7 ANALYSIS OF VARIANCE (ANOVA)

Objectives

After studying this chapter you should

- appreciate the need for analysing data from more than two samples;
- understand the underlying models to analysis of variance;
- understand when, and be able, to carry out a one way analysis of variance;
- understand when, and be able, to carry out a two way analysis of variance.

7.0 Introduction

What is the common characteristic of all tests described in Chapter 4?

Consider the following two investigations.

(a) A car magazine wishes to compare the average petrol consumption of THREE similar models of car and has available six vehicles of each model.

(b) A teacher is interested in a comparison of the average percentage marks attained in the examinations of FIVE different subjects and has available the marks of eight students who all completed each examination.

In both these investigations, interest is centred on a comparison of **more than two populations**; THREE models of car, FIVE examinations.

In (a), six vehicles of each of the three models are available so there are three **independent samples**, each of size six. This example requires an extension of the test considered in Section 4.3, which was for two normal population means using independent samples and a pooled estimate of variance.

In (b) however, there is the additional feature that the same eight students each completed the five examinations, so there are five **dependent samples** each of size eight. This example requires an extension of the test considered in Section 4.4, which was for two normal population means using dependent (paired) samples.

This chapter will show that an appropriate method for investigation (a) is a **one way** anova to test for differences between the three models of car. For (b), an appropriate method is a two way anova to test for differences between the five subjects and, if required, for differences between the eight students.

7.1 Ideas for data collection

Undertake at least one of Activities 1 and 2 AND at least one of Activities 3 and 4. You will require your data for subsequent analysis later in this chapter.

Activity 1 Estimating length

Draw a straight line of between 20 cm and 25 cm on a sheet of plain white card. (Only you should know its exact length.)

Collect 6 to 10 volunteers from each of school years 7, 10 and 13. Ask each volunteer to estimate independently the length of the line.

Do differences in year means appear to outweigh differences within years?

Activity 2 Apples

Obtain random samples of each of at least three varieties of apple. The samples should be of at least 5 apples but need not be of the same size.

Weigh, as accurately as you are able, each apple.

Compare variation within varieties with variability between varieties.

Activity 3 Shop prices

Make a list of 10 food/household items purchased regularly by your family.

Obtain the current prices of the items in three different shops; preferably a small 'corner' shop, a small supermarket and a large supermarket or hyper market.

Compare total shop prices.

Activity 4 Weighing scales

Obtain the use of at least three different models of bathroom scales, preferably one of which is electronic. Collect about 10 volunteers and record their weights on each scale. If possible, also have each volunteer weighed on more accurate scales such as those found in health centres or large pharmacies.

You will need to ensure that your volunteers are each wearing, as far as is possible, the same apparel at every weighing.

Assess the differences in total weights between the weighing devices used.

Which model of bathroom scales appears the most accurate?

7.2 Factors and factor levels

Two new terms for analysis of variance need to be introduced at this stage.

Factor – a characteristic under consideration, thought to influence the measured observations.

Level – a value of the factor.

In Activity 1, there is one factor (school year) at three levels (7, 10 and 13).

In Activity 3, there are two factors (item and shop) at 10 and 3 levels, respectively.

What are the factors and levels in Activities 2 and 4?

7.3 One way (factor) anova

In general, one way anova techniques can be used to study the effect of $k(>2)$ levels of a single factor.

To determine if different levels of the factor affect measured observations differently, the following hypotheses are tested.

$$H_0: \mu_i = \mu \quad \text{all} \quad i = 1, 2, \ldots, k$$

$$H_1: \mu_i \neq \mu \quad \text{some} \quad i = 1, 2, \ldots, k$$

where μ_i is the population mean for level i.

Assumptions

When applying one way analysis of variance there are three key assumptions that should be satisfied. They are essentially the same as those assumed in Section 4.3 for $k = 2$ levels, and are as follows.

1. The observations are obtained independently and randomly from the populations defined by the factor levels.
2. The population at each factor level is (approximately) normally distributed.
3. These normal populations have a common variance, σ^2.

Thus for factor level i, the population is assumed to have a distribution which is $N(\mu_i, \sigma^2)$.

Example

The table below shows the lifetimes under controlled conditions, in hours in excess of 1000 hours, of samples of 60W electric light bulbs of three different brands.

Brand		
1	2	3
16	18	26
15	22	31
13	20	24
21	16	30
15	24	24

Assuming all lifetimes to be normally distributed with common variance, test, at the 1% significance level, the hypothesis that there is no difference between the three brands with respect to mean lifetime.

Solution

Here there is one factor (brand) at three levels (1, 2 and 3).
Also the sample sizes are all equal (to 5), though as you will see later this is not necessary.

$$H_0: \mu_i = \mu \quad \text{all} \quad i = 1, 2, 3$$

$$H_1: \mu_i \neq \mu \quad \text{some} \quad i = 1, 2, 3$$

The sample mean and variance (divisor $(n-1)$) for each level are as follows.

	Brand		
	1	2	3
Sample size	5	5	5
Sum	80	100	135
Sum of squares	1316	2040	3689
Mean	16	20	27
Variance	9	10	11

Since each of these three sample variances is an estimate of the common population variance, σ^2, a pooled estimate may be calculated in the usual way as follows.

$$\hat{\sigma}_W^2 = \frac{(5-1)\times 9 + (5-1)\times 10 + (5-1)\times 11}{5+5+5-3} = 10$$

This quantity is called the **variance within samples**. It is an estimate of σ^2 based on $v = 5+5+5-3 = 12$ degrees of freedom. This is irrespective of whether or not the null hypothesis is true, since differences between levels (brands) will have no effect on the within sample variances.

