n = 5
8	3·50	12·25	
3	−1·50	2·25	
	$\Sigma D = 6{\cdot}50$	$\Sigma D^2 = 35{\cdot}25$	

$$\text{Variance estimate} = \frac{\Sigma D^2}{n} = \frac{35{\cdot}25}{5} = 7{\cdot}05$$

Comment
Using the population mean (4·50) we estimate the population variance to be 7·05. (It is in fact 8·25.)

TABLE 3.5 Estimating the population variance using the population mean (TRUE MEAN VARIANCE).

true population mean (true mean variance). Our average estimate was 8·62. This yields a much better estimate than the method using the sample mean. Fifty-six per cent of our new estimates were excessive (over 8·25) and 44% underestimates. There is no strong bias present.

We cannot generalize from the sample variance to the population variance because it is an underestimate. So much we know. We also know that this is caused by the use of a value for the population mean which is slightly in error. The variance estimate can be corrected by the use of the true population mean. This procedure produces a higher estimate of the population variance. This can be checked in table 3.6, where the values in column 3 are always higher than the corresponding value in column 4.

SAMPLE SCORES	MEAN	(Using *sample* mean) BIASED VARIANCE	(Using mean = 4·5) TRUE MEAN VARIANCE	(Using N − 1) UNBIASED ESTIMATE
0,3,9,1,8	4·20	13·36	13·45	16·70
8,6,4,9,5	6·40	3·44	7·05	4·30
4,7,3,7,2	4·60	4·24	4·25	5·30
2,1,8,7,0	3·60	10·64	11·45	13·30
2,8,5,2,2	3·80	5·76	6·25	7·20
9,9,4,4,5	6·20	5·36	8·25	6·70
3,8,7,8,3	5·80	5·36	7·05	6·70
8,3,3,0,7	4·20	8·56	8·65	10·70
1,0,0,4,1	1·20	2·16	13·05	2·70
3,5,0,9,5	4·40	8·64	8·65	10·80
6,6,3,5,7	5·40	1·84	2·65	2·30
6,4,5,6,9	6·00	2·80	5·05	3·50
0,8,9,9,3	5·80	13·36	15·05	16·70
2,0,4,2,9	3·40	9·44	10·65	11·80
2,8,5,6,9	6·00	6·00	8·25	7·50
6,3,8,0,9	5·20	10·96	11·45	13·70
4,3,5,3,7	4·40	2·24	2·25	2·80
5,8,2,6,8	5·80	4·96	6·65	6·20
8,0,2,3,7	4·00	9·20	9·45	11·50
1,7,4,0,7	3·80	8·56	9·05	10·70
8,9,6,8,0	6·20	10·56	13·45	13·20
0,4,6,5,5	4·00	4·40	4·65	5·50
2,4,6,7,8	5·40	4·64	5·45	5·80
6,1,9,3,2	4·20	8·56	8·65	10·70
6,4,3,0,1	2·80	4·56	7·45	5·70
4,7,2,0,1	2·80	6·16	9·05	7·70
3,1,9,0,5	3·60	10·24	11·05	12·80
6,0,4,1,0	2·20	5·76	11·05	7·20
8,0,1,0,1	2·00	9·20	15·45	11·50
3,3,3,8,2	3·80	4·56	5·05	5·70
9,5,2,5,5	5·20	4·96	5·45	6·20
1,0,3,5,3	2·40	3·04	7·45	3·80
4,3,8,5,7	5·40	3·44	4·25	4·30
4,2,1,8,6	4·20	6·56	6·65	8·20
7,7,0,1,1	3·20	9·76	11·45	12·20
9,3,8,3,9	6·40	7·84	11·45	9·80
2,8,3,8,0	4·20	10·56	10·65	13·20
4,9,2,9,6	6·00	7·60	9·85	9·50
6,3,3,1,1	2·80	3·36	6·25	4·20
4,9,7,1,3	4·80	8·16	8·25	10·20
9,1,8,2,3	4·60	10·64	10·65	13·30
3,9,7,9,4	6·40	6·24	9·85	7·80
4,7,0,4,6	4·20	5·76	5·85	7·20
7,8,5,6,3	5·80	2·96	4·65	3·70
8,9,3,2,8	6·00	8·40	10·65	10·50
3,9,4,7,8	6·20	5·36	8·25	6·70
0,4,1,2,3	2·00	2·00	8·25	2·50
1,9,2,8,7	5·40	10·64	11·45	13·30
3,4,0,1,7	3·00	6·00	8·25	7·50
8,7,8,7,8	7·60	·24	9·85	·30
3,5,6,7,4	5·00	2·00	2·25	2·50
3,9,2,1,2	3·40	8·24	9·45	10·30
9,8,2,4,6	5·80	6·56	8·25	8·20
2,9,7,3,5	5·20	6·56	7·05	8·20
0,9,9,2,4	4·80	13·36	13·45	16·70
2,7,8,9,3	5·80	7·76	9·45	9·70
4,9,1,0,5	3·80	10·16	10·65	12·70

SAMPLE SCORES	MEAN	(Using *sample* mean) BIASED VARIANCE	(Using mean = 4·5) TRUE MEAN VARIANCE	(Using N−1) UNBIASED ESTIMATE
0,0,7,5,5	3·40	8·24	9·45	10·30
3,9,2,4,2	4·00	6·80	7·05	8·50
5,0,4,7,2	3·60	5·84	6·65	7·30
3,9,5,8,1	5·20	8·96	9·45	11·20
4,4,0,3,6	3·40	3·84	5·05	4·80
5,4,5,1,8	4·60	5·04	5·05	6·30
4,6,8,6,5	5·80	1·76	3·45	2·20
7,2,4,7,9	5·80	6·16	7·85	7·70
0,2,7,4,1	2·80	6·16	9·05	7·70
7,5,7,3,2	4·80	4·16	4·25	5·20
9,9,8,0,8	6·80	11·76	17·05	14·70
0,2,3,8,2	3·00	7·20	9·45	9·00
7,7,2,0,1	3·40	9·04	10·25	11·30
4,4,9,3,2	4·40	5·84	5·85	7·30
8,8,9,7,8	8·00	·40	12·65	·50
8,4,2,8,1	4·60	8·64	8·65	10·80
4,5,6,5,0	4·00	4·40	4·65	5·50
2,8,0,1,6	3·40	9·44	10·65	11·80
7,7,7,5,3	5·80	2·56	4·25	3·20
3,9,4,9,5	6·00	6·40	8·65	8·00
4,1,8,4,7	4·80	6·16	6·25	7·70
1,9,6,3,4	4·60	7·44	7·45	9·30
8,2,6,8,1	5·00	8·80	9·05	11·00
6,1,5,8,9	5·80	7·76	9·45	9·70
3,5,4,8,6	5·20	2·96	3·45	3·70
5,9,5,0,0	3·80	11·76	12·25	14·70
2,0,0,6,0	1·60	5·44	13·85	6·80
8,9,7,6,9	7·80	1·36	12·25	1·70
5,4,8,7,0	4·80	7·76	7·85	9·70
7,5,5,8,6	6·20	1·36	4·25	1·70
0,7,8,5,3	4·60	8·24	8·25	10·30
2,5,3,1,8	3·80	6·16	6·65	7·70
0,1,9,9,5	4·80	14·56	14·65	18·20
8,7,7,8,9	7·80	·56	11·45	·70
4,1,2,1,2	2·00	1·20	7·45	1·50
7,4,9,0,7	5·40	9·84	10·65	12·30
9,0,7,3,4	4·60	9·84	9·85	12·30
3,1,9,4,6	4·60	7·44	7·45	9·30
2,4,9,2,1	3·60	8·24	9·05	10·30
4,0,1,1,3	1·80	2·16	9·45	2·70
3,7,3,9,5	5·40	5·44	6·25	6·80
5,1,4,0,6	3·20	5·36	7·05	5·70
9,8,0,9,9	7·00	12·40	18·65	15·50
Average	4·62	6·59	8·62	8·23

TABLE 3.6 Three different estimates of population variance (8·25) based on random samples. See text.

An unbiased estimate of variance

So far, we have taken part in our sampling experiment from a privileged position. We know what the true population mean is (4·50). We know this because we have access to the total survey. There are very few experimental conditions which allow this, however, and we must seek another method of estimating the population variance from the available sample.

Theoretical considerations have led statisticians to a very simple and elegant solution. Since the sample variance is typically an underestimate, we can attempt to improve it by increasing the value slightly. There is a simple correction which works very well. The standard formula involves dividing by n to obtain the variance:

$$\text{Sample variance} = \frac{\Sigma D^2}{n}$$

We could increase our estimate slightly by dividing by $n-1$. This turns out to yield an unbiased estimate of the population variance:

$$\text{Unbiased estimate of population variance} = \frac{\Sigma D^2}{n-1}$$

The final column in table 3.6 gives the unbiased estimate of the population variance. You can see that the unbiased estimate is always greater than the sample variance (biased variance estimate), for obvious reasons. The average unbiased estimate is given as 8·23 which is an excellent approximation of the true population variance of 8·25. In future, whenever we wish to estimate the population variance from only a restricted sample *we shall always use the formula for the unbiased estimate.*

Table 3.7 gives the layout for the calculation of unbiased mean and variance estimates. The figures used are those from chapter 2, the teacher group. In this case, we assume that the 10 schoolboys are only a small sample of the total population of the kind of schoolboys we are interested in. We intend our sample to be representative of this population and we shall use our sample to generate unbiased estimates of the mean and variance of the more general population. Complete the example and check your results with table 3.8.

X (TEACHER GROUP)	D (DEVIATE SCORES)	D^2
2		
2		
2		$n = 10$
3		
4		$n-1 = 9$
4		
4		
6		
6		
7		
$\Sigma X =$	$\Sigma D =$	$\Sigma D^2 =$

$$M = \frac{\Sigma X}{n} =$$

$$\text{Unbiased variance estimate} = \frac{\Sigma D^2}{n-1} = \text{——} =$$

Comment

TABLE 3.7 Calculation of unbiased estimates of the mean and variance of a general population from a small sample.

X	D	D^2	
2	−2	4	
2	−2	4	
2	−2	4	
3	−1	1	$n = 10$
4	0	0	
4	0	0	$n-1 = 9$
4	0	0	
6	2	4	
6	2	4	
7	3	9	
$\Sigma X = 40$	$\Sigma D = 0$	$\Sigma D^2 = 30$	

$$M = \frac{\Sigma X}{n} = \frac{40}{10} = 4{\cdot}0$$

$$\text{Unbiased variance estimate} = \frac{\Sigma D^2}{n-1} = \frac{30}{9} = 3{\cdot}33$$

Comment
Our unbiased estimate of the mean is 4·0. Our estimate of the variance is 3·33.

TABLE 3.8 Completed version of table 3.7.

This small alteration to the variance formula will need to be applied to the one-stage formula too. This was:

$$\text{Sample variance} = \frac{\Sigma X^2 - \dfrac{(\Sigma X)^2}{n}}{n}$$

which becomes:

$$\text{Unbiased variance estimate} = \frac{\Sigma X^2 - \dfrac{(\Sigma X)^2}{n}}{n-1}$$

Table 3.9 gives an opportunity to use this formula in calculating the variance estimate for the teaching machine group. In future, all variance estimates will be unbiased. Accordingly, it will be convenient to drop the term unbiased. During calculations it will also be convenient to contract the expression 'variance estimate' to 'var. est.'

X TEACHING MACHINE GROUP	X^2	
0		$n =$
0		
0		$\text{Variance estimate} = \dfrac{\Sigma X^2 - \dfrac{(\Sigma X)^2}{n}}{n-1}$
2		
2		
3		$= \dfrac{\qquad\qquad}{\qquad\qquad}$
3		
5		$= \dfrac{\qquad\qquad}{\qquad\qquad} =$
7		
8		
—	—	
$\Sigma X =$	$\Sigma X^2 =$	

Comment

TABLE 3.9 Calculation of an unbiased estimate of variance using the one-stage formula.

Degrees of freedom

The change in the variance formula may seem slight enough, a change from n to $n-1$, but it is symptomatic of a very important concept in statistics. This concept deals with *degrees of freedom.*

Calculation of variance depends upon an estimate of the value of the mean but not vice versa. We say that the mean is free, but the variance is restricted by the sampling fluctuation of estimates of the mean. In this case the variance has lost one degree of freedom. There are occasions when additional restraints are imposed upon the variance and more degrees of freedom are lost.

In the calculation of variance estimates, we subtract one from the sample size and divide the sum of the squared deviates by $n-1$. It is not a coincidence that we should subtract one. If for some reason we had lost two degrees of freedom then we would have subtracted two. Under normal circumstances, a variance estimate has $n-1$ degrees of freedom. If there are ten scores in the sample, then there are $10-1$ or 9 degrees of freedom. Our formula for variance estimates can now be written in a more general form.

$$\text{Variance estimate} = \frac{\Sigma D^2}{n-1} = \frac{\Sigma D^2}{\text{Degrees of freedom}} = \frac{\Sigma D^2}{\text{df.}}$$

where df. = degrees of freedom. We shall see a great deal of this formula in the future.

There is another helpful way of looking at degrees of freedom. Consider the sample scores from the teacher group of examination results. Their mean value is 4·0, so the 10 scores must add up to a total of 40. If we know the first nine scores then we can predict the tenth score. In this case, the

tenth score must be 7 because the other scores add up to 33. As soon as we know the grand total, one of the scores becomes fixed. It is no longer free to vary; we have lost one degree of freedom and are left with only nine.

$$2,2,2,3,4,4,4,6,6, \textcircled{7} \text{ TOTAL} = 40.$$

X	X^2
0	0
0	0
0	0
2	4
2	4
3	9
3	9
5	25
7	49
8	64
$\Sigma X = 30$	$\Sigma X^2 = 164$

$$n = 10$$

$$\text{Variance estimate} = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n-1}$$

$$= \frac{164 - \frac{30 \times 30}{10}}{9}$$

$$= \frac{164 - 90}{9} = 8{\cdot}22$$

Comment
The unbiased estimate of variance for scores in the teaching machine group is 8·22.

TABLE 3.10 Completed version of table 3.9.

4

Comparing two variances

In chapter 2, we discussed an experiment comparing the effectiveness of two types of instruction (teachers and teaching machines). There were 10 boys in each experimental group and their examination results were scored along a scale from 0 to 10. In chapter 3, we arrived at final unbiased estimates of the mean and variance of the populations from which they were presumably selected. The results were as follows:

	TEACHER GROUP	MACHINE GROUP
Mean	4·0	3·0
Unbiased variance estimate	3·33	8·22

There was a small difference between the *means* of the two groups but a very large difference between the two *variance estimates*. This final difference (which had been predicted in chapter 1) may make us want to test the hypothesis that teaching machine instruction is more variable in its effectiveness than normal teaching methods.

There is a straightforward statistical procedure for assessing the idea that two variances are significantly different, i.e., they differ more than can be accounted for by trivial factors such as sampling fluctuation or random experimental error. This procedure is called the F-test in honour of R. A. Fisher who pioneered it.

The first step in this test is to construct the F-ratio by dividing the expected larger variance estimate by the expected smaller:

$$\text{F-ratio} = \frac{\text{Predicted larger variance estimate}}{\text{Predicted smaller variance estimate}}$$

In the case of our example:

$$\text{F-ratio} = \frac{\text{Machine group variance estimate}}{\text{Teacher group variance estimate}} = \frac{8{\cdot}22}{3{\cdot}33} = 2{\cdot}47$$

The second step is to decide whether the value 2·47 is large enough to justify a decision in favour of the hypothesis. It must be remembered that we cannot expect the two variance estimates to be exactly equal, even if machine

and teacher instruction are broadly similar in their effect. They will undoubtedly differ to a small degree.

Table 3.6 in chapter 3 shows that two variance estimates can often be very different even when based on samples drawn from the same population. *Thus we must expect variance estimates to differ for purely trivial reasons*, that is to say, for reasons essentially unrelated to what the experiment is about. In table 3.6 the variance estimates may be said to result from sampling fluctuations despite the fact that all the samples were drawn from the same population.

Whenever we calculate an F-ratio we are faced with two mutually exclusive possibilities or hypotheses. The first possibility reads as follows:

Null hypothesis The two samples have been drawn from the same population. Any apparent differences have been caused by trivial factors such as sampling fluctuations.

The second possibility reads as follows:

Alternative hypothesis The observed difference between the two variance estimates is due to a real effect. The samples can be said to be drawn from two different populations.

A decision has to be made. We must reject one hypothesis and accept the other. It is a healthy practice to assume that the null hypothesis is true until there is strong evidence to cause us to reject it. In this respect, Fisher's tables or F-tables are vital in helping us to assess the evidence.

Fisher's F-tables

These tables are given in the appendix. The tables make one simple assumption, that the two samples are drawn from the same population. Under these circumstances, we expect the F-ratio to equal one but from time to time it could be much higher or lower depending upon random fluctuations in sampling.

For example, if we had two independent random samples each of 10 scores, the F-ratio has a 1 in 20 chance of being greater than or equal to 3·18. In other words, if we took a hundred pairs of such samples we would expect 5% of them to have an F-ratio greater than 3·18. This figure the author was able to derive from the F-tables in the appendix, in the manner described below.

In the example above, we had two samples (teacher vs. machine) with 10 scores in each. The F-ratio we obtained was 2·47 which is much smaller than 3·18. This means that we would get an F-ratio of 2·47 more often than on 5% of occasions even when the samples were drawn from the same population. Accordingly, the evidence is not strong enough to cause us to abandon the null hypothesis. The F-ratio is too small; it is statistically not significant.

The F-tables are easy to use. There is one table for each of the main significance levels (5%, 2·5%, 1% and 0·1%). The 5% tables gives us those extreme values of F which we expect to obtain on approximately 5% of occasions even when the null hypothesis is true, i.e., even when the difference in variance estimates is due to trivial random factors.

The 1% tables gives us that value of F which is exceeded on at least 1% of occasions even when there is no real difference between the groups. Most scientists insist that the F-value obtained in an experiment should exceed the value given in the 1% tables before it can be considered as statistically significant. In certain difficult research areas, this criterion is often relaxed and of course these criteria are arbitrary standards. The level of 'significance' which has to be achieved before you, personally, are convinced is your own affair.

In order to locate the critical F-ratio for different sizes of samples, we use the degrees of freedom associated with each sample. In our case, we had 10 scores in each group and 9 degrees of freedom were associated with each sample. The degrees of freedom for the sample on the top of the ratio is given along the top of the table, and this determines the column we are to use. The row to be used is determined by the degrees of freedom associated with the sample on the bottom of the ratio.

The following shorthand symbols refer to the degrees of freedom:

v_1 symbolizes the degrees of freedom associated with the upper variance estimate.

v_2 symbolizes the degree of freedom associated with the lower variance estimate.

We write the F-ratio thus:

$$F_{v1,\,v2} = \frac{\text{Predicted larger variance estimate (1)}}{\text{Predicted smaller variance estimate (2)}}$$

In our example:

$$F_{9,9} = \frac{8{\cdot}22}{3{\cdot}33} = 2{\cdot}47$$

We should now consult the F-tables for 9 and 9 degrees of freedom to find which values of F need to be exceeded to be significant at the different percentage levels. Consulting the tables for $F_{9,9}$, at

5·0% level F = 3·1
2·5% level F = 4·0
1·0% level F = 5·3
0·1% level F = 10·1

Our obtained F-value of 2·47 is smaller than all of these. Accordingly, we describe the result as statistically non-significant and, consequently, we retain

the null hypothesis. We are not convinced that teaching machine instruction produces more variable results than teacher instruction. Whatever the scores on this occasion, the difference between the variances is simply not large enough to convince us that we would always get the same result.

Example 4.1 *Anonymity in opinion polls.* An investigator was studying the kinds of bias which affected the kind of replies given to public opinion pollsters in the street. He felt sure that answers varied according to whether or not the interviewee was to remain anonymous. Accordingly, he went out to the main street and interviewed 28 passers-by whom he had selected at random. He asked each person 20 questions which he had taken from a 'tough-mindedness' scale. Each question was an invitation to agree or disagree with a tough-minded statement such as the following:

'Children ought to be seen and not heard.'
'Sexual offenders ought to be flogged.'
'Foreign immigrants ought to be sent home.'

The interviewee scored one point for each statement he agreed with. His total score was called his tough-mindedness rating.

Thirteen subjects were interviewed under strictly anonymous conditions. These subjects constitute the 'anonymous' group. The remaining fifteen subjects belonged to the 'open-identity' group. They were asked to give their names and addresses before the questioning took place. The scores for both groups are given in table 4.1.

GROUP A OPEN-IDENTITY Sample size = 15 Degrees of freedom = 15−1 = 14			GROUP B ANONYMOUS Sample size = 13 Degrees of freedom = 13−1 = 12		
TOUGH-MINDEDNESS SCORE			TOUGH-MINDEDNESS SCORE		
X	D	D^2	X	D	D^2
6	−4	16	2	−8	64
8	−2	4	6	−4	16
8	−2	4	6	−4	16
8	−2	4	6	−4	16
8	−2	4	8	−2	4
10	0	0	10	0	0
10	0	0	10	0	0
10	0	0	10	0	0
10	0	0	12	2	4
12	2	4	14	4	16
12	2	4	14	4	16
12	2	4	16	6	36
12	2	4	16	6	36
12	2	4			
12	2	4			
$\Sigma X = 150$	$\Sigma D = 0$	$\Sigma D^2 = 56$	$\Sigma X = 130$	$\Sigma D = 0$	$\Sigma D^2 = 224$
MEAN = 150/15 = 10 points VARIANCE ESTIMATE = $\Sigma D^2/df.$ $= \frac{56}{14} = 4$ points			MEAN = 130/13 = 10 points VARIANCE ESTIMATE = $\Sigma D^2/df.$ $= \frac{224}{12} = 18{\cdot}7$ points		

TABLE 4.1 Tough-mindedness scores for two groups of anonymous and open-identity interviewees.

Clearly, there is little difference between the average score of the two groups. There is, however, a considerable difference between the variance estimates. It looks as if the open-identity group is conforming more and refusing to give extreme tough- or tender-minded scores. The two variance scores, 4 points and 18·8 points, are very different but are they significantly different? Perhaps we would get a similar discrepancy quite often merely as the result of sampling fluctuations. Fisher's F-test is most appropriate here:

$$F_{12,\,14} = \frac{\text{Larger variance estimate}}{\text{Smaller variance estimate}} = \frac{18{\cdot}7}{4} = 4{\cdot}7$$

Consulting the tables for $F_{12,\,14}$ at:

5% level F = 2·5
2·5% level F = 3·0
1·0% level F = 3·8
0·1% level F = 6·1

Comment

Our calculated value for F (4·7) lies between the 1% and 0·1% levels of significance.

