MAS353 Experimental Design

M. Farrow
University of Newcastle upon Tyne

Semester 2, 2005-6

1 Principles of Experimental Design

1.1 Why plan experiments?

- Planned experiments save time and money and help make the best use of limited resources.

  - Experiment too large? — waste of resources
  - Experiment too small? — waste of time *and* resources

- In a well-planned experiment *practical significance* will be very likely to lead to *statistical significance*.

1.2 Communication

- Experimentation often involves both an *investigator* and a *statistician*. They need to work closely together and communicate well.

- A good experimenter will recognize the importance of statistics in providing a valid interpretation of results. A good statistician will try to understand the context of the experiment and will attempt to explain statistical concepts in non-technical language.

Stages of an experiment

1. The idea.

2. The plan.

3. The execution.

4. The analysis of data.

5. The interpretation.

6. The report.

7. The next idea.

The statistician's contribution

1. Avoidance of bias

2. Choice of design

3. Choice of sample size (to meet stated objectives)

4. Variability reduction techniques

5. Use of controls, choice of treatment levels

6. Data checking (assumptions), detection of "outliers"

7. Analysis of data

Also, valuable cross-disciplinary experience.

1.3 What do we need to know

Before starting to plan an experiment, it is essential to collect certain information. Often, earlier experiments can help. We need to know:

- The expected scale and variablity of the response.

- The size of effect worth detecting: *practical significance.*

- Any factors which may affect the response or limit the validity of the conclusions.

- Any limitations on time, cost, space or other resources.

1.4 Will it do the job?

Finally, we must ensure that the proposed design will be capable of answering the questions posed and meeting the purpose of the experiment. we require:

- Certainty of interpretation (avoidance of *confounding*).

- Avoidance of bias or subjectivity (or even any suspicion of these).

- Simplicity of structure and analysis within the constraints of the investigation.

Often a trial analysis, perhaps using data from a previous or pilot study, can help to check for possible pitfalls.

1.5 Observational studies and planned experiments

In many investigations, e.g. in social sciences, in economics, in ecological field studies, there is no control over the way that data arise. No interventions are made. we simply observe. These are *observational studies.* The standard methods of analysis (ANOVA, regression etc.) are valid but the wider applicability of conclusions is limited.

In experiments, such as laboratory experiments, we can deliberately vary the conditions and observe the changing responses. Such *planned* experiments allow conclusions to be drawn which can not be drawn from observational studies, e.g. about cause and effect.

1.6 Features of planned experiments

Because these methods are used in a wide range of areas of application, a standard terminology is used.

Experimental Unit : The basic set of conditions/materials on which a single element of the experiment is carried out (e.g. a plot of ground, plant pot, patient, blood sample, reaction vessel, piece of material etc.).

Treatment : What we do to the experimental unit. In various situations a treatment can be a type of plant, a watering regime, a drug, a catalyst, a type of fertiliser, a methodology etc.

Trial : A single, independent application of a treatment to an experimental unit.

Response : The outcome of a trial - also known as an observation, or collectively as data.

Block : A set of experimental units with some common property, e.g. neighbouring plots, patients with similar characteristics, time of day, samples from a common source.

Factor : A common, underlying aspect of a set of treatments, e.g. in an experiment to find out how to achieve the best yield in a chemical reaction, we could vary temperature, pressure, duration of reaction, quality of catalyst, etc. Each of these is a factor. A particular set of values for the factors would be a *treatment* (or *treatment combination*).

(Sometimes we use "factor" for a categorical explanatory variable rather than a quantitative explanatory variable).

Nuisance factor : As above but outside the control of the experimenter, e.g. environmental variables.

Replication : A single replication is the application of every treatment (or combination of treatments) once. Most experiments have several replications. To be genuine replications, each replication must be carried out *independently* on distinct sets of experimental units.

False replication : If the replications are not carried out on independent experimental units, we do not have genuine replication and error estimates may be misleading. This can happen if, for example, observations are made on the same unit over a period of time or when repeat measurements are taken to improve accuracy.

Unless care is taken, false replication can lead to spurious sensitivity and an increase in the probability of rejecting the null hypothesis when it is true (Type I error).

1.7 Randomisation

At various stages during an experiment, there are decisions to be made which are essentially arbitrary. these include, for example, which order to carry out the trials and how to allocate a particular treatment to an experimental unit.

It is considered to be a key principle of good design that any arbitrary choice of action should be determined by a process of *randomisation.* This ensures:

1. Objectivity (i.e. lack of subjective bias or even the suspicion of it).

2. Some protection against unexpected variability in the experimental units.

3. Avoidance of false replication.

4. That a valid analysis can be performed.

2 Student's t-test and the One-Way Analysis of Variance (Revision)

2.1 The one-sample t-test

2.1.1 Example 1

The mean amount of a certain impurity in an industrial mineral is supposed to be not greater than 50 ppm. Samples are taken from twenty wagonloads and analysed. The data are as follows (ppm):

| 55.1 | 49.4 | 54.5 | 55.3 | 53.4 | 55.8 | 50.9 | 49.3 | 55.2 | 53.5 |
| 54.9 | 53.1 | 47.8 | 55.3 | 51.9 | 51.7 | 51.6 | 51.7 | 50.3 | 48.2 |

Is there any evidence that the mean impurity concentration is greater than 50 ppm?

2.1.2 Theory

Let us suppose that we can make the following assumptions about the data in the example:

- They are **independent** observations.

- They are taken from a **normal** distribution.

Suppose we call the mean of this normal distribution μ. We can then test the null hypothesis that $\mu = 50$ against the alternative hypothesis $\mu > 50$. (This is a one-sided alternative. The two-sided alternative would be $\mu \neq 50$). We call this a one-sample test because we are comparing the sample mean, $\bar{y}$, from one sample with a theoretical population mean, $\mu_0 = 50$.

We can write our model in this case as

$$Y_i = \mu + \varepsilon_i$$

where Y_i is obervation i and $\varepsilon_i \sim N(0, \sigma^2)$ with ε_i independent of ε_j for $i \neq j$.

To find the significance of the difference we must standardise it by dividing by the standard deviation (called the *standard error*) of $\bar{y}$. The actual standard error of $\bar{y}$ is $\sqrt{\sigma^2/n}$ (where n is the number of observations in the sample, in this case 20, and σ^2 is the variance of the distribution from which the observations were taken) but the value of σ^2 is unknown so we use instead an estimate $\sqrt{s^2/n}$. Because of this the distribution of our test statistic, t, is not normal but *Student's t on $n-1$ degrees of freedom* or t_{n-1} for short. We compare our value for t with figures given in a t-table for the appropriate number of degrees of freedom.

2.1.3 Example 1 continued

We calculate

$$
\begin{aligned}
\sum y &= 1048.9 \\
\sum y^2 &= 55132.33 \\
\bar{y} &= 52.445 \\
s^2 &= 6.4615526
\end{aligned}
$$

To test the null hypothesis $H_0 : \mu = 50$ against the alternative $H_A : \mu > 50$ we calculate the test statistic

$$
t_{19} = \frac{52.445 - 50}{\sqrt{6.4615526/20}} = 4.30.
$$

The upper 0.1 % point of the t_{19} distribution is 3.579 so the result is significant at the 0.1 % level. We therefore reject H_0 and conclude that $\mu > 50$.

We can calculate a 95 % lower confidence bound for μ as

$$
\mu > \bar{y} - 1.729\sqrt{6.4615526/20}.
$$

That is $\mu > 51.5$.

Note that for a two-sided test the 0.1 % critical value is 3.883 so the result would still be significant at the 0.1 % level. A two-sided 95 % confidence interval would be $\bar{y} \pm 2.093\sqrt{s^2/20}$. That is $51.3 < \mu < 53.6$.

2.1.4 General one-sample t-test

Suppose $Y_1, \ldots, Y_n$ are independent and identically distributed observations from the $N(\mu, \sigma^2)$ distribution. That is they are a random sample from $N(\mu, \sigma^2)$. Our model is $Y_i \sim N(\mu, \sigma^2)$ or, equivalently $Y_i = \mu + \varepsilon_i$ where $\varepsilon_i \sim N(0, \sigma^2)$.

Under $H_0 : \mu = \mu_0$,

$$
\frac{\bar{Y} - \mu_0}{\sqrt{S^2/n}} \sim t_{n-1},
$$

where

$$
\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i
$$

and

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

Hence we can test H_0 by comparing the observed value of the test statistic

$$
t = \frac{\bar{y} - \mu_0}{\sqrt{s^2/n}}
$$

with critical points of the t_{n-1} distribution. Similarly we can construct confidence intervals.

2.1.5 Example 2

The amounts of milk in 20 bottles from a dairy are measured and the amounts in pints are such that the sample mean is $\bar{x} = 0.9978$ and the sample variance is $s^2 = 5.34 \times 10^{-4}$. Test the hypothesis that the mean amount is one pint (assuming that the observations are independent and identically distributed and normal).

The test statistic is

$$t_{19} = \frac{0.9978 - 1}{\sqrt{5.34 \times 10^{-4}/20}} = -0.426.$$

For a two-tailed test (against the alternative $\mu \neq 1$) the 5 % point of t_{19} is ± 2.093 so the result is not significant and we do not reject the null hypothesis that $\mu = 1$.

For a one-tailed test against the alternative $\mu < 1$ the 5 % point of t_{19} is -1.729 so again the result is not significant.

2.2 Example 3

In ten tests involving a new type of car the following miles per gallon figures were achieved.

$$30.5 \quad 29.7 \quad 35.3 \quad 29.1 \quad 37.8 \quad 32.3 \quad 39.8 \quad 25.4 \quad 29.8 \quad 30.1$$

Assuming that these may be thought of as independent and identically distributed normal observations, give a symmetric 95 % confidence interval for the population mean miles per gallon.

The summary statistics are $\bar{x} = 31.98$ and $s^2 = 19.27$ on 9 degrees of freedom. The appropriate critical value from the t_9 distribution is 2.262 so our confidence interval is given by

$$\bar{x} - 2.262\sqrt{\frac{19.27}{10}} < \mu < \bar{x} + 2.262\sqrt{\frac{19.27}{10}}.$$

That is

$$28.84 < \mu < 35.12.$$

2.3 The two-sample t-test

2.3.1 Introduction

We have already looked at making inferences about the mean of a normal distribution. Now we turn our attention to the difference between the means of two normal distributions. For example, in an experiment to examine whether a food additive affects the rate of absorption of sugar, we may wish to compare the mean blood sugar concentrations in blood samples from two groups of volunteers where one group has been given the additive and the other has not.

Suppose we have two samples of observations. The n_1 observations, $Y_{11}, \ldots, Y_{n_1 1}$, in the first sample are independent and identically distributed from a $N(\mu_1, \sigma^2)$ distribution. The n_2 observations, $Y_{12}, \ldots, Y_{n_2 2}$, in the second sample are independent and identically distributed from a $N(\mu_2, \sigma^2)$ distribution. The observations in the two sample are independent of each other. The three parameters, μ_1, μ_2 and σ^2, are all unknown. Notice that we are assuming:

- Independence,

- Normality,

- The two distributions have the same variance.

We can use the t distribution to test, for example, the hypothesis that $\mu_1 = \mu_2$.

(For what to do when we can not assume that the variances are equal see, for example, Wetherill, G.B. (1967) "Elementary Statistical Methods", Section 6.6.)

Our model is

$$Y_{ij} = \mu_j + \varepsilon_{ij},$$

where

$$\varepsilon_{ij} \sim N(0, \sigma^2).$$

We can test the null hypothesis $H_0 : \mu_1 = \mu_2$ by comparing $\bar{y}_1 - \bar{y}_2$ with zero. (Here $\bar{y}_1$ and $\bar{y}_2$ are the sample means for the two samples.

Alternatively, we can write

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}.$$

We then impose the constraint that $\alpha_1 + \alpha_2 = 0$ so that we staill have only two independent mean parameters. We call α_1 and α_2 *effects* and we test the null hypothesis that $\alpha_1 = \alpha_2 = 0$.

The actual variance of $\bar{y}_1 - \bar{y}_2$ is

$$\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

where σ^2 is the variance of the normal distributions and n_1 and n_2 are the numbers of observations in the samples. We need an estimator for σ^2. We use the following which pools information from both samples.

$$S^2 = \frac{\sum_{i=1}^{n_1}(Y_{i1} - \bar{Y}_1)^2 + \sum_{i=1}^{n_2}(Y_{i2} - \bar{Y}_2)^2}{n_1 + n_2 - 2},$$

where $\bar{Y}_1$ is the sample mean of the first sample etc. This estimator is unbiassed, it has $n_1 + n_2 - 2$ degrees of freedom and

$$(n_1 + n_2 - 2)\frac{S^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}.$$

To test the hypothesis $H_0 : \mu_1 = \mu_2$ we use the following test statistic:

$$T_{n_1+n_2-2} = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{S^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

which is compared with critical points of the $t_{n_1+n_2-2}$ distribution which it follows under H_0. Observe that $S^2(1/n_1 + 1/n_2)$ is an unbiased estimator of the variance of $\bar{Y}_1 - \bar{Y}_2$.

Similarly we can construct confidence intervals.

2.3.2 Example 4

A new supplier of the industrial mineral in Example 1 (2.1.1) is tried. Fifteen samples are analysed giving the following amounts in ppm.

$$
\begin{array}{cccccccccc}
50.8 & 48.6 & 51.9 & 52.1 & 52.2 & 53.0 & 49.8 & 47.5 & 52.0 & 53.8 \\
53.4 & 54.1 & 49.5 & 50.6 & 49.6 & & & & &
\end{array}
$$

Is there any evidence of a difference in the mean amount of impurity between the two suppliers?