The variability between samples may be estimated from the three sample means as follows.

	Brand		
	1	2	3
Sample mean	16	20	27
Sum		63	
Sum of squares		1385	
Mean		21	
Variance		31	

This variance (divisor $(n-1)$), denoted by $\hat{\sigma}_B^2$ is called the **variance between sample means**. Since it calculated using sample means, it is an estimate of

$$\frac{\sigma^2}{5} \text{ (that is } \frac{\sigma^2}{n} \text{ in general)}$$

based upon $(3-1)=2$ degrees of freedom, but only if the null hypothesis is true. If H_0 is false, then the subsequent 'large' differences between the sample means will result in $5\hat{\sigma}_B^2$ being an inflated estimate of σ^2.

The two estimates of σ^2, $\hat{\sigma}_W^2$ and $5\hat{\sigma}_B^2$, may be tested for equality using the *F*-test of Section 4.1 with

$$F = \frac{5\hat{\sigma}_B^2}{\hat{\sigma}_W^2}$$

as lifetimes may be assumed to be normally distributed.

Recall that the *F*-test requires the two variances to be independently distributed (from independent samples). Although this is by no means obvious here (both were calculated from the same data), $\hat{\sigma}_W^2$ and $\hat{\sigma}_B^2$ are in fact independently distributed.

The test is always one-sided, upper-tail, since if H_0 is false, $5\hat{\sigma}_B^2$ is inflated whereas $\hat{\sigma}_W^2$ is unaffected.

Thus in analysis of variance, the convention of placing the larger sample variance in the numerator of the F statistic is NOT applied.

The solution is thus summarised and completed as follows.

H_0: $\mu_i = \mu$ all $i = 1, 2, 3$

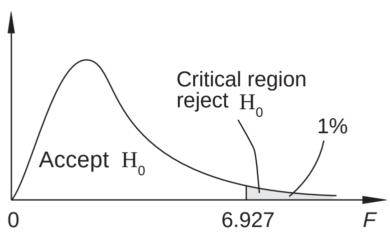
H_1: $\mu_i \neq \mu$ some $i = 1, 2, 3$

Significance level, $\alpha = 0.01$

Degrees of freedom, $v_1 = 2$, $v_2 = 12$

Critical region is $F > 6.927$

Test statistic is $F = \frac{5\hat{\sigma}_B^2}{\hat{\sigma}_W^2} = \frac{155}{10} = 15.5$



This value does lie in the critical region. There is evidence, at the 1% significance level, that the true mean lifetimes of the three brands of bulb do differ.

What is the value of the variance (divisor $(n-1)$), $\hat{\sigma}_T^2$, of the lifetimes, if these are considered simply as one sample of size 15?

What is the value of $14\hat{\sigma}_T^2$?

What is the value of $12\hat{\sigma}_W^2 + 10\hat{\sigma}_B^2$?

At this point it is useful to note that, although the above calculations were based on (actual lifetimes - 1000), the same value would have been obtained for the test statistic (F) using actual lifetimes. This is because F is the ratio of two variances, both of which are unaffected by subtracting a working mean from all the data values. Additionally, in analysis of variance, data values may also be scaled by multiplying or dividing by a constant without affecting the value of the F ratio. This is because each variance involves the square of the constant which then cancels in the ratio. Scaling of data values can make the subsequent analysis of variance less cumbersome and, sometimes, even more accurate.

Notation and computational formulae

The calculations undertaken in the previous example are somewhat cumbersome, and are prone to inaccuracy with non-integer sample means. They also require considerable changes when the sample sizes are unequal. Equivalent computational formulae are available which cater for both equal and unequal sample sizes.

First, some notation.

Number of samples (or levels)	$= k$
Number of observations in ith sample	$= n_i, \quad i = 1, 2, \ldots, k$
Total number of observations	$= n = \sum_i n_i$
Observation j in ith sample	$= x_{ij}, \quad j = 1, 2, \ldots, n_i$
Sum of n_i observations in ith sample	$= T_i = \sum_j x_{ij}$
Sum of all n observations	$= T = \sum_i T_i = \sum_i \sum_j x_{ij}$

The computational formulae now follow.

Total sum of squares,	$SS_T = \sum_i \sum_j x_{ij}^2 - \frac{T^2}{n}$
Between samples sum of squares,	$SS_B = \sum_i \frac{T_i^2}{n_i} - \frac{T^2}{n}$
Within samples sum of squares,	$SS_W = SS_T - SS_B$

A mean square (or unbiased variance estimate) is given by

(sum of squares) $\div$ (degrees of freedom)

e.g. $$\hat{\sigma}^2 = \frac{\sum (x-\bar{x})^2}{n-1}$$

Hence

Total mean square,	$MS_T = \frac{SS_T}{n-1}$
Between samples mean square,	$MS_B = \frac{SS_B}{k-1}$
Within samples mean square,	$MS_W = \frac{SS_W}{n-k}$

Note that for the degrees of freedom: $(k-1)+(n-k)=(n-1)$

Activity 5

For the previous example on 60W electric light bulbs, use these computational formulae to show the following.

