Chapter 13

Analysis of Variance and Experimental Design

Learning objectives

After reading this chapter and doing the exercises, you should be able to:

1. Understand how the analysis of variance procedure can be used to determine if the means of more than two populations are equal.
2. Know the assumptions necessary to use the analysis of variance procedure.
3. Understand the use of the *F* distribution in performing the analysis of variance procedure.
4. Know how to set up an ANOVA table and interpret the entries in the table.
5. Use output from computer software packages to solve analysis of variance problems.
6. Know how to use Fisher's least significant difference (LSD) procedure and Fisher's LSD with the Bonferroni adjustment to conduct statistical comparisons between pairs of population means.
7. Understand the difference between a completely randomized design, a randomized block design and factorial experiments.
8. Know the definition of the following terms:
   comparisonwise Type I error rate
   experimentwise Type I error rate
   factor
   level
   treatment
   partitioning
   blocking
   main effect
   interaction
   replication

In this chapter we introduce a statistical procedure called *analysis of variance* (ANOVA).

First, we show how ANOVA can be used to test for the equality of three or more population means using data obtained from an observational study. Then, we discuss the use of ANOVA for analyzing data obtained from three types of experimental studies: a completely randomized design, a randomized block design and a factorial experiment. In the following chapters we will see that ANOVA plays a key role in analyzing the results of regression analysis involving both experimental and observational data.

13.1 An introduction to analysis of variance

National Computer Products (NCP) manufactures printers and fax machines at plants located in Ayr, Dusseldorf and Stockholm. To measure how much employees at these plants know about total quality management, a random sample of six employees was selected from each plant and given a quality awareness examination. The examination scores obtained for these 18 employees are listed in Table 13.1. The sample means, sample variances and sample standard deviations for each group are also provided. Managers want to use these data to test the hypothesis that the mean examination score is the same for all three plants.

We will define population 1 as all employees at the Ayr plant, population 2 as all employees at the Dusseldorf plant, and population 3 as all employees at the Stockholm plant. Let

μ_1 = mean examination score for population 1
μ_2 = mean examination score for population 2
μ_3 = mean examination score for population 3

Statistics in Practice

Product customization and manufacturing trade-offs

The analysis of variance technique was used recently in a study to investigate trade-offs between product customization and other manufacturing priorities. A total of 102 UK manufacturers from eight industrial sectors were involved in the research. Three levels of customization were considered: full customization where customer input was incorporated at the product design or fabrication stages; partial customization with customer input incorporated into product assembly or delivery stages and standard products which did not incorporate any customer input at all.

Interior of a car manufacturing plant. © George Clerk.

The impact of customization was considered against four competitive imperatives – cost, quality, delivery and volume flexibility.

It was found that customization had a significant effect on delivery (both in terms of speed and lead times); also on manufacturer's costs (though not design, component, delivery and servicing costs).

The findings suggest that customization is not cost-free and that the advent of mass customization is unlikely to see the end of trade-offs with other key priorities.

Source: Squire, B., Brown, S., Readman, J. and Bessant J. (2005) The impact of mass customization on manufacturing trade-offs. *Production and Operations Management* Journal 15(1) 10–21

Although we will never know the actual values of μ_1, μ_2 and μ_3, we want to use the sample results to test the following hypotheses.

H_0: $\mu_1 = \mu_2 = \mu_3$
H_1: Not all population means are equal

As we will demonstrate shortly, analysis of variance is a statistical procedure that can be used to determine whether the observed differences in the three sample means are large enough to reject H_0.

Table 13.1 Examination scores for 18 employees

Observation	Plant 1 Ayr	Plant 2 Dusseldorf	Plant 3 Stockholm
1	85	71	59
2	75	75	64
3	82	73	62
4	76	74	69
5	71	69	75
6	85	82	67
Sample mean	79	74	66
Sample variance	34	20	32
Sample standard deviation	5.83	4.47	5.66

In the introduction to this chapter we stated that analysis of variance can be used to analyse data obtained from both an observational study and an experimental study.

To provide a common set of terminology for discussing the use of analysis of variance in both types of studies, we introduce the concepts of a response variable, a factor and a treatment.

The two variables in the NCP example are plant location and score on the quality awareness examination. Because the objective is to determine whether the mean examination score is the same for plants located in Ayr, Dusseldorf and Stockholm, examination score is referred to as the dependent or *response variable* and plant location as the independent variable or *factor*. In general, the values of a factor selected for investigation are referred to as levels of the factor or *treatments*. Thus, in the NCP example the three treatments are Ayr, Dusseldorf and Stockholm. These three treatments define the populations of interest in the NCP example. For each treatment or population, the response variable is the examination score.

Assumptions for analysis of variance

Three assumptions are required to use analysis of variance.

1. **For each population, the response variable is normally distributed**. Implication: In the NCP example, the examination scores (response variable) must be normally distributed at each plant.
2. **The variance of the response variable, denoted σ^2, is the same for all of the populations**. Implication: In the NCP example, the variance of examination scores must be the same for all three plants.
3. **The observations must be independent**. Implication: In the NCP example, the examination score for each employee must be independent of the examination score for any other employee.

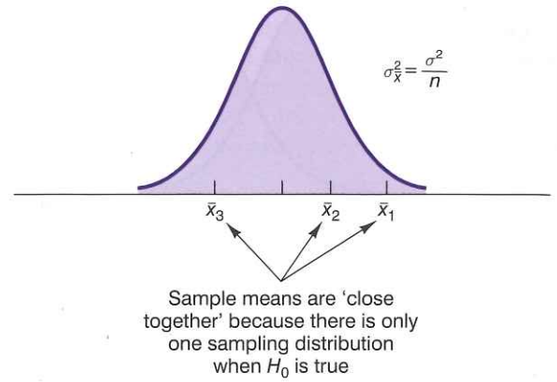
A conceptual overview

If the means for the three populations are equal, we would expect the three sample means to be close together. In fact, the closer the three sample means are to one another, the more evidence we have for the conclusion that the population means are equal. Alternatively, the more the sample means differ, the more evidence we have for the conclusion that the population means are not equal. In other words, if the variability among the sample means is 'small', it supports H_0; if the variability among the sample means is 'large', it supports H_1.

If the null hypothesis, H_0: $\mu_1 = \mu_2 = \mu_3$, is true, we can use the variability among the sample means to develop an estimate of σ^2. First, note that if the assumptions for analysis of variance are satisfied, each sample will have come from the same normal distribution with mean μ and variance σ^2. Recall from Chapter 7 that the sampling distribution of the sample mean for a simple random sample of size n from a normal population will be normally distributed with mean μ and variance σ^2/n. Figure 13.1 illustrates such a sampling distribution.

Therefore, if the null hypothesis is true, we can think of each of the three sample means, $\bar{x}_1 = 79$, $\bar{x}_2 = 74$, and $\bar{x}_3 = 66$, from Table 13.1 as values drawn at random from the sampling distribution shown in Figure 13.1. In this case, the mean and variance of the three values can be used to estimate the mean and variance of the sampling distribution. When the sample sizes are equal, as in the NCP example, the best estimate of the mean of the sampling distribution of $\bar{X}$ is the mean or average of the sample means. Thus, in the NCP example, an estimate of the mean of the sampling

Figure 13.1 Sampling distribution of $\bar{X}$ given H_0 is true



distribution of $\bar{X}$ is $(79 + 74 + 66)/3 = 73$. We refer to this estimate as the *overall sample mean*. An estimate of the variance of the sampling distribution of $\bar{X}$, $\sigma^2_{\bar{x}}$ is provided by the variance of the three sample means.

$$s^2_{\bar{x}} = \frac{(79 - 73)^2 + (74 - 73)^2 + (66 - 73)^2}{3 - 1} = \frac{86}{2} = 43$$

Because $\sigma^2_{\bar{x}} = \sigma^2 / n$, solving for σ^2 gives

$$\sigma^2 = n\sigma^2_{\bar{x}}$$

Hence,

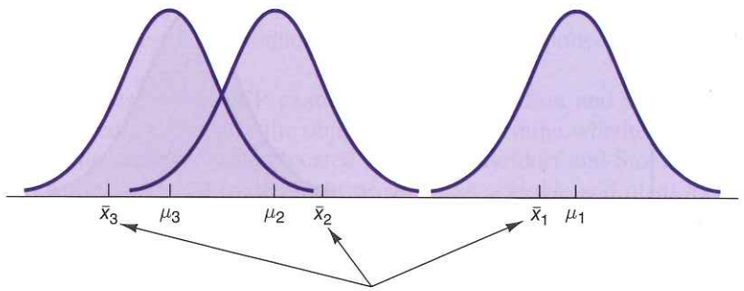
$$\text{Estimate of } \sigma^2 = n \text{ (Estimate of } \sigma^2_{\bar{x}}) = ns^2_{\bar{x}} = 6(43) = 258$$

The result, $ns^2_{\bar{x}} = 258$, is referred to as the *between-treatments* estimate of σ^2.

The between-treatments estimate of σ^2 is based on the assumption that the null hypothesis is true. In this case, each sample comes from the same population, and there is only one sampling distribution of $\bar{X}$. To illustrate what happens when H_0 is false, suppose the population means all differ. Note that because the three samples are from normal populations with different means, they will result in three different sampling distributions. Figure 13.2 shows that in this case, the sample means are not as close together as they were when H_0 was true. Thus, $s^2_{\bar{x}}$ will be larger, causing the between-treatments estimate of σ^2 to be larger. In general, when the population means are not equal, the between-treatments estimate will overestimate the population variance σ^2.

The variation within each of the samples also has an effect on the conclusion we reach in analysis of variance. When a simple random sample is selected from each population, each of the sample variances provides an unbiased estimate of σ^2. Hence, we can combine or pool the individual estimates of σ^2 into one overall estimate. The estimate of σ^2 obtained in this way is called the *pooled* or *within-treatments* estimate of σ^2. Because each sample variance provides an estimate of σ^2 based only on the variation within each sample, the within-treatments estimate of σ^2 is not affected by whether the population means are equal.

Figure 13.2 Sampling distributions of $\bar{x}$ given H_0 is false



When the sample sizes are equal, the within-treatments estimate of σ^2 can be obtained by computing the average of the individual sample variances. For the NCP example we obtain

$$\text{Within-treatments estimate of } \sigma^2 = \frac{34 + 20 + 32}{3} = \frac{86}{3} = 28.67$$

In the NCP example, the between-treatments estimate of σ^2 (258) is much larger than the within-treatments estimate of σ^2 (28.67). In fact, the ratio of these two estimates is 258/28.67 = 9.00. Recall, however, that the between-treatments approach provides a good estimate of σ^2 only if the null hypothesis is true; if the null hypothesis is false, the between treatments approach overestimates σ^2. The within-treatments approach provides a good estimate of σ^2 in either case. Thus, if the null hypothesis is true, the two estimates will be similar and their ratio will be close to 1. If the null hypothesis is false, the between-treatments estimate will be larger than the within-treatments estimate, and their ratio will be large. In the next section we will show how large this ratio must be to reject H_0.

In summary, the logic behind ANOVA is based on the development of two independent estimates of the common population variance σ^2. One estimate of σ^2 is based on the variability among the sample means themselves, and the other estimate of σ^2 is based on the variability of the data within each sample. By comparing these two estimates of σ^2, we will be able to determine whether the population means are equal.

13.2 Analysis of variance: testing for the equality of *k* population means

Analysis of variance can be used to test for the equality of k population means. The general form of the hypotheses tested is

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$
H_1: Not all population means are equal

where

μ_j = mean of the jth population

We assume that a simple random sample of size n_j has been selected from each of the k populations or treatments. For the resulting sample data, let

x_{ij} = value of observation i for treatment j
n_j = number of observations for treatment j
$\bar{x}_j$ = sample mean for treatment j
s_j^2 = sample variance for treatment j
s_j = sample standard deviation for treatment j

The formulae for the sample mean and sample variance for treatment j are as follows.

Testing for the Equality of *k* Population means sample mean for Treatment *j*

$$\bar{x}_j = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j} \tag{13.1}$$

Sample Variance for Treatment *j*

$$s_j^2 = \frac{\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{n_j - 1} \tag{13.2}$$

The overall sample mean, denoted $\bar{\bar{x}}$, is the sum of all the observations divided by the total number of observations. That is,

Overall Sample Mean

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n_j} x_{ij}}{n_T} \tag{13.3}$$

where

$$n_T = n_1 + n_2 + \ldots + n_k \tag{13.4}$$

If the size of each sample is n, $n_T = kn$; in this case equation (13.3) reduces to

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n_j} x_{ij}}{kn} = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n_j} x_i / n}{k} = \frac{\sum_{j=1}^{k} \bar{x}_j}{k} \tag{13.5}$$

In other words, whenever the sample sizes are the same, the overall sample mean is just the average of the k sample means.

Because each sample in the NCP example consists of $n = 6$ observations, the overall sample mean can be computed by using equation (13.5). For the data in Table 13.1 we obtained the following result.

$$\bar{\bar{x}} = \frac{79 + 74 + 66}{3} = 73$$

If the null hypothesis is true ($\mu_1 = \mu_2 = \mu_3 = \mu$), the overall sample mean of 73 is the best estimate of the population mean μ.

Between-treatments estimate of population variance

In the preceding section, we introduced the concept of a between-treatments estimate of σ^2 and showed how to compute it when the sample sizes were equal. This estimate of σ^2 is called the *mean square due to treatments* and is denoted MSTR. The general formula for computing MSTR is

$$\text{MSTR} = \frac{\sum_{j=1}^{k} n_j (\bar{x}_j - \bar{\bar{x}})^2}{k - 1} \qquad (13.6)$$

The numerator in equation (13.6) is called the *sum of squares due to treatments* and is denoted SSTR. The denominator, $k - 1$, represents the degrees of freedom associated with SSTR. Hence, the mean square due to treatments can be computed by the following formula.

Mean square due to treatments

$$\text{MSTR} = \frac{\text{SSTR}}{k - 1} \qquad (13.7)$$

where

$$\text{SSTR} = \sum_{j=1}^{k} n_j (\bar{x}_j - \bar{\bar{x}})^2 \qquad (13.8)$$

If H_0 is true, MSTR provides an unbiased estimate of σ^2. However, if the means of the k populations are not equal, MSTR is not an unbiased estimate of σ^2; in fact, in that case, MSTR should overestimate σ^2.