We therefore *reject* the idea that the variance of estimates of the two groups could come from the same population.

Example 4.2 *Trunk road advertising.* Do hoardings along busy main traffic routes get read? If so by whom are they read? A researcher chose a 2-mile stretch of main road between two major towns which contained a series of 10 gigantic hoarding advertisements. He then selected two groups of passengers—car passengers and bus passengers—for study. He interviewed bus passengers who alighted at the first bus stop after the end of the 2-mile stretch and car passengers in cars which stopped for petrol in a nearby filling station. Both types of passenger were asked to recall as many of the advertised products as they could. The results given below list the total number of products remembered by only those passengers who had never made the journey before. Both groups show the same average level of recall, but the bus passenger group seems to be more variable. Is the difference in variability significant?

GROUP A CAR PASSENGERS
Sample size =
Degrees of freedom = =

NUMBER OF ADVERTISEMENTS RECALLED

X	D	D^2
4		
4		
4		
4		
1		
5		
6		
$\Sigma X =$	$\Sigma D =$	$\Sigma D^2 =$

$$\text{MEAN} = \frac{\Sigma X}{N} = \text{——} =$$

$$\text{VARIANCE ESTIMATE} = \frac{\Sigma D^2}{df.} = \text{——} =$$

GROUP B BUS PASSENGERS
Sample size =
Degrees of freedom =
=

NUMBER OF ADVERTISEMENTS RECALLED

X	D	D^2
6		
3		
1		
9		
8		
2		
2		
2		
3		
$\Sigma X =$	$\Sigma D =$	$\Sigma D^2 =$

$$\text{MEAN} = \frac{\Sigma X}{N} = \text{——} =$$

$$\text{VARIANCE ESTIMATE} = \frac{\Sigma D^2}{df.} = \text{——} =$$

$$F_{\quad,} = \frac{\text{Larger variance estimate}}{\text{Smaller variance estimate}} = \text{——} =$$

Consulting the tables for $F_{\quad,}$ at:

5% level =
2·5% level =
1·0% level =
0·1% level =

Comment

TABLE 4.2 Recall of advertisements for car and bus passengers.

GROUP A CAR PASSENGERS
Sample size = 7
Degrees of freedom = 7 − 1 = 6

NUMBER OF ADVERTISEMENTS RECALLED

X	D	D^2
4	0	0
4	0	0
4	0	0
4	0	0
1	−3	9
5	+1	1
6	+2	4
$\Sigma X = 28$	$\Sigma D = 0$	$\Sigma D^2 = 14$

$$\text{MEAN} = \frac{28}{7} = 4$$

$$\text{VARIANCE ESTIMATE} = \frac{14}{6} = 2{\cdot}33$$

$$F_{8,6} = \frac{8{\cdot}5}{2{\cdot}33} = 3{\cdot}65$$

Consulting the tables for $F_{8,6}$ at:

5% level = 4·1
2·5% level = 5·6
1·0% level = 8·1
0·1% level = 19·0

GROUP B BUS PASSENGERS
Sample size = 9
Degrees of freedom = 9 − 1 = 8

NUMBER OF ADVERTISEMENTS RECALLED

X	D	D^2
6	+2	4
3	−1	1
1	−3	9
9	+5	25
8	+4	16
2	−2	4
2	−2	4
2	−2	4
3	−1	1
$\Sigma X = 36$	$\Sigma D = 0$	$\Sigma D^2 = 68$

$$\text{MEAN} = \frac{36}{9} = 4$$

$$\text{VARIANCE ESTIMATE} = \frac{68}{8} = 8{\cdot}5$$

Comment
Our calculated value for F (3·65) is *less* than the value required at the 5% level of significance. It is therefore quite likely that the difference between the two sample variances does not represent a real effect but merely a sampling fluctuation.

We *reject* the notion that the variability of advertisement recall for car and bus passengers is different.

TABLE 4.3 Worked version of table 4.2.

5

The difference between means

In chapter 4, we saw how to test the significance of the difference between two variances. More often, we are concerned to decide on the significance of the two groups. In chapter 2, we studied the examination results of schoolboys who had been taught a topic by two different methods; standard teaching and teaching machines. Most students of educational practice will ask the questions 'Which is better?'; 'Which gives the better examination results?' In our example, the teacher group was better, giving an average examination result of 4·0 compared with an average of 3·0. Does this mean that the teaching method is superior? If we were to repeat the same experiment with a different random selection of boys, should we expect to get a similar result? In other words, how significant is the difference between the means?

This problem arises again and again in experiments. We have difficulty in knowing how confidently we may generalize our results. In this case, we have used only a small sample of boys but we would like to generalize our results to cover all similar schoolboys. We never know for sure just how well balanced our samples were. Perhaps we were lucky in having slightly brighter boys in one group than another as the result of random selection. Perhaps some of the boys in the teaching machine group were slightly off-colour on the day of the examination.

Analysis of variance is the technique which comes to the rescue at this point. Just as it allows us to test the significance of a difference between variances, so it helps us to test the differences between a set of means. It is, of course, impossible to decide in absolute terms how large a difference needs to be before it becomes significant. Instead, analysis of variance uses a relative measure which is quite easy to grasp intuitively even though the full mathematical analysis is complex. The following two examples should illustrate the basic principle.

Example 5.1 Table 5.1 gives the results of an imaginary experiment in persuasion. The two groups represent 'hard sell' and 'soft sell' encyclopedia salesmen. The total number of sales over a three-month period for each salesman is given. The 'soft sell' group has a better average record than the 'hard sell' team. Does this mean that 'soft sell' is better and that future

promotions should be based on it? Obviously not; the difference between the two groups is small relative to the large fluctuations within the groups. When the basic measure is highly variable we expect a certain amount of fluctuation in the group means. We call the variability of the scores within the groups the *within groups variance*. The difference between the means we call the *between groups variance*. We can define our relative measure as:

$$\frac{\text{Between groups variance}}{\text{Within groups variance}}$$

In this case, the relative measure is small because the within groups variance is so large.

HARD SELL	SOFT SELL
2	4
8	10
11	13
12	14
16	18
17	19
18	20
—	—
MEAN = 12	MEAN = 14

TABLE 5.1 Total number of sales over a three-month period for two groups of encyclopedia salesmen. (High within groups variance example.)

Table 5.2 gives an analogous set of scores but the significance of the difference has altered radically. Notice that the means of the two groups are the same as in the previous example. The variability of the scores within each group is considerably reduced, however. Under these new conditions, the difference between the sales record of the two groups looks all the more impressive. In this case, the relative measure is large.

HARD SELL	SOFT SELL
10	11
11	14
12	14
12	14
12	14
13	15
14	16
—	—
MEAN = 12	MEAN = 14

TABLE 5.2 Encyclopedia sales (low within groups variance examples).

Almost the whole business of analysis of variance is to reduce problems to two variance estimates which can be compared using the F-ratio. In this example, the F-ratio may be expressed simply:

$$F = \frac{\text{Between groups variance estimate}}{\text{Within groups variance estimate}}$$

If the F-ratio is significantly large then we say that the differences between the means are significant. We infer that the average scores for the groups are more different than the inherent variability of the scores would lead us to expect. In other words, the difference between the means is due to something other than random sampling fluctuations.

The rest of this chapter is devoted to procedures for calculating the between groups variance and the within groups variance. These procedures will be illustrated with data from a hypnosis experiment.

The illustrative example for this chapter concerns an experiment with three different groups which arose in connection with a hypnosis project. The investigators believed that hypnosis would be much more effective if subjects were encouraged to believe that hypnosis induction would be quick and easy. They tried to instill this belief by giving subjects a 'hypnosis' pill ten minutes before the hypnotic trance was due to be induced. The 'hypnosis' pill was nothing more than a sugar pill but subjects were unaware of this. A second group of subjects were given the 'hypnosis' pill in the same way but in this case it was accompanied by a short lecture designed to convince the recipient of the value of this newly discovered pill! A third group was used as a control; subjects in this group received neither pill nor lecture.

Fifteen male subjects were used. Each subject was allotted at random to one of the three experimental groups so that there were five subjects in each group. The hypnotic trance was induced with a specially prepared tape recording inviting the subject to relax and sleep. The subject sat alone in a warm, dimly lit room with a small recording instrument on his throat which registered each time the subject swallowed. It is commonly observed that hypnotized individuals do not swallow as spontaneously as non-hypnotized people do when sitting calmly. Accordingly, the researcher measured the time elapsing before the last swallow as the time taken for the suggestions to take effect. The results are given in table 5.3.

GROUP	PILL+TALK	PILL ONLY	CONTROL	
	2	3	5	
	4	6	6	
	6	6	7	
	5	8	8	
	8	7	9	GRAND MEAN
MEAN	5	6	7	6

TABLE 5.3 Time elapsing before last spontaneous swallowing in hypnosis induction experiment.

The detailed procedure for calculating the between and within groups variance is given on page 36. At this point, however, it will be worth while

taking a little time to indicate the general line of thinking which underlies these procedures. No new concepts are introduced, but the old one of variance will be extended somewhat.

Within groups variance

We may assess the overall variability within the various groups by averaging the variance within each individual group:

Within groups variance estimate =

$$\frac{(\text{Variance of group 1} + \text{Variance of group 2} + \text{Variance of group 3})}{\text{Number of groups}}$$

If we express each score as a deviation from its own group mean (D_W), this formula becomes:

$$\text{W.G. var. est.} = \frac{\frac{\Sigma D_W^2}{\text{df.}} + \frac{\Sigma D_W^2}{\text{df.}} + \frac{\Sigma D_W^2}{\text{df.}}}{\text{Number of groups}} = \frac{\Sigma\Sigma D_W^2}{\text{Number of groups} \times \text{df.}}$$

We call the top of this expression ($\Sigma\Sigma D_W^2$), the within groups sum of squares (W.G.S.S.). This means add up *all* the deviate scores in *all* of the groups. For convenience, we will write it as merely ΣD_W^2. We call the bottom term (Number of groups × df.) the within groups degrees of freedom. Whence

$$\text{W.G. var. est.} = \frac{\text{W.G.S.S.}}{\text{W.G. df.}} = \frac{\Sigma D_W^2}{\text{W.G. df.}}$$

Between groups variance

At this point, we ignore the variance within the groups and concentrate on the variance between the groups. We can do this by setting each score equal to its group mean and calculating the difference between this score and the grand mean. This is a between groups deviation (D_B). If we add up the squares of these values we can estimate the between groups variance.

$$\text{B.G. var. est.} = \frac{\text{B.G.S.S.}}{\text{B.G. df.}} = \frac{\Sigma D_B^2}{\text{B.G. df.}}$$

At first, this may look arbitrary but the measure works well and reflects accurately the size of the differences between the groups. We will be looking at the number of degrees of freedom later in the chapter.

Total variance

For completeness, we should also calculate the total variability of the scores. To do this, imagine that all scores belong to the same group. Express each

score as a deviation from the grand mean (D_T)—not the group mean—and then calculate the total sum of squares (T.S.S.). The total variance estimate follows in the usual way:

$$\text{Total var. est.} = \frac{\text{T.S.S.}}{\text{Total df.}} = \frac{\Sigma D_T^2}{\text{Total df.}}$$

Partitioning of variance

If we have carried through our calculations without error, then the following should be true:

$$\text{T.S.S.} = \text{W.G.S.S.} + \text{B.G.S.S.}$$

That is, the total sum of squares has been sub-divided into two complementary parts. It is the two complementary parts which are to be compared.

Step by step procedure for constructing table 5.4

1. Draw up the table of raw scores.
2. Calculate the means for each group and the grand mean M.
3. Express each score:

   (a) as a deviation from the grand mean: column 3—D_T.
   (b) as a deviation from its group mean: column 5—D_W.
   (c) as a deviation of its group mean from the grand mean: column 7—D_B.

   For example, the final score in group C is 9, the grand mean is 6; its group mean is 7.

$$\begin{aligned} D_T &= 9-6 = +3 \\ D_W &= 9-7 = +2 \\ D_B &= 7-6 = +1 \end{aligned}$$

   CHECK that, for each score:

$$\begin{aligned} D_T &= D_W + D_B \\ 3 &= 2+1 \end{aligned}$$

4. Square each of the deviation scores (D_T, D_W and D_B) and note their values in columns 4, 6 and 8 respectively.
5. By summing columns 4, 6 and 8 find

$$\Sigma D_T^2,\ \Sigma D_W^2 \text{ and } \Sigma D_B^2$$

6. CHECK that $\Sigma D_T^2 = \Sigma D_W^2 + \Sigma D_B^2$

Column 1	2	3	4	5	6	7	8
GROUP	RAW SCORE X	TOTAL DEVIATION D_T	D_T^2	WITHIN GROUP DEVIATION D_W	D_W^2	BETWEEN GROUP DEVIATION D_B	D_B^2
Group A	2	−4	16	−3	9	−1	1
	4	−2	4	−1	1	−1	1
	6	0	0	+1	1	−1	1
	5	−1	1	0	0	−1	1
	8	+2	4	+3	9	−1	1
Group totals	25	(Grand mean = 6)	$\Sigma D_T^2 = 25$	(Group mean = 5)	$\Sigma D_W^2 = 20$		$\Sigma D_B^2 = 5$
Group B	3	−3	9	−3	9	0	0
	6	0	0	0	0	0	0
	6	0	0	0	0	0	0
	8	+2	4	+2	4	0	0
	7	+1	1	+1	1	0	0
Group totals	30	(Grand mean = 6)	$\Sigma D_T^2 = 14$	(Group mean = 6)	$\Sigma D_W^2 = 14$		$\Sigma D_B^2 = 0$
Group C	5	−1	1	−2	4	+1	1
	6	0	0	−1	1	+1	1
	7	+1	1	0	0	+1	1
	8	+2	4	+1	1	+1	1
	9	+3	9	+2	4	+1	1
Group totals	35	(Grand mean = 6)	$\Sigma D_T^2 = 15$	(Group mean = 7)	$\Sigma D_W^2 = 10$		$\Sigma B_B^2 = 5$
GRAND TOTAL		$\Sigma D_T^2 = 54$		$\Sigma D_W^2 = 44$		$\Sigma D_B^2 = 10$	

WITHIN GROUPS sum of squares = 44 (W.G.S.S.)
BETWEEN GROUPS sum of squares = 10 (B.G.S.S.)

TOTAL sum of squares = 54 (T.S.S.)

Notes:

D_T = difference of a raw score from the GRAND MEAN.
D_W = difference of a raw score from its own GROUP MEAN.
D_B = difference of a score's GROUP MEAN from its GRAND MEAN.

In any row,

$$D_W + D_B = D_T$$

TABLE 5.4 Dividing the total squared deviation into two components, within groups and between groups deviation.

Degrees of freedom

Having calculated the appropriate sums of squares, we will soon be in a position to calculate the variance estimates:

$$\text{Total variance estimate} = \frac{\text{Total s.s.}}{\text{Total df.}}$$

$$\text{Within groups variance estimate} = \frac{\text{W.G.S.S.}}{\text{W.G. df.}}$$

$$\text{Between groups variance estimate} = \frac{\text{B.G.S.S.}}{\text{B.G. df.}}$$

But first we need to know how many degree of freedom apply to each sum of squares. You may remember that we need to divide by degrees of freedom to produce an unbiased estimate of variance. The present examples are no exception. As a general rule, we avoid bias by dividing the sum of squares by the number of scores which are free to vary.

We will look at the three variance estimates one at a time to assess the appropriate degrees of freedom. In each case, we express the scores as deviate scores (D_T, D_W and D_B respectively).

(a) Total variance All of the deviates from the grand mean (D_T) are free to vary except one. This one is fixed by the necessity to sum to zero (see chapter 2 page 22).

group A	−4	−2	0	−1	+2	
group B	−3	0	0	+2	+1	
group C	−1	0	+1	+2	(+3)	$\Sigma D_T = 0$

In this case, there are (15−1) = 14 degrees of freedom.
General rule: Total variance has N−1 df.
(where N is the total number of scores).

(b) Within groups variance *Within each group*, one of the deviations from the group mean is fixed by the necessity to sum to zero.

A	−3	−1	+1	0	(+3)	$\Sigma D_W = 0$
B	−3	0	0	+2	(+1)	$\Sigma D_W = 0$
C	−2	−1	0	+1	(+2)	$\Sigma D_W = 0$

So we lose three degrees of freedom leaving, in this case, 12 df.
General rule: Within groups variance df. = N − Number of groups
In this case, we have (15−3) = 12 df.

(c) Between groups variance There are even fewer degrees of freedom here since all values within a group must be the same as the first one. Furthermore, they must all sum to zero so the values in the last group are all fixed

A	−1	(−1)	(−1)	(−1)	(−1)	
B	0	(0)	(0)	(0)	(0)	
C	(+1)	(+1)	(+1)	(+1)	(+1)	$\Sigma D_B = 0$

leaving only 2 df!

General rule: Between group variance df. = Number of groups − 1
In this case, we have (3 − 1) = 2 df.
Finally, CHECK Total df. = W.G. df. + B.G. df.

$$14 = 12 + 2$$

$$\text{So B.G. var. est.} = \frac{\Sigma D_B{}^2}{\text{df.}} = \frac{10}{2} = 5{\cdot}00$$

$$\text{W.G. var. est.} = \frac{\Sigma D_W{}^2}{\text{df.}} = \frac{44}{12} = 3{\cdot}67$$

At last we are ready to compare the variance estimates:

$$F_{v1,\,v2} = \frac{\text{Between groups variance estimate}}{\text{Within groups variance estimate}}$$

In our case, the degrees of freedom are:

$$v_1 = 2 \text{ and}$$
$$v_2 = 12$$

$$\text{So, } F_{2,12} = \frac{5{\cdot}00}{3{\cdot}67} = 1{\cdot}36$$

Consulting the F-tables for 2 and 12 degrees of freedom, we have:

$$F_{2,\,12} \text{ at } 5{\cdot}0\% \text{ level} = 3{\cdot}9$$
$$F_{2,\,12} \text{ at } 2{\cdot}5\% \text{ level} = 5{\cdot}1$$
$$F_{2,\,12} \text{ at } 1{\cdot}0\% \text{ level} = 6{\cdot}9$$
$$F_{2,\,12} \text{ at } 0{\cdot}1\% \text{ level} = 13{\cdot}0$$

Clearly, our value for F is too small. Such a value is quite likely to occur when there is no treatment effect. We may not reject the null hypothesis, so we reject the idea that 'suggestion' was effective in speeding onset of the hypnotic trance. The results are summarized in a *summary table* (table 5.5).

SOURCE OF VARIANCE	SUM OF SQUARES	d.f.	VARIANCE ESTIMATE	F
Between groups (suggestion)	10	2	5·00	1·36
Within groups (error)	44	12	3·67	
TOTAL	54	14		

TABLE 5.5 Summary table for one-way analysis of variance. The effect of suggestion on delay to sleep onset.

If our results had been quite different, yielding a significant F-ratio, we might conclude that suggestion did affect delay to sleep onset. It would not specify which type of suggestion was effective nor whether the two types of suggestion were significantly different from each other. We would need an additional test if we wished to be more specific. Analysis of variance merely assesses the differences between the groups in terms of whether such a set of differences could be attributable to random error or not.

Example 5.2 Example 5.2 is an experiment to investigate the influence of sex of jury members on their speed in coming to a decision. The experiment involved setting up a mock trial where the defendant—who is charged with theft—is clearly guilty. The jury is them timed to see how long it takes to form a unanimous verdict. The results are compared for two main conditions. In the first condition, all members of the jury are of the same sex as the defendant. In the second condition, all member of the jury are of the opposite sex to the defendant. The times for 12 juries are given in table 5.6.

Same-sex jury
5, 6, 0, 3, 2, 2
Opposite-sex jury
9, 7, 7, 5, 9, 5

TABLE 5.6 Time in minutes taken by each jury to reach a unanimous verdict of 'guilty'.

The results suggest that juries are much quicker to convict members of their own sex. We must, however, test these results using analysis of variance. Carry out the tests, filling in the blanks. Check your results against the worked examples. Follow carefully the procedure laid down in the previous pages.