(a) $SS_T = 430$ (b) $SS_B = 310$

(c) $MS_B = 155 \; \left(5\hat{\sigma}_B^2\right)$ (d) $MS_W = 10 \; \left(\hat{\sigma}_W^2\right)$

Note that $F = \frac{MS_B}{MS_W} = \frac{155}{10} = 15.5$ as previously.

Anova table

It is convenient to summarise the results of an analysis of variance in a table. For a one factor analysis this takes the following form.

Source of variation	Sum of squares	Degrees of freedom	Mean square	*F* ratio
Between samples	SS_B	$k-1$	MS_B	$\frac{MS_B}{MS_W}$
Within samples	SS_W	$n-k$	MS_W	
Total	SS_T	$n-1$		

Example

In a comparison of the cleaning action of four detergents, 20 pieces of white cloth were first soiled with India ink. The cloths were then washed under controlled conditions with 5 pieces washed by each of the detergents. Unfortunately three pieces of cloth were 'lost' in the course of the experiment. Whiteness readings, made on the 17 remaining pieces of cloth, are shown below.

Detergent

A	B	C	D
77	74	73	76
81	66	78	85
61	58	57	77
76		69	64
69		63	

Assuming all whiteness readings to be normally distributed with common variance, test the hypothesis of no difference between the four brands as regards mean whiteness readings after washing.

Solution

H_0: no difference in mean readings $\mu_i = \mu$ all i

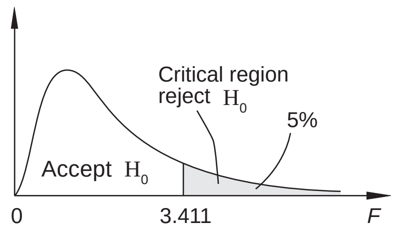
H_1: a difference in mean readings $\mu_i \neq \mu$ some i

Significance level, $\alpha = 0.05$ (say)

Degrees of freedom, $v_1 = k - 1 = 3$

and $v_2 = n - k = 17 - 4 = 13$

Critical region is $F > 3.411$



	A	B	C	D	Total
n_i	5	3	5	4	$17 = n$
T_i	364	198	340	302	$1204 = T$

$$\sum\sum x_{ij}^2 = 86362$$

$$SS_T = 86362 - \frac{1204^2}{17} = 1090.47$$

$$SS_B = \left(\frac{364^2}{5} + \frac{198^2}{3} + \frac{340^2}{5} + \frac{302^2}{4}\right) - \frac{1204^2}{17} = 216.67$$

$$SS_W = 1090.47 - 216.67 = 873.80$$

The anova table is now as follows.

Source of variation	Sum of squares	Degrees of freedom	Mean square	*F* ratio
Between detergents	216.67	3	72.22	1.07
Within detergents	873.80	13	67.22	
Total	1090.47	16		

The F ratio of 1.07 does not lie in the critical region.

Thus there is no evidence, at the 5% significance level, to suggest a difference between the four brands as regards mean whiteness after washing.

Activity 6

Carry out a one factor analysis of variance for the data you collected in either or both of Activities 1 and 2.

Model

From the three assumptions for one factor anova, listed previously,

$$x_{ij} \sim N\left(\mu_i, \sigma^2\right) \quad \text{for } j = 1, 2, \ldots, n_i \text{ and } i = 1, 2, \ldots, k$$

Hence $\qquad x_{ij} - \mu_i = \varepsilon_{ij} \sim N\left(0, \sigma^2\right)$

where ε_{ij} denotes the variation of x_{ij} about its mean μ_i and so represents the inherent random variation in the observations.

If $\mu = \frac{1}{k}\sum_{i=1}^{k}\mu_i$, then $\sum_i\left(\mu_i - \mu\right) = 0$.

Writing $\mu_i - \mu = L_i$ results in $\mu_i = \mu + L_i$ where $\sum_i L_i = 0$.

Hence L_i can be interpreted as the mean effect of factor level i relative to an overall mean μ.

Combining $x_{ij} - \mu_i = \varepsilon_{ij}$ with $\mu_i - \mu = L_i$ results in

$$x_{ij} = \mu + L_i + \varepsilon_{ij} \quad \text{for } j = 1, 2, \ldots, n_i \text{ and } i = 1, 2, \ldots, k$$

This formally defines a model for one way (factor) analysis of variance, where

x_{ij} = jth observation at ith level (in ith sample),

μ = overall factor mean,

L_i = mean effect of ith level of factor relative to μ, where $\sum_i L_i = 0$,

ε_{ij} = inherent random variation $\sim N(0, \sigma^2)$.

Note that as a result,

$$H_0: \mu_i = \mu \text{ (all } i) \Rightarrow H_0: L_i = 0 \text{ (all } i)$$

Estimates of μ, L_i and ε_{ij} can be calculated from observed measurements by

$$\frac{T}{n}, \left(\frac{T_i}{n_i} - \frac{T}{n}\right) \text{ and } \left(x_{ij} - \frac{T_i}{n_i}\right), \text{ respectively.}$$

Thus for the example on 60W electric light bulbs for which the observed measurements (x_{ij}) were

	Brand	
1	**2**	**3**
16	18	26
15	22	31
13	20	24
21	16	30
15	24	24

with $n = 15$, $n_1 = n_2 = n_3 = 5$, $T = 315$, $T_1 = 80$, $T_2 = 100$ and $T_3 = 135$.

Hence, estimates of μ, L_1, L_2 and L_3 are 21, –5, –1 and +6, respectively.

Estimates of the ε_{ij} are best tabulated as shown.