For the NCP data in Table 13.1, we obtain the following results.

$$\text{SSTR} = \sum_{j=1}^{k} n_j (\bar{x}_j - \bar{\bar{x}})^2 = 6(79 - 73)^2 + 6(74 - 73)^2 + 6(66 - 73)^2 = 516$$

$$\text{MSTR} = \frac{\text{SSTR}}{k - 1} = \frac{516}{2} = 258$$

Within-treatments estimate of population variance

Earlier, we introduced the concept of a within-treatments estimate of σ^2 and showed how to compute it when the sample sizes were equal. This estimate of σ^2 is called the *mean square due to error* and is denoted MSE. The general formula for computing MSE is

$$\text{MSE} = \frac{\sum_{j=1}^{k} (n_j - 1) s_j^2}{n_T - k} \qquad (13.9)$$

The numerator in equation (13.9) is called the *sum of squares due to error* and is denoted SSE. The denominator of MSE is referred to as the degrees of freedom associated with SSE. Hence, the formula for MSE can also be stated as follows.

Mean square due to error

$$\text{MSE} = \frac{\text{SSE}}{n_T - k} \qquad (13.10)$$

where

$$\text{SSE} = \sum_{j=1}^{k} (n_j - 1) s_j^2 \qquad (13.11)$$

Note that MSE is based on the variation within each of the treatments; it is not influenced by whether the null hypothesis is true. Thus, MSE always provides an unbiased estimate of σ^2.

For the NCP data in Table 13.1 we obtain the following results.

$$\text{SSE} = \sum_{j=1}^{k} (n_j - 1) s_j^2 = (6 - 1)34 + (6 - 1)20 + (6 - 1)32 = 430$$

$$\text{MSE} = \frac{\text{SSE}}{n_T - k} = \frac{430}{18 - 3} = \frac{430}{15} = 28.67$$

Comparing the variance estimates: the *F* test

If the null hypothesis is true, MSTR and MSE provide two independent, unbiased estimates of σ^2. Based on the material covered in Chapter 11 we know that for normal populations, the sampling distribution of the ratio of two independent estimates of σ^2 follows an F distribution. Hence, if the null hypothesis is true and the ANOVA assumptions are valid, the sampling distribution of MSTR/MSE is an F distribution with numerator degrees of freedom equal to $k - 1$ and denominator degrees of freedom equal to $n_T - k$. In other words, if the null hypothesis is true, the value of MSTR/MSE should appear to have been selected from this F distribution.

However, if the null hypothesis is false the value of MSTR/MSE will be inflated because MSTR overestimates σ^2. Hence, we will reject H_0 if the resulting value of MSTR/MSE appears to be too large to have been selected from an F distribution with $k - 1$ numerator degrees of freedom and $n_T - k$ denominator degrees of freedom.

Because the decision to reject H_0 is based on the value of MSTR/MSE, the test statistic used to test for the equality of k population means is as follows.

Test statistic for the equality of *k* population means

$$F = \frac{\text{MSTR}}{\text{MSE}} \qquad (13.12)$$

The test statistic follows an F distribution with $k - 1$ degrees of freedom in the numerator and $n_T - k$ degrees of freedom in the denominator.

Returning to the National Computer Products example we use a level of significance $\alpha = 0.05$ to conduct the hypothesis test. The value of the test statistic is

$$F = \frac{MSTR}{MSE} = \frac{258}{28.67} = 9$$

The numerator degrees of freedom is $k - 1 = 3 - 1 = 2$ and the denominator degrees of freedom is $n_T - k = 18 - 3 = 15$. Because we will only reject the null hypothesis for large values of the test statistic, the p-value is the upper tail area of the F distribution to the right of the test statistic $F = 9$. Figure 13.3 shows the sampling distribution of $F = MSTR/MSE$, the value of the test statistic, and the upper tail area that is the p-value for the hypothesis test.

From Table 4 of Appendix B we find the following areas in the upper tail of an F distribution with two numerator degrees of freedom and 15 denominator degrees of freedom.

Area in upper tail	0.10	0.05	0.025	0.01
F value ($df_1 = 2$, $df_2 = 15$)	2.70	3.68	4.77	6.36

$F = 9$

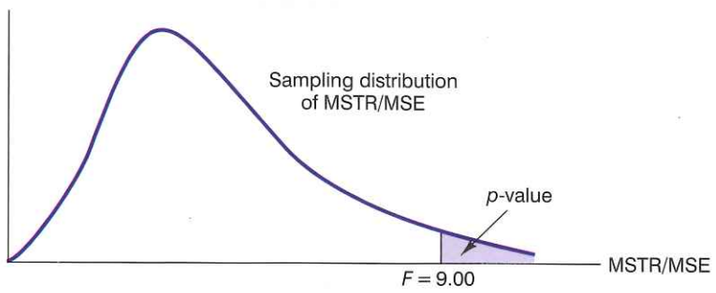
Because $F = 9$ is greater than 6.36, the area in the upper tail at $F = 9$ is less than 0.01. Thus, the p-value is less than 0.01. With a p-value $\leq \alpha = 0.05$, H_0 is rejected. The test provides sufficient evidence to conclude that the means of the three populations are not equal. In other words, analysis of variance supports the conclusion that the population mean examination scores at the three NCP plants are not equal.

Because the F table only provides values for upper tail areas of 0.10, 0.05, 0.025 and 0.01, we cannot determine the exact p-value directly from the table. MINITAB, EXCEL or PASW provide the p-value as part of the standard ANOVA output. The software section at the end of the chapter shows the procedures that can be used. For the NCP example, the exact p-value corresponding to the test statistic $F = 9$ is 0.003.

As with other hypothesis testing procedures, the critical value approach may also be used. With $\alpha = 0.05$, the critical F value occurs with an area of 0.05 in the upper tail of an F distribution with 2 and 15 degrees of freedom. From the F distribution table, we find $F_{0.05} = 3.68$. Hence, the appropriate upper tail rejection rule for the NCP example is

Reject H_0 if $F \geq 3.68$

Figure 13.3 Computation of p-value using the sampling distribution of MSTR/MSE



With $F = 9$, we reject H_0 and conclude that the means of the three populations are not equal. A summary of the overall procedure for testing for the equality of k population means follows.

Test for the equality of k population means

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$

H_1: Not all population means are equal

Test statistic

$$F = \frac{MSTR}{MSE}$$

Rejection rule

p-value approach: Reject H_0 if p-value $\leq \alpha$

Critical value approach: Reject H_0 if $F \geq F_\alpha$

where the value of F_α is based on an F distribution with $k - 1$ numerator degrees of freedom and $n_T - k$ denominator degrees of freedom

ANOVA table

The results of the preceding calculations can be displayed conveniently in a table referred to as the analysis of variance or **ANOVA table**. Table 13.2 is the analysis of variance table for the National Computer Products example. The sum of squares associated with the source of variation referred to as 'total' is called the total sum of squares (SST). Note that the results for the NCP example suggest that SST = SSTR + SSE, and that the degrees of freedom associated with this total sum of squares is the sum of the degrees of freedom associated with the between-treatments estimate of σ^2 and the within-treatments estimate of σ^2.

We point out that SST divided by its degrees of freedom $n_T - 1$ is nothing more than the overall sample variance that would be obtained if we treated the entire set of 18 observations as one data set. With the entire data set as one sample, the formula for computing the total sum of squares, SST, is

Total sum of squares

$$SST = \sum_{j=1}^{k} \sum_{i=1}^{n} (x_{ij} - \bar{\bar{x}})^2 \qquad (13.13)$$

It can be shown that the results we observed for the analysis of variance table for the NCP example also apply to other problems. That is,

Partitioning of sum of squares

$$SST = SSTR + SSE \qquad (13.14)$$

Table 13.2 Analysis of variance table for the NCP example

Source of variation	Degrees of freedom	Sum of squares	Mean square	F
Treatments	2	516	258.00	9.00
Error	15	430	28.67	
Total	17	946		

In other words, SST can be partitioned into two sums of squares: the sum of squares due to treatments and the sum of squares due to error. Note also that the degrees of freedom corresponding to SST, $n_T - 1$, can be partitioned into the degrees of freedom corresponding to SSTR, $k - 1$, and the degrees of freedom corresponding to SSE, $n_T - k$. The analysis of variance can be viewed as the process of **partitioning** the total sum of squares and the degrees of freedom into their corresponding sources: treatments and error. Dividing the sum of squares by the appropriate degrees of freedom provides the variance estimates and the F value used to test the hypothesis of equal population means.

Computer results for analysis of variance

Because of the widespread availability of statistical computer packages, analysis of variance computations with large sample sizes or a large number of populations can be performed easily. In Figure 13.4 we show output for the NCP example obtained from the MINITAB computer package. The first part of the computer output contains the familiar ANOVA table format. Comparing Figure 13.4 with Table 13.2, we see that the same information is available, although some of the headings are slightly different. The heading Source is used for the source of variation column, and Factor identifies the treatments row. A p-value is provided for the F test. Thus, at the $\alpha = 0.05$ level of significance, we reject H_0 because the p value $= 0.003 < \alpha = 0.05$.

Figure 13.4 MINITAB output for the NCP analysis of variance

One-way ANOVA: Ayr, Dusseldorf, Stockholm

```
Source  DF     SS     MS     F      P
Factor   2  516.0  258.0  9.00  0.003
Error   15  430.0   28.7
Total   17  946.0

S = 5.354   R-Sq = 54.55%   R-Sq(adj) = 48.48%


                              Individual 95% CIs For Mean Based on
                              Pooled StDev
Level       N    Mean  StDev  --------+---------+---------+---------+-
Ayr         6  79.000  5.831                    (-------*------)
Dusseldorf  6  74.000  4.472           (------*-------)
Stockholm   6  66.000  5.657  (-------*-------)
                              --------+---------+---------+---------+-
                                   66.0      72.0      78.0      84.0

Pooled StDev = 5.354
```


Note that following the ANOVA table the computer output contains the respective sample sizes, the sample means, and the standard deviations. In addition, MINITAB provides a figure that shows individual 95 per cent confidence interval estimates of each population mean.

In developing these confidence interval estimates, MINITAB uses MSE as the estimate of σ^2. Thus, the square root of MSE provides the best estimate of the population standard deviation σ. This estimate of σ on the computer output is Pooled StDev; it is equal to 5.354. To provide an illustration of how these interval estimates are developed, we will compute a 95 per cent confidence interval estimate of the population mean for the Ayr plant.

From our study of interval estimation in Chapter 8, we know that the general form of an interval estimate of a population mean is

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \qquad \textbf{(13.15)}$$

where s is the estimate of the population standard deviation σ. In the analysis of variance the best estimate of σ is provided by the square root of MSE or the Pooled StDev, therefore we use a value of 5.354 for s in expression (13.15). The degrees of freedom for the t value is 15, the degrees of freedom associated with the within-treatments estimate of σ^2. Hence, with $t_{0.025} = 2.131$ we obtain

$$79 \pm 2.131 \frac{5.354}{\sqrt{6}} = 79 \pm 4.66$$

Therefore, the individual 95 per cent confidence interval for the Ayr plant goes from $79 - 4.66 = 74.34$ to $79 + 4.66 = 83.66$. Because the sample sizes are equal for the NCP example, the individual confidence intervals for the Dusseldorf and Stockholm plants are also constructed by adding and subtracting 4.66 from each sample mean. Thus, in the figure provided by MINITAB we see that the widths of the confidence intervals are the same.

Exercises

Methods

1 Five observations were selected from each of three populations. The data obtained follow.

Observation	Sample 1	Sample 2	Sample 3
1	32	44	33
2	30	43	36
3	30	44	35
4	26	46	36
5	32	48	40
Sample mean	30	45	36
Sample variance	6.00	4.00	6.50

a. Compute the between-treatments estimate of σ^2.
b. Compute the within-treatments estimate of σ^2.

c. At the $\alpha = 0.05$ level of significance, can we reject the null hypothesis that the means of the three populations are equal?
d. Set up the ANOVA table for this problem.

2 Four observations were selected from each of three populations. The data obtained follow.

Observation	Sample 1	Sample 2	Sample 3
1	165	174	169
2	149	164	154
3	156	180	161
4	142	158	148
Sample mean	153	169	158
Sample variance	96.67	97.33	82.00

a. Compute the between-treatments estimate of σ^2.
b. Compute the within-treatments estimate of σ^2.
c. At the $\alpha = 0.05$ level of significance, can we reject the null hypothesis that the three population means are equal? Explain.
d. Set up the ANOVA table for this problem.

3 Samples were selected from three populations. The data obtained follow.

	Sample 1	Sample 2	Sample 3
	93	77	88
	98	87	75
	107	84	73
	102	95	84
	85	75	82
$\bar{x}_j$	100	85	79
s_j^2	35.33	35.60	43.50

a. Compute the between-treatments estimate of σ^2.
b. Compute the within-treatments estimate of σ^2.
c. At the $\alpha = 0.05$ level of significance, can we reject the null hypothesis that the three population means are equal? Explain.
d. Set up the ANOVA table for this problem.

4 A random sample of 16 observations was selected from each of four populations. A portion of the ANOVA table follows.

Source of variation	Degrees of freedom	Sum of squares	Mean square	F
Treatments			400	
Error				
Total		1500		

a. Provide the missing entries for the ANOVA table.
b. At the $\alpha = 0.05$ level of significance, can we reject the null hypothesis that the means of the four populations are equal?

5 Random samples of 25 observations were selected from each of three populations. For these data, SSTR = 120 and SSE = 216.

a. Set up the ANOVA table for this problem.
b. At the $\alpha = 0.05$ level of significance, can we reject the null hypothesis that the three population means are equal?

Applications

6 To test whether the mean time needed to mix a batch of material is the same for machines produced by three manufacturers, the Jacobs Chemical Company obtained the following data on the time (in minutes) needed to mix the material. Use these data to test whether the population mean times for mixing a batch of material differ for the three manufacturers. Use $\alpha = 0.05$.