For the new data

$$
\begin{aligned}
\sum y &= 768.9 \\
\sum y^2 &= 39468.33 \\
\bar{y}_2 &= 51.26
\end{aligned}
$$

The value of $\sum(y - \bar{y})^2$ for this new sample is thus

$$39468.33 - (768.9)^2/15 = 54.526.$$

The corresponding value for Sample 1 is 122.7695. We calculate the *pooled* variance estimate

$$s^2 = \frac{122.7695 + 54.516}{33} = 5.3722879.$$

Our test statistic for the test against the two-sided alternative $H_A : \mu_1 \neq \mu_2$ is

$$t_{33} = \left| \frac{52.445 - 51.26}{\sqrt{5.3722879(1/20 + 1/15)}} \right| = 1.497.$$

6

For a 5 % two-sided test the critical value of t_{33} is 2.034 so the result is not significant. We do not reject H_0. There is no strong evidence of a difference.

A 95 % confidence interval for $\mu_1 - \mu_2$ is given by $52.445 - 51.26 \pm 2.034\sqrt{s^2(1/20 + 1/15)}$. That is $-0.43 < \mu_1 - \mu_2 < 2.80$.

(We can also have one-sided tests and confidence intervals).

2.3.3 Example 5

In an experiment to see whether there was evidence that an additive in the lubricant affected the amount of wear on certain bearings, twenty five bearings were subjected to the same amount of use and then weighed. Ten of the bearings were tested with the additive and the remainder without. The results were as follows (weight loss):

With additive	10.77	9.47	10.96	11.02	10.30	11.19	10.24	10.26	11.02	8.88
Without additive	11.28	11.34	9.21	10.53	11.79	11.90	11.58	11.78	11.16	12.04
	11.17	11.34	10.61	10.41	10.92					

Assuming that the observations are independent and approximately normally distributed and that the variance is the same under the two conditions, test the hypothesis that the mean amount of wear is unchanged and give a 95 % confidence interval for the difference between the means under the two conditions.

In sample 1:

$$\text{we have} \quad \sum\nolimits_1 y = 104.11 \quad \text{and} \quad \sum\nolimits_1 y^2 = 1088.9619$$
$$\text{so} \quad \bar{y}_1 = 10.411 \quad \text{and} \quad \sum\nolimits_1 (y - \bar{y})^2 = 5.07269 .$$

In sample 2:

$$\text{we have} \quad \sum\nolimits_2 y = 167.06 \quad \text{and} \quad \sum\nolimits_2 y^2 = 1868.0762$$
$$\text{so} \quad \bar{y}_2 = 11.137 \quad \text{and} \quad \sum\nolimits_2 (y - \bar{y})^2 = 7.47329 .$$

This leads to

$$s^2 = \frac{5.07269 + 7.47329}{23} = 0.54548$$

and hence

$$t_{23} = \frac{10.411 - 11.137}{\sqrt{0.54548\left(\frac{1}{10} + \frac{1}{15}\right)}} = -2.408.$$

Critical points (2-tailed) of t_{23} are ± 2.069 at 5 % and ± 2.807 at 1 % so the result is significant at the 5 % level. At this level we would reject the null hypothesis and conclude that there is evidence that the mean weight loss is less for bearings with the additive.

The confidence interval is:

$$10.411 - 11.137 \pm 2.069 s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

That is

$$-1.350 < \mu_1 - \mu_2 < -0.102.$$

2.4 The paired-sample t-test

2.4.1 Introduction

Sometimes, when we wish to compare the effects of two "treatments" by observing samples of individuals, or experimental units, subjected to the two treatments, each member of one sample can be paired, in some natural way, with a corresponding member of the other sample. In cases where the variation between observations in the same sample can be expected to be greater than the variation between members of the same pair the pairing can be used to increase the power of the test or, equivalently, reduce the number of observations necessary to achieve a given power.

2.4.2 Analysis

Let the first sample be $Y_{11}, \ldots, Y_{n1}$ and the second sample be $Y_{12}, \ldots, Y_{n2}$ so that Y_{i1}, Y_{i2} form a pair for $i = 1, \ldots, n$.

The model is

$$Y_{ij} = \mu_j + \tau_i + \eta_{ij},$$

where $\eta_{ij} \sim N(0, \sigma_j^2)$.

The usual null hypothesis is $H_0 : \mu_1 = \mu_2$.

In a paired sample test we eliminate the between-pair variation by looking at the differences $D_i = Y_{i1} - Y_{i2}$. We now have, in our model:

$$D_i = \mu_1 - \mu_2 + \varepsilon_i,$$

where $\varepsilon_i = (\eta_{i1} - \eta_{i2}) \sim N(0, \sigma^2)$ and $\sigma^2 = \sigma_1^2 + \sigma_2^2$. That is:

$$D_i \sim N(\mu_1 - \mu_2, \sigma^2).$$

We can write the usual null hypothesis as $H_0 : \mu_1 - \mu_2 = 0$ and apply a one-sample t test to the differences.

Alternatively we can write

$$Y_{ij} = \mu + \alpha_j + \tau_i + \eta_{ij},$$

with $\alpha_1 + \alpha_2 = 0$.. Then

$$D_i = Y_{i1} - Y_{i2} = 2\alpha_1 + \varepsilon_i$$

and our null hypothesis is that $\alpha_1 = \alpha_2 = 0$.

2.4.3 When to use the paired test

The following conditions govern the use of the paired sample test:

- The pairing is done before the experiment is carried out, as part of the design of the experiment.

- The pairing is, in some way, "natural" (if possible), for example "before" and "after" on some individual.

- The within-pair variation is expected to be much less than the between-pair variation.

If these conditions are satisfied then the paired-sample test can be much more sensitive than the two-sample test because of the elimination of between-pair variation, even though some information is lost in just looking at the differences.

2.4.4 Example 6

Samples of blood taken from eight patients were used in a test to compare two methods of measuring the amount of a certain blood constituent in the samples. Each sample was divided into two parts which were then allocated at random to the two methods. Assuming that the observations can be taken to be normally distributed with a variance which is constant between patients, test whether there is evidence that one method tends to give greater values than the other.

Patient number	1	2	3	4	5	6	7	8
Method 1	13.6	12.0	12.4	12.8	13.3	12.7	12.2	13.0
Method 2	13.0	11.3	11.6	12.5	12.4	12.2	12.7	11.7

The eight differences (Method 1 - Method 2) are

$$d_i \quad 0.6 \quad 0.7 \quad 0.8 \quad 0.3 \quad 0.9 \quad 0.5 \quad \text{-0.5} \quad 1.3$$

Hence $\sum d_i = 4.6$ and $\sum d_i^2 = 4.58$.

We find the sample mean difference $\bar{d} = 4.6/8 = 0.575$.

We estimate the variance of the differences

$$s_d^2 = \frac{1}{7} \left[4.58 - \frac{(4.6)^2}{8} \right] = 0.276.$$

Our test statistic is

$$t_7 = \frac{\bar{d} - 0}{\sqrt{s_d^2/8}} = \frac{0.575}{\sqrt{0.276/8}} = 3.09.$$

Critical points of t_7 (two-sided) are 2.36 (5 %) and 3.50 (1 %) so our test statistic is significant at the 5 % level so we reject $H_0 : \mu_1 = \mu_2$ and conclude that the first method tends to give greater values.

(Analysis of the same data as a two-sample t test gives $t_{14} = 2.025$ which is not significant).

2.5 Use of Minitab for t-tests

The t-test functions can be found from the `Stat` menu by selecting `Basic Statistics`.

One-sample Use `1-sample t`.

Two-sample Use `2-sample t`. You can either have each of the two samples in a separate column or have all of the observations in one column with the subscript, i.e. the sample number, in a second column. By default, Minitab does an analysis where the variances are *not* assumed equal. To use the equal-variance analysis, select the `Assume variances equal` box.

Paired Use `Paired t`. You can either have the two observations in separate columns of just supply summary statistics. If you just have a column of differences then you can use a one-sample test.

2.6 The One-Way Analysis of Variance

2.6.1 Introduction

In the two-sample t-test we compared the means of two normal samples. In the one-way analysis of variance we can compare the means of several normal samples. The analysis of variance (ANOVA) works by splitting up the variation between the data into components which are assigned to various *sources*. In the one-way ANOVA these sources include differences among the population means and random *errors*.

Suppose we have k samples. Sample j contains n_j independent observations. Our model for Y_{ij}, the ith observation in sample j, is

$$Y_{ij} = \mu_j + \varepsilon_{ij}$$

or, equivalently,

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij}.$$

In each case the errors ε_{ij} are independent and identically distributed $N(0, \sigma^2)$. In the second parameterisation μ is an overall mean and τ_j is a *treatment effect*. In order not to have too many parameters we apply the constraint

$$\sum_{j=1}^{k} \tau_j = 0.$$

(Alternatively we may make μ represent the mean for the first population and set $\tau_1 = 0$. In this case $\tau_2, \ldots, \tau_k$ represent differences of the other population means from the first). Notice that the *error variance* σ^2 is assumed to be the same in all populations.

We thus have three key assumptions:

- normality,

- independence,

- equal variances.

The usual null hypothesis is $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$ or, equivalently, $H_0 : \tau_1 = \tau_2 = \cdots = \tau_k = 0$.

2.6.2 Analysis

Whether or not H_0 is true, we can estimate σ^2 by an estimator which pools information from all samples:

$$s^2 = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2}{\sum_{j=1}^{k} (n_j - 1)}$$

on

$$\sum_{j=1}^{k} (n_j - 1) = N - k$$

degrees of freedom, where N is the total number of observations and $\bar{Y}_j$ is the sample mean for sample j. (Compare with the two-sample t-test).

If H_0 is true then $\bar{Y}_j \sim N(\mu, \sigma^2/n_j)$ so

$$B = \frac{\sum_{j=1}^{k} n_j (\bar{Y}_j - \bar{Y})^2}{k - 1},$$

where $\bar{Y}$ is the overall mean of the data, gives an unbiassed estimator of σ^2 on $k - 1$ degrees of freedom.

Now $(N - k)s^2/\sigma^2 \sim \chi^2_{N-k}$ and, when H_0 is true, $(k - 1)B/\sigma^2 \sim \chi^2_{k-1}$. We call $(N - k)s^2$ the *within samples sum of squares* or the *residual sum of squares*. We call $(k - 1)B$ the *between samples sum of squares*. Moreover B and s^2 are independent so, when H_0 is true,

$$\frac{B}{s^2} \sim F_{k-1,N-k}.$$

When H_0 is false B is increased by the variation between the population means while s^2 remains unaffected. Therefore a large value of B/s^2, when compared with critical points of $F_{k-1,N-k}$, would lead to rejection of the null hypothesis.

Now

$$
\begin{aligned}
(N - k)s^2 + (k - 1)B &= \sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2 + \sum_{j=1}^{k} \sum_{i=1}^{n_j} (\bar{Y}_j - \bar{Y})^2 \\
&= \sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j + \bar{Y}_j - \bar{Y})^2 - 2 \sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)(\bar{Y}_j - \bar{Y}) \\
&= \sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y})^2
\end{aligned}
$$

This is called the *total sum of squares.*

We usually calculate the total sum of squares

$$\sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y})^2 = \sum_{j=1}^{k} \sum_{i=1}^{n_j} Y_{ij}^2 - G^2/N,$$

where G is the grand total

$$G = \sum_{j=1}^{k} \sum_{i=1}^{n_j} Y_{ij},$$

10

and the between samples sum of squares

$$\sum_{j=1}^{k} n_j (\bar{Y}_j - \bar{Y})^2 = \sum_{j=1}^{k} T_j^2 / n_j - G^2 / N.$$

where T_j is the sample total

$$T_j = \sum_{i=1}^{n_j} Y_{ij},$$

and find the residual sum of squares by subtraction.

We present the results in an analysis of variance table.

Source of variation	Degrees of freedom	Sum of squares	Mean square	Variance ratio
Between samples	$k-1$	$\sum_{j=1}^{k} T_j^2 / n_j - G^2 / N$	B	B/s^2
Residual	$N-k$	(by subtraction)	s^2	
Total	$N-1$	$\sum_{j=1}^{k} \sum_{i=1}^{n_j} Y_{ij}^2 - G^2 / N$		

The sums of squares are divided by the degrees of freedom to obtain the mean squares. The variance ratio test statistic is compared with critical points of $F_{k-1,N-k}$.

Note that subtracting a constant from every observation will not affect either B or s^2 and multiplying every observation by a constant will not affect the ratio B/s^2. This can be useful to simplify calculations.

2.6.3 Example

The following data are from an experiment to compare the effects of four catalysts on the yield of a chemical process.

Catalyst	observations					n_j	$T_j = \sum_{i=1}^{n_j} y_{ij}$	$\sum_{i=1}^{n_j} y_{ij}^2$
A	4.85	4.89	5.60	5.08	3.42	5	23.84	116.2974
B	9.43	6.77	5.39	7.39		4	28.98	218.4220
C	4.67	5.49	5.96	3.71	5.51	5	25.34	131.5948
D	5.50	4.90	6.00	5.81	4.68	5	26.89	145.9185
						$N = 19$	$G = 105.05$	612.2327

$$\frac{G^2}{N} = 580.81592$$

- Total sum of squares $612.2327 - 580.81592 = 31.417$.

- Between samples sum of squares $(23.84^2/5) + (28.98^2/4) + (25.34^2/5) + (26.89^2/5) - 580.81592 = 15.851$.

- Residual sum of squares $31.417 - 15.851 = 15.566$.

Source of variation	Degrees of freedom	Sum of squares	Mean square	Variance ratio
Between catalysts	3	15.851	5.284	5.09
Residual	15	15.566	1.038	
Total	18	31.417		

The upper 5% point of $F_{3,15}$ is 3.287 so 5.09 is significant at the 5% level and, at that level, we would reject the hypothesis that the population means are the same for all four catalysts.

2.6.4 Minitab

You can find the ANOVA functions by slecting `ANOVA` from the `Stat` menu. There are two functions specifically for the one-way ANOVA.

- `One-Way`: You need two columns of data. One contains the response Y and the other contains the level j of the factor.