	Brand (estimates of ε_{ij})	
1	**2**	**3**
0	–2	–1
–1	2	4
–3	0	–3
5	–4	3
–1	4	–3

Notice that, relative to the original measurements, these values representing inherent random variation are quite small.

What is the sum of these values?

What is sum of squares of these values and how was it found earlier?

Activity 7

Calculate estimates of μ, L_i and ε_{ij} for the data you collected in either of Activities 1 and 2.

Exercise 7A

1. Four treatments for fever blisters, including a placebo (A), were randomly assigned to 20 patients. The data below show, for each treatment, the numbers of days from initial appearance of the blisters until healing is complete.

Treatment	Number of days				
A	5	8	7	7	8
B	4	6	6	3	5
C	6	4	4	5	4
D	7	4	6	6	5

Test the hypothesis, at the 5% significance level, that there is no difference between the four treatments with respect to mean time of healing.

2. The following data give the lifetimes, in hours, of three types of battery.

	I	50.1	49.9	49.8	49.7	50.0
Type	II	51.0	50.8	50.9	50.9	50.6
	III	49.5	50.1	50.2	49.8	49.3

Analyse these data for a difference between mean lifetimes. (Use a 5% significance level.)

3. Three different brands of magnetron tubes (the key component in microwave ovens) were subjected to stress testing. The number of hours each operated before needing repair was recorded.

	A	36	48	5	67	53
Brand	B	49	33	60	2	55
	C	71	31	140	59	224

Although these times may not represent lifetimes, they do indicate how well the tubes can withstand stress.

Use a one way analysis of variance procedure to test the hypothesis that the mean lifetime under stress is the same for the three brands.

What assumptions are necessary for the validity of this test? Is there reason to doubt these assumptions for the given data?

4. Three special ovens in a metal working shop are used to heat metal specimens. All the ovens are supposed to operate at the same temperature. It is known that the temperature of an oven varies, and it is suspected that there are significant mean temperature differences between ovens. The table below shows the temperatures, in degrees centigrade, of each of the three ovens on a random sample of heatings.

Oven	Temperature (^{o}C)					
1	494	497	481	496	487	
2	489	494	479	478		
3	489	483	487	472	472	477

Stating any necessary assumptions, test for a difference between mean oven temperatures.

Estimate the values of μ (1 value), L_i (3 values) and ε_{ij} (15 values) for the model

(temperature)$_{ij} = x_{ij} = \mu + L_i + \varepsilon_{ij}$. Comment on what they reveal.

5. Eastside Health Authority has a policy whereby any patient admitted to a hospital with a suspected coronary heart attack is automatically placed in the intensive care unit. The table below gives the number of hours spent in intensive care by such patients at five hospitals in the area.

Hospital				
A	**B**	**C**	**D**	**E**
30	42	65	67	70
25	57	46	58	63
12	47	55	81	80
23	30	27		
16				

Use a one factor analysis of variance to test, at the 1% level of significance, for differences between hospitals. (AEB)

6. An experiment was conducted to study the effects of various diets on pigs. A total of 24 similar pigs were selected and randomly allocated to one of the five groups such that the control group, which was fed a normal diet, had 8 pigs and each of the other groups, to which the new diets were given, had 4 pigs each. After a fixed time the gains in mass, in kilograms, of the pigs were measured. Unfortunately by this time two pigs had died, one which was on diet A and one which was on diet C. The gains in mass of the remaining pigs are recorded below.

Diets	**Gain in mass** (kg)			
Normal	23.1	9.8	15.5	22.6
	14.6	11.2	15.7	10.5
A	21.9	13.2	19.7	
B	16.5	22.8	18.3	31.0
C	30.9	21.9	29.8	
D	21.0	25.4	21.5	21.2

Use a one factor analysis of variance to test, at the 5% significance level, for a difference between diets.

What further information would you require about the dead pigs and how might this affect the conclusions of your analysis? (AEB)

7.4 Two way (factor) anova

This is an extension of the one factor situation to take account of a second factor. The levels of this second factor are often determined by groupings of subjects or units used in the investigation. As such it is often called a **blocking factor** because it places subjects or units into homogeneous groups called **blocks**. The design itself is then called a **randomised block design**.

Example

A computer manufacturer wishes to compare the speed of four of the firm's compilers. The manufacturer can use one of two experimental designs.

(a) Use 20 similar programs, randomly allocating 5 programs to each compiler.

(b) Use 4 copies of any 5 programs, allocating 1 copy of each program to each compiler.

Which of (a) and (b) would you recommend, and why?

Solution

In (a), although the 20 programs are similar, any differences between them may affect the compilation times and hence perhaps any conclusions. Thus in the 'worst scenario', the 5 programs allocated to what is really the fastest compiler could be the 5 requiring the longest compilation times, resulting in the compiler appearing to be the slowest! If used, the results would require a one factor analysis of variance; the factor being compiler at 4 levels.

In (b), since all 5 programs are run on each compiler, differences between programs should not affect the results. Indeed it may be advantageous to use 5 programs that differ markedly so that comparisons of compilation times are more general. For this design, there are two factors; compiler (4 levels) and program (5 levels). The factor of principal interest is compiler whereas the other factor, program, may be considered as a blocking factor as it creates 5 blocks each containing 4 copies of the same program.

Thus (b) is the better designed investigation.

The actual compilation times, in milliseconds, for this two factor (randomised block) design are shown in the following table.