Manufacturer		
1	2	3
20	28	20
26	26	19
24	31	23
22	27	22

7 Managers at all levels of an organization need adequate information to perform their respective tasks. One study investigated the effect the source has on the dissemination of information. In this particular study the sources of information were a superior, a peer and a subordinate. In each case, a measure of dissemination was obtained, with higher values indicating greater dissemination of information. Use $\alpha = 0.05$ and the following data to test whether the source of information significantly affects dissemination. What is your conclusion, and what does it suggest about the use and dissemination of information?

Superior	Peer	Subordinate
8	6	6
5	6	5
4	7	7
6	5	4
6	3	3
7	4	5
5	7	7
5	6	5

8 A study investigated the perception of corporate ethical values among individuals specializing in marketing. Use $\alpha = 0.05$ and the following data (higher scores indicate higher ethical values) to test for significant differences in perception among the three groups.

Marketing managers	Marketing research	Advertising
6	5	6
5	5	7
4	4	6
5	4	5
6	5	6
4	4	6

STRESS

9 A study reported in the *Journal of Small Business Management* concluded that self-employed individuals experience higher job stress than individuals who are not self-employed. In this study job stress was assessed with a 15-item scale designed to measure various aspects of ambiguity and role conflict. Ratings for each of the 15 items were made using a scale with 1–5 response options ranging from strong agreement to strong disagreement. The sum of the ratings for the 15 items for each individual surveyed is between 15 and 75, with higher values indicating a higher degree of job stress. Suppose that a similar approach, using a 20-item scale with 1–5 response options, was used to measure the job stress of individuals for 15 randomly selected property agents, 15 architects and 15 stockbrokers. The results obtained follow.

Property agent	Architect	Stockbroker
81	43	65
48	63	48
68	60	57
69	52	91
54	54	70
62	77	67
76	68	83
56	57	75
61	61	53
65	80	71
64	50	54
69	37	72
83	73	65
85	84	58
75	58	58

Use $\alpha = 0.05$ to test for any significant difference in job stress among the three professions.

SHIPS

10 *Condé Nast Traveler* conducts an annual survey in which readers rate their favourite cruise ships. Ratings are provided for small ships (carrying up to 500 passengers), medium ships (carrying 500 to 1500 passengers) and large ships (carrying a minimum of 1500 passengers). The following data show the service ratings for eight randomly selected small ships, eight randomly selected medium ships and eight randomly selected large ships. All ships are rated on a 100-point scale, with higher values indicating better service (*Condé Nast Traveler*, February 2003).

Small ships		Medium ships		Large ships	
Name	Rating	Name	Rating	Name	Rating
Hanseactic	90.5	Amsterdam	91.1	Century	89.2
Mississippi Queen	78.2	Crystal Symphony	98.9	Disney Wonder	90.2
Philae	92.3	Maasdam	94.2	Enchantment of the Seas	85.9
Royal Clipper	95.7	Noordam	84.3	Grand Princess	84.2
Seabourn Pride	94.1	Royal Princess	84.8	Infinity	90.2
Seabourn Spirit	100.0	Ryndam	89.2	Legend of the Seas	80.6
Silver Cloud	91.8	Statendam	86.4	Paradise	75.8
Silver Wind	95.0	Veendam	88.3	Sun Princess	82.3

Use $\alpha = 0.05$ to test for any significant difference in the mean service ratings among the three sizes of cruise ships.

13.3 Multiple comparison procedures

When we use analysis of variance to test whether the means of k populations are equal, rejection of the null hypothesis allows us to conclude only that the population means are *not all equal*. In some cases we will want to go a step further and determine where the differences among means occur. The purpose of this section is to introduce two **multiple comparison procedures** that can be used to conduct statistical comparisons between pairs of population means.

Fisher's LSD

Suppose that analysis of variance provides statistical evidence to reject the null hypothesis of equal population means. In this case, Fisher's least significant difference (LSD) procedure can be used to determine where the differences occur. To illustrate the use of Fisher's LSD procedure in making pairwise comparisons of population means, recall the NCP example introduced in Section 13.1. Using analysis of variance, we concluded that the population mean examination scores are not the same at the three plants. In this case, the follow-up question is: We believe the plants differ, but where do the differences occur? That is, do the means of populations 1 and 2 differ? Or those of populations 1 and 3? Or those of populations 2 and 3?

In Chapter 10 we presented a statistical procedure for testing the hypothesis that the means of two populations are equal. With a slight modification in how we estimate the population variance, Fisher's LSD procedure is based on the t test statistic presented for the two-population case. The following table summarizes Fisher's LSD procedure.

Fisher's LSD Procedure

$$H_0: \mu_i = \mu_j$$
$$H_1: \mu_i \neq \mu_j$$

Test statistic for Fisher's LSD procedure

$$t = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} \qquad \textbf{(13.16)}$$

Rejection rule

p-value approach: Reject H_0 if p-value $\leq \alpha$
Critical value approach: Reject H_0 if $t \leq -t_{\alpha/2}$ or $t \geq t_{\alpha/2}$
where the value of $t_{\alpha/2}$ is based on a t distribution with $n_T - k$ degrees of freedom.

Let us now apply this procedure to determine whether there is a significant difference between the means of population 1 (Ayr) and population 2 (Dusseldorf) at the $\alpha = 0.05$ level of significance. Table 13.1 shows that the sample mean is 79 for the Ayr plant and 74 for the Dusseldorf plant. Table 13.2 shows that the value of MSE is 28.67; it is the estimate of σ^2 and is based on 15 degrees of freedom. For the NCP data the value of the test statistic is

$$t = \frac{79 - 74}{\sqrt{28.67\left(\frac{1}{6} + \frac{1}{6}\right)}} = 1.62$$

The t distribution table (Table 2 in Appendix B) shows that with 15 degrees of freedom $t = 1.341$ for an area of 0.10 in the upper tail and $t = 1.753$ for an area of 0.05 in the upper tail. Because the test statistic $t = 1.62$ is between 1.341 and 1.753, we know that the area in the upper tail must be between 0.05 and 0.10. Because this test is a two-tailed test, we double these values to conclude that the p-value is between 0.10 and 0.20. MINITAB or EXCEL can be used to show that the p-value corresponding to $t = 1.62$ is 0.1261. Because the p-value is greater than $\alpha = 0.05$, we cannot reject the null hypothesis. Hence, we cannot conclude that the population mean score at the Ayr plant is different from the population mean score at the Dusseldorf plant.

Many practitioners find it easier to determine how large the difference between the sample means must be to reject H_0. In this case the test statistic is $\bar{x}_i - \bar{x}_j$ and the test is conducted by the following procedure.

Fisher's LSD procedure based on the test statistic $\bar{x}_i - \bar{x}_j$

$$H_0: \mu_i - \mu_j$$
$$H_1: \mu_i \neq \mu_j$$

Test statistic

$$\bar{x}_i - \bar{x}_j$$

Rejection rule at a level of significance α

Reject H_0 if $|\bar{x}_i - \bar{x}_j| > \text{LSD}$

where

$$\text{LSD} = t_{\alpha/2}\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \qquad \textbf{(13.17)}$$

For the NCP example the value of LSD is

$$\text{LSD} = 2.131\sqrt{28.67\left(\frac{1}{6} + \frac{1}{6}\right)} = 6.59$$

Note that when the sample sizes are equal, only one value for LSD is computed. In such cases we can simply compare the magnitude of the difference between any two sample means with the value of LSD. For example, the difference between the sample means for population 1 (Ayr) and population 3 (Stockholm) is $79 - 66 = 13$. This difference is greater than 6.59, which means we can reject the null hypothesis that the population mean examination score for the Ayr plant is equal to the population mean score for the Stockholm plant.

Similarly, with the difference between the sample means for populations 2 and 3 of $74 - 66 = 8 > 6.59$, we can also reject the hypothesis that the population mean examination score for the Dusseldorf plant is equal to the population mean examination score for the Stockholm plant. In effect, our conclusion is that the Ayr and Dusseldorf plants both differ from the Stockholm plant.

Fisher's LSD can also be used to develop a confidence interval estimate of the difference between the means of two populations. The general procedure follows.

Confidence interval estimate of the difference between two Population means using Fisher's LSD procedure

$$\bar{x}_i - \bar{x}_j \pm \text{LSD} \qquad \textbf{(13.18)}$$

where

$$\text{LSD} = t_{\alpha/2}\sqrt{\text{MSE}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \qquad \textbf{(13.19)}$$

and $t_{\alpha/2}$ is based on a t distribution with $n_T - k$ degrees of freedom. If the confidence interval in expression (13.18) includes the value zero, we cannot reject the hypothesis that the two population means are equal. However, if the confidence interval does not include the value zero, we conclude that there is a difference between the population means. For the NCP example, recall that LSD = 6.59 (corresponding to $t_{0.025} = 2.131$). Thus, a 95 per cent confidence interval estimate of the difference between the means of populations 1 and 2 is $79 - 74 \pm 6.59 = 5 \pm 6.59 = -1.59$ to 11.59; because this interval includes zero, we cannot reject the hypothesis that the two population means are equal.

Type I error rates

We began the discussion of Fisher's LSD procedure with the premise that analysis of variance gave us statistical evidence to reject the null hypothesis of equal population means.

We showed how Fisher's LSD procedure can be used in such cases to determine where the differences occur. Technically, it is referred to as a *protected* or *restricted* LSD test because it is employed only if we first find a significant F value by using analysis of variance.

To see why this distinction is important in multiple comparison tests, we need to explain the difference between a *comparisonwise* Type I error rate and an *experimentwise* Type I error rate.

In the NCP example we used Fisher's LSD procedure to make three pairwise comparisons.

Test 1	Test 2	Test 3
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$	$H_0: \mu_1 = \mu_3$ $H_1: \mu_1 \neq \mu_3$	$H_0: \mu_2 = \mu_3$ $H_1: \mu_2 \neq \mu_3$

In each case, we used a level of significance of $\alpha = 0.05$. Therefore, for each test, if the null hypothesis is true, the probability that we will make a Type I error is $\alpha = 0.05$; hence, the probability that we will not make a Type I error on each test is $1 - 0.05 = 0.95$. In discussing multiple comparison procedures we refer to this probability of a Type I error ($\alpha = 0.05$) as the **comparisonwise Type I error rate**; comparisonwise Type I error rates indicate the level of significance associated with a single pairwise comparison.

Let us now consider a slightly different question. What is the probability that in making three pairwise comparisons, we will commit a Type I error on at least one of the three tests? To answer this question, note that the probability that we will not make a Type I error on any of the three tests is $(0.95)(0.95)(0.95) = 0.8574$.* Therefore, the probability of making at least one Type I error is $1 - 0.8574 = 0.1426$. Thus, when we use Fisher's LSD procedure to make all three pairwise comparisons, the Type I error rate associated with this approach is not 0.05, but actually 0.1426; we refer to this error rate as the *overall* or **experimentwise Type I error rate**. To avoid confusion, we denote the experimentwise Type I error rate as α_{EW}.

The experimentwise Type I error rate gets larger for problems with more populations. For example, a problem with five populations has ten possible pairwise comparisons. If we tested all possible pairwise comparisons by using Fisher's LSD with a comparisonwise error rate of $\alpha = 0.05$, the experimentwise Type I error rate would be $1 - (1 - 0.05)^{10} = 0.40$. In such cases, practitioners look to alternatives that provide better control over the experimentwise error rate.

One alternative for controlling the overall experimentwise error rate, referred to as the *Bonferroni adjustment*, involves using a smaller comparisonwise error rate for each test. For example, if we want to test C pairwise comparisons and want the maximum probability of making a Type I error for the overall experiment to be α_{EW}, we simply use a comparisonwise error rate equal to α_{EW}/C. In the NCP example, if we want to use Fisher's LSD procedure to test all three pairwise comparisons with a maximum experimentwise error rate of $\alpha_{EW} = 0.05$, we set the comparisonwise error rate to be $\alpha = 0.05/3 = 0.017$. For a problem with five populations and ten possible pairwise comparisons, the Bonferroni adjustment would suggest a comparisonwise error rate of $0.05/10 = 0.005$. Recall from our discussion of hypothesis testing in Chapter 9 that for a fixed sample size, any decrease in the probability of making a Type I error will result in an increase in the probability of making a Type II error, which corresponds to accepting the hypothesis that the two population means are

*The assumption is that the three tests are independent, and hence the joint probability of the three events can be obtained by simply multiplying the individual probabilities. In fact, the three tests are not independent because MSE is used in each test; therefore, the error involved is even greater than that shown.

equal when in fact they are not equal. As a result, many practitioners are reluctant to perform individual tests with a low comparisonwise Type I error rate because of the increased risk of making a Type II error.

Several other procedures, such as Tukey's procedure and Duncan's multiple range test, have been developed to help in such situations. However, there is considerable controversy in the statistical community as to which procedure is 'best'. The truth is that no one procedure is best for all types of problems.

Exercises

Methods

11 In exercise 1, five observations were selected from each of three populations. For these data, $\bar{x}_1 = 30, \bar{x}_2 = 45, \bar{x}_3 = 36$ and MSE = 5.5. At the $\alpha = 0.05$ level of significance, the null hypothesis of equal population means was rejected. In the following calculations, use $\alpha = 0.05$.

a. Use Fisher's LSD procedure to test whether there is a significant difference between the means of populations 1 and 2, populations 1 and 3, and populations 2 and 3.
b. Use Fisher's LSD procedure to develop a 95 per cent confidence interval estimate of the difference between the means of populations 1 and 2.

12 Four observations were selected from each of three populations. The data obtained are shown. In the following calculations, use $\alpha = 0.05$.

	Sample 1	Sample 2	Sample 3
	63	82	69
	47	72	54
	54	88	61
	40	66	48
$\bar{x}_j$	51	77	58
s_j^2	96.67	97.34	81.99

a. Use analysis of variance to test for a significant difference among the means of the three populations.
b. Use Fisher's LSD procedure to see which means are different.

Applications

13 Refer to exercise 6. At the $\alpha = 0.05$ level of significance, use Fisher's LSD procedure to test for the equality of the means for manufacturers 1 and 3. What conclusion can you draw after carrying out this test?

14 Refer to exercise 13. Use Fisher's LSD procedure to develop a 95 per cent confidence interval estimate of the difference between the means of population 1 and population 2.

15 Refer to exercise 8. At the $\alpha = 0.05$ level of significance, we can conclude that there are differences in the perceptions for marketing managers, marketing research specialists and advertizing specialists. Use the procedures in this section to determine where the differences occur. Use $\alpha = 0.05$.