- `One-Way (Unstacked)`: In this case each sample is in a separate column.

Problems 1

2.7 Problems

1. The actual voltages produced by ten nominally 6-volt batteries were measured. The observations were as follows.

$$
\begin{array}{ccccc}
5.81 & 6.33 & 6.15 & 6.07 & 6.10 \\
6.44 & 6.14 & 6.03 & 5.88 & 6.06
\end{array}
$$

Assuming a normal model, give a symmetric 95% confidence interval for the true mean voltage. Are the data consistent with the nominal voltage?

2. Ten observations on the time (in minutes) taken to perform a certain task gave the following data.

$$
\begin{array}{ccccc}
13.5 & 19.0 & 18.0 & 14.0 & 17.0 \\
20.5 & 13.5 & 15.5 & 19.0 & 15.0
\end{array}
$$

Assuming normality, test the hypothesis that the population mean time is 15 minutes.

Give a 95% confidence interval for the population mean.

3. Measurements were taken on various characteristics of two populations of the common starling *(sturnus vulgaris)*, the resident birds and continental migrants. They include the following on wing length in cm.

Residents	10.7	10.9	11.4	10.9	11.1	11.7
	11.0	10.9	11.1	10.6	10.7	
Migrants	11.0	11.8	11.9	11.2	10.9	11.5
	11.8	12.0	11.6			

Assuming normality and that the population variances are equal, test the hypothesis that the mean wing lengths for the two populations are equal. Give a 95% confidence interval for the difference in mean wing length for the two populations.

4. The marks attained by a class of twelve students on two assignments were as follows. (The marks are real but anonymised).

Student	1	2	3	4	5	6	7	8	9	10	11	12
Assignment 1	50	70	69	64	63	64	69	58	54	70	17	56
Assignment 2	61	62	54	61	61	60	64	51	53	71	33	50

Assuming normality, test the hypothesis that the population mean mark for students is the same in each assignment. Give a 95% confidence interval for the difference in population mean marks between the assignments.

5. The blood concentrations of C-peptide in a sample of fifteen elderly diabetic subjects before, and thirty minutes after, being given some glucose were as follows.

Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Before	0.5	2.6	6.0	0.9	6.0	2.2	5.4	1.6	2.1	0.8	2.6	2.0	0.3	1.5	0.9
After	0.8	2.7	6.1	1.0	6.6	2.6	5.9	1.6	2.4	0.9	3.8	2.7	0.4	1.8	1.0

Assuming normality, give a 95% confidence interval for the population mean increase in C-peptide concentration and test the hypothesis that this mean is zero.

6. The following data come from an experiment to study the gain in weight of rats fed on four different diets, distinguished by amount of protein (low and high) and by source of protein (beef and cereal). Each diet was fed to a separate sample of ten rats.

Use a one-way analysis of variance to test the hypothesis that the mean weight gain is the same for each diet.

Protein source	Beef	Beef	Cereal	Cereal
Protein amount	Low	High	Low	High
	90	73	107	98
	76	102	95	74
	90	118	97	56
	64	104	80	111
	86	81	98	95
	51	107	74	88
	72	100	74	82
	90	87	67	77
	95	117	89	86
	78	111	58	92

Assignment 1

Please submit solutions to questions 1, 3, 5, 6, in the homework box, by 5.00pm on Wednesday February 8th.

3 Completely Randomised Design

3.1 The design

The completely randomised design is very simple. It is also known as the *one-way classification*.

We have N independent and identical (interchangeable) experimental units. We wish to compare k different treatments. Very often we want the sample sizes to be as nearly equal as possible (i.e. $n_j \approx N/k$), but this is not essential and there may be good reasons not to do this.

We choose $n_1, n_2, \ldots, n_k$ with

$$\sum_{j=1}^{k} n_j = N.$$

Because allocation of the treatments to units is arbitrary, we use randomisation.

A single observation is taken on each unit and (subject to appropriate distributional assumptions) we use the one-way ANOVA to analyse the resulting data. Multiple comparison methods may also be used. (See later).

3.2 Example

In an experiment on the effect of various treatments on the growth of plants, ten plants were randomly assigned to each of four treatments, as follows.

A Control.

B Addition of dilute boric acid to irrigation water.

C Addition of nutrient supplement to irrigation water.

D Addition of both dilute boric acid and nutrient supplement to irrigation water.

After 20 weeks, the heights of the plants were measured as follows, in cm.

Treatment A		Treatment B		Treatment C		Treatment D	
23.8	25.3	26.8	27.9	21.1	23.3	25.9	26.8
24.4	25.5	27.3	26.8	21.8	24.6	26.3	21.2
24.5	20.3	25.5	27.0	20.0	25.5	24.2	25.3
22.7	24.4	27.0	26.1	25.3	25.3	24.1	27.2
25.1	24.8	27.0	24.7	22.9	22.8	27.3	25.9

3.3 The design matrix

We can write the model as

$$Y_{ij} = \mu_j + \varepsilon_{ij}$$

where Y_{ij} is observation i of sample j and $\varepsilon_{ij} \sim N(0, \sigma^2)$ with ε_{ij} independent of $\varepsilon_{i'j'}$ unless $i = i'$ and $j = j'$.

Alternatively, we can number all of the observations consecutively with a single subscript i. In the example $Y_i, \ldots, Y_{10}$ receive treatment A, $Y_{11}, \ldots, Y_{20}$ receive treatment B and so on. We can then write

$$Y_i = x_{i1}\mu_1 + x_{i2}\mu_2 + \cdots + x_{ip}\mu_p + \varepsilon_i$$

if there are p treatments, where $x_{ij} = 1$ if observation i receives treatment j and $x_{ij} = 0$ otherwise. In the example there are four treatments so

$$Y_i = x_{i1}\mu_1 + x_{i2}\mu_2 + x_{i3}\mu_3 + x_{i4}\mu_4 + \varepsilon_i.$$

Now we can collect the observations into a vector $\underline{Y} = (Y_1, Y_2, \ldots, Y_n)^T$ and collect the errors into a vector $\underline{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^T$. In the example, $n = 40$. We can also collect the parameters into a vector $\underline{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)^T$. In the example $p = 4$. Finally we can arrange the coefficients x_{ij}

into a matrix X in which row i corresponds to the observation Y_i and column j corresponds to the parameter μ_j. We can then write the model as

$$\underline{Y} = X\underline{\mu} + \underline{\varepsilon}.$$

The matrix X is called the *design matrix*.

In the example, the design matrix is as follows. (To save space only three rows are written in full for each sample).

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

It is sometimes convenient to write the model in a different way. For example, we could write

$$Y_i = \mu + x_{i2}\alpha_2 + \cdots + x_{ip}\alpha_p + \varepsilon_i$$

where $\mu = \mu_1$, and, for $j = 2, \ldots, p$, $\alpha_j = \mu_j - \mu_1$. Then we can write

$$\underline{Y} = \tilde{X}\underline{\beta} + \underline{\varepsilon}$$

where $\underline{\beta} = (\mu, \alpha_2, \ldots, \alpha_p)^T$. In the example we have

$$\tilde{X} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

3.4 Analysis

The analysis of the results of the experiment is just the usual one-way analysis of variance. This is easily done by hand or using Minitab.

After fitting the model, the assumptions can be checked. Minitab provides some graphs which can be requested and used for this purpose.

Histogram of the residuals : This can be used to check the normality of the residuals.

Normal probability plot of the residuals : This can also be used to check the normality of the residuals.

Plot of residuals against fitted values : The fitted values are just the sample means. The variance of the residuals should be the same for each treatment. The sample variances will not be exactly the same, of course.

Examination of these plots might suggest the need for transformation of the response variable.

We can make a confidence interval for the mean of a treatment using

$$\bar{y}_j \pm t_\nu \sqrt{s^2/n_j}$$

where $\bar{y}_j$ is the sample mean for the treatment in question, t_ν is a suitable critical value of the t-distribution on ν degrees of freedom, the number of degrees of freedom, ν, is the number of residual degrees of freedom in the analysis of variance, s^2 is the residual mean square and n_j is the number of observations in sample j.

Similarly a confidence interval for the difference between two means would have the form

$$\bar{y}_j - \bar{y}_k \pm t_\nu \sqrt{s^2 \left(\frac{1}{n_j} + \frac{1}{n_k} \right)}.$$

In the example, the analysis is as follows.

4 Multiple Comparisons

4.1 Introduction

In the one-way analysis of variance we get an overall test of the null hypothesis that $\mu_1 = \mu_2 = \cdots = \mu_k$, where μ_j is the mean for population j. If we reject this hypothesis and conclude that there is at least one difference, it is natural to want to go a bit further and say something about which means differ from which others.

There are difficulties with this, however:

Multiple testing : If we carry out several pairwise comparisons then, even if the null hypothesis is true, the probability that we get at least one "significant" result is greater than the apparent significance level of the tests.

Post hoc testing : If we carry out tests suggested by the data themselves then the significance level of this procedure, seen as a whole, is not the apparent significance level of the test. For example, suppose that we have k samples and we decide to compare the two samples with the largest and smallest means. Clearly the distribution of the difference between the largest and smallest means in a collection of k means is not the same as the distribution of the difference between the means of two pre-specified samples.

The *post-hoc* problem does not arise if we decide on the tests in advance. Such pre-planned tests usually use *contrasts*. (See Lecture 5).

4.2 Studentised range: Tukey's test

This test allows for the distribution of the range of a set of k sample means, for equally sized samples of size n. Let s^2 be an estimate of the error variance, on ν degrees of freedom (usually the residual mean square from the Anova table).

The studentized range, $Q_{k,\nu}$, is a generalisation of Student's t_ν. Tables of critical values $q_{k,\nu}(\alpha\%)$ can be found in Biometrika tables. An abridged table is provided as a handout. If the range of the sample means exceeds

$$R = q_{k,\nu}(\alpha\%) \frac{s}{\sqrt{n}}$$

then we have a result, significant at the $\alpha\%$ level, indicating that differences exist between the population means. In fact, *any* pair of sample means differing by more than this amount can be taken as a significant result at the $\alpha\%$ level indicating that the corresponding population means are different.

Example (3.2)

4.3 Least significant difference

The least significant difference (LSD) is simply based on the two sample t-test. It is

$$L = t_\nu(\alpha\%)s\sqrt{\frac{2}{n}}.$$

The LSD does not take into account the number of comparisons to be made and, unlike Tukey's test, it does not imply significance. It does, however, give a lower bound to the difference between sample means required for significance. Any pair of means which do not differ by more than this amount will not correspond to a significant difference.

Example (3.2)

5 Contrasts

5.1 Introduction

In section 3.3 we saw an example of *reparameterising*. That is we wrote the model in terms of a different, though equivalent set of parameters. This is often useful. In particular we can reparameterise using *contrasts* and this allows us to test for differences between particular means or groups of means. That is we can *contrast* these means.

Consider a one-way ANOVA where the sample size in sample j is $n_j = n$ for all j. (We can also form contrasts when the n_j are not all equal but this is a little more complicated). Let the number of treatments be k and the total sample size be $N = kn$.

For example, suppose there are $k = 5$ treatments. We can use the parameters

$$\underline{\beta}^T = (\mu, \beta_2, \beta_3, \beta_4, \beta_5).$$

The design matrix might be as follows. (It is written as the transpose to save space).

$$X^T = \begin{pmatrix} 1 & 1 & \ldots & 1 & 1 & 1 & \ldots & 1 & 1 & 1 & \ldots & 1 & 1 & 1 & \ldots & 1 & 1 & 1 & \ldots & 1 \\ 1 & 1 & \ldots & 1 & -1 & -1 & \ldots & -1 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 1 & 1 & \ldots & 1 & -1 & -1 & \ldots & -1 & 0 & 0 & \ldots & 0 \\ 1 & 1 & \ldots & 1 & 1 & 1 & \ldots & 1 & -1 & -1 & \ldots & -1 & -1 & -1 & \ldots & -1 & 0 & 0 & \ldots & 0 \\ 1 & 1 & \ldots & 1 & 1 & 1 & \ldots & 1 & 1 & 1 & \ldots & 1 & 1 & 1 & \ldots & 1 & -4 & -4 & \ldots & -4 \end{pmatrix}.$$

- Columns 2 to $k = 5$ of X here are examples of *orthogonal contrasts.*

- Each is a *contrast* because the sum of the elements in each case is zero.

- The contrasts are orthogonal because each column $\underline{c}_j$ of X is orthogonal to every other column. That is $\underline{c}_j^T \underline{c}_l = 0$ for $j \neq l$. Thus

$$
X^T X = \begin{pmatrix}
N & 0 & 0 & 0 & 0 \\
0 & \sum c_{i2}^2 & 0 & 0 & 0 \\
0 & 0 & \sum c_{i3}^2 & 0 & 0 \\
0 & 0 & 0 & \sum c_{i4}^2 & 0 \\
0 & 0 & 0 & 0 & \sum c_{i5}^2
\end{pmatrix}
$$

where $\underline{c}_j^T = (c_{1j}, \ldots, c_{nj})$.

Notice that, for example, the means for observations in samples 1 and 2 are

$$
\begin{aligned}
\mu_1 &= \mu + \beta_2 + \beta_4 + \beta_5 \\
\mu_1 &= \mu - \beta_2 + \beta_4 + \beta_5
\end{aligned}
$$

So $\mu_1 - \mu_2 = 2\beta_2$ and the hypothesis that $\mu_1 = \mu_2$ is equivalent to the hypothesis that $\beta_2 = 0$. So we can test $H_0 : \ \beta_2 = 0$ to compare the means for treatments 1 and 2.