	Compiler			
	1	**2**	**3**	**4**
Program A	29.21	28.25	28.20	28.62
Program B	26.18	26.02	26.22	25.56
Program C	30.91	30.18	30.52	30.09
Program D	25.14	25.26	25.20	25.02
Program E	26.16	25.14	25.26	25.46

Assumptions and interaction

The three assumptions for a two factor analysis of variance when there is only one observed measurement at each combination of levels of the two factors are as follows.

1. The population at each factor level combination is (approximately) normally distributed.
2. These normal populations have a common variance, σ^2.
3. The effect of one factor is the same at all levels of the other factor.

Hence from assumptions 1 and 2, when one factor is at level *i* and the other at level *j*, the population has a distribution which is $N(\mu_{ij}, \sigma^2)$.

Assumption 3 is equivalent to stating that there is no interaction between the two factors.

Now interaction exists when the effect of one factor depends upon the level of the other factor. For example consider the effects of the two factors:

sugar (levels none and 2 teaspoons),

and stirring (levels none and 1 minute),

on the sweetness of a cup of tea.

Stirring has no effect on sweetness if sugar is not added but certainly does have an effect if sugar is added. Similarly, adding sugar has little effect on sweetness unless the tea is stirred.

Hence factors sugar and stirring are said to interact.

Interaction can only be assessed if more than one measurement is taken at each combination of the factor levels. Since such situations are beyond the scope of this text, it will always be assumed that interaction between the two factors does not exist.

Thus, for example, since it would be most unusual to find one compiler particularly suited to one program, the assumption of no interaction between compilers and programs appears reasonable.

Is it likely that the assumption of no interaction is valid for the data you collected in each of Activities 3 and 4?

Notation and computational formulae

As illustrated earlier, the data for a two factor anova can be displayed in a two-way table. It is thus convenient, in general, to label the factors as

a **row factor** and a **column factor**.

Notation, similar to that for the one factor case, is then as follows.

Number of levels of row factor	$= r$
Number of levels of column factor	$= c$
Total number of observations	$= rc$
Observation in $(i\,j)$ th cell of table	$= x_{ij}$
(ith level of row factor and	$i = 1, 2, \ldots, r$
jth level of column factor)	$j = 1, 2, \ldots, c$

Sum of c observations in ith row $= T_{Ri} = \sum_j x_{ij}$

Sum of r observations in jth column $= T_{Cj} = \sum_i x_{ij}$

Sum of all rc observations $= T = \sum_i \sum_j x_{ij} = \sum_i T_{Ri} = \sum_j T_{Cj}$

These lead to the following computational formulae which again are similar to those for one factor anova except that there is an additional sum of squares, etc for the second factor.

Total sum of squares, $SS_T = \sum_i \sum_j x_{ij}^2 - \frac{T^2}{rc}$

Between rows sum of squares, $SS_R = \sum_i \frac{T_{Ri}^2}{c} - \frac{T^2}{rc}$

Between columns sum of squares, $SS_C = \sum_j \frac{T_{Cj}^2}{r} - \frac{T^2}{rc}$

Error (residual) sum of squares, $SS_E = SS_T - SS_R - SS_C$

What are the degrees of freedom for SS_T, SS_R and SS_C when there are 20 observations in a table of 5 rows and 4 columns?

What is then the degrees of freedom of SS_E?

Anova table and hypothesis tests

For a two factor analysis of variance this takes the following form.

Source of variation	Sum of squares	Degrees of freedom	Mean square	*F* ratio
Between rows	SS_R	$r-1$	MS_R	$\frac{MS_R}{MS_E}$
Between columns	SS_C	$c-1$	MS_C	$\frac{MS_C}{MS_E}$
Error (residual)	SS_E	$(r-1)(c-1)$	MS_E	
Total	SS_T	$rc-1$		

Notes:

1. The three sums of squares, SS_R, SS_C and SS_E are independently distributed.

2. For the degrees of freedom:

$$(r-1)+(c-1)+(r-1)(c-1)=(rc-1).$$

Using the F ratios, tests for significant row effects and for significant column effects can be undertaken.

H_0: no effect due to row factor

H_1: an effect due to row factor

Critical region,

$$F > F^{(\alpha)}_{[(r-1),\,(r-1)(c-1)]}$$

Test statistic,

$$F_R = \frac{MS_R}{MS_E}$$

H_0: no effect due to column factor

H_1: an effect due to column factor

Critical region,

$$F > F^{(\alpha)}_{[(c-1),\,(r-1)(c-1)]}$$

Test statistic,

$$F_C = \frac{MS_C}{MS_E}$$

Example

Returning to the compilation times, in milliseconds, for each of five programs, run on four compilers.

Test, at the 1% significance level, the hypothesis that there is no difference between the performance of the four compilers.

Has the use of programs as a blocking factor proved worthwhile? Explain.

The data, given earlier, are reproduced below.

	Compiler			
	1	**2**	**3**	**4**
Program A	29.21	28.25	28.20	28.62
Program B	26.18	26.02	26.22	25.56
Program C	30.91	30.18	30.52	30.09
Program D	25.14	25.26	25.20	25.02
Program E	26.16	25.14	25.26	25.46

Solution

To ease computations, these data have been transformed (coded) by

$$x = 100 \times (\text{time} - 25)$$

to give the following table of values and totals.