Group	I	II
	5	9
	6	7
	0	7
	3	5
	2	9
	2	5
	—	—
$M_1 =$		$M_2 =$
GRAND MEAN =		

		TOTAL DEVIATION		WITHIN GROUP DEVIATION		BETWEEN GROUP DEVIATION	
Group	X	D_T	D_T^2	D_W	D_W^2	D_B	D_B^2
I	5						
	6						
	0						
	3						
	2						
	2						
Group totals							
II	9						
	7						
	7						
	5						
	9						
	5						
Group totals							
GRAND TOTALS		$\Sigma D_T^2 =$		$\Sigma D_W^2 =$		$\Sigma D_B^2 =$	

CHECK: $\Sigma D_T^2 = \Sigma D_W^2 + \Sigma D_B^2$

$= \quad +$

Degrees of freedom

TOTAL df. =

WITHIN GROUPS S.S. df. =

BETWEEN GROUPS S.S. df. =

Variance estimates

WITHIN GROUP variance $= \dfrac{\Sigma D_W^2}{\text{W.G. df.}} =$

BETWEEN GROUP variance $= \dfrac{\Sigma D_B^2}{\text{B.G. df.}} =$

$$F = \frac{\text{BETWEEN GROUP variance estimate}}{\text{WITHIN GROUP variance estimate}} =$$

Analysis of variance summary

SOURCE	SUM OF SQUARES	df.	VARIANCE ESTIMATE	F
Between groups				
Within groups				
Total				

Consulting the tables for F , at:

5% level F = 2·5% level F =

1·0% level F = 0·1% level F =

Comment

Group I	II
5	9
6	7
0	7
3	5
2	9
2	5
18	42

$M_1 = 3 \quad M_2 = 7$

GRAND MEAN = 5

		TOTAL DEVIATION		WITHIN GROUP DEVIATION		BETWEEN GROUP DEVIATION	
Group	X	D_T	D_T^2	D_W	D_W^2	D_B	B_B^2
I	5	0	0	+2	4	−2	4
$M_1 = 3$	6	+1	1	+3	9	−2	4
	0	−5	25	−3	9	−2	4
	3	−2	4	0	0	−2	4
	2	−3	9	−1	1	−2	4
	2	−3	9	−1	1	−2	4
Group totals	18		48		24		24
II	9	+4	16	+2	4	+2	4
$M_2 = 7$	7	+2	4	0	0	+2	4
	7	+2	4	0	0	+2	4
	5	0	0	−2	4	+2	4
	9	+4	16	+2	4	+2	4
	5	0	0	−2	4	+2	4
Group totals	42		40		16		24
GRAND TOTALS	60	$\Sigma D_T^2 = 88$		$\Sigma D_W^2 = 40$		$\Sigma D_B^2 = 48$	

CHECK: $\Sigma D_T^2 = \Sigma D_W^2 + \Sigma D_B^2$

$88 = 40 + 48$

Degrees of freedom

$$\begin{aligned} \text{TOTAL} \quad & \text{df.} = 12-1 = 11 \\ \text{WITHIN GROUPS} \quad & \text{df.} = 12-2 = 10 \\ \text{BETWEEN GROUPS} \quad & \text{df.} = 2-1 = 1 \end{aligned}$$

Variance estimates

$$\text{WITHIN GROUP variance} = \frac{\Sigma D_W{}^2}{\text{W.G. df.}} = \frac{40}{10} = 4$$

$$\text{BETWEEN GROUP variance} = \frac{\Sigma D_B{}^2}{\text{B.G. df.}} = \frac{48}{1} = 48$$

$$F_{1,\,10} = \frac{\text{BETWEEN GROUP variance}}{\text{WITHIN GROUP variance}} = \frac{48}{4} = 12$$

Analysis of variance summary

SOURCE	SUM OF SQUARES	df.	VARIANCE ESTIMATE	F
Between groups	48	1	48	12
Within groups	40	10	4	
Total	88	11		

Consulting the tables for $F_{1,\,10}$ at:

$$\begin{aligned} 5\% \text{ level } F &= 4{\cdot}7 \\ 2{\cdot}5\% \text{ level } F &= 6{\cdot}9 \\ 1{\cdot}0\% \text{ level } F &= 10{\cdot}0 \\ 0{\cdot}1\% \text{ level } F &= 21{\cdot}0 \end{aligned}$$

Comment

Our value of F is significant at better than the 1·0% level. We reject the null hypothesis and accept that the difference in scores between the two groups is representative of a real effect.

6

One factor analysis of variance (quick method)

The procedures involved in the two-stage analysis of variance are too cumbersome for frequent use with large sample sizes. For these purposes, I recommend the standard quick method for one factor analysis of variance. This procedure produces *exactly the same results*, but much more quickly. It does not involve the lengthy conversion of raw scores to deviate scores; it uses, instead, raw scores throughout. The only new feature is the calculation of a correction term (C), where

$$C = \frac{(\text{Grand total of all scores})^2}{\text{Total number of scores}} = \frac{T^2}{N}$$

The procedure is quite mechanical and is outlined below; follow each of the steps in the example which uses the same scores as example 5.1 in chapter 5. The number of scores in each group is n_1, n_2, n_3, etc., (depending on the number of groups). Follow the procedure carefully using the example overleaf.

Example 6.1 The procedure here is:

1. Calculate the group totals; t_1, t_2, t_3, etc.
2. Calculate the grand total; $T = t_1 + t_2 + t_3 + \ldots +$, etc.
3. Calculate the correction term C;

$$C = T^2/N$$

(where N is the total number of scores in the whole experiment).

4. Calculate the sum of squares of the raw scores (ΣX^2).
5. Calculate the total sum of squares (T.S.S.);

$$\text{T.S.S.} = \Sigma X^2 - C$$

(Put this in the summary table.)

6. Calculate the between group sum of squares (B.G.S.S.);

$$\text{B.G.S.S.} = \frac{t_1^2}{n_1} + \frac{t_2^2}{n_2} + \cdots +, \text{ etc.} - C$$

(Put this in the summary table.)

7. Calculate the within group sum of squares (W.G.S.S.);

$$\text{W.G.S.S.} = \text{T.S.S.} - \text{B.G.S.S.}$$

(Put this in the summary table.)

8. Find the degrees of freedom and put them in the summary table. Where k is the number of groups:

   (a) Between groups df.
   $= \text{Number of groups} - 1 = k - 1$
   (b) Within groups df.
   $= \text{Number of scores} - \text{Number of groups} = N - k$
   (c) Total df.
   $= \text{Number of scores} - 1 = N - 1$

   CHECK B.G. df. + W.G. df. = Total df.
   (Put these values in the summary table.)

9. Calculate the variance estimates for between and within groups.

   Variance estimate = (Sum of squares)/(Degrees of freedom)

   (Put these values in the summary table.)

10. Calculate F; F = (B.G. var.)/(W.G. var.).
11. Consult the F-tables and note the F-values required at the 5, 2·5, 1 and 0·1% levels of significance.
12. *Comment*

Can we reject the idea that the group totals represent samples drawn from the same population?

Group	I	II	III		
	2	3	5	k	$= 3$
	4	6	6	n_1	$= 5$
	6	6	7	n_2	$= 5$
PROCEDURE	5	8	8	n_3	$= 5$
STEP	8	7	9		
1.	$t_1 = 25$	$t_2 = 30$	$t_3 = 35$	N	$= 15$
2.				T	$= 90$
3. $C = 90 \times 90/15$				C	$= 540$
4. $\Sigma X_1^2 = 145$; $\Sigma X_2^2 = 194$; $\Sigma X_3^2 = 255$; $\Sigma X^2 = 145+194+255$				ΣX^2	$= 594$
5. Total Sum of Squares (T.S.S.) $= \Sigma X^2 - C = 594-540$					$= 54$
6. Between Groups Sum of Squares (B.G.S.S.) $= \frac{t_1^2}{n_1}+\frac{t_2^2}{n_2}+\frac{t_3^2}{n_3}-C = \frac{25^2}{5}+\frac{30^2}{5}+\frac{35^2}{5}-540$					$= 10$
7. Within Group Sum of Squares (W.G.S.S.) = T.S.S. $-$ B.G.S.S. $= 54-10$					$= 44$
8. Degrees of Freedom:					
(a) B.G. degrees of freedom $= k-1 = 3-1$					$= 2$
(b) W.G. degrees of freedom $= N-k = 15-3$					$= 12$
(c) Total degrees of freedom $= N-1 = 15-1$					$= 14$
CHECK Total df. = B.G. df. + W.G. df. $14 = 12 + 2$					
9. Variance Estimate:					
Between groups variance = B.G.S.S./B.G. df. $= 10/2$					$= 5{\cdot}0$
Within groups variance = W.G.S.S./W.G. df. $= 44/12$					$= 3{\cdot}67$
10. F: (B.G. df.), (W.G. df.) = (B.G. var.)/(W.G. var.) $F_{2,\,12} = 5{\cdot}0/3{\cdot}67 = 1{\cdot}36$					

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F
Between groups	10	2	5·00	1·36
Within groups	44	12	3·67	
Total	54	14		

11. Consulting the tables for $F_{2,\,12}$ at:
5% level F = 3·9 2·5% level F = 5·1
1·0% level F = 6·9 0·1% level F = 13·0

12. *Comment*
Our value of F(1·36) is less than that required at the 5% level of significance. We therefore *reject* the idea that there is a significant difference between the means.

Note that the summary table is EXACTLY the same as the corresponding table in chapter 5.

Example 6.2 Now you carry out example 2, following exactly the procedure for one stage analysis of variance. Your *summary table* should be exactly like the table obtained from the two-stage method in chapter 5 (page 45).

Group I	II	
5	9	k =
6	7	n_1 =
0	7	n_2 =
3	5	N =
2	9	
2	5	

1. t_1 = t_2 =
2. T = ———
3. $C = T^2/N$ = ———
4. ΣX_1^2 = ; ΣX_2^2 = ; ΣX_3^2 = ΣX^2 = ———
5. Total Sum of Squares = $\Sigma X^2 - C$ = = ———
6. B.G.S.S. = $\frac{t_1^2}{n_1} + \frac{t_2^2}{n_2} - C$ = = ———
7. W.G.S.S. = T.S.S. − B.G.S.S. = ———
8. Degrees of Freedom
   B.G. df. = k − 1 = ———
   W.G. df. = N − k = ———
   Total df. = N − 1 = ———

   CHECK
   Total df. = B.G. df. + W.G. df.
   = +
9. Variance Estimates
   Between groups variance = B.G.S.S./B.G. df. = ———
   Within groups variance = W.G.S.S./W.G. df. = ———
10. F: (B.G. var.)/(W.G. var.)
    F , =

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F
Between groups				
Within groups				
Total				

11. Consulting the tables for F , at:
    level F = level F =
    level F = level F =

12. *Comment*

Simplifying the calculations

The examples in the body of this book have been especially constructed to allow for rapid and trouble-free calculation. In general, you will find that the results of real experiments are not so tractable. This is no problem if you have a calculating machine near at hand. If you have no machine, then you may take advantage of a simple rule which is used to reduce the absolute size of all the scores to be dealt with.

The results of analysis of variance computations are in no way affected if a constant is subtracted from all of the raw scores before the calculations begin. It is important to reduce all scores by the same amount but the exact value of the number you subtract is a matter of personal choice. Sometimes the values become negative, if too large a constant is removed. This will not affect the result but you should take special care not to get mixed up. If you are not happy with negative signs subtract a smaller constant. By way of illustration we are repeating example 6.1 of this chapter. The modest constant of 2 has been subtracted from each score. It does not make much difference to an already easy problem but it demonstrates that the same result is obtained in both cases.

Group	I	II	III	
	0	1	3	$k = 3$
	2	4	4	$n_1 = 5$
	4	4	5	$n_2 = 5$
	3	6	6	$n_3 = 5$
	6	5	7	

1. $t_1 = 15;\ t_2 = 20;\ t_3 = 25$ $\qquad N = 15$
2. $\qquad T = 60$
3. $C = 60 \times 60/15$ $\qquad C = 240$
4. $\Sigma x_1^2 = 65;\ \Sigma x_2^2 = 94;\ \Sigma x_3^2 = 135$
   $\Sigma x^2 = 65+94+135 = 294$
5. T.S.S. $= \Sigma x^2 - C = 294-240$ $\qquad$ T.S.S. $= 54!$
6. B.G.S.S. $= \dfrac{15^2}{5}+\dfrac{20^2}{5}+\dfrac{25^2}{5}-C = 45+80+125-240$

   $\qquad$ B.G.S.S. $= 10!$
7. W.G.S.S. = T.S.S. B.G.S.S.
   $= 54 - 10$ $\qquad$ W.G.S.S. $= 44!$

The rest of the computations are exactly as on page 49.

7

Two factor analysis of variance

One factor analysis of variance has a single very important limitation; it is designed for situations involving only one source of variance. It is often difficult or undesirable to design experiments which meet this restriction. If we wish to test the effects of different drugs on susceptibility to hypnosis, we will usually use both male and female subjects. Accordingly, we now have two factors which are being varied—drugs *and* sex. We will need two factor analysis of variance to analyse the results.

Example 7.1 The illustrative example for this chapter concerns the difficulty of designing the layout of manual controls in space modules. Clearly, a set of controls which may be manipulated easily under normal gravity could well prove inadequate during takeoff or in space when gravitational conditions are abnormal. Table 7.1 below gives the results of an (imaginary) experiment using five different control layouts under three different gravitational conditions. These conditions were produced in a special aircraft which simulated the gravitational properties of (a) EARTH (b) TAKEOFF and (c) SPACE. Fifteen different test pilots were allotted at random, one to each condition. Each pilot was expected to carry out an exacting series of manoeuvres. The number of errors he made was noted and is shown in table 7.1.

MACHINE LAYOUT	GRAVITATIONAL CONDITION			TOTALS
	EARTH	TAKEOFF	SPACE	
1	0	1	5	6
2	2	4	6	12
3	2	4	6	12
4	3	5	7	15
5	3	6	6	15
TOTALS	10	20	30	60

TABLE 7.1 Number of errors made by test pilots using different control layouts under different gravitational conditions.

At first glance, the results suggest that errors are determined largely by the gravitational conditions and to a lesser extent by the type of layout used. Of course, we would not expect all of the row totals to be exactly the same even if the layouts were equally good. We would have anticipated some differences due to the calibre of the pilots assigned to that layout and other uncontrolled influences on performances. The problem for analysis of variance is this: given that the pilots (of differing skill) were allotted at random to the different conditions, how likely is it that the row totals will vary as much as they in fact did? In other words, we have to decide between two possibilities:

(a) some layouts are better than others.
(b) the layouts are equally good but the differences in the row totals are due to an uneven distribution of pilot skill from one layout to another.

The same arguments apply to gravitational conditions. Perhaps we can explain the unevenness of the results by the possibility that all the skilled pilots were fortuitously put in the EARTH condition while the less skilled pilots were put in the other two conditions. It looks from the results that this is an unlikely explanation. Analysis of variance will tell us just how unlikely.

The general approach will be outlined now while the details of the procedure are given below on page 56. The basic idea is this: the bigger the differences between the row totals, the less likely it is that they are caused by chance factors alone. Thus, the bigger the variance of the row totals, the more significant are the differences between the control layouts. A similar argument applies to the column totals (gravitational conditions). Accordingly, two of the most important values which we need to compute are:

(a) the BETWEEN ROWS VARIANCE.
(b) the BETWEEN COLUMNS VARIANCE.

In order to assess these two values, we need a yardstick. This is provided by a quantity called the RESIDUAL VARIANCE. Briefly, this refers to variability among individual scores which cannot be accounted for by examination of the row and column totals. For example, we find that on average, subjects perform worse under SPACE conditions than TAKEOFF conditions (see the column totals). Layout one, however, gives us an unexpected result; the pilot under SPACE conditions performed much worse than the pilot under TAKEOFF conditions. This kind of deviation from the general pattern is attributed to RESIDUAL VARIANCE.

We may now compare our two main sources of variance (gravity conditions and layouts) against the RESIDUAL VARIANCE using the F-ratio

$$(1)\ \text{Column effects } F = \frac{\text{Between Columns Variance est.}}{\text{Residual Variance est.}}$$

$$\text{(2) Row effects } F = \frac{\text{Between Rows Variance est.}}{\text{Residual Variance est.}}$$

We assess the significance of F by consulting the F-tables. Before we may do so we need to know the degrees of freedom associated with our three sources of variance.

In general, the number of degrees of freedom associated with the BETWEEN COLUMNS VARIANCE is equal to one less than the number of column totals. Similarly, the number of degrees of freedom associated with the BETWEEN ROWS VARIANCE is equal to one less than the number of rows.

A similar argument applies to the RESIDUAL VARIANCE. The row and column totals fix some of the scores as can be seen here:

	0	1	⑤	= 6
	2	4	⑥	= 12
8 degrees of freedom	2	4	⑥	= 12
	3	5	⑦	= 15
	③	⑥	⑥	= 15
	10	20	30	60

In general, we may compute the residual df. by multiplying the df. for row and column variances:

$$4\text{ (Row df.)} \times 2\text{(Column df.)} = 8\text{ (Residual df.)}$$

There is another way of estimating the degrees of freedom for the residual variance.

We know that the degrees of freedom of the TOTAL VARIANCE are one less than the number of scores (N − 1), in this case 14 (= 15 − 1). It is also true that the degrees of freedom associated with the BETWEEN COLUMNS, BETWEEN ROWS and RESIDUAL VARIANCES must sum to equal the degrees of freedom for the TOTAL VARIANCE:

$$\text{df. (TOTAL)} = \text{df. (ROWS)} + \text{df. (COLUMNS)} + \text{df. (RESIDUAL)}$$

So it follows that:

$$\text{df. (RESIDUAL)} = \text{df. (TOTAL)} - \text{df. (ROWS)} - \text{df. (COLUMNS)}$$
$$8 = 14 - 4 - 2$$

The computational details of two-way analysis of variance are set out below. First of all some notation:

n_c = Number of scores in any column.
n_r = Number of scores in any row.
N = Total number of scores.

It is important to note that there are the same number of scores in each row and the same number of scores in each column. This is essential at this stage.

Follow the procedure carefully with the aid of the completed example on the next page.

1. Calculate the column totals: t_1, t_2, t_3, etc.
2. Calculate the row totals: r_1, r_2, r_3, etc.
3. Calculate the grand total: T
4. Calculate the correction term: $C = T^2/N$
5. Calculate the sum of squares of the raw scores: Σx^2
6. Calculate the TOTAL SUM OF SQUARES: (T.S.S.)

$$\text{T.S.S.} = \Sigma x^2 - C$$

7. Calculate the BETWEEN COLUMN SUM OF SQUARES (Columns S.S.)

$$\text{Columns S.S.} = \frac{t_1^2 + t_2^2 + t_3^2 + \text{etc.}}{n_c} - C$$

8. Calculate the BETWEEN ROW SUM OF SQUARES (Rows S.S.)

$$\text{Rows S.S.} = \frac{r_1^2 + r_2^2 + r_3^2 + \text{etc.}}{n_r} - C$$

9. Calculate the RESIDUAL SUM OF SQUARES (Residual S.S.)
   Residual S.S. = T.S.S. − Columns S.S. − Rows S.S.
10. Enter the appropriate degrees of freedom into the summary table.

    Columns df. = Number of columns − 1
    Rows df. = Number of rows − 1
    Total df. = Number of scores − 1
    Residual df. = Total df. − Columns df. − Rows df.

11. Calculate the variance estimates by dividing the sums of squares by the appropriate degrees of freedom.
12. Calculate the F-ratios:

    $F_{(\text{Columns})}$ = (Column variance)/(Residual variance)
    $F_{(\text{Rows})}$ = (Row variance)/(Residual variance)

13. Consult the F-distribution tables using the appropriate degrees of freedom. Find the *best* level of significance *exceeded* by your F-value and mark this in the extreme right hand column of the summary table.
14. Comment on your results. Draw a distinct conclusion for both row and column effects.

MACHINE LAYOUT	GRAVITATIONAL CONDITION			
	EARTH	TAKEOFF	SPACE	TOTALS
1	0	1	5	6
2	2	4	6	12
3	2	4	6	12
4	3	5	7	15
5	3	6	6	15
TOTALS	10	20	30	60

$n_c = 5$
$n_r = 3$
$N = 15$

TABLE 7.2 (See text).

1. Column totals: 10, 20, 30.
2. Row totals: 6, 12, 12, 15, 15.
3. Grand total: 60.
4. Correction term: $C = T^2/N = 60^2/15 = 3600/15 = 240$.
5. Sum of squares of Raw scores: $\Sigma x^2 = 302$.
6. Total sum of squares: T.S.S. $= \Sigma x^2 - C = 302 - 240 = 62$.
7. Between Column Sum of Squares: Columns S.S.

$$= \frac{t_1^2 + t_2^2 + t_3^2 - C}{n_c} = \frac{10^2 + 20^2 + 30^2}{5} - 240 = 40$$

8. Between Rows Sum of Squares: Rows S.S.

$$= \frac{r_1^2 + r_2^2 + r_3^2 + r_4^2 + r_5^2}{n_r} - C$$

$$= \frac{6^2 + 12^2 + 12^2 + 15^2 + 15^2}{3} - 240 = 18$$

9. Residual Sum of Squares = T.S.S. − Columns S.S. − Rows S.S.
   = 62 − 40 − 18 = 4
10. Degrees of freedom: Total df. = N − 1 = 15 − 1 = 14
    Columns df. = Number of columns − 1 = 3 − 1 = 2
    Rows df. = Number of rows. 1 = 5 − 1 = 4
    Residual df. = total df. − Rows df. − Columns
    = 14 − 4 − 2 df. = 8

Summary table

SOURCE		S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Gravity	(Columns)	40	2	20	20	0·1%
Layout	(Rows)	18	4	4·5	9	1·0%
Residual	(error)	4	8	0·5		
TOTAL		62	14			

Comment

Both values of F are significantly large, therefore: (a) we reject the idea that the different gravitational conditions produce the same number of errors. (b) we reject the idea that the five layouts are equally easy to use.

Example 7.2 Using exactly the same procedure, carry out an analysis of variance on the example on the following page.

This example concerns an (imaginary) experiment which compared the effects of four kinds of treatment on four kinds of neurotic patients. The treatments were: (a) behaviour therapy (B.Th.), (b) electroconvulsive shock therapy (E.C.T.), (c) drug therapy (drug), and (d) group therapy (group). The illnesses were (a) phobias, (b) depression, (c) anxiety states and (d) adolescent disturbances.

Four patients from each illness category were allotted at random, one each, to each treatment category. At the end of treatment each patient was assessed for improvement on a rating scale from 0–5 where 0 means no improvement and 5 means complete recovery.

Do the results support the idea that the different illnesses have significantly different recovery rates?

Do the results support the idea that the different treatments are significantly different in their overall effectiveness?