Each contrast has a *contrast sum of squares*. Because the contrasts are orthogonal we can split the between-treatments sum of squares into $k - 1$ independent components, each on one degree of freedom, and test each separately if we wish. For example, suppose that $n = 3$, $k = 4$ and the first two samples correspond to one group of treatments and the other two to another. We can contrast the two groups as follows.

$$
X^T = \begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 \\
1 & 1 & 1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & -1 & -1 & -1
\end{pmatrix}.
$$

Here

- $\underline{c}_1$ is for μ,

- $\underline{c}_2$ contrasts the two groups of treatments,

- $\underline{c}_3$ and $\underline{c}_4$ are two other contrasts, orthogonal to $\underline{c}_2$, for the remaining between-treatments variation (on 2 d.f.), that is within groups.

Source of variation	D.f.
Group 1 v Group 2	1
Within group 1,2	2
Between treatments	3
Residual	8
Total	11

The sum of squares associated with $\underline{c}_j$ is easily shown to be

$$
\frac{(\sum c_{ij} y_i)^2}{\sum c_{ij}^2}
$$

on one degree of freedom. We can divide this by the residual mean square to obtain a variance ratio to test the contrast.

Usually, $c_{1j} = c_{2j} = \cdots = c_{nj} = c_j$. In this case we can write the contrast sum of squares as

$$
\frac{(\sum c_j T_j)^2}{n \sum c_j^2}
$$

where T_j is the total for sample j.

The two-way ANOVA etc. can also be derived using these ideas.

5.2   Example (3.2)

5.3   Post-experiment contrasts: Scheffé's method

If the contrasts are decided after the experiment, that is they are suggested by the data, then we can use an adjusted test which makes allowance for this. We compare

$$\frac{(\sum c_j T_j)^2}{n(k-1)s^2 \sum c_j^2}$$

with critical values of $F_{k-1,\nu}$, where k is the number of samples and s^2 is an estimate of the error variance on ν degrees of freedom. Usually s^2 would be the residual mean square from the ANOVA table.

Scheffé's test is more widely applicable than Tukey's test (see 4.2) but is less efficient if used purely for paired comparisons. It is a much weaker test than that for pre-experiment contrasts described in 5.1 above.

Example (3.2)

6 The Two-Way Analysis of Variance

6.1 Introduction

In the one-way ANOVA we look for the effect of one "factor". In the two-way ANOVA we look for the effects of two factors. The second factor may also be of interest or it may be a "nuisance factor". (Compare with the paired sample t-test).

A further complication arises because of the possibility of "interactions" between the factors. The different values of a factor are called "levels". Both the "one-way ANOVA" and "two-way ANOVA" refer to examples of *factorial experiments*. A factorial experiment is one where all of the explanatory variables are qualitative factors rather than quantitative variables.

Suppose we have two factors and Factor 1 has five levels while Factor 2 has six levels. There are thus 30 means in the model, one for each combination of a level of Factor 1 and a level of Factor 2. Let Y_{ijk} be observation i of the observations made at level j of Factor 1 and level k of Factor 2. Then

$$Y_{ijk} = \mu_{jk} + \varepsilon_{ijk}.$$

Sometimes we make a simplifying assumption and write

$$\mu_{jk} = \alpha_j + \beta_k.$$

This is called an *additive* model. The assumption is that there are no *interactions* between the factors. So the mean for, say, the combination of level 2 of Factor 1 with level 4 of Factor 2 could be written $\alpha_2 + \beta_4$. We would appear to need only $5 + 6 = 11$ parameters, $\alpha_1, \ldots, \alpha_5$ and $\beta_1, \ldots, \beta_6$, to specify all 30 means. In fact we only need 10. This is because we could add a fixed amount δ to each α_j and subtract the same amount from each β_k and still have the same μ_{jk}. So, if we have 11 parameters, they are not well defined and, without loss of generality, we can reduce the number of parameters to 10. We could do this, for example, by fixing one of the β_k to be 0. In fact, instead, we often constrain the β_k to sum to zero. Furthermore we constrain the α_j to sum to zero as well and add an extra parameter μ in compensation. Thus $\mu_{jk} = \mu + \alpha_j + \beta_k$ and

$$Y_{ijk} = \mu + \alpha_j + \beta_k + \varepsilon_{ijk}$$

with

$$\sum_{j=1}^{5} \alpha_j = \sum_{k=1}^{6} \beta_k = 0.$$

If we allowed interaction between the factors then we would not impose this simplifying structure and each of the 30 combinations of levels of the factors could have a separate parameter for its mean. In fact we tend to write the model in terms of other parameters. See 6.3.

6.2 Single replication

Suppose factor 1 has t_1 levels and factor 2 has t_2 levels.

Suppose first that there is only one observation per treatment combination (i.e. one observation in each *cell*). We say that there is a single *replication*. Suppose that we can assume additivity. Then the model is

$$y_{jk} = \mu + \alpha_j + \beta_k + \varepsilon_{jk}$$

where
y_{jk}	is the observation at level j of factor 1 and level k of factor 2
μ	is an overall mean or constant term
α_j	is a factor 1 effect (for level j)
β_k	is a factor 2 effect (for level k)
ε_{jk}	is the random "error" such that $\varepsilon_{jk} \sim N(0, \sigma^2)$

In order to avoid over-parameterising we need to put some restrictions on the parameters.

$$\sum_{j=1}^{t_1} \alpha_j = \sum_{k=1}^{t_2} \beta_k = 0.$$

We can now calculate sums of squares as follows. Let

$$A_j = \sum_{k=1}^{t_2} y_{jk}$$

be the total for level j of factor 1, formed from t_2 observations,

$$B_k = \sum_{j=1}^{t_1} y_{jk}$$

be the total for level k of factor 2, formed from t_1 observations, and

$$G = \sum_{j=1}^{t_1} \sum_{k=1}^{t_2} y_{jk}$$

be the overall total, formed from $t_1 t_2$ observations.

Then let

$$
\begin{aligned}
S_T &= \sum_{j=1}^{t_1} \sum_{k=1}^{t_2} y_{jk}^2 - G^2/(t_1 t_2) \\
S_1 &= \sum_{j=1}^{t_1} \{A_j^2\}/t_2 - G^2/(t_1 t_2) \\
S_2 &= \sum_{k=1}^{t_2} \{B_k^2\}/t_1 - G^2/(t_1 t_2) \\
S_R &= S_T - S_1 - S_2
\end{aligned}
$$

Here
S_T is the total sum of squares on $t_1 t_2 - 1$ degrees of freedom
S_1 is the factor 1 sum of squares on $t_1 - 1$ degrees of freedom
S_2 is the factor 2 sum of squares on $t_2 - 1$ degrees of freedom
S_R is the residual sum of squares on $(t_1 - 1)(t_2 - 1)$ degrees of freedom.
No degrees of freedom are left for an interaction sum of squares.

6.3   Multiple replications

Suppose now that we have more than one observation per cell. Provided that there is the same number of observations in each cell then the calculations are again straightforward. If we wish to allow the possibility of an interaction then we adopt the following model.

$$y_{ijk} = \mu + \alpha_j + \beta_k + \gamma_{jk} + \varepsilon_{ijk}$$

where

- y_{ijk} is observation i on level j of factor 1 and level k of factor 2,

- α_j is the *effect* of level j of factor 1,

- β_k is the *effect* of level k of factor 2,

- γ_{jk} is an *interaction effect*.

We impose the additional constraints

$$\sum_{j=1}^{t_1} \gamma_{jk} = 0$$

for all k and

$$\sum_{k=1}^{t_2} \gamma_{jk} = 0$$

for all j.

(If we assume additivity then we simply omit the interaction terms and the analysis is similar to that for the one-observation-per-cell case above, in an obvious way).

6.4   Example

Suppose we have 4 levels of factor 1, 3 levels of factor 2 and 2 *replications* in each cell so that there are 24 observations.

Total sum of squares	$S_T = \sum\sum\sum y_{ijk}^2 - G^2/24$
Total between treatments sum of squares	$S_F = (1/2)\{T_{11}^2 + \cdots + T_{43}^2\} - G^2/24$
Factor 1 sum of squares	$S_1 = (1/6)\{A_1^2 + A_2^2 + A_3^2 + A_4^2\} - G^2/24$
Factor 2 sum of squares	$S_2 = (1/8)\{B_1^2 + B_2^2 + B_3^2\} - G^2/24$
Interaction sum of squares	$S_I = S_F - S_1 - S_2$
Residual sum of squares	$S_R = S_T - S_F$

The total for the observations at level j of factor 1 and level k of factor 2 is

$$T_{jk} = y_{1jk} + y_{2jk}.$$

The total for level j of factor 1 is

$$A_j = \sum_{k=1}^{3}\sum_{i=1}^{2} y_{ijk}.$$

The total for level k of factor 2 is

$$B_k = \sum_{j=1}^{4}\sum_{i=1}^{2} y_{ijk}.$$

Source of variation	Degrees of freedom	Sum of squares	Mean square	Variance ratio
Factor 1	3	S_1		
Factor 2	2	S_2		
Interaction	6	S_I		
Total between treatments	11	S_F		
Residual	12	S_R		
Total	23	S_T		

If the interaction variance ratio is significant then compare the main effects (factor 1, factor 2) mean squares with the interaction mean square to see which is the biggest effect. If the interaction variance ratio is not significant then test factors individually in the usual way.

6.5   Minitab

You can do a two-way ANOVA using Minitab as follows. Put all of the response (y) values in a single column. Put the levels of factor 1 in another column and the levels of factor 2 in another. It is often useful in doing this to use select `Make Patterned Data` then `Simple Set of Numbers` from the `Calc` menu. Then, form the `Stat` menu, selct `ANOVA` then `Two-Way` .

6.6 Numerical example

See Problems 1, Question 6.

See Problems 1, Question 6.

Problems 2

1. A measure of the yield of a certain industrial electrolytic process was observed for one week's operation of 30 electrolytic cells. Fifteen of the cells were operated using the standard control scheme (A) while eight of the others were operated under a modified control scheme (B) and the remaining seven were operated under a third control scheme (C). The results were as follows.

Control Scheme	Yields							
A	81.4	82.8	80.9	84.6	83.0	81.3	82.9	80.8
	80.3	86.4	81.5	79.2	83.4	81.6	84.9	
B	79.8	85.3	84.8	86.0	81.2	82.9	81.9	85.4
C	85.0	80.3	81.0	79.5	82.1	80.9	81.5	

Making the usual assumptions for an analysis of variance, test whether there is evidence in the results to suggest that there are differences between the mean yields given by these three control schemes.

Use a suitable contrast to compare the standard scheme (A) with the new schemes (B,C) and hence also compare B with C.

2. The following data (from R. Till, 1974, "Statistical Methods for the Earth Scientist") are salinity values (parts per thousand) for the three separate water masses in the Bimini Lagoon, Bahamas.

Water Mass 1			Water Mass 2			Water Mass 3		
37.54	37.01	36.71	40.17	40.80	39.76	39.04	39.21	39.05
37.03	37.32	37.01	39.70	40.79	40.44	38.24	38.53	38.71
37.03	37.70	37.36	39.79	39.38		38.89	38.66	38.51
36.75	37.65	38.85				40.08		

Test the hypothesis that the mean salinity is the same in all three water masses.

3. The following data (from Amess et al., 1978) are the red cell folate levels (μg/l) in three groups of cardiac bypass patients given different levels of nitrous oxide ventilation.

Group 1	Group 2	Group 3
243	206	241
251	210	258
275	226	270
291	249	293
347	255	328
354	273	
380	285	
392	295	
	309	

(a) Compare the three using an analysis of variance.

(b) Give a 95% confidence interval for the population mean for treatment 1.

(c) Give a 95% confidence interval for the difference in population means between group 1 and group 2.

4. The following table (from Till) gives the strontium content (% Sr) of a set of carbonate sediments obtained by three different analytical methods:

- I – flame photometry.
- II – spectrographic analysis.

- III – atomic absorption.

Sediment Specimen no.	Method I	Method II	Method III
1	0.96	0.94	0.98
2	0.96	0.98	1.01
3	0.85	0.87	0.86
4	0.86	0.84	0.90
5	0.86	0.87	0.89
6	0.89	0.93	0.92

(a) Analyse these data using an analysis of variance.

(b) Compare the first two methods using a suitable contrast. Comment on how this comparison differs from a t-test.

5. The following data are from an experiment to compare the effects of four catalysts on the yield of a chemical process.

Catalyst	observations				
A	4.85	4.89	5.60	5.08	3.42
B	9.43	6.77	5.39	7.39	6.98
C	4.67	5.49	5.96	3.71	5.51
D	5.50	4.90	6.00	5.81	4.68

(a) Analyse the data using and analysis of variance to test for differences between the effects of the catalysts.

(b) Compare the sample means using the Studentized range.

(c) Compare the sample means used the method of least significant difference.

References:

Amess, J.A.L., Burman, J.F., Rees, G.M. et al. (1978), Megaloblastic haemopoiesis in patients receiving nitrous oxide. *Lancet*, **ii**, 1483-6.

Till, R. (1974), "Statistical Methods for the Earth Scientist".

Assignment 2

Please submit solutions to questions 1, 3, 5 in the homework box, by 5.00pm on Wednesday February 22nd.

7 Randomized Blocks I

7.1 Blocking

Often we can not assume that all experimental units are identical. There may be a "nuisance factor," that is some factor or variable which is associated with the response and where we can not avoid differences between units but which is not of interest in itself. There are two approaches to this:

Regression approach : Measure an auxiliary variable which is known to be correlated with the response variable. Use this to adjust the responses. This leads to the "analysis of covariance."

Design approach : Group the units into relatively homogeneous blocks. Plan the experiment to balance the treatments across blocks and hence cancel out the "nuisance factor." This is "blocking."

7.2 Randomised blocks design

Suppose that there are not enough homogeneous experimental units for a complete experiment but we can group the units into relatively homogeneous "blocks" – e.g. litters of mice. The variation between blocks can be eliminated and, if this is large compared to the between-treatments variation, this will lead to a more powerful test.

Each treatment occurs once within each block. Within the blocks the treatments are allocated at random to the experimental units. This is an example of "restricted randomisation."