	Compiler 1	2	3	4	Row totals (T_{Ri})
Program A	421	325	320	362	1428
Program B	118	102	122	56	398
Program C	591	518	552	509	2170
Program D	14	26	20	2	62
Program E	116	14	26	46	202
Column totals (T_{Cj})	1260	985	1040	975	4260 $= T$

$$\sum\sum x_{ij}^2 = 1757768$$

The sums of squares are now calculated as follows.

(Rows = Programs, Columns = Compilers)

$$SS_T = 1757768 - \frac{4260^2}{20} = 850388$$

$$SS_R = \frac{1}{4}\left(1428^2 + 398^2 + 2170^2 + 62^2 + 202^2\right) - \frac{4260^2}{20} = 830404$$

$$SS_C = \frac{1}{5}\left(1260^2 + 985^2 + 1040^2 + 975^2\right) - \frac{4260^2}{20} = 10630$$

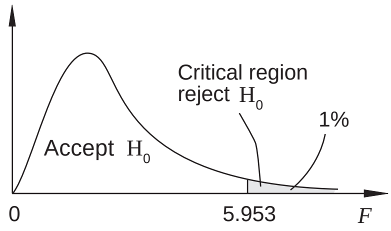
$$SS_E = 850388 - 830404 - 10630 = 9354$$

Anova table

Source of variation	Sum of squares	Degrees of freedom	Mean square	*F* ratio
Between programs	830404	4	207601.0	266.33
Between compilers	10630	3	3543.3	4.55
Error (residual)	9354	12	779.5	
Total	850388	19		

H_0: no effect on compilation times due to compilers

H_1: an effect on compilation times due to compilers



Significance level, $\alpha = 0.01$

Degrees of freedom, $v_1 = c - 1 = 3$

and $v_2 = (r-1)(c-1) = 4 \times 3 = 12$

Critical region is $F > 5.953$

Test statistic $F_C = 4.55$

This value does not lie in the critical region. Thus there is no evidence, at the 1% significance level, to suggest a difference in compilation times between the four compilers.

The use of programs as a blocking factor has been very worthwhile. From the anova table

(a) SS_R accounts for $\frac{830404}{850388} \times 100 = 97.65\%$ of the total variation in the observations, much of which would have been included in SS_E had not programs been used as a blocking variable,

(b) $F_R = 266.33$ which indicates significance at any level!

Activity 8

Carry out a two factor analysis of variance for the data you collected in either or both of Activities 3 and 4.

In each case identify the blocking factor, and explain whether or not it has made a significant contribution to the analysis.

Model

With x_{ij} denoting the one observation in the ith row and jth column, (ij)th cell, of the table, then

$$x_{ij} \sim N\left(\mu_{ij}, \sigma^2\right) \quad \text{for } i = 1, 2, \ldots, r \text{ and } j = 1, 2, \ldots, c$$

or $$x_{ij} - \mu_{ij} = \varepsilon_{ij} \sim N\left(0, \sigma^2\right)$$

However, it is assumed that the two factors do not interact but simply have an additive effect, so that

$\mu_{ij} = \mu + R_i + C_i$ with $\sum_i R_i = \sum_j C_j = 0$, where

μ = overall mean

R_i = mean effect of ith level of row factor relative to μ

C_j = mean effect of jth level of column factor relative to μ

ε_{ij} = inherent random variation

As a result, when testing for an effect due to rows, the hypotheses may be written as

$$H_0: R_i = 0 \quad \text{(all } i)$$

$$H_1: R_i \neq 0 \quad \text{(some } i)$$

What are the corresponding hypotheses when testing for an effect due to columns?

If required, estimates of μ, R_i, C_j and ε_{ij} can be calculated from observed measurements by

$$\frac{T}{rc}, \quad \left(\frac{T_{Ri}}{c} - \frac{T}{rc}\right), \quad \left(\frac{T_{Cj}}{r} - \frac{T}{rc}\right), \quad \left(x_{ij} - \frac{T_{Ri}}{c} - \frac{T_{Cj}}{r} + \frac{T}{rc}\right),$$

respectively.

What are the estimates of μ, R_1, C_2 and ε_{12} in the previous example, based upon the transformed data?

Activity 9

Calculate estimates of some of μ, R_i, C_j and ε_{ij} for the data you collected in either of Activities 3 and 4.

Exercise 7B

1. Prior to submitting a quotation for a construction project, companies prepare a detailed analysis of the estimated labour and materials costs required to complete the project. A company which employs three project cost assessors, wished to compare the mean values of these assessors' cost estimates. This was done by requiring each assessor to estimate independently the costs of the same four construction projects. These costs, in £0000s, are shown in the next column.

	Assessor		
	A	B	C
Project 1	46	49	44
Project 2	62	63	59
Project 3	50	54	54
Project 4	66	68	63

Perform a two factor analysis of variance on these data to test, at the 5% significance level, that there is no difference between the assessors' mean cost estimates.

2. In an experiment to investigate the warping of copper plates, the two factors studied were the temperature and the copper content of the plates. The response variable was a measure of the amount of warping. The resultant data are as follows.

	Copper content (%)			
Temp (°C)	40	60	80	100
50	17	19	23	29
75	12	15	18	27
100	14	19	22	30
125	17	20	22	30

Stating all necessary assumptions, analyse for significant effects.

3. In a study to compare the body sizes of silkworms, three genotypes were of interest: heterozygous (HET), homozygous (HOM) and wild (WLD). The length, in millimetres, of a separately reared cocoon of each genotype was measured at each of five randomly chosen sites with the following results.