ILLNESS	PHOBIAS	DEPRESSION	ANXIETY	ADOLESCENTS	TOTALS
TREATMENT					
B.Th	4	1	0	0	
E.C.T.	2	3	1	1	
Drug	2	0	3	2	
Group	2	0	0	3	
TOTALS					

$n_c =$
$n_r =$
$N =$

TABLE 7.3

1. Column totals:
2. Row totals:
3. Grand total: $T =$
4. Correction Term: $C = T^2/N =$
5. Sum of Squares of Raw Scores: $\Sigma x^2 =$
6. Total Sum of Squares: T.S.S. =
7. Between Column Sum of Squares:

$$\text{Columns S.S.} = \frac{t_1^2+t_2^2+t_3^2+t_4^2}{n_c} - C =$$

8. Between Rows Sum of Squares.

$$\text{Rows S.S.} = \frac{r_1^2+r_2^2+r_3^2+r_4^2}{n_r} - C =$$

9. Residual Sum of Squares:

$$\text{Residual S.S.} = \text{T.S.S.} - \text{Columns S.S.} - \text{Rows S.S.}$$
$$= \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad =$$

10. Degrees of Freedom:

$$\text{Total df.} = N - 1 =$$
$$\text{Columns df.} = \text{Number of Columns} - 1 =$$
$$\text{Rows df.} = \text{Number of Rows} - 1 =$$
$$\text{Residual df.} = \text{Total df.} - \text{Columns df.} - \text{Rows df.}$$
$$= \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad =$$

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Illness (Columns)					
Treatment (Rows)					
Residual					
TOTAL					

Comment

ILLNESS	PHOBIAS	DEPRESSION	ANXIETY	ADOLESCENTS	TOTALS
TREATMENT					
B.Th	4	1	0	0	5
E.C.T.	2	3	1	1	7
Drug	2	0	3	2	7
Group	2	0	0	3	5
TOTAL	10	4	4	6	24

$n_c = 4$
$n_r = 4$
$N = 16$

TABLE 7.4

1. Column Totals: 10, 4, 4, 6
2. Row Totals: 5, 7, 7, 5
3. Grand Total $T = 24$
4. Correction Term: $C = T^2/N = 24^2/16$ $C = 36$
5. Sum of Squares of Raw Scores: $\Sigma x^2 = 62$
6. Total Sum of Squares: T.S.S. $= \Sigma x^2 - C = 62 - 36$ T.S.S $= 26$
7. Between Column Sum of Squares:

$$\text{B.C.S.S.} = \frac{10^2+4^2+4^2+6^2}{4} - C = \frac{168}{4} - 36$$

Columns S.S. = 6

8. Between Rows Sum of Squares:

$$\text{B.R.S.S.} = \frac{5^2+7^2+7^2+5^2}{4} - C = \frac{148}{4} - 36$$

Rows S.S. = 1

9. Residual Sum of Squares:
   Residual S.S. = T.S.S. − Columns S.S. − Rows S.S.
   = 26 − 6 − 1 Residual = 19
10. Degrees of Freedom:
    Total df. = N − 1 = 16 − 1 Total df. = 15
    Columns df. = Number of Columns − 1 = 4 − 1 Columns df. = 3
    Rows df. = Number of Rows − 1 = 4 − 1 Rows df. = 3
    Residual df. = Total df. − Columns df. − Rows df.
    = 15 − 3 − 3 Residual df. = 9

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Illness (Columns)	6	3	2	< 1	ns
Treatment (Rows)	1	3	0·333	< 1	ns
Residual	19	9	2·111		
TOTAL	26				

Comment

Neither of the treatments had any statistically significant effect.

8

Two factor analysis of variance with replications

Two factors analysis of variance—as described in chapter 7—makes an important assumption which is largely hidden. This assumption concerns the ADDITIVITY of factors and will be considered below. In general, we may NOT assume additivity of factors and our method of analysis must be extended in order to circumvent the problem. As we shall see, in some experiments we are primarily concerned to test the additivity assumption. Under these circumstances we need to use two factor analysis of variance WITH REPLICATIONS.

The additivity of factors

Table 8.1 gives the scores for example 7.1 in chapter 7. The results of the two factor analysis of variance suggested that there were significant differences in the error rates for different machine layouts and for different gravitational conditions. The first machine layout produced fewest errors and so did the earth condition. We may therefore expect that the pilot who worked *both* with layout one *and* under earth conditions should produce fewest errors. This is in fact the case; the two factors combined additively to produce the lowest error rate.

In table 8.2 are given the EXPECTED scores for each combination of conditions assuming that the factors are additive in their effects. Each score is calculated by adding the row mean to the column mean and then subtracting the grand mean (4). For example:

Earth × Machine layout 1; $2+2-4 = 0$
Space × Machine layout 5; $6+5-4 = 7$

These two results are the values we EXPECT under the *additive assumption*. We can see that table 8.2 corresponds very closely with table 8.1; that is, the observed results correspond very well with the results expected under the additive assumption.

Table 8.3 gives the differences between the observed and expected results.

MACHINE LAYOUT	GRAVITATIONAL CONDITIONS EARTH	TAKEOFF	SPACE	MEANS
1	0	1	5	2
2	2	4	6	4
3	2	4	6	4
4	3	5	7	5
5	3	6	6	5
MEANS	2	4	6	4

TABLE 8.1 The marginal figures are the group means (see table 7.1).

LAYOUT	EARTH	TAKEOFF	SPACE
1	0	2	4
2	2	4	6
3	2	4	6
4	3	5	7
5	3	5	7

TABLE 8.2 Expected results under the additivity assumption (see text).

LAYOUT	EARTH	TAKEOFF	SPACE
1	0	−1	+1
2	0	0	0
3	0	0	0
4	0	0	0
5	0	+1	−1

$\Sigma X = 0 \quad \Sigma X^2 = 4$

TABLE 8.3 Residual scores (see text).

We call this the table of RESIDUALS. The residual part of any score is that part which is not accounted for under the additive assumption. In this example, the residuals are small and we have no reason to challenge the assumption that the two factors are combining additively to produce individual scores.

We can find the residual sum of squares directly from table 8.3 by simply summing the squares of the individual scores:

$$\text{RESIDUAL Sum of scores} = \Sigma((+1)^2+(-1)^2+(+1)^2+(-1)^2) = 4$$

You will notice that this is the same value as that computed in chapter 7 by a much more roundabout route.

We get a somewhat different picture if we look at example 7.2 from chapter 7. These scores are given again in table 8.4. Using the same procedure as in the previous example we may generate the scores EXPECTED under the ADDITIVE assumption. These scores are given in table 8.5. For example, we calculate the expected recovery score of the phobic patient undergoing behaviour therapy (B.Th.) by:

adding (a) the mean recovery score for phobics:	+2·50
and (b) the mean recovery score for behaviour therapy:	+1.25
and subtracting (c) the grand mean recovery score:	−1·50
yielding the expected recovery score:	2·25

Table 8.6 gives the residuals, that is the difference between what we observed and what we expected. Thus we observed a recovery score of 4 for the phobic undergoing behaviour therapy, whereas we expect a recovery score of only 2·25. The residual (i.e., unexpected) part of the score is 1·75 and is given in table 8.6. The sum of squares of the residuals is 19 and this corresponds to the value for residual s.s. obtained in chapter 7.

The residual scores are so large that we may be excused if we conclude that the additive assumption is not borne out in this case. Clearly, the success of behaviour therapy in curing the phobia is not repeated with any other form of illness. Behaviour therapy is therefore not *generally* successful but only successful in specific cases. Similarly, group therapy was particularly useful with the adolescent problem. Accordingly, we must not ask which treatments are the best overall or which illnesses have the best overall recovery rate. Instead, we must ask which illnesses respond best to which treatments. In such cases, we say that individual scores are not determined by the *additivity* of the factors but by the *interaction* of factors.

TREATMENT	ILLNESS				MEAN
	PHOBIAS	DEPRESSION	ANXIETY	ADOLESCENTS	
B.Th.	4	1	0	0	1·25
E.C.T.	2	3	1	1	1·75
Drug	2	0	3	2	1·75
Group	2	0	0	3	1·25
MEANS	2·50	1·00	1·00	1·50	1·50

TABLE 8.4 Recovery scores for 16 patients (5=complete recovery) undergoing different kinds of treatment.

	PHOBIAS	DEPRESSION	ANXIETY	ADOLESCENTS
B.Th.	2·25	0·75	0·75	1·25
E.C.T.	2·75	1·25	1·25	1·75
Drug	2·75	1·25	1·25	1·75
Group	2·25	0·75	0·75	1·25

TABLE 8.5 Expected scores under the additivity assumption (see text).

	PHOBIAS	DEPRESSION	ANXIETY	ADOLESCENTS
B.Th.	+1·75	+0·25	−0·75	−1·25
E.C.T.	−0·75	+1·75	−0·25	−0·75
Drug	−0·75	−1·25	+1·75	+0·25
Group	−0·25	−0·75	−0·75	+1·75

$\Sigma x = 0$
$\Sigma x^2 = 19 =$ residual s.s.

TABLE 8.6 Residuals (see text).

Interaction and error

In table 8.3, the residual scores are very small. In such a case, it seems reasonable to attribute the residual scores to random error in, say, allotting pilots of varying skill to the different conditions.

In table 8.6, the residual scores are very large. Under these circumstances, we attribute the residual scores to a combination of two sources:

(a) random error.
(b) interaction of factors.

Ideally, we would like to separate the random error from the interaction of factors. Firstly, because we need a pure error term for use in our F-ratios. Secondly, because we need to test the interaction term to see if it is significantly large. We cannot do this with the scores at hand. We can do it, however, if we are prepared to repeat the experiment one or more times. Hence the title of this chapter, 'Two factor analysis of variance with replications.'

Table 8.7 gives the results for the first experiment with three replications. Forty-five pilots were used in all and 3 different pilots were used for each combination of factors. None of the pilots using layout 1 under earth gravitational conditions scores any errors. These three zero scores constitute a *cell* and the cell total (0) is given in brackets. Accordingly, we have 15 cell totals, 5 row totals and 3 column totals.

The broad outline of the procedure is as follows:

1. Calculate the TOTAL SUM OF SQUARES (as usual).
2. Calculate the BETWEEN COLUMNS S.S. (as usual).
3. Calculate the BETWEEN ROWS S.S. (as usual).
4. Calculate the RESIDUAL S.S. (as usual).
5. Calculate the ERROR S.S. This is sometimes called the WITHIN CELLS S.S. and is directly analogous to the WITHIN GROUPS S.S. in one factor analysis of variance except that we use cell totals instead of group totals.
6. Calculate the INTERACTION S.S. = RESIDUAL − ERROR.

You should note that the residual is only an intermediate figure in this case and it does not appear in the summary table.

The procedure will be outlined in the usual way, but first of all some notation:

N = Total number of scores.
n_r = Number of scores in any row.
n_c = Number of scores in any column.
n_{cell} = Number of scores in any cell.
t_1 = A column total.
r_1 = A row total.
c_1 = A cell total.
C = Correction term.

Follow the procedure carefully with the aid of the example in table 8.7.

1. Calculate the cell totals: c_1, c_2, c_3, etc.
2. Calculate the column totals: t_1, t_2, t_3, etc.
3. Calculate the row totals: r_1, r_2, r_3, etc.
4. Calculate the grand total: T.
5. Calculate the correction term: $C = T^2/N$.
6. Calculate the sum of the squares of the raw scores: Σx^2.
7. Calculate the TOTAL SUM OF SQUARES: (T.S.S.) T.S.S. $= \Sigma x^2 - C$
8. Calculate the BETWEEN COLUMNS SUM OF SQUARES: (Columns S.S.)

$$\text{Columns S.S.} = \frac{t_1^2 + t_2^2 + \text{etc.}}{n_c} - C$$

9. Calculate the BETWEEN ROWS SUM OF SQUARES: (Rows S.S.)

$$\text{Rows S.S.} = \frac{r_1^2 + r_1^2 + \text{etc.}}{n_r} - C$$

10. Calculate the RESIDUAL sum of squares: (Residual S.S.)
    Residual S.S. = T.S.S. − Columns S.S. − Rows S.S.
11. Calculate the ERROR (within cells): (Error S.S.)

$$\text{Error S.S.} = \Sigma x^2 - \frac{\Sigma c^2}{n_{cell}}$$

12. Calculate the INTERACTION S.S.
    Interaction S.S. = Residual S.S. − Error S.S.
13. Enter the appropriate degrees of freedom in the summary table:
    Total df. = N − 1
    Rows df. = number of rows − 1
    Columns df. = number of columns − 1
    Inter df. = rows df. × cols df.
    Error df. = Total df. − Rows df. − Columns df. − Inter df.
14. Calculate the variance estimates in the usual way.
15. Calculate the F-ratios for rows and columns, and interaction by dividing each variance estimate by the *Error* (within cells) variance estimates.
16. Assess the significance of your F-ratios and comment clearly on the result.

MACHINE LAYOUT	GRAVITATIONAL CONDITIONS EARTH	TAKEOFF	SPACE	ROW TOTALS
1	0,0,0(0)	0,0, 3(3)	4,5,6(15)	18
2	1,1,4(6)	2,5, 5(12)	4,7,7(18)	36
3	1,2,3(6)	3,4, 5(12)	6,6,6(18)	36
4	2,2,5(9)	2,5, 8(15)	6,7,8(21)	45
5	0,3,6(9)	2,6,10(18)	6,6,6(18)	45
COLUMN TOTAL	30	60	90	180

TABLE 8.7 Error scores for 45 pilots under different gravitational conditions using different machine layouts.

1. Cell totals: 0,3,15; 6,12,18; 6,12,18; 9,15,21; 9,18,18;
2. Column totals: 30, 60, 90
3. Row totals: 18, 36, 36, 45, 45
4. Grand total: $T = 180$
5. Correction term: $C = T^2/N = 180^2/45 = 32{,}400/45 = 720$
6. Sum of squares of raw scores: $\Sigma x^2 = 1012$
7. Total sum of squares: T.S.S. $= \Sigma x^2 - C = 1012 - 720 = 292$
8. Between columns sum of squares:

$$\text{Columns S.S.} = \frac{30^2+60^2+90^2}{15} - C = \frac{12600}{15} - 720 = 120$$

9. Between Rows Sum of Squares:

$$\text{Rows S.S.} = \frac{18^2+36^2+36^2+45^2+45^2}{9} - C = \frac{6966}{9} - 720 = 54$$

10. Residual sum of squares: Residual S.S. $= 292 - 120 - 54 = 118$
11. Error sum of squares:

$$\text{Error S.S.} = 1012 - \frac{0^2+3^2+15^2+6^2+12^2+18^2+6^2+12^2+18^2+9^2+15^2+21^2+9^2+18^2+18^2}{3}$$

$$= 1012 - 2718/3 = 1012 - 906 = 106$$

12. Interaction sum of squares:
    Interaction S.S. $= 118 - 106 = 12$
13. Degrees of freedom:

Total df.	$= N-1 = 45-1$	$= 44$
Rows df.	$=$ Rows $-1 = 5-1$	$= 4$
Column df.	$=$ Columns $-1 = 3-1$	$= 2$
Inter df.	$=$ (Rows df. $\times$ Columns df.) $= 4 \times 2$	$= 8$
Error df.	$= 44-4-2-8$	$= 30$

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Cols (Gravity)	120	2	60	17.0	<0·1%
Rows (Layout)	54	4	13·5	3·8	<2·5%
Interaction (Rows × Columns)	12	8	1·5	<1	ns
Error	106	30	3·5		
TOTAL	292	44			

Comment

1. The difference in error scores for the three gravitational conditions is highly significant.
2. The difference in error scores for the five machine layouts is less significant at the 2·5% level.
3. There is no significant indication that the two factors interact.

Using exactly the same procedure, perform a two-way analysis of variance on the following data. The scores were derived from a sleep experiment involving twelve subjects. Half of the subjects were chosen at random from a psychiatrist's register of neurotic patients. The other half were chosen at random from a neighbouring G.P.'s list of patients.

Three neurotics and three non-neurotics were given a bottle of ten sugar pills, one to be taken before retiring on each of 10 nights. The remaining subjects were given no pills at all. During the experiment period of 10 nights all subjects were asked to count the number of disturbed nights. The results are given in table 8.8.

The neurotic subjects report more disturbed nights than the control group. Is the difference significant?

The sugar pill produced fewer disturbed nights than the no pill treatment. Is this difference significant?

The results suggest that neurotics sleep very badly without a pill but very well with a pill. Non-neurotics do not seem to be greatly affected. Is this interaction effect significant?

	SUGAR PILL	NO PILL	TOTALS
Neurotic	0,1,2	4,5,2	
Non-neurotic	1,2,1	1,2,3	
TOTALS			

$N =$
$n_c =$
$n_r =$
$n_{cell} =$

TABLE 8.8 Twelve subjects rated for number of nights with reported sleep disturbance.

1. Cell totals:
2. Column totals:
3. Row totals:
4. Grand total: $T =$
5. Correction term: $C = T^2/N =$
6. Sum of Squares of raw scores: $\Sigma x^2 =$
7. Total sum of squares: T.S.S. $= \Sigma x^2 - C =$
8. Between Columns sum of squares:

$$\text{Columns S.S.} = \frac{t_1^2 + t_2^2 + \text{etc.}}{n_c} - C$$

$=$ $=$

9. Between Rows sum of squares:

$$\text{Rows S.S.} = \frac{r_1^2 + r_2^2 + \text{etc.}}{n_r} - C$$

$=$ $=$

10. Residual sum of squares:
    Residual S.S. $=$ T.S.S. $-$ Columns S.S. $-$ Rows S.S.
    $=$
11. Error sum of squares:
    Error S.S. $= \Sigma x^2 - (\Sigma c^2)/n_{cell}$
    $=$
12. Interaction sum of squares:
    Interaction S.S. $=$ Residual S.S. $-$ Error S.S.
    $=$ $=$
13. Degrees of Freedom

    Total df. $= N - 1$ $=$
    Rows df. $=$ Rows $- 1$ $=$
    Columns df. $=$ Columns $- 1$ $=$
    Inter df. $=$ Rows df. $\times$ Columns df. $=$
    Within cells $=$ Total df. $-$ Rows df. $-$ Columns df. $-$ Inter df. $=$

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Columns (pill)					
Rows (neurosis)					
Interaction					
Within cells (Error)					
TOTAL					

Comment

	SUGAR PILL	NO PILL	TOTALS
Neurotic	0,1,2(3)	4,5,2(11)	14
Non-neurotic	1,2,1(4)	1,2,3(6)	10
TOTALS	7	17	24

$N = 12$
$n_c = 6$
$n_r = 6$
$n_{cell} = 3$

TABLE 8.9 Twelve subjects rated for number of nights with reported sleep disturbance.

1. Cell totals: 3, 11, 4, 6
2. Column totals: 7, 17
3. Row totals: 14, 10
4. Grand total: $T = 24$
5. Correction term: $C = T^2/N = 24^2/12 = 48$
6. Sum of squares of raw scores: $\Sigma x^2 = 70$
7. Total sum of squares: T.S.S. $= \Sigma x^2 - C - 48 = 22$
8. Between Columns sum of squares:

$$\text{Columns S.S.} = \frac{7^2+17^2}{6} - 48 = 56{\cdot}33 - 48 = 8{\cdot}33$$

9. Between Rows sum of squares:

$$\text{Rows S.S.} = \frac{14^2+10^2}{6} - 48 = 49{\cdot}33 - 48 = 1{\cdot}33$$

10. Residual sum of squares: Residual S.S. $= 22 - 8{\cdot}33 - 1{\cdot}33 = 12{\cdot}34$
11. Error sum of squares:

$$\text{Error S.S.} = 70 - \frac{(3^2+11^2+4^2+6^2)}{3} = 70 - 60{\cdot}66 = 9{\cdot}33$$

12. Interaction Sum of squares: Interaction S.S. $= 12{\cdot}33 - 9{\cdot}33 = 3$
13. Degrees of Freedom

Total df.	$= N-1$	$= 12-1$	$= 11$
Rows df.	$=$ Rows -1	$= 2-1$	$= 1$
Columns df.	$=$ Columns -1	$= 2-1$	$= 1$
Inter df.	$=$ Rows df. $\times$ Columns df.	$= 1\times 1$	$= 1$
Within cells	$= 11-1-1-1$		$= 8$

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Columns (pill)	8·333	1	8·333	7·14	$<2{\cdot}5\%$
Rows (neurosis)	1·333	1	1·333	1·14	ns
Interaction	3	1	3	2·57	ns
Within cells (Error)	9·333	8	1·166		
TOTAL					

Comment

The differences between the pill and no pill groups are significant at the 2·5% level. The differences between neurotic and non-neurotic groups are not significant. There is no significant evidence for an interaction effect between the two factors.

So far in this chapter we have calculated the interaction and error sum of squares by partitioning the residual sum of squares. This technique had the advantage of illustrating how replications alter the nature of the residual term. We may now alter the procedure slightly and generate the formulae which are more commonly used by experienced statisticians. Assume that we have already calculated the row and column sum of squares. The next steps are:

10. $$\text{Cells s.s.} = \frac{\Sigma c^2}{n_{cell}} - C$$

11. Interaction s.s. = Cells s.s. − Rows s.s. − Columns s.s.
12. Error s.s. = Total s.s. − cells s.s.

Of course, these formulae generate the same results as the old ones. Some feel that this procedure is quicker and more elegant. Certainly, when we come to three factor analysis of variance, we will have to use this procedure to avoid confusion. Use the new procedure to analyse the data in table 8.10. You will have to get along without assistance this time but your final summary table should be the same as that in table 8.11.

ILLNESS	PHOBIAS	DEPRESSION	ANXIETY	ADOLESCENTS
TREATMENT				
B.Th	4,5,3	0,0,1	0,1,1	0,2,1
E.C.T.	0,2,2	4,3,2	1,2,1	0,0,1
Drug	3,0,0	1,2,1	3,3,3	0,2,0
Group	2,1,1	2,0,0	0,0,0	4,1,4

TABLE 8.10 Recovery ratings for 48 psychiatric patients. See chapter 7 for further explanation.

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Treatment	3·73	3	1·2	1.3	ns
Illness	0·56	3	0·2	< 1	ns
Interaction	57·52	9	6·4	6·8	0·1%
Error	30·00	32	0·94		
TOTAL	91·81	47			

Comment
Only the interaction term reaches significance. This indicates clearly that the effectiveness of the treatment is closely dependent on the type of patient receiving the treatment.

TABLE 8.11 Summary table for two factor analyses of variance applied to the data given in table 8.10.

9

Linear trend test

An important application of analysis of variance comes in the study of trends. The following example is a case in point (table 9.1). Subjects experience various degrees of gravitational stress whilst doing simple mental arithmetic tasks. The results indicate that as stress increases so does the number of errors. The group totals are plotted against degree of stress in Fig. 9.1. There is an almost straight line relationship between them. We could say that the results may be described in terms of a *linear trend.*

GRAVITATIONAL CONDITION	1_g	2_g	3_g	4_g
	0	3	4	4
	0	1	2	3
	1	1	2	3
TOTALS	1	5	8	10

TABLE 9.1 Error scores for 12 subjects performing mental arithmetic tasks under different gravitational conditions.

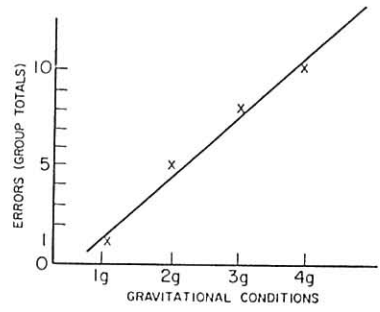


FIGURE 9.1 Errors (group totals) plotted against gravitational conditions (see table 9.1).

Very broadly our procedure for testing the significance of the linear trend will be as follows:

1. One factor analysis of variance. We must test first of all the significance of the differences between the group totals.
2. We divide the BETWEEN GROUP S.S. into two parts; the LINEAR S.S. and the NON-LINEAR S.S.
   (a) LINEAR; the more obvious the linear trend, the larger is the LINEAR S.S.
   (b) NON-LINEAR; if the NON-LINEAR S.S. is significant then the trend is COMPLEX, even if the LINEAR S.S. is significant.