16 To test for any significant difference in the number of hours between breakdowns for four machines, the following data were obtained.

Machine 1	Machine 2	Machine 3	Machine 4
6.4	8.7	11.1	9.9
7.8	7.4	10.3	12.8
5.3	9.4	9.7	12.1
7.4	10.1	10.3	10.8
8.4	9.2	9.2	11.3
7.3	9.8	8.8	11.5

a. At the $\alpha = 0.05$ level of significance, what is the difference, if any, in the population mean times among the four machines?
b. Use Fisher's LSD procedure to test for the equality of the means for machines 2 and 4. Use a 0.05 level of significance.

17 Refer to exercise 16. Use the Bonferroni adjustment to test for a significant difference between all pairs of means. Assume that a maximum overall experimentwise error rate of 0.05 is desired.

18 Refer to exercise 10. At the 0.05 level of significance, we can conclude that there are differences between the mean service ratings of small ships, medium ships, and large ships. Use the procedures in this section to determine where the differences occur. Use $\alpha = 0.05$.

13.4 An introduction to experimental design

Statistical studies can be classified as being either experimental or observational. In an experimental study, variables of interest are identified. Then, one or more factors in the study are controlled so that data can be obtained about how the factors influence the variables. In *observational* or *non-experimental* studies, no attempt is made to control the factors. A survey (see Chapter 22) is perhaps the most common type of observational study.

The NCP example that we used to introduce analysis of variance is an illustration of an observational statistical study. To measure how much NCP employees knew about total quality management, a random sample of six employees was selected from each of NCP's three plants and given a quality-awareness examination. The examination scores for these employees were then analyzed by analysis of variance to test the hypothesis that the population mean examination scores were equal for the three plants.

As an example of an experimental statistical study, let us consider the problem facing the Chemietech company. Chemietech developed a new filtration system for municipal water supplies.

The components for the new filtration system will be purchased from several suppliers, and Chemietech will assemble the components at its plant in North Saxony. The industrial engineering group is responsible for determining the best assembly method for the new filtration system. After considering a variety of possible approaches, the group narrows the alternatives to three: method A, method B and method C. These methods differ in the sequence of steps used to assemble the product. Managers at Chemietech want to determine which assembly method can produce the greatest number of filtration systems per week.

In the Chemietech experiment, assembly method is the independent variable or **factor**. Because three assembly methods correspond to this factor, we say that

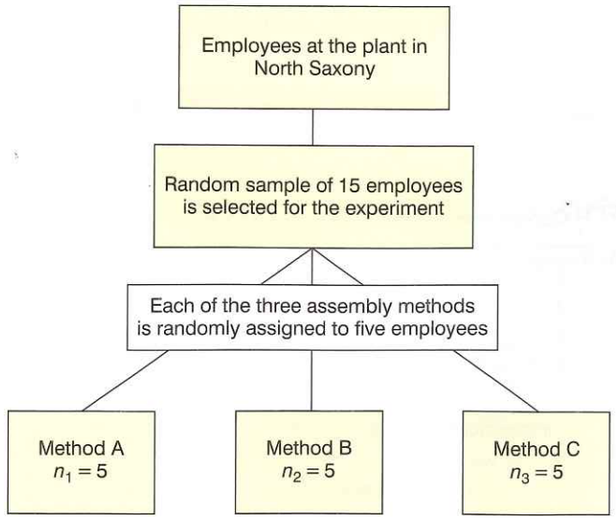
three treatments are associated with this experiment; each **treatment** corresponds to each of the three assembly methods. The Chemietech problem is an example of a **single-factor experiment** involving a qualitative factor (method of assembly). Other experiments may consist of multiple factors; some factors may be qualitative and some may be quantitative.

The three assembly methods or treatments define the three populations of interest for the Chemietech experiment. One population is all Chemietech employees who use assembly method A, another is those who use method B and the third is those who use method C. Note that for each population the dependent or response variable is the number of filtration systems assembled per week, and the primary statistical objective of the experiment is to determine whether the mean number of units produced per week is the same for all three populations.

Suppose a random sample of three employees is selected from all assembly workers at the Chemietech production facility. In experimental design terminology, the three randomly selected workers are the **experimental units**. The experimental design that we will use for the Chemietech problem is called a **completely randomized design**. This type of design requires that each of the three assembly methods or treatments be assigned randomly to one of the experimental units or workers. For example, method A might be randomly assigned to the second worker, method B to the first worker and method C to the third worker. The concept of *randomization*, as illustrated in this example, is an important principle of all experimental designs.

Note that this experiment would result in only one measurement or number of units assembled for each treatment. To obtain additional data for each assembly method, we must repeat or replicate the basic experimental process. Suppose, for example, that instead of selecting just three workers at random we selected 15 workers and then randomly assigned each of the three treatments to five of the workers. Because each method of assembly is assigned to five workers, we say that five replicates have been obtained. The process of *replication* is another important principle of experimental design. Figure 13.5 shows the completely randomized design for the Chemietech experiment.

Figure 13.5 Completely randomized design for evaluating the Chemietech assembly method experiment



Data collection

Once we are satisfied with the experimental design, we proceed by collecting and analysing the data. In the Chemietech case, the employees would be instructed in how to perform the assembly method assigned to them and then would begin assembling the new filtration systems using that method. After this assignment and training, the number of units assembled by each employee during one week is as shown in Table 13.3. The sample mean number of units produced with each of the three assembly methods is reported in the following table.

Assembly method	Mean number produced
A	62
B	66
C	52

From these data, method B appears to result in higher production rates than either of the other methods.

The real issue is whether the three sample means observed are different enough for us to conclude that the means of the populations corresponding to the three methods of assembly are different. To write this question in statistical terms, we introduce the following notation.

μ_1 = mean number of units produced per week for method A
μ_2 = mean number of units produced per week for method B
μ_3 = mean number of units produced per week for method C

Although we will never know the actual values of μ_1, μ_2 and μ_3, we want to use the sample means to test the following hypotheses.

H_0: $\mu_1 = \mu_2 = \mu_3$
H_1: Not all population means are equal

The problem we face in analysing data from a completely randomized experimental design is the same problem we faced when we first introduced analysis of variance as a method for testing whether the means of more than two populations are equal. In the next section we will show how analysis of variance is applied in problem situations such as the Chemietech assembly method experiment.

Table 13.3 Number of units produced by 15 workers

	Method		
Observation	A	B	C
1	58	58	48
2	64	69	57
3	55	71	59
4	66	64	47
5	67	68	49
Sample mean	62	66	52
Sample variance	27.5	26.5	31.0
Sample standard deviation	5.24	5.15	5.57

13.5 Completely randomized designs

The hypotheses we want to test when analysing the data from a completely randomized design are exactly the same as the general form of the hypotheses we presented in Section 13.2.

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$
H_1: Not all means are equal

Hence, to test for the equality of means in situations where the data are collected in a completely randomized experimental design, we can use analysis of variance as introduced in Sections 13.1 and 13.2. Recall that analysis of variance requires the calculation of two independent estimates of the population variance σ^2.

Between-treatments estimate of population variance

The between-treatments estimate of σ^2 is referred to as the *mean square due to treatments* and is denoted MSTR. The formula for computing MSTR follows:

Completely randomized designs
Mean square due to treatments

$$\text{MSTR} = \frac{\sum_{j=1}^{k} n_j(\bar{x}_j - \bar{\bar{x}})^2}{k-1} \qquad (13.20)$$

The numerator in equation (13.20) is called the *sum of squares between* or *sum of squares due to treatments* and is denoted SSTR. The denominator $k - 1$ represents the degrees of freedom associated with SSTR.

For the Chemietech data in Table 13.3, we obtain the following results (note: $\bar{\bar{x}} = 60$).

$$\text{SSTR} = \sum_{j=1}^{k} n_j(\bar{x}_j - \bar{\bar{x}})^2 = 5(62-60)^2 + 5(66-60)^2 + 5(52-60)^2 = 520$$

$$\text{MSTR} = \frac{\text{SSTR}}{k-1} = \frac{520}{3-1} = 260$$

Within-treatments estimate of population variance

The within-treatments estimate of σ^2 is referred to as the *mean square due to error* and is denoted MSE. The formula for computing MSE follows.

Mean square due to error

$$\text{MSE} = \frac{\sum_{j=1}^{k}(n_j - 1)s_j^2}{n_T - k} \qquad (13.21)$$

The numerator in equation (13.21) is called the *sum of squares within* or *sum of squares due to error* and is denoted SSE. The denominator of MSE is referred to as the degrees of freedom associated with SSE.

For the Chemietech data in Table 13.3, we obtain the following results.

$$\text{SSE} = \sum_{j=1}^{k} (n_j - 1)s_j^2 = 4(27.5) + 4(26.5) + 4(31) = 340$$

$$\text{MSE} = \frac{\text{SSE}}{n_T - k} = \frac{340}{15 - 3} = 28.33$$

Comparing the variance estimates: the *F* test

If the null hypothesis is true and the ANOVA assumptions are valid, the sampling distribution of MSTR/MSE is an F distribution with numerator degrees of freedom equal to $k - 1$ and denominator degrees of freedom equal to $n_T - k$. Recall also that if the means of the k populations are not equal, the value of MSTR/MSE will be inflated because MSTR overestimates σ^2. Hence we will reject H_0 if the resulting value of MSTR/MSE appears to be too large to have been selected at random from an F distribution with degrees of freedom $k - 1$ in the numerator and $n_T - k$ in the denominator.

Let us return to the Chemietech problem and use a level of significance $\alpha = 0.05$ to conduct the hypothesis test. The value of the test statistic is

$$F = \frac{\text{MSTR}}{\text{MSE}} = \frac{260}{28.33} = 9.18$$

The numerator degrees of freedom is $k - 1 - 3 - 1 = 2$ and the denominator degrees of freedom is $n_T - k = 15 - 3 = 12$. Because we will only reject the null hypothesis for large values of the test statistic, the p-value is the area under the F distribution to the right of $F = 9.18$. From Table 4 of Appendix B we find that the F value with an area of 0.01 in the upper tail is 6.93. Because the area in the upper tail for an F value of 9.18 must be less than 0.01, the p-value for the Chemietech hypothesis test is less than 0.01. Alternatively, we can use MINITAB, PASW or EXCEL to show that the exact p-value corresponding to $F = 9.18$ is 0.0038. With p-value $\leq \alpha = 0.05$, H_0 is rejected. The test gives us sufficient evidence to conclude that not all the population means are equal.

ANOVA table

We can now write the result that shows how the total sum of squares, SST, is partitioned.

$$\text{SST} = \text{SSTR} + \text{SSE} \qquad (13.22)$$

This result also holds true for the degrees of freedom associated with each of these sums of squares; that is, the total degrees of freedom is the sum of the degrees of freedom associated with SSTR and SSE. The general form of the ANOVA table for a completely randomized design is shown in Table 13.4; Table 13.5 is the corresponding ANOVA table for the Chemietech problem.

Table 13.4 ANOVA table for a completely randomized design

Source of variation	Degrees of freedom	Sum of squares	Mean square	F
Treatments	$k - 1$	SSTR	$\text{MSTR} = \frac{\text{SSTR}}{k - 1}$	$\frac{\text{MSTR}}{\text{MSE}}$
Error	$n_T - k$	SSE	$\text{MSE} = \frac{\text{SSE}}{n_T - k}$	
Total	$n_T - 1$	SST		

Table 13.5 ANOVA table for the chemietech problem

Source of variation	Degrees of freedom	Sum of squares	Mean square	F
Treatments	2	520	260.00	9.18
Error	12	340	28.33	
Total	14	860		

Pairwise comparisons

We can use Fisher's LSD procedure to test all possible pairwise comparisons for the Chemietech problem. At the 5 per cent level of significance, the t distribution table shows that with $n_T - k = 15 - 3 = 12$ degrees of freedom, $t_{0.025} = 2.179$. Using MSE = 28.33 in equation (13.17), we obtain Fisher's least significant difference.

$$\text{LSD} = t_{\alpha/2}\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} = 2.179\sqrt{28.33\left(\frac{1}{5} + \frac{1}{5}\right)} = 7.34$$

If the magnitude of the difference between any two sample means exceeds 7.34, we can reject the hypothesis that the corresponding population means are equal. For the Chemietech data in Table 13.3, we obtain the following results.

Sample differences	significant?
Method A − Method B = 62 − 66 = −4	No
Method A − Method C = 62 − 52 = 10	Yes
Method B − Method C = 66 − 52 = 14	Yes

Thus, the difference in the population means is attributable to the difference between the means for method A and method C and the difference between the means for method B and method C. Methods A and B therefore are preferred to method C. However, more testing should be done to compare method A with method B. The current study does not provide sufficient evidence to conclude that these two methods differ.

Exercises

Methods

19 The following data are from a completely randomized design.

	Treatment		
Observation	A	B	C
1	162	142	126
2	142	156	122
3	165	124	138
4	145	142	140
5	148	136	150
6	174	152	128
$\bar{x}_j$	156	142	134
s_j^2	164.4	131.2	110.4

a. Compute the sum of squares between treatments.
b. Compute the mean square between treatments.
c. Compute the sum of squares due to error.
d. Compute the mean square due to error.
e. At the $\alpha = 0.05$ level of significance, test whether the means for the three treatments are equal.

20 Refer to exercise 19.

a. Set up the ANOVA table.
b. At the $\alpha = 0.05$ level of significance, use Fisher's least significant difference procedure to test all possible pairwise comparisons. What conclusion can you draw after carrying out this procedure?

21 In a completely randomized experimental design, seven experimental units were used for each of the five levels of the factor. Complete the following ANOVA table.

Source of variation	Degrees of freedom	Sum of squares	Mean square	F
Treatments		300		
Error				
Total		**460**		

22 Refer to exercise 21.

a. What hypotheses are implied in this problem?
b. At the $\alpha = 0.05$ level of significance, can we reject the null hypothesis in part (a)? Explain.

23 In an experiment designed to test the output levels of three different treatments, the following results were obtained: SST = 400, SSTR = 150, $n_T = 19$. Set up the ANOVA table and test for any significant difference between the mean output levels of the three treatments. Use $\alpha = 0.05$.