E.g. 4 treatments, A, B, C, D, 3 blocks.

This design requires b blocks of t units, where t is the number of treatments. This can often be achieved easily when the material is "continuous" (batches of liquid or powder, strips of material, time units etc.). There will be more of a constraint with "discrete" material and other designs (e.g. balanced incomplete blocks) may be needed.

We can think of the usual situation in which a paired-sample t-test is used as a randomized blocks experiment in which there are just two treatments and each treatment is applied once within each block.

7.3 Illustrative examples

Medicine : Match patients according to medical history, age, sex etc.

Split blood samples into t subsamples for analysis.

Multi-centre trial where experiment takes place at several hospitals.

Agriculture : Divide a field into zones of fertility. Split each zone into t plots.

Compare dietary additives on pigs from different litters.

Industry : Blocks might be, for example, machines or operators.

Paint traffic signs using t types of paint and expose these at b locations.

7.4   Pros and cons

Disadvantages of randomised blocks designs:

1. Limitation of block size to t units can cause problems when the natural block size is $k \neq t$.

2. Estimation of block effects uses up a few degrees of freedom and reduces the number of residual degrees of freedom. This might lead to a reduction in sensitivity in some cases.

Advantages of randomised blocks designs:

1. Variability in units caused by the blocking factor is identified and removed, often resulting in large gains in sensitivity.

2. Opportunity to vary the experimental units *deliberately* and allow a wider comparison of the treatments and greater generality of the results.

7.5   Model

The model associated with the randomised blocks design is

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

where

- y_{ij} is the response to treatment i in block j.

- μ is the overall mean.

- τ_i is the treatment effect for treatment i, $(i = 1, \ldots, t)$.

- β_j is the block effect for block j, $(j = 1, \ldots, b)$.

- ε_{ij} is the error in the observation for treatment i and block j.

- $N = bt$ is the total number of observations.

There are constraints:

$$\sum_{i=1}^{t} \tau_i = 0 \qquad \text{and} \qquad \sum_{j=1}^{b} \beta_j = 0$$

and assumptions:

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

with ε_{ij} independent of $\varepsilon_{i'j'}$ unless $i = i'$ and $j = j'$. (Normality, independence, equal variance).

There is also an implicit assumption that τ_i and β_j are *additive* effects. That is, the effect of changing from one treatment to another is the same in each block. This is sometimes described as the assumption of *no interactions* or *independence* of blocks and treatments.

If the additivity assumption is not valid, the residual sum of squares will be inflated and the sensitivity of the test will be reduced. Interaction is discussed in more detail later.

7.6 Analysis

We can test the treatment effects using a two-way ANOVA. There are no degrees of freedom for interaction so the additive model is assumed.

7.7 Example

An experiment is conducted to investigate the effects of various treatments on the storage properties of apples. There are four treatments, A, B, C, D. Each treatment is applied to one batch of apples. Each batch of apples is then divided into three sub-batches for storage. Three "blocks" of apples are formed, each consisting of one sub-batch from each of the four batches. The apples in each block are stored together for two months. At the end of this period, one observation on the concentration of a chemical in the skin of the apples is made on each sub-batch. The data are as follows.

	Treatment A	Treatment B	Treatment C	Treatment D
Block 1	72.4	53.1	72.2	57.1
Block 2	51.4	38.5	49.3	47.7
Block 3	65.9	52.1	64.6	51.6

8 Randomized Blocks II

8.1 Deriving estimates in a randomized blocks design

8.2   Example

We continue Example 7.7.

9 Latin Squares I

9.1 Introduction

In a randomised blocks design we allow for one nuisance factor by "blocking." In a *Latin square* design we allow for *two* nuisance factors by a more complicated system of blocking. The two nuisance factors must be *balanced* with respect to each other. We often represent the nuisance factors as rows and columns in a diagram of the design. Each treatment then occurs exactly one in each row and exactly once in each column.

Suppose that there are t treatments. Then we need t^2 experimental units and the t treatments are balanced across both factors using a Latin square layout.

<u>Example</u>

The design is called a "Latin square design" and it could represent, for example, the layout of plots in a field but it need not necessarily have such a direct physical interpretation.

9.2 Advantages and disadvantages

The advantages and disadvantages are broadly the same as for the randomized blocks design but the restriction to t^2 units can be a difficulty.

9.3 Planning the layout – randomisation

For a given value of t, several designs may be constructed. Some are simple permutations of the rows and columns and treatments, others are "essentially" different. In order to regularise this, we define a *standard square* in which the treatments come in alphabetical order in the first row and first column. If two standard squares are different then they must be essentially different. Other squares may be produced from standard squares by permuting the rows and columns.

<u>Example:</u> Two essentially different squares.

The number of different standard squares increases rapidly with t. The number of permutations of each standard square is $t! \times (t-1)!$, as in the following table.

Hence there are only two 2 Latin squares but the number of 6×6 Latin squares is 812,851,200.

9.4 Randomisation

To randomise, first choose a standard square at random from a list. Then randomise by permuting the rows and columns. It is sufficient to choose one of the $t!$ permutations of the columns and then one of the $(t-1)!$ permutations of rows 2 to t.

9.5 Analysis

We can write the model as

$$y_{jk} = \mu + \tau_{i(jk)} + \rho_j + \kappa_k + \varepsilon_{jk}$$

where

- y_{jk} is the observation in row j and column k.

- μ is the overall mean.

- τ_i is the treatment effect for treatment i. The treatment assigned to row j and column k is $i(jk)$.

- ρ_j is the effect of the "row" nuisance factor in row j.

- κ_k is the effect of the "column" nuisance factor in column k.

- $\varepsilon_{jk} \sim N(0, \sigma^2)$ is the error in the observation in row j and column k.

As usual the various effects sum to zero

$$\sum_{i=1}^{t} \tau_i = \sum_{j=1}^{t} \rho_j = \sum_{k=1}^{t} \kappa_k = 0$$

and the errors are independent with ε_{jk} independent of $\varepsilon_{j'k'}$ unless $j = j'$ and $k = k'$.
Let

T_i	be the total of the observations for treatment i.
R_j	be the total of the observations for row j.
C_k	be the total of the observations for column k.
G	be the overall total of the observations

10 Latin Squares II

10.1 Example

The effects of four different "recipes" for making cement on the compressive strength of cement cubes was investigated. The cement could be "gauged" (i.e. mixed) by each of four gaugers and the strength measured by each of four "breakers." A Latin square design was used with results as follows.

	Breaker 1	Breaker 2	Breaker 3	Breaker 4
Gauger 1	A 503	B 478	D 411	C 536
Gauger 2	D 485	C 606	A 586	B 531
Gauger 3	C 452	A 490	B 370	D 467
Gauger 4	B 432	D 510	C 510	A 547

10.2   Interpretation

This is similar to the randomised blocks design. We test treatments formally but comment informally on the effectiveness of the row and column nuisance factors.

10.3   Residuals

It can be shown (Exercise) that a residual (Observed-Fitted) is

$$\hat{\varepsilon}_{jk} = y_{jk} - \bar{y}_{T,i)jk)} - \bar{y}_{R,j} - \bar{y}_{C,k} + 2\bar{y}$$

where $\bar{y}_{T,i(jk)}$ is the mean for treatment $i(jk)$,
   $\bar{y}_{R,j}$ is the mean for row j,
   $\bar{y}_{C,k}$ is the mean for column k,
   $\bar{y}$ is the overall mean of the data,
   $i(jk)$ is the treatment applied in row i and column j.
   Residuals can be plotted by rows, columns and treatments and against fitted values.

10.4   Minitab

In Minitab you can use the `General Linear Model` function. Select the `Stat` menu then `ANOVA` and then `General Linear Model`.
   To use this function you need four columns of data. In one column, put the response, y. You then need three columns for the factor levels. One of these is for the row factor, one for the column factor and one for the treatments. Assuming that you enter the data in a straightforward order, you can get Minitab to generate the levels of the row and column factors by selecting `Make Patterned Data` and then `Simple Set of Numbers` from the `Calc` menu. You will need to type the treatment levels (using numbers: 1, 2, 3, ...) into another column.
   You can request residual plots, multiple comparisons (e.g. Tukey) etc.

10.5   Replicated Latin squares

For small t there are few residual degrees of freedom. In some cases it may be possible to improve matters by repeating the whole Latin square (re-randomised).

- Calculate a Rows sum of squares for each replication and add them.

- Calculate a Columns sum of squares for each replication and add them.

- "Replications" becomes a third nuisance factor.

   Example: A field experiment, to compare three varieties of wheat, is carried out at four sites, each using an independent randomisation. This is called a *replicated* Latin square.

	Site 1			Site 2	
B 59.8	A 58.9	C 56.8	C 54.3	B 53.8	A 57.6
A 64.4	C 58.1	B 63.0	A 60.8	C 47.4	B 55.7
C 52.3	B 57.2	A 56.9	B 58.0	A 49.7	C 49.2

	Site 3			Site 4	
A 67.0	B 61.6	C 53.2	C 60.9	A 65.4	B 63.0
C 50.3	A 53.4	B 52.9	A 64.4	B 64.7	C 56.0
B 58.3	C 51.2	A 58.3	B 58.9	C 55.1	A 59.6

Solution:

Treatment and site totals:

	Treat A	Treat B	Treat C	Total
Site 1	180.2	180.0	167.2	537.4
Site 2	168.1	167.5	150.9	486.5
Site 3	178.7	172.8	154.7	506.2
Site 4	189.4	186.6	172.0	548.0
Total	716.4	706.9	644.8	1068.1

Row totals:

	Row 1	Row 2	Row 3	Total	Sum of squares
Site 1	175.5	185.5	166.4	527.4	60.84666
Site 2	165.7	163.9	156.9	486.5	14.40889
Site 3	181.8	156.6	167.8	506.2	106.37556
Site 4	189.3	185.1	173.6	548.0	44.04222
				Total	225.57333

Column totals:

	Col 1	Col 2	Col 3	Total	Sum of squares
Site 1	176.5	174.2	176.7	527.4	1.28667
Site 2	173.1	150.9	162.5	486.5	82.19556
Site 3	175.6	166.2	164.4	506.2	24.11556
Site 4	184.2	185.2	178.6	548.0	8.43556
				Total	116.03335

10.6   Tied Latin squares

In some cases it may be possible to use the same levels of one nuisance factor.

Example: An experiment in the chemical industry uses three reaction vessels at a time. each run lasts several days. Three catalysts are compared over twelve runs in the form of four 3×3 Latin squares. These are called *tied* Latin squares. In this case they are tied by the factor "Reaction Vessel."

			Run											
		1	2	3	4	5	6	7	8	9	10	11	12
Reaction	1	A	B	C	B	A	C	C	A	B	B	C	A
Vessel		87	82	98	81	91	96	89	78	71	69	93	84
	2	B	C	A	A	C	B	B	C	A	A	B	C
		81	99	100	95	99	88	80	86	85	85	86	94
	3	C	A	B	C	B	A	A	B	C	C	A	B
		82	86	80	87	78	88	79	65	78	77	86	71

Solution:
Totals:

Vessel			Run	
1	1019		1	250
2	1078		2	267
3	957		3	278
Total	3054		4	263
			5	268
Catalyst			6	272
A	1044		7	248
B	932		8	229
C	1078		9	234
Total	3054		10	231
			11	265
			12	249
			Total	3054

10.7 Graeco-Latin squares

It is often possible to construct designs which will eliminate further nuisance factors by introducing them in an orthogonal way.

Consider the following design.

A α 1	B β 2	C γ 3	D δ 4
B γ 4	A δ 3	D α 2	C β 1
C δ 2	D γ 1	A β 4	B α 3
D β 3	C α 5	B δ 1	A γ 2

There are 4 treatments: A, B, C, D.

There are 4 nuisance factors: rows, columns, Greek letters, numbers. If we ignored the numbers, this would be called a "Graeco-Latin square."

N.B. There are some sizes of square, e.g. 6, 10, for which this is impossible.

Problems 3

1. The data are taken from Snee (1985). A randomized blocks experiment was carried out to investigate the effect of a drug added to the feed of chickens in an attempt to promote growth. The comparison is between three treatments: standard feed (control), standard feed plus low dose of drug, standard feed plus high dose of drug. The experimental unit is a group of chickens, reared together in the birdhouse. The experimental units are grouped three to a block, with physically adjacent units being in the same block. The response is the average weight, in pounds, per bird at maturity for the group of birds in the experimental unit.

   (I wonder why the weights of individual birds were not used).

Block	Control	Low dose	High dose
1	3.93	3.99	3.96
2	3.78	3.96	3.94
3	3.88	3.96	4.02
4	3.93	4.03	4.06
5	3.84	4.10	3.94
6	3.75	4.02	4.09
7	3.98	4.06	4.17
8	3.84	3.92	4.12

   Analyse these data to test for evidence of differences between the effects of the treatments. What assumptions are you making about the distributions of the data?

2. In a fertiliser experiment on potatoes there were six treatments, labelled A-F, defined as follows.

A:	Control	B:	Sulphate of ammonia (N)
C:	N+phosphate (P)	D:	N+Lime
E:	N+P+potash (K)	F:	Special NPK mixture

   There were four sites, each having six plots. A randomised blocks design was used with the following layout and results.

Site 1			Site 2		
F: 331	B: 286	E: 312	D: 292	F: 323	A: 185
C: 311	D: 280	A: 177	C: 294	B: 278	E: 322

Site 3			Site 4		
F: 313	C: 266	A: 182	C: 291	B: 253	E: 328
E: 319	D: 284	B: 258	A: 193	D: 233	F: 319

(a) Carry out a basic randomised blocks analysis on these data and draw up the ANOVA table. Test the null hypothesis that there are no differences between treatments and comment on the effectiveness of the blocking on the experiment.

(b) Use a suitable set of independent pre-experiment contrasts to investigate differences among treatments.

(c) Evaluate residuals for all of the observations and use them to check the assumptions made in the analysis.