The first step in the trend test is to complete the analysis of variance. The trend test is always the second step. Here is the one way analysis of variance:

GRAVITATIONAL CONDITIONS	1_g	2_g	3_g	4_g	
	0	3	4	4	$n_{group} = 3$
	0	1	2	3	$N = 12$
	1	1	2	3	

1. Group totals: $t_1 = 1 \quad t_2 = 5 \quad t_3 = 8 \quad t_4 = 10$
2. Grand total: $T = 24$
3. Correction term: $C = T^2/N = 24^2/12 = 48$ $C = 48$
4. Sum of squares of raw scores: $\Sigma x^2 = 70$
5. Total sum of squares: T.S.S. $= \Sigma x^2 - C = 70 - 48 = 22$
6. Between group sum of squares:

$$\text{B.G.S.S.} = \frac{1^2+5^2+8^2+10^2}{3} - 48 = \frac{190}{3} - 48 = 15{\cdot}334$$

7. Within group sum of squares: W.G.S.S. = T.S.S. − B.G.S.S.
   $= 22 - 15{\cdot}334 = 6{\cdot}67$
8. Degrees of freedom:
   (a) B.G. degrees of freedom $= 4-1 = 3.$
   (b) W.G. degrees of freedom $= 12-4 = 8.$
   (c) Total degrees of freedom $= 12-1 = 11.$

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Between groups (incentive)	15·334	3	5·111	6·13	<2·5%
Within groups (error)	6·67	8	0·833		
TOTAL	22·00	11			

Comment

The differences between the group totals is significant at better than the 2·5% level of significance.

The analysis of variance shows that the differences between the gravitational conditions are significant. Inspection of Fig. 9.1 suggests that much of the between group variance can be explained by a linear trend. Some of the variance, however, cannot and this causes kinks in the line. In a linear trend test, we divide the between group sum of squares into two parts:

Between group sum of squares = LINEAR sum of squares
+ NON-LINEAR sum of squares.

The linear s.s. is that proportion of the group total variance which is predicted from a straight line increase. The non-linear s.s. is the remainder.

If the group totals were 2, 5, 8, 11, the non-linear s.s. would be zero because this is a perfect linear trend (equal steps of 3 between each total).

If the group totals were 2, 6, 6, 2, the linear s.s. would be zero because these scores are perfectly non-linear.

Usually, there is a compromise between these two extreme situations.

Caution The trend test, as you shall see, is a very simple device. But to keep it so a number of simplifying restrictions must be strictly observed:

1. Equal number of scores in each group.
2. Equal steps between the group treatments.

(In our experiment we have 1_g steps between the treatments.)

If these restrictions cannot be observed, then more complex procedures are available, but these are best avoided.

The general conceptual approach is straightforward. We compare the group totals to a *standard linear trend line*. If the comparison is good we have a large linear sum of squares. If the comparison is poor we have a small linear sum of squares. Here is a standard linear trend line:

$$-3 \qquad -1 \qquad 1 \qquad 3$$

The trend is linear because the values rise in equal steps (of two), just like a straight line graph. You should notice that all trend line values must sum to zero.

$$(-3)+(-1)+(1)+(3) = 0$$

This restriction is observed *because it greatly simplifies the calculations* during the trend test.

Trend line values are usually called *polynominal coefficients.*

There is one coefficient for each group total. In this case, there are 4 but there may be many more.

The comparison of the trend line with the data is easily performed. Multiply each coefficient by the corresponding group total.

Group total:	1	5	8	10
coefficient:	−3	−1	1	3
product (c) × (t):	−3	−5	8	30 Sum (Σct) = 30.

The sum of the products is large (30), which indicates that a linear trend is a good fit. If we rearrange the group totals so that there is no trend you will see that Σct becomes relatively small. For example:

t:	5	10	1	8	
c:	−3	−1	1	3	
ct:	−15	−10	1	24	Σct = 0.

So, the better the trend the larger the value of Σct

We convert Σct to LINEAR SUM OF SQUARES by squaring it and dividing by the sum of the squares of the coefficients and by the number of scores in each group.

So that LINEAR S.S. $= \dfrac{(\Sigma ct)^2}{n\Sigma c^2}$

In our example,

$$\begin{aligned} (\Sigma ct)^2 &= (30)^2 = 900 \\ \Sigma c^2 &= (-3)^2+(-1)^2+(1)^2+(3)^2 \\ &= 9+1+1+9 = 20 \\ n_{group} &= 3 \end{aligned}$$

So LINEAR S.S. $= \dfrac{(\Sigma ct)^2}{n\Sigma c^2} = \dfrac{900}{3\times 20} = \dfrac{900}{60} = 15$

The S.S *for a trend has only one degree of freedom.*
The S.S. *for the non-linear* S.S. *is one less than the between groups df.*

Hence,

Linear df. = 1
Non-linear df. = B.G. df. − 1

An example of the full procedure is given below. Notice that the sum of squares is treated in the usual way once it has been entered into the summary table.

1. Group		(1g)	(2g)	(3g)	(4g)	
Group total	(t)	1	5	8	10	$n = 3$
linear coefficients	(c)	−3	−1	1	3	
2. ct		−3	−5	8	30	
3. Σct						$\Sigma ct = 30$
4. $(\Sigma ct)^2$						$(\Sigma ct)^2 = 900$
5. c^2		9	1	1	9	
6. Σc^2						$\Sigma c^2 = 20$

7. Linear sum of squares:

$$\text{Linear s.s.} = \frac{(\Sigma ct)^2}{n\Sigma c^2} = \frac{900}{3 \times 20} = \frac{900}{60} = 15$$

8. Non-linear sum of squares:

Since Between Group s.s. = Linear s.s. + Non-linear s.s.

Non-linear s.s. = B.G.S.S. − Linear s.s.

= 15·333 — 15 = 0·333

9. Degrees of freedom:

Between Groups df. = 3

Linear df. = 1

Non-linear df. = 3 − 1 = 2

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Between groups	15·333	3	5·111	6·13	< 2·5%
Linear	15·000	1	15·000	18·01	< 1·0%
Non-linear	·333	2	0·166	< 1	ns
Within group	6·666	8	0·833		
TOTAL	22·000	11			

Comment

The linear trend is significant at better than the 1% level of significance. Deviations from the linear trend are non-significant. The differences between the group totals may reasonably be described in terms of a linear trend.

Consider the following experiment and carry out a one-way analysis of variance on the results. Check your results on page 80. Then carry out a test for linear trend on page 81. There are five experimental treatments so a different set of coefficients is required:

Linear trend coefficients for 5 group totals:

$$-2 \quad -1 \quad 0 \quad 1 \quad 2$$

Notice that the sum of the coefficients is zero.

A sample of 15 female rats was used in a litter retrieval experiment. Each rat was suckling a litter of 6 babies. The rats were trained over a period of a week to run a T-maze to retrieve a single baby rat which had been removed from the litter and placed in the goal box. The infant was always placed in the arm opposite to the position preference arm, i.e., the mother had to overcome a position preference in maze running in order to retrieve successfully.

On the final testing day the procedure was altered so that more than one infant could be removed to vary the incentive level. Five treatments were used which involved the removal of 1, 2, 3, 4 and 5 infants. Three rats were allotted to each group. Twenty maze runs were observed for each rat. The total number of correct runs in excess of 10 were noted. The results are given in table 9.2.

NUMBER OF MISSING BABIES	1	2	3	4	5
	0	0	2	4	3
	0	2	2	4	2
	1	1	3	4	2
TOTALS	1	3	7	12	7

TABLE 9.2 Correct maze choices for maternal rats in retrieval task. Notice that there are an equal number of scores in each group and that there are equal differences between adjacent treatments (1 additional missing infant).

GROUP	1	2	3	4	5	
	0	0	2	4	3	n_{group} =
	0	2	2	4	2	N =
	1	1	3	4	2	

Group totals:
Grand total: T =
Correction term: C =
Sum of squares of raw scores: Σx^2 =
Total sum of squares: T.S.S. =
Between Group Sum of Squares:
B.G.S.S. =
Within Group Sum of Squares:
W.G.S.S. =
Degrees of freedom:
Total df. =
B.G. df. =
W.G. df. =

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE

Comment

GROUP	1	2	3	4	5
	0	0	2	4	3
	0	2	2	4	2
	1	1	3	4	2
TOTALS	1	3	7	12	7

$n_{group} = 3$
$N = 15$

$T = 30$

$C = 30^2/15 = 60$

$\Sigma x^2 = 88$

T.S.S. $= 88 - 60 = 28$

B.G.S.S. $= \frac{252}{3} - 60 = 24$

W.G.S.S. $= 28 - 24 = 4$
Total df. $= 14$
B.G. df. $= 4$
W.G. df. $= 10$

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Between groups (incentive)	24	4	6	15	<0·1%
Within groups (error)	4	10	0·4		
TOTAL	28	14			

Comment
The differences between the group totals are significant at better than the 0·1% level of significance.

Now carry out the trend test; the coefficients are:

		−2	−2	0	1	2	
1.	Group	1	2	3	4	5	
	Group total	1	3	7	12	7	n_{group} =
	Linear coefficients						
2.	ct						
3.	Σct						Σct =
4.	$(\Sigma ct)^2$						$(\Sigma ct)^2$ =
5.	c^2						
6.	Σc^2						Σc^2 =

7. Linear sum of squares:

$$\text{Linear S.S.} = \frac{(\Sigma ct)^2}{n\Sigma c^2} =$$

8. Non-linear sum of squares:

Non-linear S.S. = B.G.S.S. − Linear S.S.

= =

9. Degrees of freedom:

Between groups df. =

Linear df. =

Non-linear df. =

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Between groups					
linear					
Non-linear					
Within groups					
TOTAL					

Comment

1. Group	1	2	3	4	5	
Group total	1	3	7	12	7	$n_{group} = 3$
Linear coefficients	−2	−1	0	1	2	
2. ct	−2	−3	0	12	14	
3. Σct						$\Sigma ct = 21$
4. $(\Sigma ct)^2$						$(\Sigma ct)^2 = 441$
5. c^2	4	1	0	1	4	
6. Σc^2						$\Sigma c^2 = 10$

7. Linear sum of squares

$$\text{Linear s.s.} = \frac{(\Sigma ct)^2}{n\Sigma c^2} = \frac{441}{3 \times 10} = \frac{441}{30} = 14{\cdot}7$$

8. Non-linear sum of squares:

$$\text{Non-linear s.s.} = \text{B.G.S.S.} - \text{Linear s.s.}$$
$$= 24 - 14{\cdot}7 = 9{\cdot}3$$

9. Degrees of freedom:

B.G. df. = 4
Linear df. = 1
Non-linear df. = 3

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Between groups	24·0	4	6·0	15	<0·1%
linear	14·7	1	14·7	36·75	<0·1%
Non-linear	9·3	3	3·1	7·75	<1%
Within groups	4·0	10	0·4		
TOTAL	28·0	14			

Comment

Both linear and non-linear-trend components are significant. An explanation of the results in terms of a simple linear trend is not tenable. The trend is complex with a strongly linear element.

10

Multiple trend tests

In the previous chapter, we examined procedures for applying a simple linear trend test. In both of the examples we used, a linear trend seemed to describe most aptly the relationship between the differing treatments and the resulting scores. This is not always the case. A linear trend is only one of a whole range of possible trends. Figure 10.1 shows a quite different trend. This is often called an inverted U-shaped curve. More specifically it belongs to the family of QUADRATIC curves. Quadratic curves have a single hump, bend or turning point. If this curve were turned upside down it would still be a quadratic curve.

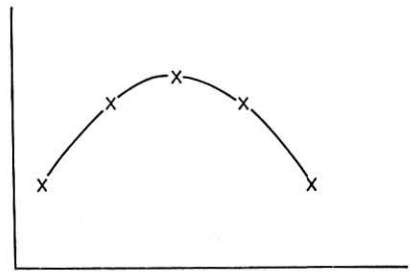


FIGURE 10.1 An example of a quadratic trend.

We can test for a quadratic trend just as easily as for a linear trend. Here are a set of quadratic trend coefficients:

$$-2 \quad 1 \quad 2 \quad 1 \quad -2$$

Notice that the coefficients form a symmetrical inverted U-shaped curve. Notice, also, that the coefficients sum to zero.

We will test the data used in the previous chapter for a quadratic trend. The group totals were:

$$1 \quad 3 \quad 7 \quad 12 \quad 7$$

and the one-way analysis of variance summary table was:

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Between groups	24	4	6	15	<0·1%
Within groups	4	10	0·4		
TOTAL	28	14			

Quadratic trend test

		I	II	III	IV	V	
1.	Group						
	Group total	1	3	7	12	7	$n = 3$
	Quadratic coefficients	−2	1	2	1	−2	
2.	ct	−2	3	14	12	−14	
3.	Σct						$\Sigma ct = 13$
4.	$(\Sigma ct)^2$						$(\Sigma ct)^2 = 169$
5.	c^2	4	1	4	1	4	
6.	Σc^2						$\Sigma c^2 = 14$

7. Quadratic sum of squares

$$\text{Quadratic S.S.} = \frac{(\Sigma ct)^2}{n\Sigma c^2} = \frac{169}{3 \times 14} = \frac{169}{42} = 4{\cdot}0238$$

8. Non-quadratic sum of squares

Non-quadratic S.S. = B.G.S.S. − Quadratic S.S.
= 24 − 4·0238
= 19·9762

9. Degrees of Freedom

Between groups df. = 4
Quadratic trend df. = 1
Non-quadratic df. = 3

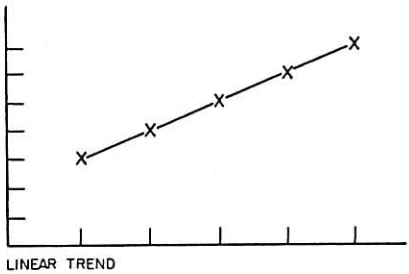
Summary tables

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Between groups	24	4	6	15	<0·1%
Quadratic trend	4·0238	1	4·0238	10·06	=0·1%
Non-quadratic	19·9762	3	6·6587	16·65	<0·01%
Within groups	4	10	0·4		
TOTAL	28	14			

Comment

Both quadratic and non-quadratic trend components are significant. An explanation of the results in terms of a single quadratic trend is not tenable. The trend is complex with a significant quadratic component.

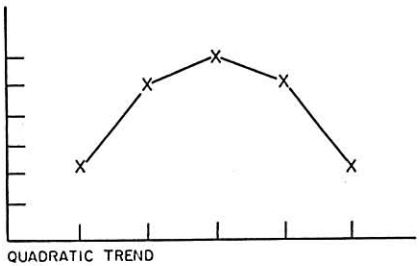
Using five points it is possible to construct four simple trends. These are illustrated in Fig. 10.2 with some sample trend coefficients.



COEFFICIENTS

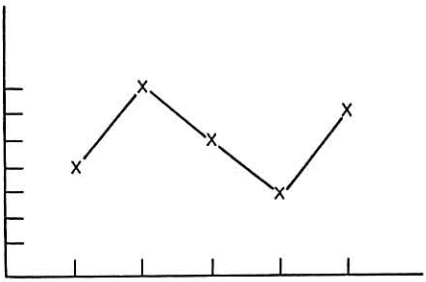
LINEAR

−2 −1 0 1 2



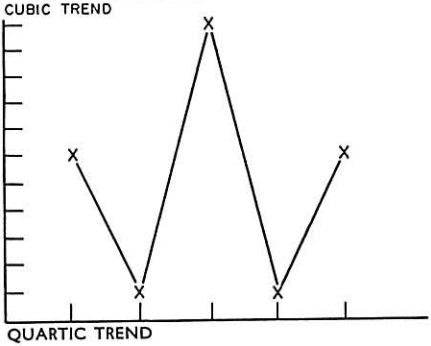
QUADRATIC

−2 1 2 1 −2



CUBIC

−1 2 0 −2 1



QUARTIC

1 −4 6 −4 1

FIGURE 10.2 Four possible simple trends using five points.

From Fig. 10.2 it will be clear that:

(a) a LINEAR trend has *NO* turning point.
(b) a QUADRATIC trend has *ONE* turning point.
(c) a CUBIC trend has *TWO* turning points.
(d) a QUARTIC trend has *THREE* turning points.

Using only five points, it is impossible to describe a trend with four turning points (QUINTIC). If we have six group totals, however, then we would be able to test for a quintic curve. Notice that the five group totals in the last example allow us four degrees of freedom and that we may test for only four possible trends. If there were ten group totals, then we might test for nine possible trends.

In any trend test, the sum of squares for a single trend has but one degree of freedom. Under normal circumstances, it is not necessary to calculate your own trend coefficients. These may be found in most books of statistical tables. The sum of squares of the coefficients is also usually included in the table to help reduce the possibility of arithmetic error.

Using the group total scores from the last example, apply a test for quartic trend. The group totals are:

1 3 7 12 7

The quartic trend coefficients are:

1 −4 6 −4 1

Note that the coefficients sum to zero.
The trend has one degree of freedom.

Quartic trend test

1. Group I II III IV V
   Group total 1 3 7 12 7 n =
   Quartic coefficients
2. ct
3. Σct
4. $(\Sigma ct)^2$
5. c^2
6. Σc^2
7. Quartic sum of squares

$$\text{Quartic s.s.} = \frac{(\Sigma ct)^2}{n\Sigma c^2} =$$

8. Non-quartic sum of squares
   Non-quartic s.s. = B.G.S.S. − Quartic s.s.
   =
9. Degrees of Freedom
   Between groups df. =
   Quartic trend df. =
   Non-quartic df. =

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE

Comment

Quartic trend test

1. Group	I	II	III	IV	V	
Group total	1	3	7	12	7	
Quartic coefficients	1	−4	6	−4	1	n = 3
2. ct	1	−12	42	−48	7	
3. Σct						$\Sigma ct = -10$
4. $(\Sigma ct)^2$						$(\Sigma ct)^2 = 100$
5. c^2	1	16	36	16	1	
6. Σc^2						$\Sigma c^2 = 70$

7. Quartic sum of squares

$$\text{Quartic s.s.} = \frac{(\Sigma ct)^2}{n\Sigma c^2} = \frac{100}{3 \times 70} = \frac{100}{210} = 0{\cdot}476$$

8. Non-quartic sum of squares

Non-quartic s.s. = B.G.S.S. − Quartic s.s.
= 24 − 0·476
= 23·524

9. Degrees of Freedom

Between groups df. = 4
Quartic trend df. = 1
Non-quartic df. = 3

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Between groups	24	4	6	15	<0·1%
Quartic trend	0·476	1	0·476	1·19	ns
Non-quartic trend	23·524	3	7·841	19·6	<0·01%
Within groups	4	10	0·4		
TOTAL	28	14			

Comment

The quartic trend sum of squares is not significant.

Hitherto, we have only applied one trend test at a time. Under certain circumstances, we may wish to test for all possible trends. This is possible if we imagine the BETWEEN GROUP SUM OF SQUARES as being made up of a number of separate components, one component for each trend. In the example, we have been dealing with there would be four components each with one degree of freedom.

Between Group Sum of Squares =

Linear s.s.
+Quadratic s.s.
+Cubic s.s.
+Quartic s.s.

This implies one important feature, *that there is no overlap among the components*. The problem now is to generate the trend sum of squares so that they sum exactly to the between group sum of squares. This problem has been solved quite neatly by insisting that the trend coefficients are independent and uncorrelated. If the sets of trend coefficients are uncorrelated there is no overlap among the trend components. The mathematical term for uncorrelated is ORTHOGONAL.

We must now look for ORTHOGONAL sets of trend coefficients. There is a simple test for orthogonality. Two sets of numbers are said to be mutually orthogonal if the sum of their cross products is zero. This is easily illustrated with the coefficients for linear and quadratic trends which we have already used:

Linear coefficients	−2	−1	0	1	2
Quadratic coefficients	−2	1	2	1	−2
Cross products	4	−1	0	1	−4

Sum of cross products = (4)+(−1)+(0)+(1)+(−4) = 0

The sum of the cross products of the coefficients is zero. Therefore, the linear and quadratic coefficients have passed the test for mutual orthogonality.

Linear and cubic coefficients must also pass the test:

Linear	−2	−1	0	1	2
Cubic	−1	2	0	−2	1
Cross products	2	−2	0	−2	2

Sum = (2)+(−2)+(0)+(−2)+(2) = 0

also quadratic and cubic coefficients pass the test:

Quadratic	−2	1	2	1	−2	
Cubic	−1	2	0	−2	1	
Cross products	2	2	0	−2	−2	Sum = 0

We may now say that linear, quadratic and cubic coefficients are mutually orthogonal. Here are the four sets of coefficients together. Check that the quartic coefficients are orthogonal to the other three sets of coefficients:

Linear	−2	−1	0	1	2
Quadratic	−2	1	2	1	−2
Cubic	−1	2	0	−2	1
Quartic	1	−4	6	−4	1

It is not a simple matter to produce sets of coefficients which all sum to zero and at the same time are mutually orthogonal. It is mainly for this reason that special tables have been prepared. It is a simple matter to look up the appropriate set of ORTHOGONAL POLYNOMIALS in table 10.1 on page 96.

Here is an example of a complete trend test. Notice how the trend component sum of squares add up exactly to the between group sum of squares figure.

Analysis of variance: complete trend analysis

Group	I	II	III	IV	V			
Group total	1	3	7	12	7	n	=	3
Linear								
Coefficients	−2	−1	0	1	2	Σc^2	=	10
ct	−2	−3	0	12	14	Σct	=	21
Quadratic								
Coefficients	2	−1	−2	−1	2	Σc^2	=	14
ct	2	−3	−14	−12	14	Σct	=	−13
Cubic								
Coefficients	−1	2	0	−2	1	Σc^2	=	10
ct	−1	6	0	−24	7	Σct	=	−12
Quartic								
Coefficients	1	−4	6	−4	1	Σc^2	=	70
ct	1	−12	42	−48	7	Σct	=	−10

Sum of Squares

Linear $= (\Sigma ct)^2/(n\Sigma c^2) = 21^2/(3 \times 10) = 441/30 = 14{\cdot}7000$

Quadratic $= (\Sigma ct)^2/(n\Sigma c^2) = (-13)^2/(3 \times 14) = 169/42 = 4{\cdot}0238$

Cubic $= (\Sigma ct)^2/(n\Sigma c^2) = (-12)^2/(3 \times 10) = 144/30 = 4{\cdot}8000$

Quartic $= (\Sigma ct)^2/(n\Sigma c^2) = (-10)^2/(3 \times 70) = 100/210 = 0{\cdot}4762$

24·0000

Check: B.G.S.S. = Sum of trend S.S.