		Site				
		1	**2**	**3**	**4**	**5**
Silkworm	HOM	29.87	28.24	32.27	31.21	29.85
	HET	32.51	30.82	34.46	34.01	32.99
	WLD	35.76	34.14	36.54	34.95	36.11

Identify the blocking factor. Has it proved useful? Explain.

Test, at the 1% significance level, for a difference in mean lengths between genotypes.

4. Four different washing solutions were being compared to study their effectiveness in retarding bacteria growth in milk containers. The study was undertaken in a laboratory, and only four trials could be run on any one day. As days could represent a potential source of variability, the experimenter decided to use days as a blocking variable. Observations were made for five days with the following (coded) results.

	Day				
Solution	**1**	**2**	**3**	**4**	**5**
A	12	21	17	38	29
B	15	23	16	43	35
C	6	11	7	32	28
D	18	27	23	43	35

Stating any necessary assumptions, analyse for significant differences between solutions.

Was the experimenter wise to use days as a blocking factor? Justify your answer.

5. The Marathon of the South West took place in Bristol in April 1982. The table below gives the times taken, in hours, by twelve competitors to complete the course, together with their type of occupation and training method used.

	Types of occupation		
Training methods	**Office worker**	**Manual worker**	**Professional sportsperson**
A	5.7	2.9	3.6
B	4.5	4.8	2.4
C	3.9	3.3	2.6
D	6.1	5.1	2.7

Carry out an analysis of variance and test, at the 5% level of significance, for differences between types of occupation and between training methods.

The age and sex of each of the above competitors are subsequently made available to you. Is this information likely to affect your conclusions and why? (AEB)

6. Information about the current state of a complex industrial process is displayed on a control panel which is monitored by a technician. In order to find the best display for the instruments on the control panel, three different arrangements were tested by simulating an emergency and observing the reaction times of five different technicians. The results, in seconds, are given below.

	Technician				
Arrangement	**P**	**Q**	**R**	**S**	**T**
A	2.4	3.3	1.9	3.6	2.7
B	3.7	3.2	2.7	3.9	4.4
C	4.2	4.6	3.9	3.8	4.5

Carry out an analysis of variance and test for differences between technicians and between arrangements at the 5% significance level.

Currently arrangement C is used and it is suggested that this be replaced by arrangement A. Comment, briefly, on this suggestion and on what further information you would find useful before coming to a definite decision. (AEB)

7.5 Miscellaneous Exercises

1. After completing a six month typing course with the Speedyfingers Institute, four people, A, B, C and D, had their typing speed measured, in words per minute, on each of five kinds of work. The results are given in the table below.

	Legal	Business	Numeric	Prose I	Prose II
A	40	47	42	45	53
B	34	32	14	36	44
C	33	40	31	48	44
D	24	26	25	27	45

Carry out an analysis of variance and test, at the 5% level of significance, for differences between the people and between kinds of work.

Subsequently it transpired that A and C used electric typewriters, whilst B and D used manual typewriters. Does this information affect your conclusions and why? (AEB)

2. A batch of bricks was randomly divided into three parts and each part was stored by a different method. After one week the percentage water content of a number of bricks stored by each method was measured.

Method of storage	% water content					
1	7.4	8.5	7.1	6.2	7.8	
2	5.5	7.1	5.6			
3	4.8	5.1	6.2	4.9	6.1	7.1

Making any necessary assumptions, use a one factor analysis of variance to test, at the 5% significance level, for differences between methods of storage.

If low water content is desirable, state which method of storage you would recommend, and calculate a 95% confidence interval for its mean percentage water content after one week. [You may assume that the estimated variance of a sample mean is given by (Within samples mean square) $\div$ (sample size).] (AEB)

3. A textile factory produces a silicone proofed nylon cloth for making into rainwear. The chief chemist thought that a silicone solution of about 12% strength would yield a cloth with a maximum waterproofing index. It was also suspected that there might be some batch to batch variation because of slight differences in the cloth.

To test this, five different strengths of solution were tested on each of three different batches of cloth. The following values of the waterproofing index were obtained.

	Strength of silicone solution (%)				
Cloth	6	9	12	15	18
A	20.8	20.6	22.0	22.6	20.9
B	19.4	21.2	21.8	23.9	22.4
C	19.9	21.1	22.7	22.7	22.1

[You may assume that the total sum of squares of the observations $\left(\sum x^2\right) = 7022.79$.]

Carry out an analysis of variance to test, at the 5% significance level, for differences between strengths of silicone solution and between cloths.

Comment on the chief chemist's original beliefs in the light of these results and suggest what actions the chief chemist might take. (AEB)

4. (a) A catering firm wishes to buy a meat tenderiser, but was concerned with the effect on the weight loss of meat during cooking. The following results were obtained for the weight loss of steaks of the same pre-cooked weight when three different tenderisers were used.

		Weight loss in grams				
Tenderiser	A	36	28	42	58	
	B	17	59	33		
	C	36	74	29	55	48

Carry out a one factor analysis of variance and test at the 5% significance level whether there is a difference in weight loss between tenderisers.

(b) Time and temperature are important factors in the weight loss during cooking. As these had not been taken account of during the first trial, a further set of results was obtained where all the steaks were cooked at the same temperature and cooking times of 20, 25 and 30 minutes were used. An analysis of these data led to the following table.