24 In a completely randomized experimental design, 12 experimental units were used for the first treatment, 15 for the second treatment and 20 for the third treatment. Complete the following analysis of variance. At a 0.05 level of significance, is there a significant difference between the treatments?

Source of variation	Degrees of freedom	Sum of squares	Mean square	F
Treatments		1200		
Error				
Total		**1800**		

25 Develop the analysis of variance computations for the following experimental design. At $\alpha = 0.05$, is there a significant difference between the treatment means?

	Treatment		
	A	B	C
	136	107	92
	120	114	82
	113	125	85
	107	104	101
	131	107	89
	114	109	117
	129	97	110
	102	114	120
		104	98
		89	106
$\bar{x}_j$	119	107	100
s_j^2	146.86	96.44	173.78

Applications

26 Three different methods for assembling a product were proposed by an industrial engineer. To investigate the number of units assembled correctly with each method, 30 employees were randomly selected and randomly assigned to the three proposed methods in such a way that each method was used by ten workers. The number of units assembled correctly was recorded, and the analysis of variance procedure was applied to the resulting data set.

The following results were obtained: SST = 10 800; SSTR = 4560.

a. Set up the ANOVA table for this problem.
b. Use $\alpha = 0.05$ to test for any significant difference in the means for the three assembly methods.

27 In an experiment designed to test the breaking strength of four types of cables, the following results were obtained: SST = 85.05, SSTR = 61.64, $n_T = 24$. Set up the ANOVA table and test for any significant difference in the mean breaking strength of the four cables. Use $\alpha = 0.05$.

28 To study the effect of temperature on yield in a chemical process, five batches were produced at each of three temperature levels. The results follow. Construct an analysis of variance table. Use a 0.05 level of significance to test whether the temperature level has an effect on the mean yield of the process.

Temperature		
50°C	60°C	70°C
34	30	23
24	31	28
36	34	28
39	23	30
32	27	31

29 Auditors must make judgments about various aspects of an audit on the basis of their own direct experience, indirect experience, or a combination of the two. In a study, auditors were asked to make judgments about the frequency of errors to be found in an audit. The judgments by the auditors were then compared with the actual results. Suppose the following data were obtained from a similar study; lower scores indicate better judgments.

AUDJUDG

Direct	Indirect	Combination
17.0	16.6	25.2
18.5	22.2	24.0
15.8	20.5	21.5
18.2	18.3	26.8
20.2	24.2	27.5
16.0	19.8	25.8
13.3	21.2	24.2

Use $\alpha = 0.05$ to test to see whether the basis for the judgment affects the quality of the judgment. What is your conclusion?

30 Four different paints are advertised as having the same drying time. To check the manufacturer's claims, five samples were tested for each of the paints. The time in minutes until the paint was dry enough for a second coat to be applied was recorded. The following data were obtained.

Paint 1	Paint 2	Paint 3	Paint 4
128	144	133	150
137	133	143	142
135	142	137	135
124	146	136	140
141	130	131	153

PAINT

At the $\alpha = 0.05$ level of significance, test to see whether the mean drying time is the same for each type of paint.

31 Details of independent random samples of average hourly output for three manufacturing plants are as follows:

PLANT

Plant		
1	2	3
83	77	81.6
86	82	83
79	82	83.6
79.8	80.6	88
81.6	81	85
83.6	80	
	79.8	

Analyze these data appropriately. Do average outputs differ significantly by plant and if so how?

32 Refer to Exercise 29. Use Fisher's least significant difference procedure to test all possible pairwise comparisons. What conclusion can you draw after carrying out this procedure? Use $\alpha = 0.05$.

33 Refer to the NCP data in Table 13.1. Use Fisher's least significant difference procedure allowing for the Bonferroni adjustment to test all pairwise comparisons. What conclusion can you draw after carrying out this procedure? Use $\alpha = 0.05$.

13.6 Randomized block design

Thus far we have considered the completely randomized experimental design. Recall that to test for a difference among treatment means, we computed an F value by using the ratio

F Test Statistic

$$F = \frac{\text{MSTR}}{\text{MSE}} \qquad (13.23)$$

A problem can arise whenever differences due to extraneous factors (ones not considered in the experiment) cause the MSE term in this ratio to become large. In such cases, the F value in equation (13.23) can become small, signalling no difference among treatment means when in fact such a difference exists.

In this section we present an experimental design known as a **randomized block design**. Its purpose is to control some of the extraneous sources of variation by removing such variation from the MSE term. This design tends to provide a better estimate of the true error variance and leads to a more powerful hypothesis test in terms of the ability to

detect differences among treatment means. To illustrate, let us consider a stress study for air traffic controllers.

Air traffic controller stress test

A study measuring the fatigue and stress of air traffic controllers resulted in proposals for modification and redesign of the controller's work station. After consideration of several designs for the work station, three specific alternatives are selected as having the best potential for reducing controller stress. The key question is: to what extent do the three alternatives differ in terms of their effect on controller stress? To answer this question, we need to design an experiment that will provide measurements of air traffic controller stress under each alternative.

In a completely randomized design, a random sample of controllers would be assigned to each work station alternative. However, controllers are believed to differ substantially in their ability to handle stressful situations. What is high stress to one controller might be only moderate or even low stress to another. Hence, when considering the within-group source of variation (MSE), we must realize that this variation includes both random error and error due to individual controller differences. In fact, managers expected controller variability to be a major contributor to the MSE term.

One way to separate the effect of the individual differences is to use a randomized block design. Such a design will identify the variability stemming from individual controller differences and remove it from the MSE term. The randomized block design calls for a single sample of controllers. Each controller in the sample is tested with each of the three work station alternatives. In experimental design terminology, the work station is the *factor of interest* and the controllers are the *blocks*. The three treatments or populations associated with the work station factor correspond to the three work station alternatives. For simplicity, we refer to the work station alternatives as system A, system B and system C.

The *randomized* aspect of the randomized block design is the random order in which the treatments (systems) are assigned to the controllers. If every controller were to test the three systems in the same order, any observed difference in systems might be due to the order of the test rather than to true differences in the systems.

To provide the necessary data, the three work station alternatives were installed at the Berlin control centre. Six controllers were selected at random and assigned to operate each of the systems. A follow-up interview and a medical examination of each controller participating in the study provided a measure of the stress for each controller on each system. The data are reported in Table 13.6.

Table 13.6 A randomized block design for the air traffic controller stress test

		Treatments		
		System A	System B	System C
Blocks	Controller 1	15	15	18
	Controller 2	14	14	14
	Controller 3	10	11	15
	Controller 4	13	12	17
	Controller 5	16	13	16
	Controller 6	13	13	13

Table 13.7 Summary of stress data for the air traffic controller stress test

		Treatments				
		System A	System B	System C	Row or block totals	Block means
Blocks	Controller 1	15	15	18	48	$\bar{x}_{1.} = 48/3 = 16.0$
	Controller 2	14	14	14	42	$\bar{x}_{2.} = 42/3 = 14.0$
	Controller 3	10	11	15	36	$\bar{x}_{3.} = 36/3 = 12.0$
	Controller 4	13	12	17	42	$\bar{x}_{4.} = 42/3 = 14.0$
	Controller 5	16	13	16	45	$\bar{x}_{5.} = 45/3 = 15.0$
	Controller 6	13	13	13	39	$\bar{x}_{6.} = 39/3 = 13.0$
Column or Treatment Totals		81	78	93	252	$\bar{\bar{x}} = \frac{252}{18} = 14.0$
Treatment Means		$\bar{x}_{.1} = \frac{81}{6} = 13.5$	$\bar{x}_{.2} = \frac{78}{6} = 13.0$	$\bar{x}_{.3} = \frac{93}{6} = 15.5$		

Table 13.7 is a summary of the stress data collected. In this table we include column totals (treatments) and row totals (blocks) as well as some sample means that will be helpful in making the sum of squares computations for the ANOVA procedure. Because lower stress values are viewed as better, the sample data seem to favour system B with its mean stress rating of 13. However, the usual question remains: do the sample results justify the conclusion that the population mean stress levels for the three systems differ? That is, are the differences statistically significant? An analysis of variance computation similar to the one performed for the completely randomized design can be used to answer this statistical question.

ANOVA procedure

The ANOVA procedure for the randomized block design requires us to partition the sum of squares total (SST) into three groups: sum of squares due to treatments, sum of squares due to blocks and sum of squares due to error. The formula for this partitioning follows.

$$\text{SST} = \text{SSTR} + \text{SSBL} + \text{SSE} \tag{13.24}$$

This sum of squares partition is summarized in the ANOVA table for the randomized block design as shown in Table 13.8. The notation used in the table is

k = the number of treatments
b = the number of blocks
n_T = the total sample size ($n_T = kb$)

Note that the ANOVA table also shows how the $n_T - 1$ total degrees of freedom are partitioned such that $k - 1$ degrees of freedom go to treatments, $b - 1$ go to blocks, and $(k - 1)(b - 1)$ go to the error term. The mean square column shows the sum of squares

Table 13.8 ANOVA table for the randomized block design with k treatments and b blocks

Source of variation	Degrees of freedom	Sum of squares	Mean square	F
Treatments	$k - 1$	SSTR	$\text{MSTR} = \dfrac{\text{SSTR}}{k-1}$	$\dfrac{\text{MSTR}}{\text{MSE}}$
Blocks	$b - 1$	SSBL	$\text{MSBL} = \dfrac{\text{SSBL}}{b-1}$	
Error	$(k-1)(b-1)$	SSE	$\text{MSE} = \dfrac{\text{SSE}}{(k-1)(b-1)}$	
Total	$n_T - 1$	SST		

divided by the degrees of freedom, and $F = \text{MSTR/MSE}$ is the F ratio used to test for a significant difference among the treatment means. The primary contribution of the randomized block design is that, by including blocks, we remove the individual controller differences from the MSE term and obtain a more powerful test for the stress differences in the three work station alternatives.

Computations and conclusions

To compute the F statistic needed to test for a difference among treatment means with a randomized block design, we need to compute MSTR and MSE. To calculate these two mean squares, we must first compute SSTR and SSE; in doing so, we will also compute SSBL and SST. To simplify the presentation, we perform the calculations in four steps. In addition to k, b and n_T as previously defined, the following notation is used.

x_{ij} = value of the observation corresponding to treatment j in block i
$\bar{x}_j$ = sample mean of the jth treatment
$\bar{x}_i$ = sample mean for the ith block
$\bar{\bar{x}}$ = overall sample mean

Step 1 Compute the total sum of squares (SST).

$$\text{SST} = \sum_{i=1}^{b}\sum_{j=1}^{k}(x_{ij} - \bar{\bar{x}})^2 \qquad (13.25)$$

Step 2 Compute the sum of squares due to treatments (SSTR).

$$\text{SSTR} = b\sum_{j=1}^{k}(\bar{x}_j - \bar{\bar{x}})^2 \qquad (13.26)$$

Step 3 Compute the sum of squares due to blocks (SSBL).

$$\text{SSBL} = k\sum_{i=1}^{b}(\bar{x}_i - \bar{\bar{x}})^2 \qquad (13.27)$$

Step 4 Compute the sum of squares due to error (SSE).

$$\text{SSE} = \text{SST} - \text{SSTR} - \text{SSBL} \qquad (13.28)$$

For the air traffic controller data in Table 13.7, these steps lead to the following sums of squares.

Step 1 $\text{SST} = (15-14)^2 + (15-14)^2 + (18-14)^2 + \ldots + (13-14)^2 = 70$

Step 2 $\text{SSTR} = 6[(13.5-14)^2 + (13.0-14)^2 + (15.5-14)^2] = 21$

Step 3 $\text{SSBL} = 3[(16-14)^2 + (14-14)^2 + (12-14)^2 + (14-14)^2 + (15-14)^2 + (13-14)^2] = 30$

Step 4 $\text{SSE} = 70 - 21 - 30 = 19$

These sums of squares divided by their degrees of freedom provide the corresponding mean square values shown in Table 13.9.

Let us use a level of significance $\alpha = 0.05$ to conduct the hypothesis test. The value of the test statistic is

$$F = \frac{\text{MSTR}}{\text{MSE}} = \frac{10.5}{1.9} = 5.53$$

The numerator degrees of freedom is $k - 1 = 3 - 1 = 2$ and the denominator degrees of freedom is $(k-1)(b-1) = (3-1)(6-1) = 10$. Because we will only reject the null hypothesis for large values of the test statistic, the p-value is the area under the F distribution to the right of $F = 5.53$. From Table 4 of Appendix B we find that with the degrees of freedom 2 and 10, $F = 5.53$ is between $F_{0.025} = 5.46$ and $F_{0.01} = 7.56$. As a result, the area in the upper tail, or the p-value, is between 0.01 and 0.025. Alternatively, we can use MINITAB, PASW or EXCEL to show that the exact p-value for $F = 5.53$ is 0.0241. With p-value $\leq \alpha = 0.05$, we reject the null hypothesis H_0: $\mu_1 = \mu_2 = \mu_3$ and conclude that the population mean stress levels differ for the three work station alternatives.

Some general comments can be made about the randomized block design. The experimental design described in this section is a *complete* block design; the word 'complete' indicates that each block is subjected to all k treatments. That is, all controllers (blocks) were tested with all three systems (treatments). Experimental designs in which some but not all treatments are applied to each block are referred to as *incomplete block* designs. A discussion of incomplete block designs is beyond the scope of this text.

Because each controller in the air traffic controller stress test was required to use all three systems, this approach guarantees a complete block design. In some cases, however, **blocking** is carried out with 'similar' experimental units in each block. For example, assume that in a pretest of air traffic controllers, the population of controllers was divided into groups ranging from extremely high-stress individuals to extremely low-stress individuals.

Table 13.9 ANOVA table for the air traffic controller stress test

F	Source of variation	Degrees of freedom	Sum of squares	Mean square
10.5/1.9 = 5.53	Treatments	2	21	10.5
	Blocks	5	30	6.0
	Error	10	19	1.9
	Total	17	70	

The blocking could still be accomplished by having three controllers from each of the stress classifications participate in the study. Each block would then consist of three controllers in the same stress group. The randomized aspect of the block design would be the random assignment of the three controllers in each block to the three systems.

Finally, note that the ANOVA table shown in Table 13.8 provides an F value to test for treatment effects but *not* for blocks. The reason is that the experiment was designed to test a single factor – work station design. The blocking based on individual stress differences was conducted to remove such variation from the MSE term. However, the study was not designed to test specifically for individual differences in stress.