24 = 24·0000

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Between groups	24	4	6·0000	15·00	<0·1%
Linear	14·7000	1	14·7000	36·75	<0·1%
Quadratic	4·0238	1	4·0238	10·06	<1·0%
Cubic	4·8000	1	4·8000	12·00	<1·0%
Quartic	0·4762	1	0·4762	1·19	ns
Within groups	4	10	0·4000		
TOTAL	28	14			

Comment

The differences between incentive levels are significant. Three of the trend components are significant (linear, quadratic, cubic). As a consequence, it is difficult to ascribe the differences in group totals wholly to any single type of trend.

Consider the following example and then carry out a complete trend analysis on the results.

Fifteen volunteer subjects were invited to attempt to solve 10 difficult mental puzzles. The total number of puzzles solved in a five minute period was noted. The subjects had been divided into five groups, each group being summoned at a fixed time of day. The five times of day were as follows: 0900 hrs, 1100 hrs, 1300 hrs, 1500 hrs, 1700 hrs. The results are given below.

TIME OF DAY	0900	1100	1300	1500	1700
	0	2	2	1	0
	0	1	2	1	1
	0	1	2	1	1
TOTALS	0	4	6	3	2

The trend coefficients to be used are as follows:

Linear	−2	−1	0	1	2
Quadratic	−2	1	2	1	−2
Cubic	−1	2	0	−2	1
Quartic	1	−4	6	−4	1

One-way analysis of variance

Group	I	II	III	IV	V	
						k =
						n =
						N =
	___	___	___	___	___	
Totals						T =
						C =
						Σx^2 =

T.S.S. =

B.G.S.S. =

W.G.S.S. =

Summary table

Comment

Analysis of variance: complete trend test

Group	I	II	III	IV	V	
Group total						n =
Linear						
Coefficients						Σc^2 =
ct						Σct =
Quadratic						
Coefficients						Σc^2 =
ct						Σct =
Cubic						
Coefficients						Σc^2 =
ct						Σct =
Quartic						
Coefficients						Σc^2 =
ct						Σct =

Sum of squares

Linear S.S. = $(\Sigma ct)^2/(n\Sigma c^2)$ =

Quadratic S.S. =

Cubic S.S. =

Quartic S.S. =

Check: B.G.S.S. = Sum of trend S.S.

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Between groups					
Linear					
Quadratic					
Cubic					
Quartic					
Within groups					
TOTAL					

Comment

One-way analysis of variance

Group	I	II	III	IV	V	
	0	2	2	1	0	
	0	1	2	1	1	n = 3
	0	1	2	1	1	N = 15
	—	—	—	—	—	
	0	4	6	3	2	
						T = 15
	$C = (15)^2/15 = 15$					C = 15
						$\Sigma x^2 = 23$

T.S.S. = 23 − 15 = 8

B.G.S.S. = 65/3 − 15 = 6·6666

W.G.S.S. = 8 − 6·67 = 1·33

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Between groups (time of day)	6·6666	4	1·6666	12·50	0·1%
Within groups (error)	1·3333	10	0·1333		
TOTAL	8·00	14			

Comment

The differences between the groups are significant.

Analysis of variance: complete trend test

Group	I	II	III	IV	V			
Group total	0	4	6	3	2	n	=	3
Linear								
Coefficients	−2	−1	0	1	2	Σc^2	=	10
ct	0	−4	0	3	4	Σct	=	3
Quadratic								
Coefficients	−2	−1	2	1	−2	Σc^2	=	14
ct	0	4	12	3	−4	Σct	=	15
Cubic								
Coefficients	−1	2	0	−2	1	Σc^2	=	10
ct	0	8	0	−6	2	Σct	=	4
Quartic								
Coefficients	1	−4	6	−4	1	Σc^2	=	70
ct	0	−16	36	−12	2	Σct	=	10

Sum of squares

Linear $= (\Sigma ct)^2/(n\Sigma c^2) = 3^2/(3 \times 10) = 9/30 = 0{\cdot}3$

Quadratic $= (\Sigma ct)^2/(n\Sigma c^2) = (15)^2/(3 \times 14) = 225/42 = 5{\cdot}3571$

Cubic $= (\Sigma ct)^2/(n\Sigma c^2) = (4)^2/(3 \times 10) = 16/30 = 0{\cdot}5333$

Quartic $= (\Sigma ct)^2/(n\Sigma c^2) = (10^2)/(3 \times 70) = 100/210 = 0{\cdot}4762$

6·6666

Check B.G.S.S. = Sum of trend S.S.

6·6666 = 6·6666

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Between groups	6·6666	4	1·6666	12·50	<0·1%
Linear	0·3000	1	0·3000	2·25	ns
Quadratic	5·3570	1	5·3570	40·18	<0·1%
Cubic	0·5333	1	0·5333	4·00	ns
Quartic	0·4762	1	0·4762	3·57	ns
Within groups	1·33	10	0·1333		
TOTAL	8·00	14			

Comment

Only the quadratic trend is significant. The data then may reasonably be described in the terms of a quadratic trend.

NUMBER OF GROUPS	TYPE OF TREND	COEFFICIENTS										Σc^2
3	Linear	−1	0	1								2
	Quadratic	1	−2	1								6
	Linear	−3	−1	1	3							20
4	Quadratic	1	−1	−1	1							4
	Cubic	−1	3	−3	1							20
	Linear	−2	−1	0	1	2						10
5	Quadratic	−2	1	2	1	−2						14
	Cubic	−1	2	0	−2	1						10
	Quartic	1	−4	6	−4	1						70
	Linear	−5	−3	−1	1	3	5					70
6	Quadratic	5	−1	−4	−4	−1	5					84
	Cubic	−5	7	4	−4	−7	5					180
	Quartic	1	−3	2	2	−3	1					28
	Linear	−3	−2	−1	0	1	2	3				28
7	Quadratic	5	0	−3	−4	−3	0	5				84
	Cubic	−1	1	1	0	−1	−1	1				6
	Quartic	3	−7	1	6	1	−7	3				154
	Linear	−7	−5	−3	−1	1	3	5	7			168
	Quadratic	7	1	−3	−5	−5	−3	1	7			168
8	Cubic	−7	5	7	3	−3	−7	−5	7			264
	Quartic	7	−13	−3	9	9	−3	−13	7			616
	Quintic	−7	23	−17	−15	15	17	−23	7			2184
	Linear	−4	−3	−2	−1	0	1	2	3	4		60
	Quadratic	28	7	−8	−17	−20	−17	−8	7	28		2772
9	Cubic	−14	7	13	9	0	−9	−13	−7	14		990
	Quartic	14	−21	−11	9	18	9	−11	−21	14		2002
	Quintic	−4	11	−4	−9	0	9	4	−11	4		468
	Linear	−9	−7	−5	−3	−1	1	3	5	7	9	330
	Quadratic	6	2	−1	−3	−4	−4	−3	−1	2	6	132
10	Cubic	−42	14	35	31	12	−12	−31	−35	−14	42	8580
	Quartic	18	−22	−17	3	18	18	3	−17	−22	18	2860
	Quintic	−6	14	−1	−11	−6	6	11	1	−14	6	780

TABLE 10.1 Orthogonal coefficients for use in analysis of variance trend test.

11

Planned comparisons

Analysis of variance techniques tell us whether different treatments produce significantly different effects. More precisely, they specify whether the variability of the group totals is greater than we might reasonably predict from inspection of the variance within the groups. Quite often, we find that it is not enough to know that the groups, taken all together, are different from each other. We may wish to compare one specific treatment with another or one set of treatments with another set. The following medical example provides an illustration.

Table 11.1 gives the results of an experiment where 12 patients, all of whom reported symptoms of giddiness, were randomly allotted 3 to each of 4 treatments. The first two treatments consisted of oral administration of ANTADIZ, a new sedative drug. Both groups received the same dosage though group 1 patients received it in one pill and group 2 patients in 2 pills. Treatment group 3 patients were given no pill at all and group 4 patients were given a harmless dummy pill (PLACEBO). Subjects were then asked to report the number of giddy spells experienced in the following 24 hrs.

	ANTADIZ		NO DRUG	
Group	1 Pill 1	2 Pills 2	No pill 3	Placebo 4
	0 1 2	1 1 2	6 4 5	3 3 4

TABLE 11.1 Number of dizziness spells reported during a 24 hr period by patients undergoing 4 types of treatment.

The variance analysis of these results is given below. We can see that the differences between the groups are significant. The analysis does not give us any indication, however, of the significance of specific effects such as the difference between sedative and non-sedative groups or placebo and 'no pill' groups. It merely informs us that the four treatments, taken together, produce different results.

One pill	Two pills	No pills	Placebo	
0	1	6	3	$n_{group} = 3$
1	1	4	3	$N = 12$
2	2	5	4	

1. Group totals: $t_1 = 3$ $t_2 = 4$ $t_3 = 15$ $t_4 = 10$
2. Grand total: $T = 32$
3. Correction term: $C = T^2/N = 32^2/12 = 1024/12$ $C = 85{\cdot}33$
4. Sum of squares of raw scores: $\Sigma x^2 = 122$
5. Total sum of squares: T.S.S. $= \Sigma x^2 - C = 122 - 85{\cdot}33$ $= 36{\cdot}67$
6. Between groups sum of squares:

$$\text{B.G.S.S.} = \frac{3^2+4^2+15^2+10^2}{3} - 85{\cdot}33 = \frac{350}{3} - 85{\cdot}33 \qquad = 31{\cdot}34$$

7. Within groups sum of squares: W.G.S.S. $=$ T.S.S. $-$ B.G.S.S.
   $= 36{\cdot}67 - 31{\cdot}34 = 5{\cdot}33$
8. Degrees of freedom:
   (a) B.G. df. $= 4-1 = 3$
   (b) W.G. df. $= 12-4 = 8$
   (c) Total df. $= 12-1 = 11$

Summary table

SOURCE	S.S.	df.	ESTIMATE VARIANCE	F	SIGNIFICANCE
Between groups (pills)	31·34	3	10·44	15·58	$<0{\cdot}1\%$
Within groups (error)	5·33	8	0·67		
TOTAL	36·67	11			

Comment

The differences between the groups taken together are significant at better than the 0·1% level.

Clearly, it is legitimate to ask specific questions concerning individual groups or sets of groups. The following questions are illustrative.

1. Does the placebo have any effect, i.e., are the scores in group 4 different from those in group 3?
2. Is it more effective to consume the same dose in two pills or just one, i.e., are scores in group 1 different from those in group 2?
3. Is the sedative effective, i.e., are scores in groups 1 and 2 different from those in groups 3 and 4?

The technique of planned comparisons has been developed to meet these requirements. Fortunately, the procedures are exactly the same as for trend tests; only the coefficients are different. In terms of calculation, then, there is nothing new to learn. We will need, merely, to consider our choice of coefficients carefully. These depend intimately upon the questions you wish to ask. Accordingly, it is not possible to produce a table of coefficients as we did for trend tests. You must create your own coefficients to suit your own needs. As you will see this can be something of an art but in most cases it is a simple matter.

The comparisons are said to be planned because they must be specified before the scores are inspected. If you make the decision to carry out comparisons after having seen the results, you will be engaging in unplanned comparisons. It cannot be overstressed that this is a different exercise, calling for different procedures. It is perfectly legitimate, of course, though somewhat more complicated and it will not be discussed in this book. The distinction centres on the difficulty of devising appropriate estimates of significance levels when the decision to test is based on the existence of a large difference between two or more groups. Plainly, if you only attempt a comparison if a clear difference is already present, then you will obtain apparent significance spuriously often. With planned comparisons, the decision to test is made before any differences have been seen. Therefore, *always plan your comparisons before inspecting the results.*

We need a set of coefficients for each comparison we make, just as we needed a set for each trend we examined. The same rules applied now as then:

1. The coefficients in each set must sum to zero.
2. If more than one set is used then all sets must be mutually orthogonal.

There is no reason why we need always attempt more than one comparison. Just as we could study a linear or quadratic trend in isolation, we may study a single comparison. In the following analysis, we shall attempt a multiple comparison, however, for the purposes of illustration. The two restrictions given above will impose a limit on the number of comparisons allowed just as the number of trends to be tested was restricted.

We shall now consider the three questions raised above. Fortunately, in all three questions we are comparing a single group against another group (group 3 vs. 4 and group 1 vs. 2) or two groups against two groups (groups 1 and 2 vs. groups 3 and 4). These may be called *even comparisons*. Even comparisons offer little difficulty in choosing coefficients. When comparing two groups we allot one a positive coefficient (say +1) and the other a negative coefficient (say −1). It does not matter which is positive and which negative. The first comparison concerns groups 3 and 4 (placebo vs. 'no pill'). The coefficients are as follows:

Group	1	2	3	4		
Coefficients	0	0	+1	−1	Σc	= 0

Notice that groups not involved in the comparison have a coefficient of zero. Notice also that the sum of the coefficients is zero.

The second comparison concerns groups 1 and 2 (one pill vs. two pills). The coefficients are accordingly:

Group	1	2	3	4		
Coefficients	+1	−1	0	0	Σc	= 0

Clearly, the second set of coefficients is orthogonal to the first, since the sum of cross products is zero.

First set	0	0	+1	−1		
Second set	+1	−1	0	0		
Cross products	0	0	0	0	Sum	= 0

The third comparison sets both groups 1 and 2 against groups 3 and 4. In order that groups 1 and 2 should have equal weight, they must have equal coefficients. The same is true for groups 3 and 4. All the coefficients must sum to zero. In addition, this new set of coefficients must be orthogonal to the first two sets. It seems to be a tall order, but in fact the answer is easy; groups 1 and 2 have the coefficient +1 and groups 3 and 4 have the coefficient −1.

Group	1	2	3	4		
Coefficients	+1	+1	−1	−1	Σc	= 0

Check for orthogonality:

First set	0	0	+1	−1		
Third set	+1	+1	−1	−1		
Cross products	0	0	−1	+1	Sum	= 0

Also:

Second set	+1	−1	0	0		
Third set	+1	+1	−1	−1		
	+1	−1	0	0	Sum	= 0

Our three sets of coefficients are accordingly:

0	0	+1	−1
+1	−1	0	0
+1	+1	−1	−1

They all sum to zero.
They are all mutually orthogonal.
It is not always so easy to hit on the correct set of coefficients. Indeed, sometimes it may not be possible. As an exercise you may try to find coefficients for the two following comparisons:

1. Group 1 vs. group 2.
2. Group 2 vs. group 3.

You will not, alas, be successful. We must resign ourselves to the fact that certain comparisons are mutually incompatible and cannot be handled by the technique of planned comparisons, though other techniques do exist. The problem centres on the fact that we are attempting to subdivide the between groups sum of squares into a number of non-overlapping parts. It is the role of the coefficients to guarantee that the parts do not overlap and that the sum of the component parts does not exceed the total between group variance. Clearly, once we have selected our first comparison, we are limited in our choice of succeeding comparisons.

The application of these coefficients to data given in table 10.1 is shown below.

Multiple planned comparison analysis

Group	One pill	Two pills	No pills	Placebo	
Group total	3	4	15	10	$n = 3$
First comparison					
Coefficients	0	0	+1	−1	$\Sigma c^2 = 2$
ct	0	0	15	−10	$\Sigma ct = 5$
Second comparison					
Coefficients	+1	−1	0	0	$\Sigma c^2 = 2$
ct	3	−4	0	0	$\Sigma ct = -1$
Third comparison					
Coefficients	+1	+1	−1	−1	$\Sigma c^2 = 4$
ct	3	4	−15	−10	$\Sigma ct = 18$

Sum of squares

First comparison $= (\Sigma ct)^2/(n\Sigma c^2) = 5^2/(3 \times 2) = 25/6 = 4{\cdot}17$
Second comparison $= (-1)^2/(3 \times 2) = 1/6 = 0{\cdot}17$
Third comparison $= 18^2/(3 \times 4) = 324/12 = 27{\cdot}00$

Sum = 31·34

Check

The sum of the comparison S.S. must *never* exceed the B.G.S.S.

In this case, they are equal (B.G.S.S. = 31·34).

Note They will always be equal if the number of comparisons is equal to the number of degrees of freedom as they are in this case.

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Between groups	31·34	3	10·44	15·58	<0·1%
1. Groups 3 vs. 4	4·17	1	4·17	6·22	ns
2. Groups 1 vs. 2	0·17	1	0·17	<1	ns
3. Groups 1 and 2 vs. 3 and 4	27.00	1	27·00	40·3	<0·1%
Within groups	5·33	8	0·67		
TOTAL	36·67	11			

Comment

The significant between groups variance is best described mainly in terms of a significant difference between the sedative groups (1 and 2) and the non-sedative groups.

Table 11.2 gives the result of an educational experiment which clearly calls for planned comparisons. The research concerns the ability of immigrant students to adjust to university life. Four students from each of five groups were selected and scored on a 10 point 'adjustment' scale from 9 points (well adjusted, very settled) to 0 points (not adjusted, very unhappy). The first group of students were nationals living with their parents while the second group were migrant nationals living far from home. Thus, groups 1 and 2 contained individuals of the same nationality. Groups 3, 4 and 5 contained Asians, West Indians and Africans, all of whom had only recently arrived in the country.

NATIONALS		IMMIGRANT		
(a) AT HOME	(b) AWAY FROM HOME	ASIAN, WEST INDIAN, AFRICAN		
Group 1	2	3	4	5
7	6	1	5	2
4	6	0	4	1
6	4	3	4	2
3	4	2	2	2

TABLE 11.2 Adjustment ratings for five groups of college students.

Before data were collected the following comparisons were planned:

1. Nationals vs. Immigrants – Group 1 and 2 vs. Groups 3, 4 and 5.
2. Nationals 'at home' vs. Nationals 'migrant' – Group 1 vs. Group 2.
3. Asian stock vs. Negroid stock (African and West Indian) – Groups 3 vs. Groups 4 and 5.
4. West Indian vs. African – Group 4 vs. Group 5.

We are dealing here with some *uneven* comparisons which, in general, offer more difficulty in the quest for coefficients. The appropriate coefficients are set out below. You should confirm that they meet the necessary restrictions of summing to zero and of mutual orthogonality. You should also check that the coefficients correspond intuitively with the spirit of the comparisons.

Group	1	2	3	4	5
Comparison 1	+3	+3	−2	−2	−2
Comparison 2	+1	−1	0	0	0
Comparison 3	0	0	+2	−1	−1
Comparison 4	0	0	0	+1	−1

Clearly, the correct coefficients are not always immediately obvious and finding them can be something of an art. Fortunately, not all examples are as difficult as this and usually the solution appears after a few moments of trial and error.

Carry out first a one way analysis of variance on the scores given in table 11.2. Then go on to attempt the multiple planned comparison exercise using the coefficients given above.

Group	(1) At home	(2) Migrant nationals	(3) Asians	(4) West Indians	(5) African
	7	6	1	5	2
	4	6	0	4	1
	6	4	3	4	2
	3	4	2	2	2
				n_{group}	=
				N	=

1. Group totals:
2. Grand totals: T =
3. Correction term: $C = T^2/N$ =
4. Sum of squares of raw scores: Σx^2 =
5. Total sum of squares: T.S.S. $= \Sigma x^2 - C$ =
6. Between groups sum of squares: =
7. Within groups sum of squares: =
8. Degrees of freedom:

B.G. df. $= k - 1 =$
W.G. df. $= N - k =$
Total df. $= N - 1 =$

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Between groups (student type)					
Within groups (error)					
TOTAL					

Comment

Multiple planned comparison analysis

Group	At home	Migrant nationals	Indian	West Indian	African	
Group total:						$n =$
First comparison						
Coefficients						$\Sigma c^2 =$
ct						$\Sigma ct =$
Second comparison						
Coefficients						$\Sigma c^2 =$
ct						$\Sigma ct =$
Third comparison						
Coefficients						$\Sigma c^2 =$
ct						$\Sigma ct =$
Fourth comparison						
Coefficients						$\Sigma c^2 =$
ct						$\Sigma ct =$

Comparison sum of squares; $(\Sigma ct)^2/(n\Sigma c^2)$

First comparison S.S. = =

Second comparison S.S. = =

Third comparison S.S. = =

Fourth comparison S.S. = =

Sum =

Check that B.G.S.S. is not less than the sum of comparison

Sum of squares ——— ≮ ———

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Between groups					
Comparisons:					
1 and 2 vs. 3, 4 and 5					
1 vs. 2					
3 vs. 4 and 5					
4 vs. 5					
Within groups					
TOTAL					

Group (1) At home	(2) Migrant nationals	(3) Asian	(4) West Indian	(5) African		
7	6	1	5	2		
4	6	0	4	1	n_{group}	$= 4$
6	4	3	4	2	N	$= 20$
3	4	2	2	2		

1. Group totals:
   $t_1 = 20 \quad t_2 = 20 \quad t_3 = 6 \quad t_4 = 15 \quad t_5 = 7$
2. Grand total: T $= 68{\cdot}0$
3. Correction term: $C = T^2/N = 68^2/20 = 4624/20$ C $= 231{\cdot}2$
4. Sum of squares of raw scores $\Sigma x^2 = 302{\cdot}0$
5. Total sum of squares: T.S.S. $= \Sigma x^2 - C = 302 - 231{\cdot}2 = 70{\cdot}8$
6. Between groups sum of squares:

$$\text{B.G.S.S.} = \frac{20^2+20^2+6^2+15^2+7^2}{4} - 231{\cdot}2 = 277{\cdot}5 - 231{\cdot}2 = 46{\cdot}3$$

7. Within groups sum of squares: W.G.S.S. $=$ T.S.S. $-$ B.G.S.S.
   $= 70{\cdot}8 - 46{\cdot}3$
   $= 24{\cdot}5$
8. Degrees of freedom:
   (a) B.G. df. $= 5 - 1 = 4$
   (b) W.G. df. $= 20 - 5 = 15$
   (c) Total df. $= 20 - 1 = 19$

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Between groups (student type)	46·3	4	11·58	7·1	$< 1{\cdot}0\%$
Within groups (error)	24·5	15	1·63		
TOTAL	70·8	19			

Comment

The differences between the student groups are significant at better than the 1·0% level.