Source of variation	Sum of squares	Degrees of freedom
Between tenderisers	321	2
Between times	697	2
Error	85	4
Total	1103	8

Test at the 5% significance level for differences between tenderisers and between times.

(c) Contrast the results obtained in (a) and (b) and comment on why the two sets of data can lead to different conclusions. (AEB)

5. A commuter in a large city can travel to work by car, bicycle or bus. She times four journeys by each method with the following results, in minutes.

Car	Bicycle	Bus
27	34	26
45	38	41
33	43	35
31	42	46

(a) Carry out an analysis of variance and test at the 5% significance level whether there are differences in the mean journey times for the three methods of transport.

(b) The time of day at which she travels to work varies. Bearing in mind that this is likely to affect the time taken for the journey, suggest a better design for her experiment and explain briefly why you believe it to be better.

(c) Suggest a factor other than leaving time which is likely to affect the journey time and two factors other than journey time which might be considered when choosing a method of transport. (AEB)

6. (a) As part of a project to improve the steerability of trucks, a manufacturer took three trucks of the same model and fitted them with soft, standard and hard front springs, respectively. The turning radius (the radius of the circle in which the truck could turn full circle) was measured for each truck using a variety of drivers, speeds and surface conditions. Use the following information to test for a difference between springs at the 5% significance level.

Source	Sum of squares	Degrees of freedom
Between springs	37.9	2
Within springs	75.6	18
Total	113.5	20

(b) A statistician suggested that the experiment would be improved if the same truck was used all the time with the front springs changed as necessary and if the speed of the truck was controlled.

The following results for turning circle, in metres, were obtained.

	Springs		
Speed	**Soft**	**Standard**	**Hard**
15 km/h	42	43	39
25 km/h	48	50	48

Carry out a two factor analysis of variance and test at the 5% significance level for differences between springs and between speeds. [You may assume that the total sum of squares about the mean (SS_T) is 92.]

(c) Compare the two experiments and suggest a further improvement to the design. (AEB)

7. A drug is produced by a fermentation process. An experiment was run to compare three similar chemical salts, X, Y and Z, in the production of the drug. Since there were only three of each of four types of fermenter A, B, C and D available for use in the production, three fermentations were started in each type of fermenter, one containing salt X, another salt Y and the third salt Z. After several days, samples were taken from each fermenter and analysed. The results, in coded form, were as follows.

Fermenter type			
A	**B**	**C**	**D**
X 67	Y 73	X 72	Z 70
Z 68	Z 65	Y 80	X 68
Y 78	X 69	Z 73	Y 69

State the type of experimental design used.

Test, at the 5% level of significance, the hypothesis that the type of salt does not affect the fermentation.

Comment on what assumption you have made about the interaction between type of fermenter and type of salt. (AEB)

8. A factory is to introduce a new product which will be assembled from a number of components. Three different designs are considered and four employees are asked to compare their speed of assembly. The trial is carried out one morning and each of the four employees assembled design A from 8.30 am to 9.30 am, design B from 10.00 am to 11.00 am and design C from 11.30 am to 12.30 pm. The number of products completed by each of the employees is shown in the following table.

	Employee			
Design	**1**	**2**	**3**	**4**
A	17	4	38	8
B	21	6	52	20
C	28	9	64	22

(a) Carry out a two factor analysis of variance and test at the 5% significance level for differences between designs and between employees. [You may assume that the total sum of squares about the mean (SS_T) is 3878.9.]

(b) Comment on the fact that all employees assembled the designs in the same order. Suggest a better way of carrying out the experiment.

(c) The two factor analysis assumes that the effects of design and employee may be added. Comment on the suitability of this model for these data and suggest a possible improvement. (AEB)

9. In a hot, third world country, milk is brought to the capital city from surrounding farms in churns carried on open lorries. The keeping quality of the milk is causing concern. The lorries could be covered to provide shade for the churns relatively cheaply or refrigerated lorries could be used but these are very expensive. The different methods were tried and the keeping quality measured. (The keeping quality is measured by testing the pH at frequent intervals and recording the time taken for the pH to fall by 0.5. A high value of this time is desirable.)

Transport method	Keeping quality (hours)					
Open	16.5	20.0	14.5	13.0		
Covered	23.5	25.0	30.0	33.5	26.0	
Refrigerated	29.0	34.0	26.0	22.5	29.5	30.5

(a) Carry out a one factor analysis of variance and test, at the 5% level, for differences between methods of transport.

(b) Examine the method means and comment on their implications.

(c) Different farms have different breeds of cattle and different journey times to the capital, both of which could have affected the results. How could the data have been collected and analysed to allow for these differences? (AEB)

10. A hospital doctor wished to compare the effectiveness of 4 brands of painkiller A, B, C and D. She arranged that when patients on a surgical ward requested painkillers they would be asked if their pain was mild, severe or very severe. The first patient who said mild would be given brand A, the second who said mild would be given brand B, the third brand C and the fourth brand D. Painkillers would be allocated in the same way to the first four patients who said their pain was severe and to the first four patients who said their pain was very severe.

The patients were then asked to record the time, in minutes, for which the painkillers were effective.

The following data were collected.

Brand	A	B	C	D
mild	165	214	173	155
severe	193	292	142	211
very severe	107	110	193	212

(a) Carry out a two factor analysis of variance and test at the 5% significance level for differences between brands and between symptoms. You may assume that the total sum of squares $(SS_T) = 28590.92$

(b) Criticise the experiment and suggest improvements.

(AEB)