Some analysts compute F = MSB/MSE and use that statistic to test for significance of the blocks. Then they use the result as a guide to whether the same type of blocking would be desired in future experiments. However, if individual stress difference is to be a factor in the study, a different experimental design should be used. A test of significance on blocks should not be performed as a basis for a conclusion about a second factor.

Exercises

Methods

34 Consider the experimental results for the following randomized block design. Make the calculations necessary to set up the analysis of variance table.

	Treatments		
Blocks	A	B	C
1	0	9	8
2	2	6	5
3	8	15	14
4	0	18	18
5	8	7	8

Use $\alpha = 0.05$ to test for any significant differences.

35 The following data were obtained for a randomized block design involving five treatments and three blocks: SST = 430, SSTR = 310, SSB = 85. Set up the ANOVA table and test for any significant differences. Use $\alpha = 0.05$.

36 An experiment has been conducted for four treatments with eight blocks. Complete the following analysis of variance table.

Source of variation	Degrees of freedom	Sum of squares	Mean Square	F
Treatments		900		
Blocks		400		
Error				
Total		**1800**		

Use $\alpha = 0.05$ to test for any significant differences.

Applications

37 A car dealer conducted a test to determine if the time in minutes needed to complete a minor engine tune-up depends on whether a computerized engine analyzer or an electronic analyzer is used. Because tune-up time varies among compact, intermediate and full-sized cars, the three types of cars were used as blocks in the experiment. The data obtained follow.

	Analyzer	
Car	Computerized	Electronic
Compact	50	42
Intermediate	55	44
Full-sized	63	46

Use $\alpha = 0.05$ to test for any significant differences.

38 A textile mill produces a silicone proofed fabric for making into rainwear. The chemist in charge thinks that a silicone solution of about 12 per cent strength should yield a fabric with maximum waterproofing-index. He also suspected there may be some batch to batch variation because of slight differences in the cloth. To allow for this possibility five different strengths of solution were used on each of the three different batches of fabric. The following values of water-proofing index were obtained:

		[Strength of silicone solution (%)]				
		6	9	12	15	18
Fabric	1	20.8	20.6	22.0	22.6	20.9
	2	19.4	21.2	21.8	23.9	22.4
	3	19.9	21.1	22.7	22.7	22.1

Using $\alpha = 0.05$, carry out an appropriate test of these data and comment on the chemist's original beliefs.

39 An important factor in selecting software for word-processing and database management systems is the time required to learn how to use the system. To evaluate three file management systems, a firm designed a test involving five word-processing operators. Because operator variability was believed to be a significant factor, each of the five operators was trained on each of the three file management systems. The data obtained follow.

	System		
Operator	A	B	C
1	6	16	24
2	9	17	22
3	4	13	19
4	3	12	18
5	8	17	22

Use $\alpha = 0.05$ to test for any difference in the mean training time (in hours) for the three systems.

13.7 Factorial experiments

The experimental designs we considered thus far enable us to draw statistical conclusions about one factor. However, in some experiments we want to draw conclusions about more than one variable or factor. **Factorial experiments** and their corresponding ANOVA computations are valuable designs when simultaneous conclusions about two or more factors are required. The term *factorial* is used because the experimental conditions include all possible combinations of the factors. For example, for a levels of factor A and b levels of factor B, the experiment will involve collecting data on ab treatments. In this section we will show the analysis for a two-factor factorial experiment. The basic approach can be extended to experiments involving more than two factors.

As an illustration of a two-factor factorial experiment, we will consider a study involving the Graduate Management Admissions Test (GMAT), a standardized test used by graduate schools of business to evaluate an applicant's ability to pursue a graduate programme in that field. Scores on the GMAT range from 200 to 800, with higher scores implying higher aptitude.

In an attempt to improve students' performance on the GMAT exam, a major Spanish university is considering offering the following three GMAT preparation programmes.

1. A three-hour review session covering the types of questions generally asked on the GMAT.
2. A one-day programme covering relevant exam material, along with the taking and grading of a sample exam.
3. An intensive ten-week course involving the identification of each student's weaknesses and the setting up of individualized programmes for improvement.

Therefore, one factor in this study is the GMAT preparation programme, which has three levels: three-hour review, one-day programme and ten-week course. Before selecting the preparation programme to adopt, further study will be conducted to determine how the proposed programmes affect GMAT scores.

The GMAT is usually taken by students from three colleges: the College of Business, the College of Engineering and the College of Arts and Sciences. Therefore, a second factor of interest in the experiment is whether a student's undergraduate college affects the GMAT score. This second factor, undergraduate college, also has three levels: business, engineering and arts and sciences. The factorial design for this experiment with three levels corresponding to factor A, the preparation programme, and three levels corresponding to factor B, the undergraduate college, will give rise to a total of $3 \times 3 = 9$ treatments. These treatments or combinations of factor levels are summarized in Table 13.10.

Assume that a sample of two students will be selected corresponding to each of the nine treatments shown in Table 13.10: two business students will take the three-hour

Table 13.10 Nine treatments for the two-factor GMAT experiment

		Factor B: College		
		Business	Engineering	Arts and sciences
Factor A: Preparation Programme	Three-hour review	1	2	3
	One-day programme	4	5	6
	Ten-week course	7	8	9

review, two will take the one-day programme and two will take the ten-week course. In addition, two engineering students and two arts and sciences students will take each of the three preparation programmes. In experimental design terminology, the sample size of two for each treatment indicates that we have two **replications**. Additional replications and a larger sample size could easily be used, but we elect to minimize the computational aspects for this illustration.

This experimental design requires that six students who plan to attend graduate school be randomly selected from *each* of the three undergraduate colleges. Then two students from each college should be assigned randomly to each preparation programme, resulting in a total of 18 students being used in the study.

Let us assume that the randomly selected students participated in the preparation programmes and then took the GMAT. The scores obtained are reported in Table 13.11.

The analysis of variance computations with the data in Table 13.11 will provide answers to the following questions.

- **Main effect (factor A):** Do the preparation programmes differ in terms of effect on GMAT scores?
- **Main effect (factor B):** Do the undergraduate colleges differ in terms of effect on GMAT scores?
- **Interaction effect (factors A and B):** Do students in some colleges do better on one type of preparation programme whereas others do better on a different type of preparation programme?

The term **interaction** refers to a new effect that we can now study because we used a factorial experiment. If the interaction effect has a significant impact on the GMAT scores, we can conclude that the effect of the type of preparation programme depends on the undergraduate college.

ANOVA procedure

The ANOVA procedure for the two-factor factorial experiment is similar to the completely randomized experiment and the randomized block experiment in that we again partition the sum of squares and the degrees of freedom into their respective sources. The formula for partitioning the sum of squares for the two-factor factorial experiments follows.

$$\text{SST} = \text{SSA} + \text{SSB} + \text{SSAB} + \text{SSE} \qquad (13.29)$$

Table 13.11 GMAT scores for the two-factor experiment

		Factor B: College		
		Business	Engineering	Arts and sciences
Factor A: Preparation Programme	Three-hour review	500	540	480
		580	460	400
	One-day programme	460	560	420
		540	620	480
	Ten-week course	560	600	480
		600	580	410

Table 13.12 ANOVA table for the two-factor factorial experiment with r replications

Source of variation	Degrees of freedom	Sum of squares	Mean square	F
Factor A	$a - 1$	SSA	$\text{MSA} = \dfrac{\text{SSA}}{a-1}$	$\dfrac{\text{MSA}}{\text{MSE}}$
Factor B	$b - 1$	SSB	$\text{MSB} = \dfrac{\text{SSB}}{b-1}$	$\dfrac{\text{MSB}}{\text{MSE}}$
Interaction	$(a - 1)(b - 1)$	SSAB	$\text{MSAB} = \dfrac{\text{SSAB}}{(a-1)(b-1)}$	$\dfrac{\text{MSAB}}{\text{MSE}}$
Error	$ab(r - 1)$	SSE	$\text{MSE} = \dfrac{\text{SSE}}{ab(r-1)}$	
Total	$n_T - 1$	SST		

The partitioning of the sum of squares and degrees of freedom is summarized in Table 13.12. The following notation is used.

a = number of levels of factor A
b = number of levels of factor B
r = number of replications
n_T = total number of observations taken in the experiment; $n_T = abr$

Computations and conclusions

To compute the F statistics needed to test for the significance of factor A, factor B, and interaction, we need to compute MSA, MSB, MSAB, and MSE. To calculate these four mean squares, we must first compute SSA, SSB, SSAB, and SSE; in doing so we will also compute SST. To simplify the presentation, we perform the calculations in five steps. In addition to a, b, r and n_T as previously defined, the following notation is used.

x_{ijk} = observation corresponding to the kth replicate taken from treatment i of factor A and treatment j of factor B
$\bar{x}_{i.}$ = sample mean for the observations in treatment i (factor A)
$\bar{x}_{.j}$ = sample mean for the observations in treatment j (factor B)
$\bar{x}_{ij}$ = sample mean for the observations corresponding to the combination of treatment i (factor A) and treatment j (factor B)
$\bar{\bar{x}}$ = overall sample mean of all n_T observations

Step 1 Compute the total sum of squares.

$$\text{SST} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(x_{ijk} - \bar{\bar{x}})^2 \qquad (13.30)$$

Step 2 Compute the sum of squares for factor A.

$$\text{SSA} = br\sum_{i=1}^{a}(\bar{x}_{i.} - \bar{\bar{x}})^2 \qquad (13.31)$$

Step 3 Compute the sum of squares for factor B.

$$\text{SSBR} = ar\sum_{i=1}^{b}(\bar{x}_{.j} - \bar{\bar{x}})^2 \qquad (13.32)$$

Step 4 Compute the sum of squares for interaction.

$$\text{SSAB} = r\sum_{i=}^{a}\sum_{j=1}^{b}(\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{\bar{x}})^2 \qquad (13.33)$$

Step 5 Compute the sum of squares due to error.

$$\text{SSE} = \text{SST} - \text{SSA} - \text{SSB} - \text{SSAB} \qquad (13.34)$$

Table 13.13 reports the data collected in the experiment and the various sums that will help us with the sum of squares computations. Using equations (13.30) through (13.34), the sums of squares for the GMAT two-factor factorial experiment can be calculated as follows.

Step 1 $\text{SST} = (500 - 515)^2 + (580 - 515)^2 + (540 - 515)^2 + \cdots + (410 - 515)^2 = 82\,450$

Step 2 $\text{SSA} = (3)(2)[(493.33 - 515)^2 + (513.33 - 515)^2 + (538.33 - 515)^2] = 6100$

Step 3 $\text{SSB} = (3)(2)[(540 - 515)^2 + (560 - 515)^2 + (445 - 515)^2] = 45\,300$

Step 4 $\text{SSAB} = 2[(540 - 493.33 - 540 - 515)^2 + (500 - 493.33 - 560 + 515)^2 + \cdots + (445 - 538.33 - 445 + 515)^2] = 11\,200$

Step 5 $\text{SSE} = 82\,450 - 6100 - 45\,300 - 11\,200 = 19\,850$

These sums of squares divided by their corresponding degrees of freedom, as shown to prepare students from the different colleges for the GMAT in Table 13.14, provide the appropriate mean square values for testing the two main effects (preparation programme and undergraduate college) and the interaction effect.

Let us use a level of significance $\alpha = 0.05$ to conduct the hypothesis tests for the two-factor GMAT study. Because of the computational effort involved in any modest- to large-size factorial experiment, the computer usually plays an important role in performing the analysis of variance computations and in the calculation of the p-values used to make the hypothesis testing decisions. Figure 13.6 shows the MINITAB output for the analysis of variance for the GMAT two-factor factorial experiment Because the p-value used to test for significant differences among the three preparation programmes (factor A) = 0.299 is greater than $\alpha = 0.05$, we deduce there is no significant difference in the mean GMAT test scores for the three preparation programmes. However, for the undergraduate college effect, the p-value = 0.005 is less than $\alpha = 0.05$; thus, there is a significant difference in the mean GMAT test scores among the three undergraduate colleges. Finally, because the p-value of 0.350 for the interaction effect is greater than $\alpha = 0.05$, there is no significant interaction effect. Therefore, the study provides no reason to believe that the three preparation programmes differ in their ability to prepare students from the different colleges for the GMAT.

Undergraduate college however was found to be a significant factor. Checking the calculations in Table 13.13, we see that the sample means are: business students $\bar{x}_{.1} = 540$, engineering students $\bar{x}_{.2} = 560$ and arts and sciences students $\bar{x}_{.3} = 445$. Tests on individual treatment means can be conducted; yet after reviewing the three sample means, we would

Table 13.13 GMAT summary data for the two-factor experiment

Treatment combination totals	Factor B: College of Business	Factor B: College of Engineering	Factor B: College of Arts and sciences	Row totals	Factor A means
Factor A: preparation programme — Three-hour review	500 580 1080	540 460 1000	480 400 880	2960	$\bar{x}_{1\cdot} = \frac{2960}{6} = 493.33$
Factor A: preparation programme — One-day programme	460 540 1000	560 620 1180	420 480 900	3080	$\bar{x}_{2\cdot} = \frac{3080}{6} = 513.33$
Factor A: preparation programme — 10-week course	560 600 1160	600 580 1180	480 410 890	3230	$\bar{x}_{3\cdot} = \frac{3230}{6} = 538.33$
Column totals	3240	3360	2670	9270 — Overall total	
Factor B means	$\bar{x}_{\cdot 1} = \frac{3240}{6} = 540$	$\bar{x}_{\cdot 2} = \frac{3360}{6} = 560$	$\bar{x}_{\cdot 3} = \frac{2670}{6} = 445$		$\bar{\bar{x}} = \frac{9270}{18} = 515$
Treatment combination means	$\bar{x}_{11} = \frac{1080}{2} = 540$	$\bar{x}_{12} = \frac{1000}{2} = 500$	$\bar{x}_{13} = \frac{880}{2} = 440$		
	$\bar{x}_{21} = \frac{1000}{2} = 500$	$\bar{x}_{22} = \frac{1180}{2} = 590$	$\bar{x}_{23} = \frac{900}{2} = 450$		
	$\bar{x}_{31} = \frac{1160}{2} = 580$	$\bar{x}_{32} = \frac{1180}{2} = 590$	$\bar{x}_{33} = \frac{890}{2} = 445$		

Table 13.14 ANOVA table for the two-factor GMAT study

Source of variation	Degrees of freedom	Sum of squares	Mean square	*F*
Factor A	2	6100	3050	3050/2206 = 1.38
Factor B	2	45 300	22 650	22 650/2206 = 10.27
Interaction	4	11 200	2800	2800/2206 = 1.27
Error	9	19 850	2206	
Total	17	82 450		

Figure 13.6 MINITAB output for the GMAT two-factor design

Two-way ANOVA: Score versus Factor A, Factor B

Source	DF	SS	MS	F	P
Factor A	2	6100	3050.0	1.38	0.299
Factor B	2	45300	22650.0	10.27	0.005
Interaction	4	11200	2800.0	1.27	0.350
Error	9	19850	2205.6		
Total	17	82450			

S = 46.96 R-Sq = 75.92% R-Sq(adj) = 54.52%

anticipate no difference in preparation for business and engineering graduates. However, the arts and sciences students appear to be significantly less prepared for the GMAT than students in the other colleges. Perhaps this observation will lead the university to consider other options for assisting these students in preparing for graduate management admission tests.