Multiple planned comparison analysis

Group	At home	Migrant nationals	Indian	West Indian	African	$n = 4$
Group total	20	20	6	15	7	
First comparison						
Coefficients	+3	+3	−2	−2	−2	$\Sigma c^2 = 30$
ct	60	60	−12	−30	−14	$\Sigma ct = 64$
Second comparison						
Coefficients	+1	−1	0	0	0	$\Sigma c^2 = 2$
ct	20	−20	0	0	0	$\Sigma ct = 0$
Third comparison						
Coefficients	0	0	+2	−1	−1	$\Sigma c^2 = 6$
ct	0	0	12	−15	−7	$\Sigma ct = -10$
Fourth comparison						
Coefficients	0	0	0	+1	−1	$\Sigma c^2 = 2$
ct	0	0	0	15	−7	$\Sigma ct = 8$

Comparison sum of squares $= (\Sigma ct)^2/(n\Sigma c^2)$

First comparison s.s. $= 64^2/(4 \times 30) = 34{\cdot}13$

Second comparison s.s. $= 0^2/(4 \times 2) = 0{\cdot}0$

Third comparison s.s. $= 10^2/(4 \times 6) = 4{\cdot}17$

Fourth comparison s.s. $= 8^2/(4 \times 2) = 8{\cdot}00$

Sum $= 46{\cdot}3$

Check that B.G.S.S. $\leqslant$ Sum of comparison sum of squares

$46{\cdot}3 \leqslant 46{\cdot}3$

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Between groups	46·3	4	11·58	7·1	< 1·0%
comparisons:					
1. 1 and 2 vs. 3, 4, and 5	34·1	1	34·1	20·94	< 0·1%
2. 1 vs. 2	0·0	1	0·0	0·0	ns
3. 3 vs. 4 and 5	4·2	1	4·2	2·58	ns
4. 4 vs. 5	8·0	1	8·0	4·9	< 5·0%
Within groups	24·5	15	1·63		
TOTAL	70·8	19			

Comment

The between group variance is best explained in terms of the large difference between nationals and immigrants and a small difference between the West Indian and African negroes.

12

Three factor analysis of variance

When two factor analysis of variance was first introduced, you probably realized that the method could be extended to three, four, five and more factors. In fact, extending the technique is quite easy. The main difficulty is that the calculations become increasingly numerous. Considerable discipline and neatness in your working will be required if mistakes are to be avoided. These procedures similarly require clarity of thinking. The three factor technique involves repeating the same exercise many times on different sets of figures. It takes no feat of memory to remember how to do it, but clear thinking is needed to pinpoint the figures to be worked.

As you already know, the basic formula involves finding a set of totals, squaring them, summing the squares, dividing the total by the number of scores which went into a single total. Finally, the correction term must be subtracted to give the sum of squares.

$$\text{Sum of squares} = \frac{\Sigma t^2}{n} - C$$

This formula is repeatedly applied to different data and is never changed. Even the total sum of squares formula:

$$\text{Total sum of squares} = \Sigma x^2 - C$$

is really the same because each score (x) is a little total with only one score contributing (n = 1). So, if x = t and n = 1,

$$\frac{\Sigma t^2}{n} - C = \frac{\Sigma x^2}{1} - C$$

We will now go through a simple example in some detail explaining the steps as we proceed. The raw scores are given in table 12.1 and the accumulated results will be found in the summary in table 12.2 which you should consult as we progress. The raw scores are taken from an experiment in person perception. Twenty-four men students each listened to a single tape recorded message and then answered a questionnaire concerning favourable and unfavourable personality features which he imagined the reader of the message to have. Half of the messages are read by men (group X) and half

by women (group Y). The message readers fall into four age groups:— group 1: 10–20 yrs; group 2: 30–40 yrs; group 3: 50–60 yrs; group 4: 70–80 yrs.

Before listening to the recorded message, subjects were asked to read a short story, supposedly unrelated to the experimental task. There are three kinds of story: A—a murder story; B—a neutral story; C—a success story. Each subject experienced one of the 24 possible conditions and his final score represents the number of favourable features he attributed to the reader of the tape recorded message.

	Age group	1	2	3	4
	Story type	A B C	A B C	A B C	A B C
SEX	X	3 5 7	2 4 5	4 4 6	2 3 5
	Y	4 6 8	2 5 7	4 5 6	1 7 6

TABLE 12.1 Number of favourable items checked in a person perception experiment.

This is a complicated experiment but the aim should be fairly clear. When people assess the personality of others on limited evidence—in this case a short tape recording—are they influenced by the sex of the speaker, or the age of the speaker, or by an immediate previous experience—the short stories? Of course, the situation may be complex. Perhaps the assessor gives favourable reports for young girls and older men. In this case, age and sex are relevant but not in a simple fashion. He does not favourably assess all females or all men; only young females and older men. Such effects are called interactions rather than main effects. Other interactions are, of course, possible and we shall discuss them as we meet them.

The first steps in the calculation are familiar:

1. Correction term: $C = T^2/N$

   $N = 24$ and $T = 111$

   $C = 111^2/24 = 513{\cdot}375$

(At each point in the calculations you should check the main figures and make sure that you know where they came from.)

2. Total sum of squares: T.S.S. $= \Sigma x^2 - C$

   T.S.S. $= 591 - 513{\cdot}375$ T.S.S. $= 77{\cdot}62$

Main effects

3. *Age*; find the age totals:

   $t_1 = 33$, $t_2 = 25$, $t_3 = 29$, $t_4 = 24$, $n = 6$ *Check*!

   $$\text{Age s.s.} = \frac{\Sigma t^2}{n} - C = \frac{3131}{6} - 513{\cdot}375 \qquad \text{Age s.s.} = 8{\cdot}46$$

   with 3 degrees of freedom.

4. *Story type*; find the totals for group A, B, C

$t_A = 22$, $t_B = 39$, $t_C = 50$, $n = 8$ *Check*!

For example, the first total covers *all* people in group A:

$$3+4+2+2+4+4+2+1$$

$$\text{Story type s.s.} = \frac{\Sigma t^2}{n} - C = \frac{4505}{8} - 513{\cdot}375 \quad \text{Story s.s.} = 49{\cdot}75$$

with 2 degrees of freedom.

5. *Sex*; find the totals for groups X and Y

$t_x = 50$, $t_y = 61$, $n = 12$ *Check*!

$$\text{Sex s.s.} = \frac{\Sigma t^2}{n} - C = \frac{6221}{12} - 513{\cdot}375 \quad \text{Sex s.s.} = 5{\cdot}04$$

with 1 degree of freedom.

Interactions

6. To explore the *age* by *story type* interaction we must draw up a two-way table of totals:

		AGE 1	2	3	4	
STORY TYPE	A	7	4	8	3	$n = 2$
	B	11	9	9	10	$\Sigma t^2 = 1155$
	C	15	12	12	11	

You should check these figures most carefully to see how each total is based on two scores only.

We calculate the sum of squares in the usual way.

$$\text{Cell s.s.} = \frac{\Sigma t^2}{n} - C = \frac{1155}{2} - 513{\cdot}375 = 64{\cdot}125$$

This 64·125 represents the variability of the totals in the above table. This variance is affected by three factors: (a) variance due to age group (b) variance due to story type and (c) variance due to neither of these factors—possibly an interaction of the first two. Accordingly, to get an estimate of the interaction we merely subtract the main effects from the cell sum of squares given above.

Age × Story interaction

$$= \text{Cell s.s.} - \text{Age s s.} - \text{Story s.s.}$$
$$= 64{\cdot}125 - 8{\cdot}46 - 49{\cdot}75$$
$$= 5{\cdot}92$$

The degrees of freedom are easily calculated. If we draw up the totals again, this time with group totals, we can see which scores are free to vary.

	1	2	3	4	TOTALS
A	7	4	8	③	22
B	11	9	9	⑩	39
C	⑮	⑫	⑫	⑪	50
TOTALS	33	25	29	24	111

Once the row and column totals are fixed, then only 6 of the scores are free to vary. There is a simple way of calculating the degrees of freedom which always works:

$$P \times Q \text{ df.} = P \text{ df.} \times Q \text{ df.}$$

In our case,

$$\text{Age} \times \text{Story df.} = \text{Age df.} \times \text{Story df.}$$
$$= 3 \times 2 = 6 \text{ df.}$$

7. *Age × Sex interaction*: first, draw up a two factor table.

		AGE 1	2	3	4	
SEX	X	15	11	14	10	n = 3
	Y	18	14	15	14	Σt^2 = 1583

Be sure you know how these totals were calculated. The top left hand value (15) covers the *three* individuals (scoring 3, 5 and 7) who listened to a tape recording of a young male voice.
Calculate the cells sum of squares.

$$\text{Cells s.s.} = \frac{\Sigma t^2}{n} - C = \frac{15^2 + 11^2 + 14^2 + \text{etc.}}{3} - C = \frac{1583}{3} - 513{\cdot}375$$
$$= 14{\cdot}29$$

Subtract the relevant main effects to estimate the interaction s.s.

Age × Sex interaction = 14·29 − 8·46 − 5·04 = 0·79

Age × Sex df. = Age df. × Sex df. = 3 × 1 = 3 df.

8. Story × Sex interaction: first, draw up the two factor table

		STORY A	B	C	
SEX	X	11	16	23	n = 4
	Y	11	23	27	Σt^2 = 2285

Calculate cells s.s.

$$\text{Cells s.s.} = \frac{\Sigma t^2}{n} - C = \frac{2285}{4} - 513{\cdot}375 = 57{\cdot}875$$

Calculate interaction

Story × Sex interaction = 57·875 − 49·75 − 5·04 = 3·08

Story × Sex df. = Story df. × Sex df. = 2 × 1 = 2

9. We have explored all possible first order interactions and now we must attempt the second order.

Age × *Story* × *Sex interaction*. We draw up the three factor table:

	1			2			3			4			
	A	B	C	A	B	C	A	B	C	A	B	C	
X	3	5	7	2	4	5	4	4	6	2	3	5	n = 1
Y	4	6	8	2	5	7	4	5	6	1	7	6	Σt^2 = 591

It just so happens that the cells sum of squares is equal to the total sum of squares, but we shall ignore this fact in the interest of demonstrating a general procedure.

$$\text{Cells s.s.} = \frac{\Sigma t^2}{n} - C = \frac{591}{1} - 513{\cdot}375 = 77{\cdot}62$$

To obtain the Age × Story × Sex interaction we must subtract all the main effects and their respective interactions with each other.

Age × Story × Sex interaction =

77·62 − 8·46 − 49·75 − 5·04 − 5·92 − 0·79 − 3·08 = 4·58

Age × Story × Sex df. = Age df. × Story df. × Sex df.

= 3 × 2 × 1 = 6 df.

10. Consult the summary table 12.2. All of the sum of squares add up to the total sum of squares and there is no error term. We met this problem first in chapter 7. Since there is only one score in each sub-group there can be no within groups (error) variance. If there had been two or more scores in each sub-group, then the Age × Story × Sex cells sum of squares would have been less than the total sum of squares leaving some within groups variance to be used as an error term.

If, however, we are prepared to assume that there is no second order interaction, then we may conclude that the Age × Story × Sex interaction sum of squares is due to error variance alone. Of course, if the interaction is a real effect then we will make a mistake, but not a serious one, since we will have an inflated error term which will reduce all our F-values accordingly. We are not, therefore, likely to obtain spuriously significant values of F. Nevertheless, it is an unhappy procedure and no substitute for having more than one score in each sub-group which yields a proper error term.

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Age	8·46	3	2·82	3·7	ns
Story type	49·75	2	24·88	32·7	<0·1%
Sex	5·04	1	5·04	6·6	<5·0%
Age × Story	5·92	6	0·99	1·3	ns
Age × Sex	0·79	3	0·27	0·3	ns
Story × Sex	3·08	2	1·54	2·0	ns
Age × Story × Sex	4·58	6	0·76		
(error)					
TOTAL	77·62	23			

TABLE 12.2 Summary table (see text). Values of F have been computed using the second order interaction variance estimate for an error term.

Discussion of results

Clearly, the biggest influence on the assessment of the personality of the message reader is the nature of the story given to the student before the test. The murder story (group A) suppressed the elicitation of favourable responses —total 22—while the success story encouraged them—total 50. This was a strong effect and clearly significant (<0·1%).

Another main effect (Sex) produced a significant difference. Female readers elicited more favourable responses—total 61—than males—total 50.

None of the interactions was significant. This is an important finding when assessing the significance of main effects. To say that the interactions were non-significant is to say that sub-group totals could be predicted with reasonable accuracy from a knowledge of main group totals—that is there were no surprises in the sub-groups. In the next example, we shall see how a significant interaction invalidates a simple interpretation of the significant main effects.

We must remember to qualify our results by repeating that we had no proper error term in this design; only a substitute based on the triple interaction term. This is a most unsatisfactory state of affairs which was only allowed because it made the exposition shorter and easier to follow. A more complete—and therefore much longer—three factor analysis of variance example follows below. You should note that each sub-classification (or cell) contains two scores. This means that we can compute the within cells variance which is the *proper* error term. The more scores we have in each cell, the more reliable the error term. We normally consider two scores per cell to be the minimum acceptable number.

The next example concerns the effect of asking a candidate to work 'as quickly as possible' in an intelligence test situation. It might seem that this instruction would improve the scoring of all candidates. Our imaginary investigator felt differently. He felt that such an instruction might fluster older candidates and cause them to make mistakes. Similarly, he felt that it would not help dull children but spur the brighter children with the spirit of competition and good hopes of success.

Accordingly, he took two groups of children; one group young (Y) the other old (O). The groups were both subdivided into high (H) and low (L) I.Q. children. Half of the children in each sub-group were told to 'work as quickly as possible' (fast or F) and the other children to 'work at their own pace' (slow or S). There were 16 children in all; 2 in each sub-classification. Each child was given the same 20 problems to solve. The number of problems correctly solved at the end of 15 minutes is given in table 12.3.

Old				Young			
Fast		Slow		Fast		Slow	
High	Low	High	Low	High	Low	High	Low
11	7	7	7	7	6	7	5
10	6	8	6	8	5	6	6

TABLE 12.3 Number of problems correctly solved within 15 minutes for 16 children.

We should remember that we expect to find that older children are better, that high I.Q. children are better and that the FAST instructions will yield a higher average score. These items are not of central interest. We are interested to know whether age or I.Q. level *interact* with the instructions. That is, do the instructions have a different effect at different ages? Do they have a different effect with children of differing intellectual capacity? This is a much more subtle issue.

To simplify the calculations, subtract 5 from all scores (see chapter 6):

Old				Young			
Fast		Slow		Fast		Slow	
H	L	H	L	H	L	H	L
6	2	2	2	2	1	2	0
5	1	3	1	3	0	1	1

N =
T =
Σx^2 =

1. Calculate the correction term: $C = T^2/N =$
2. Calculate the total sum of squares: T.S.S. $= \Sigma x^2 - C =$

Main effects

3. *Age*: draw up the age totals: $t_O =$; $t_Y =$; n =

   Age s.s. $= \dfrac{\Sigma t^2}{n} - C =$

   Age df. =
4. *Speed*: draw up the speed totals: $t_F =$; $t_S =$; n =

   Speed s.s. $= \dfrac{\Sigma t^2}{n} - C =$

   Speed df. =
5. *I.Q.*: draw up the I.Q. totals: $t_H =$; $t_L =$; n =

   I.Q. s.s. $= \dfrac{\Sigma t^2}{n} - C =$

   I.Q. df. =

Interactions

6. *Age* × *speed*: draw up the cell totals:

	F	S	
Y			n =
O			Σt^2 =

   Age × Speed cells s.s. $= \dfrac{\Sigma t^2}{n} - C =$

   Age × Speed interaction = Cells s.s. − Age s.s. − Speed s.s.

   = =

   Age × Speed df. = Age × Speed df.

   = =
7. *Age* × *I.Q.*: draw up the cell totals:

	H	L	
Y			n =
O			Σt^2 =

   Age × I.Q. cells s.s. $= \dfrac{\Sigma t^2}{n} - C =$

   Age × I.Q. interaction s.s. = Cells s.s. − Age s.s. − I.Q. s.s.

   = =

   Age × I.Q. df. = Age df. × I.Q. df.

   = =

8. *Speed* × *I.Q.*: draw up the cell totals:

	F	S	
H			n =
L			Σt^2 =

Speed × I.Q. cells S.S. $= \frac{\Sigma t^2}{n} - C =$

Speed × I.Q. interaction S.S. = Cells S.S. − Speed S.S. − I.Q. S.S.
= =

Speed × I.Q. d.f. = Speed d.f. × I.Q. d.f.
= =

9. *Age* × *Speed* × *I.Q.*: draw up the cell totals

	O		Y		
	F	S	F	S	
H					n =
L					Σt^2 =

Age × Speed × I.Q. cells S.S. $= \frac{\Sigma t^2}{n} - C =$

Age × Speed × I.Q. interaction S.S. =
Cells S.S. − Age S.S. − Speed S.S. − I.Q. S.S. − Age × Speed S.S. − Age × I.Q. S.S. − Speed × I.Q. S.S.
= =

Age × Speed × I.Q. d.f. = Age d.f. × Speed d.f. × I.Q. d.f.
= =

10. *Error*: Error = T.S.S. − (Sum of all other S.S.)
=

Error d.f. = Total d.f. − (Sum of all other d.f.)
=

Summary table

SOURCE	S.S.	d.f.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Age					
Speed					
I.Q.					
Age × Speed					
Age × I.Q.					
Speed × I.Q.					
Age × Speed × I.Q.					
Error					
TOTAL					

Comments

O				Y			
F		S		F		S	
H	L	H	L	H	L	H	L
6	2	2	2	2	1	2	0
5	1	3	1	3	0	1	1

$T = 32$
$N = 16$
$\Sigma x^2 = 104$

1. *Correction term*: $C = T^2/N = 32^2/16 = 1024/16 = 64$
2. *Total sum of squares*: T.S.S. $= \Sigma x^2 - C = 104 - 64 = 40$

Main effects

3. *Age*: age totals: $t_O = 22$; $t_Y = 10$; $n = 8$
   Age S.S. $= 1$
   Age S.S. $= \Sigma t^2/n - C = 584/8 - 64 = 73 - 64 = 9$
   Age df. $= 1$
4. *Speed*: speed totals $t_F = 20$; $t_S = 12$; $n = 8$
   Speed S.S. $= \Sigma t^2/n - C = 544/8 - 40 = 68 - 64 = 4$
   Speed df. $= 1$
5. *I.Q.*: I.Q. totals: $t_H = 24$; $t_L = 8$; $n = 8$
   I.Q. S.S. $= \Sigma t^2/n - C = 640/8 - 40 = 80 - 64 = 16$
   I.Q. df. $= 1$

Interactions

6. *Age* × *Speed*: cell totals:

	F	S
O	14	8
Y	6	4

$n = 4$
$\Sigma/t^2 = 312$

3.5 2
1.5 1

Age × Speed cells S.S. $= \Sigma t^2/n - C = 312/4 - 64 = 78 - 64 = 14$
Age × Speed interaction S.S.: Cells S.S. − Age S.S. − Speed S.S.
$= 14 - 9 - 4 = 1$
Age × Speed df. = Age df. × Speed df.
$= 1 \times 1 = 1$

7. *Age* × *I.Q.*: cell totals:

	H	L
O	16	6
Y	8	2

$n = 4$
$\Sigma t^2 = 360$

Age × I.Q. cells S.S. $= \Sigma t^2/n - C = 360/4 - 64 = 90 - 64 = 26$
Age × I.Q. interaction S.S. = cells S.S. − Age S.S. − I.Q. S.S.
$= 26 - 9 - 16 = 1$
Age × I.Q. df. = Age df. × I.Q. df.
$= 1 \times 1 = 1$

8. *Speed* × *I.Q.*: cell totals:

	F	S
H	16	8
L	4	4

$n = 4$

$\Sigma t^2 = 352$

Speed × I.Q. cells S.S. $= \Sigma t^2/n - C = 352/4 - 64 = 88 - 64 = 24$

Speed × I.Q. interactions = cells S.S. − Speed S.S. − I.Q. S.S.
$= 24 - 4 - 16 = 4$

Speed × I.Q. df. = Speed df. × I.Q. df.
$= 1 \times 1 = 1$

9. *Age* × *Speed* × *I.Q.*: cell totals:

	O		Y	
	F	S	F	S
H	11	5	5	3
L	3	3	1	1

$n = 2$

$\Sigma t^2 = 200$

Age × Speed × I.Q. cells S.S. $= 200/2 - 64 = 100 - 64 = 36$

Age × Speed × I.Q. interaction = Cells S.S. − (Sum of all others)
$= 36 - (9+4+16+1+1+4) = 1$

Age × Speed × I.Q. df. = Age df. × Speed df. × I.Q. df.
$= 1 \times 1 \times 1 = 1$

10. *Error:* Error S.S. = T.S.S. − (Sum of other S.S. terms)
$= 40 - 36 = 4$

Summary table

SOURCE	S.S.	df.	VARIANCE ESTIMATE	F	SIGNIFICANCE
Age	9	1	9	18	<1·0%
Speed	4	1	4	8	<2·5%
I.Q.	16	1	16	32	<0·1%
Age × Speed	1	1	1	2	ns
Age × I.Q.	1	1	1	2	ns
Speed × I.Q.	4	1	4	8	<2·5%
Age × Speed × I.Q.	1	1	1	2	ns
Error	4	8	0·5		
TOTAL	40	15			

Comments

All of the main effects, as we might expect, generate significant differences. Of the interaction effects, only the speed by I.Q. term reaches significance and this must be considered carefully. First of all, study the cell total on which the term is based:

		F(Fast)	S(Slow)
(High I.Q.)	H	16	8
(Low I.Q.)	L	4	4

It can be seen that the high I.Q. children perform better under instructions to work quickly. This is, however, not true for low I.Q. children. Thus, it is not true to say that, *in all cases*, instructions to hurry produce an improvement. In this respect, the significant main effects for speed and I.Q. are misleading and must be qualified with reference to the interaction of the two factors.