3. Show that the residual for the observation y_{ij} in row i and column j of a Latin square design is

$$\hat{\varepsilon}_{jk} = y_{jk} - \bar{y}_{T,i(jk)} - \bar{y}_{R,j} - \bar{y}_{C,k} + 2\bar{y}$$

where $\bar{y}_{T,i(jk)}$ is the mean for treatment $i(jk)$, $\bar{y}_{R,j}$ is the mean for row j, $\bar{y}_{C,k}$ is the mean for column k and $\bar{y}$ is the overall mean for the data.

4. An experiment was carried out to compare four varieties of wheat, A, B, C and D. A large field was divided into 16 plots and a Latin square design was used. The layout and results are given below. The yields are in kg for each plot. Analyse the data and write a brief report summarising your conclusions.

$$
\begin{array}{llll}
\text{A } 4.0 & \text{D } 3.7 & \text{C } 4.5 & \text{B } 5.1 \\
\text{D } 3.6 & \text{A } 4.4 & \text{B } 5.2 & \text{C } 4.7 \\
\text{C } 5.4 & \text{B } 6.5 & \text{A } 4.4 & \text{D } 3.9 \\
\text{B } 6.4 & \text{C } 5.6 & \text{D } 3.8 & \text{A } 4.0 \\
\end{array}
$$

You can use Minitab to analyse the data if you wish. You will need to use the *General linear model* function.

5. An experiment was conducted to measure the effect of factory music on the output of production line workers. Four different music programs (A, B, C, D) were compared with no music (E). A single replication occupied the five working days of a week, a different program being tested on each day. A Latin square design was used to rotate the programs from week to week so that, over a five-week period, each program appeared once on any specific working day of the week. The whole design was replicated in three separate but comparable production lines.

The data, in percentages of target output, are as follows. They may also be found in the file `music.txt` in the MAS353 folder on the Campus Windows system. The design, with A, B, C, D, E replaced by 1, 2, 3, 4, 5, may be found in the file `musicplan.txt`.

Line 1

Week	Monday	Tuesday	Wednesday	Thursday	Friday
1	E: 82	A:123	B:118	D:114	C:108
2	B: 88	C: 83	A: 98	E: 85	D: 78
3	D: 69	E: 81	C: 96	B:102	A: 87
4	A: 94	D:101	E:106	C:104	B:109
5	C: 80	B:109	D: 90	A:100	E: 85

Line 2

Week	Monday	Tuesday	Wednesday	Thursday	Friday
1	B:109	D:102	A:114	E:111	C: 97
2	E: 73	C: 90	B: 94	D: 74	A: 93
3	C: 90	B:102	D: 90	A: 95	E: 81
4	A: 88	E:116	C:107	B:115	D: 98
5	D: 78	A:103	E: 92	C: 95	B: 99

Week	Monday	Tuesday	Wednesday	Thursday	Friday
1	C: 96	E:112	B:122	A:124	D: 88
2	E: 69	A: 89	D: 87	B: 97	C: 74
3	D: 66	B:105	A: 94	C: 96	E: 81
4	A: 97	D:110	C:110	E:113	B:115
5	B: 88	C: 88	E: 90	D: 88	A: 88

Analyse the data to test for differences in the effects of the treatments A, B, C, D, E. (Fit main effects only, i.e. Treatments, Days, Weeks, Lines. Do not attempt to measure any interaction effects).

Use a suitable contrast to compare "No music" (E) with "Music" (A, B, C, D) and hence also test for differences between the music programs (A, B, C, D).

Write a brief report on your findings. Comment on the design.

6. In an experimental design to compare the tensile strengths of tubes manufactured by different methods there are four different methods, A, B, C, D, where A is a control. These are arranged in a *Latin square* with four different manufacturing plants as the row factor and the day of manufacture as the column factor. To provide a better estimate of experimental error, the entire experiment is repeated using the same plants but on different days.

Analyse the results (in units of tensile strength), given below, to determine whether or not there are significant differences between the methods and, in particular, to compare the control with the other three methods. Comment on your results.

		Day of manufacture							
		1	2	3	4	5	6	7	8
		A	B	C	D	C	A	B	D
	1	16.6	16.9	17.4	17.4	17.6	17.8	18.1	16.5
		D	C	B	A	A	C	D	B
Plant	2	17.1	16.8	19.2	16.6	17.1	17.5	17.4	19.6
		C	D	A	B	D	B	A	C
	3	17.4	17.0	16.8	19.2	17.6	19.4	17.1	18.3
		B	A	D	C	B	D	C	A
	4	18.6	17.4	17.4	19.2	17.6	17.8	18.9	17.1

References

Snee, R.D., 1985. Graphical display of results of three-treatment randomized block experiments. *Applied Statistics*, **47**, 71-7.

Assignment 3

Please submit solutions to questions 2 and 4 in the homework box, by 5.00pm on Wednesday 8th March.

11 Missing data and outliers

11.1 Introduction

Observations may be missing or anomalous (i.e. outliers) for variuos reasons:

Missing experimental units: In certain blocks there may be one or more missing units, e.g. two few patients in a particular category.

Failed trials: The designed experiment is complete but one or more observations are lost subsequently, e.g. a test tube is dropped, a plot is found to be contaminated, an animal dies, a patient refuses to continue a trial. We must distinguish two cases:

1. The loss is unconnected with the treatment.
2. The loss may be linked with the treatment.

Outliers: the experiment is complete and all data are recorded. It is subsequently found that one or more observations are "outliers", i.e. inconsistent with other data.

11.2 Dealing with outliers

If an outlier is found, the first step is to check for a misrecording or transcription error. (Retain original records and field notes for this possibility).

If this fails, check whether a faulty experimental unit (e.g. a blocked drain) is responsible. If this is the case, discard the observation and treat as missing (unconnected with treatment).

A third possibility is that the model is incorrect or inadequate (e.g. non-normality, non-constant variance). In this case a transformation (e.g. log, square root, power) might solve the problem.

If the outlier is unresolved, use the missing observation method (below) to reanalyse without it. In this case the influence of the doubtful observation on the conclusions must be considered and reported.

11.3 Dealing with missing observations

Missing observations will often cause a design to become unbalanced. This will mean that effects are no longer orthogonal. The data may still be analysed, for example using the Minitab GLM function, but the simple formulae for sums of squares will no longer apply and the various sums of squares will not add up to the total. (The situation is similar to a multiple regression where changes in the residual sum of squares due to adding factors to the model depend on the order in which the factors are added). If only one or two observations are missing, it might be helpful to acrry out an approximate analysis by "estimating" the missing values by least squares.

11.4 Estimation of a missing value

11.4.1 Procedure

Step 1: Let x denote the missing value.

Step 2: Write down an expression for the residual sum of squares as a function of x.

Step 3: Find the value $\hat{x}$ for x which minimises the residual sum of squares.

11.4.2 Example: Randomised blocks

Notation:

There are t treatments and b blocks. The missing observation is on treatment i' in block j'.

S_1	Treatments sum of squares
S_B	Blocks sum of squares
S_R	Residual sum of squares
S_T	Total sum of squares
$T_{i'} = T'_{i'} + x$	Treatment total for treatment i'.
$B_{j'} = B'_{j'} + x$	Block total for block j'.
$G = G' + x$	Overall total.

We know that

12   Sample size

12.1   Introduction

An important aspect of the design of experiments is the determination of sample size. How many observations should be made? Taking too many observations may lead to unnecessary cost or delay and may be unethical if human or animal subjects are involved. Taking too few observations may mean that we are unlikely to get a useful result and may also be unethical if we are involving human or animal subjects in an experiment which is likely to be a waste of time. Sample size calculations are often required by *ethics committees* in hospitals, universities etc. before approval is given for an experiment involving human or animal subjects.

The usual approach is to find the smallest sample size which will give a specified *power* to a test at a specified *significance level* under certain specified conditions. The calculations can be complicated but computer software is often available. In fact Minitab will do sample size and power calculations in some cases.

12.2   Practical significance and statistical significance

Statistical significance refers to the significance level of a test. That is the probability of rejecting the null hypothesis when it is true. This is a measure of the strength of evidence against a null hypothesis.

Practical significance refers to whether an effect, that is a difference between population means, has any real importance. Two treatments might have effects which are not exactly the same but are so similar that the difference is of no importance. A practically significant difference is one which is of real interest to the experimenter.

In determining sample size we use both of these kinds of significance. A very large experiment might give a statistically significant result even when the real difference is of no practical significance. An experiment which is too small might fail to give a statistically significant result even when a difference of practical significance exists.

12.3   Required information

Discussion between the statistician and the experimenter should establish the following.

1. What is the smallest difference of practical significance, i.e. worth detecting?

$$\Delta$$

2. What value of error standard deviation is anticipated?

$$\sigma$$

3. What significance level, i.e. Type I error probability, will be used?

$$\alpha$$

4. What power is required when the true difference is Δ, i.e. how sure do we wish to be of detecting a difference as large as Δ? The power is $1 - \beta$ where β is the Type II error probability.

$$\beta$$

5. What type of design is to be used? What nuisance factors are there? This affects the number of residual degrees of freedom.

$$\nu$$

N.B. We will assume here that sample sizes are to be equal. The size of each sample is r.

12.4 Example: Two-sample t-test

In this case $\nu = 2(r - 1)$. Note that ν depends on r.

Model:

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{2\sigma^2}{r}\right)$$

If H_0 is true,

$$\bar{X}_1 - \bar{X}_2 \sim N\left(0, \frac{2\sigma^2}{r}\right)$$

and

$$T = \frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{2/r}} \sim t_\nu$$

where S^2 is based on ν d.f.

Under $H_A : \mu_1 - \mu_2 = \Delta$, we have

$$T' = \frac{\bar{X}_1 - \bar{X}_2 - \Delta}{S\sqrt{2/r}} \sim t_\nu.$$

We reject H_0 at the $100\alpha\%$ level if $|T| > k_1$ where k_1 is the two-sided α critical point of t_ν.

We wish the probability of rejection when $\mu_1 - \mu_2 = \Delta$ to be $1 - \beta$. We reject if $|T| > k_1$. That is we reject if

$$\left| T' + \frac{\Delta}{S\sqrt{2/r}} \right| > k_1.$$

That is, we reject if

$$T' > k_1 - \frac{\Delta}{S\sqrt{2/r}}$$

or

$$T' < -k_1 - \frac{\Delta}{S\sqrt{2/r}}.$$

The latter has a very small probability so we ignore it. The rejection probability is then $1 - \beta$ if

$$k_1 - \frac{\Delta}{S\sqrt{2/r}} = -k_2,$$

where $-k_2$ is the lower β critical point of t_ν. (See diagram). Of course, in sample size calculations, we need to supply an anticipated value s for S.

So, we set

$$\frac{\Delta}{s\sqrt{2/r}} = k_1 + k_2$$

which implies

$$r = \frac{2s^2}{\Delta^2}(k_1 + k_2)^2$$

where k_1 is the two-sided α point of t_ν and k_2 is the one-sided β point of t_ν. Unfortunately the right hand side also depends on r through ν so we need to solve the equation iteratively. We guess a value for r to find ν and then use the equation to find a new value for r. We continue until the process converges.

12.5 Numerical example

A t-test is to be used to compare a drug treatment and a control. Past experiments suggest a typical mean response of 150 and standard deviation of about 10. Testing at the 5% level, the investigator wishes to be 95% sure of detecting a change of 15 units. How many observations should be taken?

Solution

Here $\Delta = 15$, $s = 10$, and $2s^2/\Delta^2 = 0.8889$.

Guess $r_0 = 10$ so $\nu = 18$, $k_1 = 2.101$, and $k_2 = 1.734$.

Then $r_1 = 0.8889(2.101 + 1.734)^2 \approx 13$ so $\nu = 24$, $k_1 = 2.064$ and $k_2 = 1.711$.

Then $r_2 = 0.8889(2.064 + 1.711)^2 \approx 13$.

So $r = 13$.

12.6 Comparing several treatments

One approach to dealing with the case when we have more than two treatments is as follows. We will wish to detect a difference greater than Δ between *any* pair of population means. So we modify the previous method by using the studentised range. This change only affects k_1, the critical value under H_0. We replace $t_\nu \times s\sqrt{2/r}$ by $q_{t,\nu} \times s\sqrt{1/r}$. Hence we replace k_1 by

$$k_1' = \frac{q_{t,\nu}(\alpha)}{\sqrt{2}}$$

where t is the number of treatments being compared.

12.7 Numerical example

12.8 Adjusting the sample size required

Sometimes the suggested sample size might seem to be too large to the experimenter. What can be done? There are various possibilities.

1. Increase Δ. Doubling Δ divides the sample size by 4.

2. Increase β or α.

3. Reduce σ. Halving σ will divide the sample size by 4.

Options 1 and 2 involve weakening the requirements. Option 3 is more desirable if it is possible, for example by improved technique or better design.

13 Factorial experiments I

13.1 Introduction

The "two-way ANOVA" is a simple example of a factorial experiment. A factorial experiment is simply an experiment where the explanatory variables are *factors*, i.e. categorical.

There is no need to stop at two factors. There is no theoretical limit to the number of factors but designs can become very complicated if there are many factors and there can be lots of interaction effects to consider.

For example, an experiment with three factors at 3, 4, 5 levels respectively is called a $3 \times 4 \times 5$ factorial experiment. It would need 60 observations for a single replication. There would be main effects, 2-way interactions between pairs of factors and a 3-way interaction between the three factors. this would leave no degrees of freedom for the residual. We would need

either to assume that some of the interactions are zero and use their sums of squares as the residual

or have more than one replication

(or both)

Instead of a single large factorial experiment with several factors we might consider a series of single-factor experiments in which one factor is varied while others are held constant. The approach using one factorial experiment is usually preferred because

1. (economy/efficiency). Fewer observations in total will be needed.

2. (flexibility). The results are more widely applicable.