Exercises

Methods

40 A factorial experiment involving two levels of factor A and three levels of factor B resulted in the following data.

		Factor B		
		Level 1	Level 2	Level 3
Factor A	Level 1	135 165	90 66	75 93
	Level 2	125 95	127 105	120 136

Test for any significant main effects and any interaction. Use $\alpha = 0.05$.

41 The calculations for a factorial experiment involving four levels of factor A, three levels of factor B, and three replications resulted in the following data: SST = 280, SSA = 26, SSB = 23, SSAB = 175. Set up the ANOVA table and test for any significant main effects and any interaction effect. Use $\alpha = 0.05$.

Applications

42 A mail-order catalogue firm designed a factorial experiment to test the effect of the size of a magazine advertisement and the advertisement design on the number of catalogue requests received (data in thousands). Three advertising designs and two different-size advertisements were considered. The data obtained follow.

		Size of advertisement	
		Small	Large
Design	A	8	12
		12	8
	B	22	26
		14	30
	C	10	18
		18	14

Use the ANOVA procedure for factorial designs to test for any significant effects due to type of design, size of advertizement, or interaction. Use $\alpha = 0.05$.

43 A factorial experiment involved measurement of average fuel consumption for 36 long journeys for three different types of vehicle by value and three different types of fuel additive. The data (km / litre) obtained follow:

	Fuel additive		
Vehicle type	1	2	3
A	7	8	8
	7	8	8
	7	7	8
	8	7	8
B	6	8	7
	6	8	7
	6	8	8
	6	8	7
C	6	8	7
	6	7	7
	6	7	7
	6	7	7

Perform an appropriate analysis of these data. Use $\alpha = 0.05$. What are your conclusions?

44 A study reported in *The Accounting Review* examined the separate and joint effects of two levels of time pressure (low and moderate) and three levels of knowledge (naïve, declarative and procedural) on key word selection behaviour in tax research. Subjects were given a tax case containing a set of facts, a tax issue and a key word index consisting of 1336 key words. They were asked to select the key words they believed would refer them to a tax authority relevant to resolving the tax case. Prior to the experiment, a group of tax experts determined that the text contained 19 relevant key words. Subjects in the naïve group had little or no declarative or procedural knowledge, subjects in the declarative group had significant declarative knowledge but little or no procedural knowledge, and subjects in the procedural group had significant declarative knowledge and procedural knowledge. Declarative knowledge consists of knowledge of both the applicable tax rules and the technical terms used to describe such rules. Procedural knowledge is knowledge of the rules that guide the tax researcher's search for relevant key words. Subjects in the low time pressure situation were told they had 25 minutes to complete the problem, an amount of time which should be 'more than adequate' to complete the case; subjects in the moderate time pressure situation were told they would have 'only' 11 minutes to complete the case. Suppose 25 subjects were selected for each of the six treatments and the sample means for each treatment are as follows (standard deviations are in parentheses).

		Knowledge		
		Naïve	Declarative	Procedural
Time pressure	Low	1.13	1.56	2.00
		(1.12)	(1.33)	(1.54)
	Moderate	0.48	1.68	2.86
		(0.80)	(1.36)	(1.80)

Use the ANOVA procedure to test for any significant differences due to time pressure, knowledge, and interaction. Use a 0.05 level of significance. Assume that the total sum of squares for this experiment is 327.5

Summary

In this chapter we showed how analysis of variance can be used to test for differences among means of several populations or treatments. We introduced the completely randomized design, the randomized block design and the two-factor factorial experiment and confirmed corresponding assumptions. The completely randomized design and the randomized block design are used to draw conclusions about differences in the means of a single factor. The primary purpose of blocking in the randomized block design is to remove extraneous sources of variation from the error term. Such blocking provides a better estimate of the true error variance and a better test for determining whether the population or treatment means of the factor differ significantly. Correspondingly factorial experiments involve conclusions being drawn about two or more factors including their potential interactions.

We showed that the basis for the statistical tests used in analysis of variance and experimental design is the development of two independent estimates of the population variance σ^2. In the single-factor case, one estimator is based on the variation between the treatments; this estimator provides an unbiased estimate of σ^2 only if the treatment means are all equal. A second estimator of σ^2 is based on the variation of the observations within each sample; this estimator will always provide an unbiased estimate of σ^2. By computing the ratio of these two estimators (the F statistic) we developed a rejection rule for determining whether to reject the null hypothesis that the population or treatment means are equal. In all the experimental designs considered, the partitioning of the sum of squares and degrees of freedom into their various sources enabled us to compute the appropriate values for the analysis of variance calculations and tests. We also showed how Fisher's LSD procedure and the Bonferroni adjustment can be used to perform pairwise comparisons to determine which means are different.

Key terms

ANOVA table
Blocking
Comparisonwise Type I error rate
Completely randomized design
Experimental units
Experimentwise Type I error rate
Factor
Factorial experiments
Interaction
Multiple comparison procedures
Partitioning
Randomized block design
Replications
Single-factor experiment
Treatment

Key formulae

Testing for the equality of *k* population means

Sample mean for treatment *j*

$$\bar{x}_j = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j} \tag{13.1}$$

Sample variance for treatment *j*

$$s_j^2 = \frac{\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{n_j - 1} \tag{13.2}$$

Overall sample mean

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n_j} x_{ij}}{n_T} \tag{13.3}$$

$$n_T = n_2 + \cdots + n_k \tag{13.4}$$

Mean square due to treatments

$$\text{MSTR} = \frac{\text{SSTR}}{k - 1} \tag{13.7}$$

Sum of squares due to treatments

$$\text{SSTR} = \sum_{j=1}^{k} n_j (\bar{x}_j - \bar{\bar{x}})^2 \tag{13.8}$$

Mean square due to error

$$\text{MSE} = \frac{\text{SSE}}{n_T - k} \tag{13.10}$$

Sum of squares due to error

$$\text{SSE} = \sum_{j=1}^{k} (n_j - 1) s_j^2 \tag{13.11}$$

Test statistic for the equality of *k* population means

$$F = \frac{\text{MSTR}}{\text{MSE}} \tag{13.12}$$

Total sum of squares

$$SST = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{\bar{x}})^2 \tag{13.13}$$

Partitioning of sum of squares

$$\text{SST} = \text{SSTR} + \text{SSE} \tag{13.14}$$

Multiple comparison procedures Test statistic for Fisher's LSD procedure

$$t = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} \tag{13.16}$$

Fisher's LSD

$$\text{LSD} = t_{\alpha/2}\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \quad (13.17)$$

Completely randomized designs
Mean square due to treatments

$$\text{MSTR} = \frac{\sum_{j=1}^{k} n_j(\bar{x}_j - \bar{\bar{x}})^2}{k-1} \quad (13.20)$$

Mean square due to error

$$\text{MSE} = \frac{\sum_{j=1}^{k} (n_j - 1)s_j^2}{n_T - k} \quad (13.21)$$

***F* test statistic**

$$\text{F} = \frac{\text{MSTR}}{\text{MSE}} \quad (13.23)$$

Randomized block designs
Total sum of squares

$$\text{SST} = \sum_{i=1}^{b}\sum_{j=1}^{k} (x_{ij} - \bar{\bar{x}})^2 \quad (13.25)$$

Sum of squares due to treatments

$$\text{SSTR} = b\sum_{j=1}^{k} (\bar{x}_j - \bar{\bar{x}})^2 \quad (13.26)$$

Sum of squares due to blocks

$$\text{SSBL} = k\sum_{i=1}^{b} (\bar{x}_i - \bar{\bar{x}})^2 \quad (13.27)$$

Sum of squares due to error

$$\text{SSE} = \text{SST} - \text{SSTR} - \text{SSBL} \quad (13.28)$$

Factorial experiments
Total sum of squares

$$\text{SST} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r} (x_{ijk} - \bar{\bar{x}})^2 \quad (13.30)$$

Sum of squares for factor A

$$\text{SSA} = br\sum_{i=1}^{a} (\bar{x}_i - \bar{\bar{x}})^2 \quad (13.31)$$

Sum of squares for factor B

$$\text{SSTR} = ar\sum_{i=1}^{b} (\bar{x}_j - \bar{\bar{x}})^2 \quad (13.32)$$

Sum of squares for interaction

$$\text{SSAB} = r\sum_{i=}^{a}\sum_{j=1}^{b} (x_{ij} - \bar{x}_i - \bar{x}_j + \bar{\bar{x}})^2 \quad (13.33)$$

Sum of squares for error

$$\text{SSE} = \text{SST} - \text{SSA} - \text{SSB} - \text{SSAB} \quad (13.34)$$

Case problem 1 Wentworth Medical Centre

As part of a long-term study of individuals 65 years of age or older, sociologists and physicians at the Wentworth Medical Centre in Britain investigated the relationship between geographic location and depression. A sample of 60 individuals, all in reasonably good health, was selected; 20 individuals were residents of Scotland, 20 were residents of England, and 20 were residents of Wales. Each of the individuals sampled was given a standardized test to measure depression. The data collected follow; higher test scores indicate higher levels of depression. These data are available on the data disk in the file Medical1.

A second part of the study considered the relationship between geographic location and depression for individuals 65 years of age or older who had a chronic health condition such as arthritis, hypertension, and/or heart ailment. A sample of 60 individuals with such conditions was identified. Again, 20 were residents of Scotland, 20 were residents of England and 20 were residents of Wales. The levels of depression recorded for this study follow.

These data are available on the CD accompanying the text in the file named Medical2.

Data from Medical1			Data from Medical2		
Scotland	England	Wales	Scotland	England	Wales
3	8	10	13	14	10
7	11	7	12	9	12
7	9	3	17	15	15
3	7	5	17	12	18
8	8	11	20	16	12
8	7	8	21	24	14
8	8	4	16	18	17
5	4	3	14	14	8
5	13	7	13	15	14
2	10	8	17	17	16
6	6	8	12	20	18

Data from Medical1			Data from Medical2		
Scotland	England	Wales	Scotland	England	Wales
2	8	7	9	11	17
6	12	3	12	23	19
6	8	9	15	19	15
9	6	8	16	17	13
7	8	12	15	14	14
5	5	6	13	9	11
4	7	3	10	14	12
7	7	8	11	13	13
3	8	11	17	11	11

An elderly lady taking part in a depression study. © Mark Papas.

Managerial Report

1. Use descriptive statistics to summarize the data from the two studies. What are your preliminary observations about the depression scores?
2. Use analysis of variance on both data sets. State the hypotheses being tested in each case. What are your conclusions?
3. Use inferences about individual treatment means where appropriate. What are your conclusions?
4. Discuss extensions of this study or other analyses that you feel might be helpful.

Case problem 2 Product Design Testing

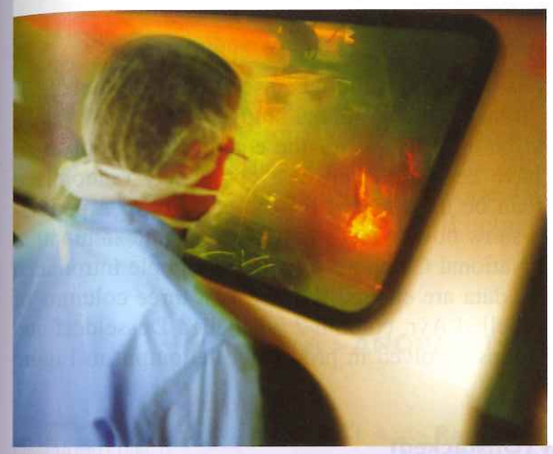
An engineering manager has been designated the task of evaluating a commercial device subject to marked variations in temperature. Three different types of component are being considered for the device. When the device is manufactured and is shipped to the field, the manager has no control over the temperature extremes that the device will encounter, but knows from experience that temperature is an important factor in relation to the component's life. Notwithstanding this, temperature can be controlled in the laboratory for the purposes of the test.

The engineering manager arranges for all three components to be tested at the temperature levels: −10°C, 20°C, and 50°C – as these temperature levels are consistent with the product end-use environment. Four components are tested for each combination of type and temperature, and all 36 tests are run in random order. The resulting observed component life data are presented in Table 1.

Table 1 Component lifetimes (000s of hours)

	Temperature(°C)					
Type	−10		20		50	
1	3.12	3.70	0.82	0.96	0.48	1.68
	1.80	4.32	1.92	1.80	1.97	1.39
2	3.60	4.51	3.02	2.93	0.60	1.68
	3.82	3.02	2.54	2.76	1.39	1.08
3	3.31	2.64	4.18	2.88	2.30	2.50
	4.03	3.84	3.60	3.34	1.97	1.44

Testing the effects of extreme temperatures on products in a laboratory. © Bartee Inc/Phototake Science.



Managerial Report

1. What are the effects of the chosen factors on the life of the component?
2. Do any components have a consistently long life regardless of temperature?
3. What recommendation would you make to the engineering manager?