13

Some theoretical considerations

Populations and samples

Consider the statement 'Americans are taller than Japanese'. Usually, this does not mean that all Americans are taller than all Japanese. It would be better expressed as 'the average height of all Americans is greater than the average height of all Japanese'. We might test this statement by measuring the height of all Americans and all Japanese. This is not, of course, a very good idea and most men of science would settle for measuring the height of a random sample of Americans and a random sample of Japanese. In this case, we are estimating population values from sample values.

Clearly, the sample values are of limited reliability. As the sample size increases, the accuracy of the sampling procedure increases too, but we can never be certain that the sample values are equal to the population values. It might happen that your sample of Japanese is taller than your sample of Americans. If your sampling is truly random, this is unlikely to happen but be assured that it will happen from time to time.

Analysis of variance has the specific purpose of dealing with uncertainty produced by using random sampling designs. It acknowledges the possibility of making errors and attempts (in the significance level statement) to assess just how likely it is that the conclusions you have made are indeed in error.

Nearly all experiments in the behavioural sciences are characterized by these sampling problems. The effects of a drug on speed of reaction can only be studied by sampling a few reaction times with and without the drug. Again, we are faced with the problem of how certain we can be that the samples are representative of the population scores. The problem is crucial if we wish to generalize on the basis of our results.

Significance levels

Random sampling has much in common with gambling, which may explain why it has been so thoroughly studied by statisticians. The algebraic formulae which they have developed allow us to estimate how likely it is that our samples are accurate within certain limits. Similarly, they predict how often we must expect samples which are positively misleading. With some additional

complication, these formulae can be applied to the analysis of variance situation.

If our experimental results indicate differences among the average scores of the groups under study, we must carefully consider the implications. It could be that our experimental results faithfully reflect reality and that repeat experiments would yield similar results. Alternatively, we might have a bad and misleading sample on this occasion.

When we see differences among the group averages we may wish to reject the null hypothesis (the idea that we expect the groups to yield equal scores). The significance level statement expresses the likelihood that we have made a mistake in rejecting the null hypothesis.

To say that we reject the null hypothesis at the 5% level of significance is to say that while we reject the null hypothesis we recognize a 5 in 100 (or 1 in 20) chance that it is really true. It may seem a little nonsensical at first but that is the real meaning of the statement.

F-tables

Formulae (for calculating how misleading your results are) are used to prepare the F-tables found in this book. Perhaps we can take it on trust that they have been carefully and accurately prepared. Nevertheless, a practical illustration of their nature may be welcome.

	GROUP	
	A	B
	6·78	23·88
	39·87	82·14
	90·59	51·76
	17·80	24·72
	81·07	23·90
	87·64	93·06
	67·96	94·08
	43·88	44·68
	98·54	40·37
	51·65	57·00
	44·09	33·41
	79·43	90·00
	31·48	73·62
	75·06	85·90
	28·25	36·82
	26·86	90·00
	12·48	36·95
	88·29	39·37
	78·80	9·43
	0·91	79·52
MEAN	52·57	55·53

F-ratio = 0·10 with 1 and 38 df.

TABLE 13.1 Results of a dummy experiment; 20 random numbers have been allocated to each group.

Imagine a dummy experiment with two groups A and B. Instead of collecting any data, we simply allot numbers at random to the two groups. Clearly, there is no basic difference between groups A and B. Whichever group gets the higher average score does so entirely by chance. Table 13.1 shows a sample set of scores and the results of a simple analysis of variance.

I went on to repeat this experiment 1000 times, with the aid of a computer, using different scores for each run. I took the resulting 1000 F values and sorted them into descending order. The list is given in table 13.2.

19·0	14·1	13·1	11·0	11·0	10·4	10·4	8·2	7·8	7·8
7·5	7·5	7·4	7·4	7·4	6·7	6·6	6·6	6·2	6·2
6·0	6·0	6·0	5·9	5·8	5·7	5·6	5·4	5·4	5·4
5·3	5·3	5·3	5·1	5·1	5·0	4·7	4·5	4·4	4·4
4·3	4·3	4·3	4·2	4·2	4·2	4·1	4·1	4·1	4·1
4·0	4·0	3·9	3·9	3·7	3·7	3·7	3·7	3·7	3·6
3·6	3·6	3·5	3·5	3·5	3·5	3·4	3·4	3·4	3·3
3·3	3·3	3·3	3·2	3·2	3·2	3·2	3·2	3·2	3·2
3·2	3·2	3·2	3·2	3·1	3·1	3·1	3·1	3·1	3·1
3·0	3·0	3·0	3·0	3·0	2·9	2·9	2·9	2·9	2·9
2·9	2·8	2·7	2·7	2·7	2·7	2·7	2·7	2·6	2·6
2·6	2·6	2·6	2·6	2·6	2·6	2·6	2·5	2·5	2·5
2·5	2·4	2·4	2·4	2·4	2·4	2·4	2·4	2·4	2·4
2·3	2·3	2·3	2·3	2·3	2·3	2·3	2·2	2·2	2·2
2·1	2·1	2·1	2·1	2·1	2·1	2·1	2·1	2·0	2·0
2·0	2·0	2·0	2·0	2·0	2·0	1·9	1·9	1·9	1·9
1·9	1·9	1·9	1·9	1·9	1·9	1·9	1·8	1·8	1·8
1·8	1·8	1·8	1·8	1·8	1·8	1·8	1·8	1·8	1·8
1·8	1·7	1·7	1·7	1·7	1·7	1·7	1·7	1·7	1·7
1·7	1·7	1·7	1·7	1·7	1·7	1·7	1·6	1·6	1·6
1·6	1·6	1·6	1·6	1·5	1·5	1·5	1·5	1·5	1·5
1·5	1·5	1·5	1·5	1·5	1·5	1·5	1·5	1·5	1·5
1·5	1·4	1·4	1·4	1·4	1·4	1·4	1·4	1·4	1·4
1·4	1·4	1·4	1·4	1·4	1·4	1·4	1·3	1·3	1·3
1·3	1·3	1·3	1·3	1·3	1·3	1·3	1·3	1·3	1·3
1·3	1·3	1·3	1·2	1·2	1·2	1·2	1·2	1·2	1·2
1·2	1·2	1·2	1·2	1·2	1·2	1·2	1·2	1·2	1·2
1·1	1·1	1·1	1·1	1·1	1·1	1·1	1·1	1·1	1·1
1·1	1·1	1·1	1·1	1·1	1·1	1·1	1·1	1·1	1·1
1·1	1·1	1·0	1·0	1·0	1·0	1·0	1·0	1·0	1·0
1·0	1·0	1·0	1·0	1·0	1·0	1·0	1·0	1·0	1·0
1·0	1·0	1·0	1·0	0·9	0·9	0·9	0·9	0·9	0·9
0·9	0·9	0·9	0·9	0·9	0·9	0·9	0·9	0·9	0·9
0·9	0·9	0·9	0·9	0·9	0·9	0·9	0·9	0·9	0·9
0·9	0·9	0·9	0·9	0·9	0·8	0·8	0·8	0·8	0·8
0·8	0·8	0·8	0·8	0·8	0·8	0·8	0·8	0·8	0·8
0·8	0·8	0·8	0·8	0·8	0·8	0·8	0·8	0·8	0·8
0·7	0·7	0·7	0·7	0·7	0·7	0·7	0·7	0·7	0·7
0·7	0·7	0·7	0·7	0·7	0·7	0·7	0·7	0·7	0·7
0·7	0·7	0·7	0·7	0·7	0·7	0·7	0·7	0·7	0·7
0·7	0·7	0·7	0·7	0·7	0·7	0·7	0·7	0·7	0·6
0·6	0·6	0·6	0·6	0·6	0·6	0·6	0·6	0·6	0·6
0·6	0·6	0·6	0·6	0·6	0·6	0·6	0·6	0·6	0·6
0·6	0·6	0·6	0·6	0·6	0·6	0·6	0·6	0·6	0·6
0·6	0·6	0·6	0·6	0·6	0·6	0·6	0·6	0·6	0·6
0·6	0·5	0·5	0·5	0·5	0·5	0·5	0·5	0·5	0·5

0·5	0·5	0·5	0·5	0·5	0·5	0·5	0·5	0·5	0·5
0·5	0·5	0·5	0·5	0·5	0·5	0·5	0·5	0·5	0·5
0·5	0·5	0·5	0·5	0·5	0·5	0·5	0·5	0·5	0·5
0·5	0·4	0·4	0·4	0·4	0·4	0·4	0·4	0·4	0·4
0·4	0·4	0·4	0·4	0·4	0·4	0·4	0·4	0·4	0·4
0·4	0·4	0·4	0·4	0·4	0·4	0·4	0·4	0·4	0·4
0·4	0·4	0·4	0·4	0·4	0·4	0·4	0·4	0·4	0·4
0·4	0·4	0·4	0·4	0·4	0·4	0·4	0·4	0·4	0·4
0·4	0·4	0·4	0·3	0·3	0·3	0·3	0·3	0·3	0·3
0·3	0·3	0·3	0·3	0·3	0·3	0·3	0·3	0·3	0·3
0·3	0·3	0·3	0·3	0·3	0·3	0·3	0·3	0·3	0·3
0·3	0·3	0·3	0·3	0·3	0·3	0·3	0·3	0·3	0·3
0·3	0·3	0·3	0·3	0·3	0·3	0·3	0·3	0·3	0·3
0·3	0·3	0·3	0·3	0·3	0·3	0·3	0·3	0·3	0·3
0·3	0·3	0·3	0·3	0·3	0·3	0·3	0·3	0·2	0·2
0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·2
0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·2
0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·2
0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·2
0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·2
0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·2
0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·2
0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·1	0·1
0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1
0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1
0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1
0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1
0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1
0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1
0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1
0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1
0·1	0·1	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0

TABLE 13.2 F-values (arranged in descending order) obtained from 1000 repeats of the experiment illustrated in table 13.1. The circled values represent the limits of the top 0·1%, 1%, 2·5% and 5% of observed values.

Notice that most scores are very small. Nevertheless, 5% of scores are above 4·0 and some values are very high indeed, high enough to cause an investigator to reject the null hypothesis! Clearly, even when the null hypothesis is true (as it is in this case) large values of F do occur from time to time.

The F-tables are supposed to predict the frequency of high values of F when the null hypothesis is true. Accordingly, we may wish to compare the figures we obtained in our experiment with those predicted in the F-tables. The critical values for both are given in table 13.3. The critical values are those values above which 5% (2·5%, etc.) of F-values should occur when the null hypothesis is true. The tables have been consulted with 1 and 38 degrees of freedom since we have 20 scores in each group. The agreement of the two sets of values is good for the 5% and 2·5% levels. It would have been better at the other levels if the experiment had been carried out a million rather than a thousand times.

SIGNIFICANCE LEVEL	CRITICAL VALUES (a) F-TABLES	(b) OBSERVED
5%	4·1 approx.	4·1
2·5%	5·4 approx.	5·8
1%	7·3 approx.	7·8
0·1%	12·6 approx.	19·0

TABLE 13.3 Summary of results from table 13.2 compared with critical value of F found in the tables for df.=1 and 38.

Decision making

Our dummy experiments have highlighted the central dilemma of statistical decision making. If you are presented with a moderately high value of F, how can you decide whether it was the result of a sampling fluke or of a genuine effect relating to the nature of the experiment. The truth is that you can never be sure. We take advantage, however, of the fact that the larger the value of F the smaller the likelihood that it was generated by chance factors alone. The significance levels are an expression of the chances that the F-value could have emerged as the result of sampling error.

Statistics merely produce the significance level statement. You produce the decision. How strong the evidence need be before you reject the null hypothesis is your own business. When you report your decision, however, you should specify the significance level so that the reader is free to come to a separate decision if he wants to.

Assumptions

When the F-tables were drawn up, the statisticians made a number of assumptions concerning the data on which the calculations would be based. These assumptions are not very onerous but should be considered carefully. An awareness of them can avoid the occasional serious mistake.

1. Random sampling In any field of measurement, sampling procedures are only effective if the objects or events to be measured are selected at random. This also means that the selection of one object or event should not prejudice the selection of others. It is difficult to be completely random but, if a serious bias creeps into your sampling, it will undermine the values of your results.

2. Homogeneity of variance You may remember (page 34) that we described the within groups variance estimate as the average value of the variances of the individual groups. We do this in an attempt to estimate the overall variability of individual scores. It is only meaningful to do this if the variance is fairly constant from group to group. If the variance in one group is three times as high as that in another then the homogeneity assumption is likely to be violated.

3. Normal distribution of scores We have not discussed the 'normality' of data in this book, although it is a topic of central importance to statisticians. A set of raw scores is said to be roughly *normally distributed* both if there are as many scores above the mean as below it and if the bulk of the scores should cluster about the mean, with extreme scores becoming rarer as they become more extreme.

Data collected from real experiments are never perfectly normal though this may be due simply to the small sample sizes used. Cause for concern, however, is only given if the scores are severely non-normal. For example, the scores given earlier in our random experiment score were not normally distributed and yet the spread of our F-values was very much as we might expect from the F-tables.

Violating the assumptions

Every experiment violates one or more of these assumptions to some degree. This is not a problem unless the violations are severe. When they are severe the significance statements may be subject to considerable error. When the violations are minor then this error is also small. In most cases, the standard procedure of taking the 'next lower' significance level compensates in a conservative manner small violations of the assumptions.

Analysis of variance, under ideal conditions, is an exact tool. Under difficult conditions it becomes less exact but a useful tool none the less. Even when the above assumptions are only barely met, the result of your calculations will give you a rough idea of the significance to attach to your data. In the circumstances of real experimental work, you should never expect anything better than an approximation. It should be noted that all the statistics in the world will never improve upon a repetition of the same experiment. Your analysis should help you to decide whether a repeat performance would be worth while.

Appendix: Instructions for use of F-tables

Degrees of freedom: Your F-value is the ratio of two variance estimates: var (1)/var (2). Each estimate has degrees of freedom associated with it: ν_1, ν_2. To find the required critical F-value, look in the column for ν_1 and the row for ν_2. The critical F-value is found at the intersection of this row and column.

To be significant at the specified significance level your experimentally determined F-ratio must be *larger* than that found in the table. If your value is significant at the 5% level, you should look at the next table and so on until you find the *best* level at which your value is still significant.

If your value for ν_1 or ν_2 is not present in the table, take the *next lower* values which are present and use those for the purpose of finding the critical value of F.

Examples

You have an F-ratio of 9·03 with degrees of freedom $\nu_1 = 1$, $\nu_2 = 20$. Consulting all four tables, we find the following critical F-ratios:

5·0% significance level F = 4·35
2·5% significance level F = 5·87
1·0% significance level F = 8·10
0·1% significance level F = 14·82

Your F-ratio is bigger than the critical values at the 5·0%, 2·5% and 1·0% significance levels, but smaller than the critical value at the 0·1% level. We therefore say that the F-value is significant at better than the 1·0% level of significance. Most experimenters would accept this result as indicating a real experimental effect.

F-TABLES
5% significance level

ν_1		1	2	3	4	5	6	8	9	12
ν_2	2	18·5	19·0	19·2	19·2	19·3	19·3	19·4	19·4	19·4
	3	10·1	9·5	9·2	9·1	9·0	8·9	8·8	8·8	8·7
	4	7·7	6·9	6·6	6·4	6·2	6·1	6·0	6·0	5·9
	5	6·6	5·8	5·4	5·2	5·0	4·9	4·8	4·8	4·6
	6	6·0	5·1	4·7	4·5	4·4	4·3	4·1	4·1	4·0
	7	5·6	4·7	4·3	4·1	4·0	3·9	3·7	3·7	3·6
	8	5·3	4·4	4·0	3·8	3·7	3·6	3·4	3·4	3·3
	9	5·1	4·2	3·9	3·6	3·5	3·4	3·2	3·2	3·0
	10	5·0	4·1	3·7	3·5	3·3	3·2	3·0	3·0	2·9
	12	4·7	3·9	3·5	3·2	3·1	3·0	2·8	2·8	2·7
	14	4·6	3·7	3·3	3·1	3·0	2·8	2·7	2·7	2·5
	16	4·5	3·6	3·2	3·0	2·8	2·7	2·6	2·6	2·4
	18	4·4	3·5	3·1	2·9	2·7	2·6	2·5	2·5	2·3
	20	4·3	3·5	3·1	2·9	2·7	2·6	2·4	2·4	2·3
	22	4·3	3·4	3·0	2·8	2·6	2·5	2·4	2·4	2·2
	24	4·2	3·4	3·0	2·8	2·6	2·5	2·3	2·3	2·2
	30	4·2	3·3	2·9	2·7	2·5	2·4	2·3	2·3	2·1
	40	4·1	3·2	2·8	2·6	2·4	2·3	2·2	2·2	2·0
	60	4·0	3·1	2·7	2·5	2·4	2·2	2·1	2·1	1·9

F-TABLES
$2\frac{1}{2}$% significance level

ν_1		1	2	3	4	5	6	8	9	12
ν_2	2	38·5	39·0	39·2	39·2	39·3	39·3	39·4	39·4	39·4
	3	17·4	16·0	15·4	15·1	14·9	14·7	14·5	14·4	14·3
	4	12·2	10·7	10·0	9·6	9·4	9·2	9·0	8·9	8·7
	5	10·0	8·4	7·8	7·4	7·1	7·0	6·8	6·7	6·5
	6	8·8	7·3	6·6	6·2	6·0	5·8	5·6	5·5	5·4
	7	8·0	6·5	5·9	5·5	5·3	5·1	4·9	4·8	4·7
	8	7·6	6·1	5·4	5·0	4·8	4·6	4·4	4·4	4·2
	9	7·2	5·7	5·1	4·7	4·5	4·3	4·1	4·1	3·9
	10	6·9	5·5	4·8	4·5	4·2	4·1	3·8	3·8	3·6
	12	6·5	5·1	4·5	4·1	3·9	3·7	3·5	3·5	3·3
	14	6·3	4·9	4·2	3·9	3·7	3·5	3·3	3·2	3·0
	16	6·1	4·7	4·1	3·7	3·5	3·3	3·1	3·1	2·9
	18	6·0	4·6	3·9	3·6	3·4	3·2	3·0	3·0	2·8
	20	5·9	4·4	3·9	3·5	3·3	3·1	2·9	2·9	2·7
	22	5·8	4·4	3·8	3·4	3·2	3·0	2·8	2·8	2·6
	24	5·7	4·3	3·7	3·4	3·1	3·0	2·8	2·8	2·5
	30	5·6	4·2	3·6	3·2	3·0	2·9	2·6	2·6	2·4
	40	5·4	4·0	3·5	3·1	2·9	2·7	2·5	2·5	2·3
	60	5·3	3·4	3·3	3·0	2·8	2·6	2·4	2·4	2·2

F-TABLES
1% significance level

ν_1		1	2	3	4	5	6	8	9	12
ν_2	2	98·5	99·0	99·2	99·2	99·3	99·3	99·4	99·4	99·4
	3	34·1	30·8	29·5	28·7	28·2	27·9	27·5	27·5	27·1
	4	21·2	18·0	16·7	16·0	15·5	15·2	14·8	14·7	14·4
	5	16·3	13·3	12·1	11·4	10·1	10·1	10·3	10·2	9·9
	6	13·7	10·9	9·8	9·1	8·7	8·5	8·1	8·0	7·7
	7	12·2	9·5	8·4	7·8	7·5	7·2	6·8	6·7	6·5
	8	11·3	8·6	7·6	7·0	6·6	6·4	6·0	5·9	5·7
	9	10·6	8·0	7·0	6·4	6·0	5·8	5·5	5·4	5·1
	10	10·0	7·6	6·5	6·0	5·6	5·4	5·1	5·0	4·7
	12	9·3	6·9	5·9	5·4	5·1	4·8	4·5	4·4	4·2
	14	8·9	6·5	5·6	5·0	4·7	4·5	4·1	4·0	3·8
	16	8·5	6·2	5·3	4·8	4·4	4·2	3·9	3·8	3·5
	18	8·3	6·0	5·1	4·6	4·2	4·0	3·7	3·6	3·4
	20	8·1	5·8	4·9	4·4	4·1	3·9	3·6	3·5	3·2
	22	7·9	5·7	4·8	4·3	4·0	3·8	3·4	3·3	3·1
	24	7·8	5·6	4·7	4·2	3·9	3·7	3·3	3·3	3·0
	30	7·6	5·4	4·5	4·0	3·7	3·5	3·2	3·2	2·8
	40	7·3	5·2	4·3	3·8	3·5	3·3	3·0	3·0	2·7
	60	7·1	5·0	4·1	3·6	3·3	3·1	2·8	2·8	2·5

F-TABLES
0·1% significance level

ν_1		1	2	3	4	5	6	8	9	12
ν_2	2	998·5	999·0	999·2	999·2	999·3	999·3	999·4	999·4	999·4
	3	167·0	148·5	141·1	137·1	134·6	132·8	130·6	130·0	128·3
	4	74·1	61·2	56·2	53·4	51·7	50·5	49·0	49·0	47·4
	5	47·2	37·1	33·2	31·1	29·8	28·8	27·6	27·3	26·4
	6	35·5	27·0	23·7	21·9	20·8	20·0	19·0	18·8	18·0
	7	29·2	21·7	18·8	17·2	16·2	15·5	14·6	14·4	13·7
	8	25·4	18·5	15·8	14·4	13·5	12·9	12·0	11·8	11·2
	9	22·9	16·4	13·9	12·6	11·7	11·1	10·4	10·2	9·6
	10	21·0	14·9	12·5	11·3	10·5	9·9	9·2	9·0	8·4
	12	18·6	13·0	10·8	9·6	8·9	8·4	7·7	7·5	7·0
	14	17·1	11·8	9·7	8·6	7·9	7·4	6·8	6·7	6·1
	16	16·1	11·0	9·0	7·9	7·3	6·8	6·2	6·0	5·5
	18	15·4	10·4	8·5	7·5	6·8	6·3	5·8	5·6	5·1
	20	14·8	9·9	8·1	7·1	6·5	6·0	5·4	5·2	4·8
	22	14·4	9·6	7·8	6·8	6·2	5·8	5·2	5·0	4·6
	24	14·0	9·3	7·5	6·6	6·0	5·5	5·0	4·9	4·4
	30	13·3	8·8	7·0	6·1	5·5	5·1	4·6	4·5	4·0
	40	12·6	8·2	6·6	5·7	5·1	4·7	4·2	4·1	3·6
	60	12·0	7·8	6·2	5·3	4·8	4·4	3·9	3·8	3·3

ν_1 = df. for numerator
ν_2 = df. for denominator
F = var. est. (1)/var. est. (2)