3. (more information). *Interaction* can only be measured in multi-factor experiments.

Main effect : The *main effect* of a factor is its direct effect, as opposed to any interactions.

Interaction : An *interaction* occurs when the difference in the response to levels of one factor depends on the levels of other factors. It is characterised by a *non-additive* response to combinations of levels of factors.

Examples of interaction :

- The effect of a fertiliser may vary with the type of soil.
- The effect of a drug may vary with the activity of a patient.
- The effect of a catalysts may vary with the temperature of a reaction.

13.2   Some terminology and notation

Levels of factors : The various values which a factor can take are called *levels*.

Factor names : The names of factors are denoted by upper case letters: A, B, C, ...

Factor levels : the levels of factors are denoted by lower case letters: $a_1, a_2, a_3, \ldots, b_1, b_2, b_3, \ldots$.

Treatment combinations : A treatment combination can be represented by, e.g., $a_1 b_3 c_2$.

Cross-classification : A cross-classification is a full set of treatment combinations.

Replications : One observation taken at each possible treatment combination makes a single replication. An experiment with 3 factors, A, B, C, with 2, 4, 3 levels respectively (a $2 \times 4 \times 3$ factorial experiment) would require 24 units per replication. Several replications can be arranged in a completely randomised, randomised blocks or Latin square design.

   N.B. Designs can become very large. To get round this problem, techniques such as *confounding* and *fractional replication* can be used. These will be described in a later lecture.

13.3   Analysis

We illustrate the principles using the two-factor case.
   The model now incorporates an interaction term. Observation k of those taken at level i of factor A and level j of factor B is written

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

The overall mean is μ.

The factor A effect is α_i.

The factor B effect is β_j.

The interaction effect is γ_{ij}.

The treatment combination effect is $\tau_{ij} = \alpha_i + \beta_j + \gamma_{ij}$.

   Here
   $i = 1, 2, \ldots, a$ (a levels)
   $j = 1, 2, \ldots, b$ (b levels)
   $k = 1, 2, \ldots, r$ (r replications)
   Unless $r > 1$, it is impossible to distinguish between γ_{ij} and ε_{ijk}. that is we can not measure interaction.
   The usual constraints and assumptions are

$$\sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{b} \beta_j = \sum_{i=1}^{a} \gamma_{ij} = \sum_{j=1}^{b} \gamma_{ij} = 0$$

$$\varepsilon_{ijk} \sim N(0, \sigma^2)$$

(independently).

   For a randomised blocks design, the model also includes a block term δ_l with $\sum_{l=1}^{d} \delta_l = 0$. We assume that block effects are additive. That is they do not interact with other factors.

13.4 Analysis of variance (two-factor case)

There are two stages:

1. First regard each treatment combination as a separate treatment. Analyse the data as in a completely randomised, randomised blocks or Latin square design as appropriate. This leads to a between-treatments sum of squares on $t-1$ degrees of freedom, where $t=ab$, and a residual sum of squares.

   In the completely randomised case:

   The total sum of squares is

$$S_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r} y_{ijk}^2 - \frac{G^2}{abr}$$

   on $abr-1$ degrees of freedom.

   The total between-treatments sum of squares is

$$S_F = \frac{1}{r}\sum_{i=1}^{a}\sum_{j=1}^{b} T_{ij}^2 - \frac{G^2}{abr}$$

   on $ab-1$ degrees of freedom.

   Here T_{ij} is the total for observations at level i of factor A and level j of factor B. That is

$$T_{ij} = \sum_{k=1}^{r} y_{ijk}.$$

   The overall total of the observations is

$$G = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r} y_{ijk} = \sum_{i=1}^{a}\sum_{j=1}^{b} T_{ij}.$$

   The residual sum of squares is then $S_R = S_T - S_F$ on $(abr-1)-(ab-1) = ab(r-1)$ degrees of freedom.

2. Next calculate the sums of squares for main effects and interactions.

   The factor A sum of squares is

$$S_A = \frac{1}{br}\sum_{i=1}^{a} A_i^2 - \frac{G^2}{abr}$$

   on $a-1$ degrees of freedom, where the total for level i of factor A is

$$A_i = \sum_{j=1}^{b}\sum_{k=1}^{r} y_{ijk} = \sum_{j=1}^{b} T_{ij}.$$

   The factor B sum of squares is

$$S_B = \frac{1}{ar}\sum_{j=1}^{b} B_j^2 - \frac{G^2}{abr}$$

   on $b-1$ degrees of freedom, where the total for level j of factor B is

$$B_j = \sum_{i=1}^{a}\sum_{k=1}^{r} y_{ijk} = \sum_{i=1}^{a} T_{ij}.$$

   The interaction sum of squares is $S_I = S_F - S_A - S_B$ on $(ab-1)-(a-1)-(b-1) = (a-1)(b-1)$ degrees of freedom.

Source of variation	Degrees of freedom	Sum of squares	Mean square	Variance ratio
Factor A	$a-1$	S_A		
Factor B	$b-1$	S_B		
Interaction	$(a-1)(b-1)$	S_I		
Between treatments	$ab-1$	S_F		
Residual	$ab(r-1)$	S_R		
Total	$abr-1$	S_T		

13.5   Example

Volunteers are given one of three diets (Factor A) and one of three drug treatments (Factor B). The levels of Factor A are a_1 : Food supplement 1, a_2 : Food supplement 2, a_3 : Control (normal diet). The levels of Factor B are b_1 : Drug 1, b_2 : Drug 2, B_3 : Control (no drug). There are four replications. The data are as follows.

Level of Factor A	Level of Factor B b_1		b_2		b_3	
a_1	3.4	3.8	4.0	4.2	2.8	2.8
	4.4	3.9	3.3	2.7	3.8	3.4
a_2	6.5	6.7	4.9	4.9	6.1	5.2
	6.0	5.7	5.4	4.8	6.1	5.5
a_3	4.6	5.8	3.9	3.7	6.5	6.2
	4.9	5.0	3.5	3.9	6.4	6.4

Problems 4

1. Derive a formula for estimating a single missing value in a Latin square design.

2. An ecologist wishes to compare the diameters of shells on adult snails of the same species bred in captivity from specimens caught in two different locations. The ecologist will use a two-sample t-test, with the assumption of equal variances. If the standard deviation of shell diameters is expected to be about 3mm, find the sample size required to give a power of 90% for a two-sided test at the 5% significance level when the difference in population means is 2mm.

3. Suppose now that the ecologist in question 2 wants to use a one-way analysis of variance to compare the shell diameters in four groups of snails. If the standard deviation is expected to be about 3mm, find the sample size required to be 90% sure of detecting a difference of 2mm between any pair of population means, using the 5% significance level.

Assignment 4

Please submit solutions to questions 1, 2, 3 in the homework box, by 5.00pm on Wednesday 22nd March.

14 Factorial experiments II

14.1 Hypothesis testing

Stage 1 Test the overall treatment effects using the between-treatment variance ratio. Test informally for block effects if applicable. Provided that it is not clear that there are no significant effects present, proceed to stage 2.

Stage 2 Test for interaction (H_0 : $\gamma_{ij} = 0$, $i = 1, \ldots, a$, $j = 1, \ldots, b$) using the interaction variance ratio which is the interaction mean square divided by the residual mean square. If the result is not significant, go to stage 3a. If the result is significant, go to stage 3b.

Stage 3a If the test for interaction does not give a significant result, test the main effect of each factor independently by comparing the each mean square to the residual mean square in a separate variance ratio.

Multiple comparisons may be applied separately to each factor.

Stage 3b If the test for interaction gives a significant result then the main effects tests may be applied as *formal* procedures but we are more interested in comparing the factor mean squares *informally* with the interaction mean square. This will tell us whether main effects or interactions dominate. In this case, use of multiple comparisons is more awkward. If we wish to apply multiple comparisons, we would usually apply them to the treatment combinations. However, if a main effect is dominant, it might still be informative to compare the mean levels for that factor.

14.2 Example (13.5)

14.3 More than two factors

In a model with three factors, A, B, C, there are seven effects:

- Three main effects: A, B, C.

- Three two-way interactions: A × B, A ×C, B × C.

- One three-way interaction: A × B × C. The three-way interaction means that the interaction of A and B varies with the level of C etc. In practice three-way (and higher order) interactions often turn out not to be significant.

The various sums of squares can be calculated a follows.

Stage 1 Calculate the total sum of squares, a between-treatment-combinations sum of squares and, if applicable, a blocks sum of squares. The residual sum of squares is obtained by subtraction.

Stage 2 Split the between-treatment-combinations sum of squares into main effects and interaction sums of squares.

1. Main effects sums of squares are calculated from the factor level totals as usual. The sum of squares for factor A is

$$S_A = \frac{1}{bcr} \sum_{i=1}^{a} A_i^2 - \frac{G^2}{abcr}$$

on $a - 1$ degrees of freedom, where a, b, c are the numbers of levels of the factors, r is the number of replications and A_i is the total for level i of factor A.

2. Calculate the sums of squares for two-way interactions. For example, to calculate the sum of squares for the A × B interaction we first calculate a sum of squares between combinations of levels of A and B:

$$S_{BT(AB)} = \frac{1}{cr} \sum_{i=1}^{a} \sum_{j=1}^{b} T_{(AB)ij}^2 - \frac{G^2}{abcr}$$

on $ab - 1$ degrees of freedom and then subtract the main effects sums of squares for A and B from this to obtain the interaction sum of squares

$$S_{AB} = S_{BT(AB)} - S_A - S_B$$

on $(a - 1)(b - 1)$ degrees of freedom, where $T_{(AB)ij}$ is the total of observations made with A at level i and B at level j. Alternatively this can be seen as the interaction sum of squares in a two-way analysis of variance obtained by ignoring factor C.

3. Calculate the three-way A × B × C interaction sum of squares. This can be found by subtracting the main effects and two-way interaction sums of squares from the between-treatment-combinations sum of squares. It has $(a - 1)(b - 1)(c - 1)$ degrees of freedom (which can also be found by subtraction).

The sums of squares can, of course, be found using computer software such as Minitab. See 14.5.

Test the hypotheses starting with the three-way interaction, then, if this is not significant, the two-way interactions, then, if these are not significant, the main effects.

Example: a 2 × 4 × 5 experiment with two replications.

	a_1b_1	a_1b_2	a_1b_3	a_1b_4	a_2b_1	a_2b_2	a_2b_3	a_2b_4
c_1	72.4	75.2	71.6	73.3	47.1	59.7	54.8	75.1
	55.2	74.5	45.6	81.2	53.7	59.4	54.5	64.5
c_2	56.8	72.3	75.7	76.3	57.3	45.5	64.8	52.1
	57.4	84.3	69.9	80.9	55.3	65.3	69.2	41.4
c_3	58.6	71.1	76.8	72.5	54.3	57.5	53.9	56.8
	65.7	74.3	82.7	82.3	72.5	55.8	54.2	53.8
c_4	53.5	73.7	62.2	82.0	53.9	59.8	44.3	70.9
	63.7	79.9	66.1	72.1	50.3	57.2	63.0	52.8
c_5	63.6	71.3	59.2	75.7	52.7	44.4	44.6	60.8
	65.7	79.3	72.7	72.1	56.5	61.5	54.8	55.4

There are 40 treatment combination totals: $72.4 + 55.2 = 127.6, \ldots, 60.8 + 55.4 = 116.2$.

	a_1b_1	a_1b_2	a_1b_3	a_1b_4	a_2b_1	a_2b_2	a_2b_3	a_2b_4
c_1	127.6	149.7	117.2	154.5	100.8	119.1	109.3	139.6
c_2	114.2	156.6	145.6	157.2	112.6	110.8	134.0	93.5
c_3	124.3	145.4	159.5	154.8	126.8	113.3	108.1	110.6
c_4	117.2	153.6	128.3	154.1	104.2	117.0	107.3	123.7
c_5	129.3	150.6	131.9	147.8	109.2	105.9	99.4	116.2

The sum of squares for treatment combinations is:

There are 8 totals for combinations of levels of A and B (column totals in the data table):

	b_1	b_2	b_3	b_4
a_1	612.6	755.9	682.5	768.4
a_2	553.6	566.1	558.1	583.6

There are 10 totals for combinations of levels of A and C:

	a_1	a_2
c_1	549.0	468.8
c_2	573.6	450.9
c_3	584.0	458.8
c_4	553.2	452.2
c_5	559.6	430.7

There are 20 totals for combinations of levels of B and C:

	b_1	b_2	b_3	b_4
c_1	228.4	268.8	226.5	294.1
c_2	226.8	267.4	279.6	250.7
c_3	251.1	258.7	267.6	265.4
c_4	221.4	270.6	235.6	277.8
c_5	238.5	256.5	231.3	264.0

Source of variation	Degrees of freedom	Sum of squares	Mean square	Variance ratio
A × B × C	12			
A × B	3			
A × C	4			
B × C	12			
A	1			
B	3			
C	4			
Between treatments	39			
Residual	40			
Total	79			

The extension of this to four or more factors is exactly as you would expect.

14.4   Use of a single replication

In designs with three or more factors, the number of observations per complete replication can be large. However the three-way and higher order interactions are often unimportant. So, if it can be decided in advance that these interactions will have no real effect, their sums of squares can be used as a residual sum of squares and this makes it possible to test the main effects and two-way interactions even if only one replication is observed.

Example: a $2 \times 4 \times 5$ experiment with a single replication.

Source of variation	Degrees of freedom	Sum of squares	Mean square	Variance ratio
A × B × C	12			
A × B	3			
A × C	4			
B × C	12			
A	1			
B	3			
C	4			
Between treatments	39			
Residual	0			
Total	39			

However, if we are willing to assume that the (true) three-way interaction effects are zero then any sum of squares associated with this effect can only be due to error and can be treated as a residual.

Example: a $2 \times 4 \times 5$ experiment with a single replication, assuming that the three-way interaction effects are zero.