Software Section for Chapter 13

Analysis of variance and experimental design using MINITAB

Single factor observational studies and completely randomized designs

In Section 13.2 we showed how analysis of variance could be used to test for the equality of k population means using data from an observational study. In Section 13.5 we showed how the same approach could be used to test for the equality of k population means in situations where the data have been collected in a completely randomized design. To illustrate how MINITAB can be used to test for the equality of k population means for both of these cases, we show how to test whether the mean examination score is the same at each plant in the National Computer Products example introduced in Section 13.1. The examination score data are entered into the first three columns of a MINITAB worksheet; column 1 is labelled Ayr, column 2 is labelled Dusseldorf and column 3 is labelled Stockholm. The steps involved in producing the output in Figure 13.4 in MINITAB follow.

Step 1 Select **Stat > ANOVA > One-way (Unstacked)** [Main menu bar]

Step 2 Enter C1-C3 in the **Responses (in separate columns)** box [**One-way (Unstacked)** panel]
Click **OK**

Randomized block designs

In Section 13.6 we showed how analysis of variance could be used to test for the equality of k population means using data from a randomized block design. To illustrate how MINITAB can be used for this type of experimental design, we show how to test whether the mean stress levels for air traffic controllers is the same for three work stations. The stress level scores shown in Table 13.6 are entered into column 1 of a MINITAB worksheet. Coding the treatments as 1 for System A, 2 for System B and 3 for System C, the coded values for the treatments are entered into column 2 of the worksheet. Finally, the corresponding number of each controller (1, 2, 3, 4, 5, 6) is entered into column 3. Thus, the values in the first row of the worksheet are 15, 1, 1; the values in row 2 are 15, 2, 1; the values in row 3 are 18, 3, 1; the values in row 4 are 14, 1, 2 and so on. In particular, the steps involved in producing the MINITAB output corresponding to the ANOVA table shown in Table 13.9 follow.

Step 1 Select **Stat > ANOVA Two-way** [Main menu bar]

Step 2 Enter C1 in the **Response** box
Enter C2 in the **Row factor** box
Enter C3 in the **Column factor** box
Select **Fit additive model**
Click **OK**
[**ANOVA Two-way** panel]

Factorial experiments

In Section 13.7 we showed how analysis of variance could be used to test for the equality of k population means using data from a factorial experiment. To illustrate how MINITAB can be used for this type of experimental design, we show how to analyse the data for the two-factor GMAT experiment introduced in that section. The GMAT scores shown in Table 13.11 are entered into column 1 of a MINITAB worksheet; column 1 is labelled Score, column 2 is labelled Factor A, and column 3 is labelled Factor B. Coding the factor A preparation programmes as 1 for the three-hour review, 2 for the one-day programme, and 3 for the ten-week course, the coded values for factor A are entered into column 2 of the worksheet. Coding the factor B colleges as 1 for Business, 2 for Engineering, and 3 for Arts and Sciences, the coded values for factor B are entered into column 3. Thus, the values in the first row of the worksheet are 500, 1, 1; the values in row 2 are 580, 1, 1; the values in row 3 are 540, 1, 2; the values in row 4 are 460, 1, 2 and so on. In particular, the steps involved in producing the MINITAB output corresponding to the ANOVA table shown in Figure 13.6 follow.

Step 1 **Stat > ANOVA > Two-way** [Main menu bar]

Step 2 Enter C1 in the **Response** box
Enter C2 in the **Row factor** box
Enter C3 in the **Column factor** box
Click **OK**
[**ANOVA Two-way** panel]

Analysis of variance and experimental design using EXCEL

Single-factor observational studies and completely randomized designs

In Section 13.2 we showed how analysis of variance could be used to test for the equality of k population means using data from an observational study. In Section 13.5 we showed how the same approach could be used to test for equality of k population means in situations where the data are collected in a completely randomized design. To illustrate how EXCEL can be used to test for the equality of k population means for both of these cases, we show how to test whether the mean examination score is the same at each plant in the National Computer Products example introduced in Section 13.1. The examination

score data are entered into worksheet rows 2 to 7 of columns B, C and D as shown in Figure 13.7. Note that cells A2:A7 are used to identify the observations at each of the plants. The steps involved in using EXCEL to produce the output shown in cells A9:G23 follow; the ANOVA portion of this output corresponds to the ANOVA table shown in Table 13.2.

Step 1 Select **Data > Data Analysis > Anova: Single-Factor** [Main menu bar]
Click **OK**

Step 2 **[Anova: Single-Factor** panel]
Enter B1:D7 in **Input Range** box
Select **Columns**
Select **Labels in First Row**
Select **Output Range** and enter A9 in the box
Click **OK**

Randomized block designs

In Section 13.6 we showed how analysis of variance could be used to test for the equality of k population means using data from a randomized block design. To illustrate how EXCEL can be used for this type of experimental design, we show

Figure 13.7 EXCEL solution for the NCP analysis of variance example

	A	B	C	D	E	F	G
1	**Observation**	**Ayr**	**Dusseldorf**	**Stockholm**			
2	1	85	71	59			
3	2	75	75	64			
4	3	82	73	62			
5	4	76	74	69			
6	5	71	69	75			
7	6	85	82	67			
8							
9	Anova: Single Factor						
10							
11	SUMMARY						
12	*Groups*	*Count*	*Sum*	*Average*	*Variance*		
13	Ayr	6	474	79	34		
14	Dusseldorf	6	444	74	20		
15	Stockholm	6	396	66	32		
16							
17							
18	ANOVA						
19	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
20	Between Groups	516	2	258	9	0.0027	3.68
21	Within Groups	430	15	28.666667			
22							
23	Total	946	17				
24							

how to test whether the mean stress levels for air traffic controllers are the same for three work stations. The stress level scores shown in Table 13.6 are entered into worksheet rows 2 to 7 of columns B, C and D as shown in Figure 13.8. The cells in rows 2 to 7 of column A contain the number of each controller (1, 2, 3, 4, 5, 6). The steps involved in using EXCEL to produce output corresponding to the ANOVA table shown in Table 13.9, follow.

Step 1 **Data > Data Analysis > Anova: Two-Factor Without Replication**
[Main menu bar]
Click **OK**

Step 2 Enter A1:D7 in **Input Range** box
[Anova: Two-Factor Without Replication panel]
Select **Labels**.
Select **Output Range** and enter A9 in the box
Click **OK**

Figure 13.8 EXCEL solution for the air traffic controller stress test

	A	B	C	D	E	F	G
1	**Controller**	**System A**	**System B**	**System C**			
2	1	15	15	18			
3	2	14	14	14			
4	3	10	11	15			
5	4	13	12	17			
6	5	16	13	16			
7	6	13	13	13			
8							
9	Anova: Two-Factor Without Replication						
10							
11	*SUMMARY*	*Count*	*Sum*	*Average*	*Variance*		
12	1	3	48	16	3		
13	2	3	42	14	0		
14	3	3	36	12	7		
15	4	3	42	14	7		
16	5	3	45	15	3		
17	6	3	39	13	0		
18							
19	System A	6	81	13.5	4.3		
20	System B	6	78	13	2		
21	System C	6	93	15.5	3.5		
22							
23							
24	ANOVA						
25	*urce of Variat*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
26	Rows	30	5	6	3.16	0.0574	3.33
27	Columns	21	2	10.5	5.53	0.0242	4.10
28	Error	19	10	1.9			
29							
30	Total	70	17				
31							

Factorial experiments

In Section 13.7 we showed how analysis of variance could be used to test for the equality of k population means using data from a factorial experiment. To illustrate how EXCEL can be used for this type of experimental design, we show how to analyse the data for the two-factor GMAT experiment introduced in that section. The GMAT scores shown in Table 13.11 are entered into worksheet rows 2 to 7 of columns B, C, and D as shown in Figure 13.9. The steps involved in using EXCEL to produce output shown in cells A10:G45 follows; the ANOVA portion of this output corresponds to the ANOVA table shown in Table 13.14.

Step 1 **Data > Data Analysis > Anova: Two-Factor With Replication** [Main menu bar]

Click **OK**

Figure 13.9 EXCEL solution for the two-factor GMAT experiment

	A	B	C	D	E	F	G
1		**Business**	**Engineering**	**Arts and Sciences**			
2	**3-hour review**	500	540	480			
3		580	460	400			
4	**1-day program**	460	560	420			
5		540	620	480			
6	**10-week course**	560	600	480			
7		600	580	410			
8							
9							
10	Anova: Two-Factor With Replication						
11							
12	SUMMARY	Business	Engineering	Arts and Sciences	Total		
13	*3-hour review*						
14	Count	2	2	2	6		
15	Sum	1080	1000	880	2960		
16	Average	540	500	440	493.33333		
17	Variance	3200	3200	3200	3946.6667		
18							
19	*1-day program*						
20	Count	2	2	2	6		
21	Sum	1000	1180	900	3080		
22	Average	500	590	450	513.33333		
23	Variance	3200	1800	1800	5386.6667		
24							
25	*10-week course*						
26	Count	2	2	2	6		
27	Sum	1160	1180	890	3230		
28	Average	580	590	445	538.33333		
29	Variance	800	200	2450	5936.6667		
30							
31	*Total*						
32	Count	6	6	6			
33	Sum	3240	3360	2670			
34	Average	540	560	445			
35	Variance	2720	3200	1510			
36							
37							
38	ANOVA						
39	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
40	Sample	6100	2	3050	1.38	0.2994	4.26
41	Columns	45300	2	22650	10.27	0.0048	4.26
42	Interaction	11200	4	2800	1.27	0.3503	3.63
43	Within	19850	9	2205.5556			
44							
45	Total	82450	17				
46							

Step 2 Enter A1:D7 in **Input Range** box [**Anova: Two-Factor With Replication** panel]
Enter 2 in **Rows per sample** box
Select **Labels**
Select **Output Range** and enter A10 in the box
Click **OK**

Analysis of variance and experimental design using PASW

Single-factor observational studies and completely randomized designs

To illustrate how PASW can be used to test for the equality of k population means, we show how to test whether the mean examination score is the same at each plant in the National Computer Products example introduced in Section 13.1. First, the data must be entered in a PASW worksheet. In 'Data View' mode, the examination score data are entered into the leftmost column of a PASW worksheet; the six values for Ayr, followed by the six for Dusseldorf and then the six for Stockholm. This is automatically labelled by the system V1. In the adjacent column to the right the code 1 (corresponding to the Ayr plant) is entered six times followed by the code 2 (corresponding to the Dusseldorf plant) six times and the code 3 (corresponding to the Stockhom plant) six times. Thus, the values in the first row of the worksheet are 85, 1; the values in row 2 are 75, 1; the values in row 3 are 82, 1; the values in row 4 are 76, 1; the values in row 5 are 71, 1; the values in row 6 are 85, 1; the values in row 7 are 71, 2 and so on.

The latter variable names can then be changed to score and plant in 'Variable View' mode. The codes used for the plant variable can also be relabelled by following the steps below.

Step 1 **Data > Define Variable Properties** [Main menu bar]

Step 2 Select plant [**Define Variable Properties** panel]
Click on **Continue**
Select plant
Attach the **Value Labels** Ayr to code 1, Dusseldorf to code 2 and Stockholm to code 3.
Click **OK**

The following steps show how PASW generates the ANOVA results shown in Figure 13.4.

Step 1 **Analyze > Compare Means > One-Way ANOVA** [Main menu bar]

Step 2 Enter score in the **Dependent List** box [**One-Way ANOVA** panel]
Enter plant in the **Factor** box
Click on **Options**
Select **Descriptive Statistics**
Click **Continue**
Click **OK**

Randomized block designs

In Section 13.6 we showed how analysis of variance could be used to test for the equality of k population means using data from a randomized block design. To illustrate how PASW can be used for this type of experimental design, we show how to test whether the mean stress levels for air traffic controllers is the same for three work stations. The stress level scores shown in Table 13.6 are entered into the leftmost column of an PASW worksheet. Coding the treatments as 1 for System A, 2 for System B, and 3 for System C, the coded values for the treatments are entered into the adjacent column to the right in the worksheet. Finally, the corresponding number of each controller (1, 2, 3, 4, 5, 6) is entered into the next adjacent column to the right. The columns are automatically labelled by the system V1, V2 and V3 but can be relabelled in Variable View mode as stress, system and controller respectively. Thus, the values in the first row of the worksheet are 15, 1, 1; the values in row 2 are 15, 2, 1; the values in row 3 are 18, 3, 1; the values in row 4 are 14, 1, 2 and so on. The following steps show how PASW generates the ANOVA results shown in Table 13.9.

Step 1 **Analyze > General Linear Model > Univariate** [Main menu bar]

Step 2 [**Univariate** panel]
Enter stress in the **Dependent Variable** box
Enter system and controller in the **Fixed Factors** box
Click on **Model**
Click on **Custom**
Select system and controller
Select Main Effects in **Build Terms** box
Click on **Continue**
Click **OK**

(Note that the F test provided in the output for the controller (blocking) factor can be effectively ignored.)

Factorial experiments

In Section 13.7 we showed how analysis of variance could be used to test for the equality of k population means using data from a factorial experiment. To illustrate how PASW can be used for this type of experimental design, we show how to analyse the data for the two-factor GMAT experiment introduced in that section. The GMAT scores shown in Table 13.11 are entered into leftmost column of a PASW worksheet in Data View mode. Coding the factor A preparation programmes as 1 for the three-hour review, 2 for the one-day programme, and 3 for the ten-week course, the coded values for factor A are entered in Data View mode into the next adjacent column to the right in the worksheet. Coding the factor B colleges as 1 for Business, 2 for Engineering and 3 for Arts and Sciences, the coded values for factor B are entered into the next rightmost column of the worksheet. The columns are automatically labelled by the system, V1, V2 and V3 but can be relabelled in Variable View mode as score, factorA and factorB respectively. (Note that variable names in PASW are not allowed to contain spaces.) Thus, the values in the first row of the worksheet are 500, 1, 1; the values in row 2 are 580, 1, 1; the values in row 3 are 540, 2, 1; the values in row 4 are 460, 2, 1 and so on. The following steps show how to produce the PASW output corresponding to the ANOVA table shown in Figure 13.6.

Step 1 [Main menu bar]
Analyze > General Linear Model > Univariate

Step 2 [**Univariate** panel]
Enter stress in the **Dependent Variable** box
Enter system and controller in the **Fixed Factors** box
Click on **Model**
Click on **Full factorial**
Click on **Continue**
Click **OK**