Source of variation	Degrees of freedom	Sum of squares	Mean square	Variance ratio
A × B	3			
A × C	4			
B × C	12			
A	1			
B	3			
C	4			
Residual	12			
Total	39			

14.5   Using Minitab

You can use

```
Stat  -->  ANOVA  -->  Balanced ANOVA
```

or

```
Stat  -->  ANOVA  -->  General Linear Model
```

In the `Model` box you have to specify the effects in the model.

Examples

- `A B` — main effects of A and B.

- `A B A*B` — main effects of A and B and the A × B interaction.

- `A|B` — this is the same as `A B A*B`.

- `A B C A*B A*C B*C A*B*C` — all main effects and two-way and three-way interactions with three factors, A, B, C.

- `A|B|C` — this is the same as `A B C A*B A*C B*C A*B*C`.

In `Balanced ANOVA`, A, B etc. will be assumed to be factors.

In `General Linear Model`, A, B etc. will be assumed to be factors unless you declare them to be covariates.

You can generate levels of factors using

```
Calc  -->  Make Patterned Data --> Simple Set of Numbers
```

15 2^n factorial experiments I

15.1 Introduction

Experiments in which there are n factors, each with just two levels, ar known as 2^n factorial experiments. They are useful for investigating a large number of factors of potential interest, particularly in the pilot stage of a study.

Although they are a special case of multi-factor experiments, their usefulness has led to a special notation.

15.2 Notation

Factor names are in upper case: A, B, C, . . .

Main effects are also in upper case: A, B, C, . . .

Interactions are written: AB, ACD, etc.

Levels of factors are written in lower case. Since there are only two levels of each factor (typically absent/present or lower/upper or standard/new or . . .), we could write, e.g., ā, a for the two levels of A. A *treatment combination* could then be written (e.g. in a 2^4 experiment) as

$$\bar{a}b\bar{c}d$$

in which A and C are athe their lower levels (or absent) and B and D are at their upper levels (or present).

However there is usually no ambiguity caused by missing out the symbols for the factors which are at their lower levels and this convention is usually adopted. So we write

$$bd$$

instead of āb̄cd. This requires a special notation for the treatment combination with all treatments at their lower level (āb̄c̄d̄). We write this as

$$(1)$$

Thus the complete set of treatment combinations in a 2^4 factorial experiment is, in *standard order*:

(1)	a	b	ab	c	ac	bc	abc
d	ad	bd	abd	cd	acd	bcd	$abcd$

(In practice, of course, the treatment combinations would be allocated to units by randomization).

IMPORTANT: Do not confuse the symbol, ac, for a treatment combination with the symbol, AC, for an interaction effect. The notation should be used with care.

15.3 Analysis

15.3.1 Outline

If we have $r > 1$ replication we can proceed as follows. (Other cases will be described later).

Stage 1 : Calculate the total sum of squares, on $2^n r - 1$ degrees of freedom, and a between-treatment-combinations sum of squares, on $2^n - 1$ degrees of freedom. Find the residual sum of squares, on $2^n(r - 1)$ degrees of freedom, by subtraction.

Stage 2 : Break down the between-treatment-combinations sum of squares into sums of squares for the various main effects and interactions. Because each factor has two levels, each sum of squares has one degree of freedom. Each effect can be represented by a *contrast*.

15.3.2 Main effects

The main effect of A can be thought of as a contrast between A present and A absent. E.g. 2^3 experiment: write the treatment combination totals (summed over r replications) as

$$T_{(1)}, \ T_a, \ T_b, \ T_{ab}, \ T_c, \ T_{ac}, \ T_{bc}, \ T_{abc}.$$

Then we use the contrast:

$$-1, \ +1, \ -1, \ +1, \ -1, \ +1, \ -1, \ +1$$

giving a contrast total $T_A = -T_{(1)} + T_a - T_b + \cdots + T_{abc}$.
Hence the contrast sum of squares is

$$S_A = \frac{T_A^2}{r \sum c_i^2} = \frac{T_A^2}{2^n r} = \frac{T_A^2}{N}.$$

(Exercise: Show that the same result is obtained from factor level totals).
The other main effect sums of squares are found similarly, of course.

15.3.3 Two-way interactions

Example: AB in a 2^3 experiment with r replications. We define an interaction as the difference in response between

- A and B at the same level (whether upper or lower),

- A and B at opposite levels.

A suitable contrast is thus:

(1)	a	b	ab	c	ac	bc	abc
+1	−1	−1	+1	+1	−1	−1	+1

Note that this is independent of all main effects.
From this we obtain a contrast total T_{AB} and a sum of squares

$$S_{AB} = \frac{T_{AB}^2}{N}.$$

Note that the contrast for AB could be obtained by multiplying corresponding coefficients in the contrasts for A and B.

A	−	+	−	+	−	+	−	+
B	−	−	+	+	−	−	+	+
AB	+	−	−	+	+	−	−	+

This generalises to more factors. E.g., for ABC,

$$
\begin{array}{rllllllll}
C & - & - & - & - & + & + & + & + \\
ABC & - & + & + & - & + & - & - & +
\end{array}
$$

From this we obtain T_{ABC} and

$$
S_{ABC} = \frac{T_{ABC}^2}{N}.
$$

15.3.4 Sign table

The calculations of the contrasts can be set out in a *sign table*.
Example: 2^3 factorial with 3 replications.

Treatment Combination	Total	A	B	C	AB	AC	BC	ABC
(1)	30.5	-	-	-	+	+	+	-
a	32.3	+	-	-	-	-	+	+
b	32.4	-	+	-	-	+	-	+
ab	26.4	+	+	-	+	-	-	-
c	23.9	-	-	+	+	-	-	+
ac	32.4	+	-	+	-	+	-	-
bc	25.9	-	+	+	-	-	+	-
abc	27.2	+	+	+	+	+	+	+
Total T	231.0	5.6	-7.2					
Sum of squares T^2/N	2223.375	1.307	2.160					

In practice, of course, we can use computer software such as Minitab to generate contrasts and do numerical calculations.

16 2^n factorial experiments II

16.1 Testing hypotheses

The usual practice is to calculate variance ratios for all effects and test their significance. the interpretation is done hierarchically, starting with the highest-order interactions. When an interaction is found to be significant, all of its component effects are regarded as established, whatever the significance of their individual tests. For example, a significant result in the test of the ABC interaction is considered to establish the existence of the ABC, AB, AC and BC interactions and the A, B and C main effects.

16.2 Choosing the best combination of treatments

Both the Minitab `Balanced ANOVA` and `General Linear Model` functions will store fitted values if requested. (Other statistical software will also do this). This makes it easy to select the treatment combination which gives the best predicted response if we consider all effects to be "real."

However often we will only wish to take into account some of the possible effects, usually those effects the existence of which is established, either

- because the experimenter specifies in advance that this effect is to be included in the model,

- by a significant result in the test of the effect itself

- or by a significant result in the test of an interaction which involves the effect.

Therefore we need first to determine which effects are established and then fit the model involving only these effects. We can then determine the treatment combination with the best predicted response by looking at the fitted values.

If we wish to do the calculations by hand we can determine the estimate of the parameter for a particular factor by dividing the contrast total by N, the overall total number of observations. The effect of a factor is often regarded as the difference in mean response between the upper and lower levels of the factor. Here we define the parameter to be half of this difference. Consider an effect E. This could be a main effect or an interaction. Its parameter is τ_E. This parameter is added in the model in cases where the effect is at its upper level and subtracted in cases where it is at its lower level (following the signs in the sign table). So (ignoring the errors) the total for the contrast for effect E is

$$T_E = \frac{N}{2}\tau_E - \frac{N}{2}(-\tau_E) = N\tau_E$$

since everything else cancels out. So we estimate it by

$$\hat{\tau}_E = \frac{T_E}{N}.$$

(Note that, using the definition of an effect size mentioned above, the effect size would be $2\tau_E$.)

So, for example, consider the 2^3 factorial in 14.3.4. We modify the sign table adding an extra column, labelled "I", for the overall mean.

Treatment Combination	I	A	B	C	AB	AC	BC	ABC
(1)	+	-	-	-	+	+	+	-
a	+	+	-	-	-	-	+	+
b	+	-	+	-	-	+	-	+
ab	+	+	+	-	+	-	-	-
c	+	-	-	+	+	-	-	+
ac	+	+	-	+	-	+	-	-
bc	+	-	+	+	-	-	+	-
abc	+	+	+	+	+	+	+	+

Using this sign table, we can see that the mean response for treatment combination ac is

$$\mu + \tau_A - \tau_B + \tau_C - \tau_{AB} + \tau_{AC} - \tau_{BC} - \tau_{ABC}.$$

So the estimate of the mean response (i.e. the predicted response) is

$$p_{ac} = \frac{1}{N}(G + T_A - T_B + T_C - T_{AB} + T_{AC} - T_{BC} - T_{ABC}).$$

In fact, in the case where we fit *all* main effects and interactions, this is just the mean for observations on treatment combination ac. However, if we omit some effects, this simplification does not apply and we have to use the formula. For example, if we omit the three-way interaction in the example then we obtain

$$p_{ac} = \frac{1}{N}(G + T_A - T_B + T_C - T_{AB} + T_{AC} - T_{BC}).$$

It is easy to see how to apply this method to other treatment combinations, other selections of fitted effects and other 2^n factorial designs.

16.3 Confidence intervals

The variance of a contrast total, which is made up of N observations, is $N\sigma^2$, where σ^2 is the error variance and $N = 2^n r$ is the total number of observations. Therefore a confidence interval for a parameter estimate, corresponding to a contrast, is

$$\frac{T}{N} \pm t_\nu \sqrt{\frac{s^2}{N}}$$

where $t_\nu(\alpha)$ is an appropriate critical point from the Student's t-distribution on $\nu = 2^n(r-1)$ degrees of freedom and s^2 is the residual mean square.

If we fit all possible effects then the predicted response for a treatment combination is just the sample mean for the observations on that treatment combination. This has variance σ^2/r, when there are r replications. So, in a 2^n factorial with r replications, a confidence interval for the response for a particular treatment combination (if all effects are fitted) is

$$\bar{y}_{tc} \pm t_\nu(\alpha)\sqrt{\frac{s^2}{r}}.$$

where $t_\nu(\alpha)$ is an appropriate critical point from the Student's t-distribution on $\nu = 2^n(r-1)$ degrees of freedom and s^2 is the residual mean square.

When less than the full model is fitted, the estimate of the mean response for a treatment combination is not so simple but is made up of a number of parameter estimates, as shown in 16.2 above. However the contrast totals are independent since the contrasts are orthogonal. This can be seen from the fact that the variance matrix of the vector of parameter estimates is $\sigma^2(X^T X)^{-1}$, where X is the design matrix, which is just the sign table with $+1$ and -1 for $+$ and $-$. It is easy to see that $X^T X$ is diagonal.

The variance of G and of each contrast total is $N\sigma^2$. So, if an estimated mean response involves G and k contrast totals, then the variance of the estimator of the mean response is

$$\frac{1}{N^2}(1+k)N\sigma^2 = \frac{(1+k)\sigma^2}{N}.$$

Note that, if all main effects and interactions are used, then $1 + k = 2^n$ and the variance simplifies to σ^2/r as above.

Thus a confidence interval for the mean response for a treatment combination, when k effects have been fitted, is

17 2^n factorial experiments III

17.1 Using a single replication

17.2 Example

Problems 5

1. Show that the same expression as given in the notes (15.3.2) for the sum of squares for the main effect of factor A can be obtained using the factor level totals.

2. An experiment on the growth of tomato plants involved three substances, A, B, C, added to the irrigation water. A 2^3 factorial experiment was used with each factor corresponding to absence or presence of a substance so, for example, ab is the treatment where A and B are present but C is absent. Four replications were carried out. The growth data are as follows.

Treatment	Observations				Total
(1)	47.7	36.4	44.8	51.2	180.1
a	48.6	38.1	51.2	48.2	186.1
b	45.0	49.3	31.2	42.3	167.8
ab	62.8	58.6	51.7	62.1	235.2
c	44.7	37.2	48.4	46.5	176.8
ac	54.7	51.9	51.8	51.7	210.1
bc	44.6	48.7	48.6	50.7	192.6
abc	68.0	70.1	61.2	63.2	262.5

   Analyse the data, testing for the interaction and, if appropriate, the main effects.

3. The following data represent the results of a 2^2 factorial design with four replications, arranged as a 4×4 Latin square. Analyse the data, testing for the interaction and, if appropriate, the main effects and allowing for row and column nuisance factors.

(1)	9	b	7	ab	13	a	10
a	11	ab	12	(1)	8	b	9
b	10	(1)	6	a	8	ab	10
ab	12	a	12	b	7	(1)	9

Assignment 5

Please submit solutions to questions 1, 2, 3 in the homework box, by 5.00pm on Wednesday 3rd May.

18 Confounding

18.1 Introduction

In a 2^n experiment, we may have a natural block size which is too small to hold a complete replication.

For example, consider a 2^3 experiment on 32 experimental units arranged in 8 blocks of size 4. There are 4 replications, each of which is split across 2 blocks. E.g., replication 1:

Block 1	(1)	ab	ac	bc
Block 2	a	b	c	abc

(Replications 2, 3 and 4 are similarly split between blocks but randomized again).

In each block, the contrast for A has two + signs and two − signs. Hence block effects cancel in the A contrast and the effect of A can be measured. This is also true for B, C, AB, AC, BC. However the three-way interaction ABC has − everywhere in Block 1 and + everywhere in Block 2.

Thus ABC is measured by the same contrast as the block effect and these effects can not be distinguished. We say that ABC is *confounded* with blocks.

Effects are confounded when they can not be distinguished in the analysis of an experiment. Whilst it is usually undesirable to have effects confounded, by careful design we can often make sure that the effects which are confounded are the unimportant ones so that important effects can be measured.

18.2 Design aspects

18.2.1 Case 1: Two blocks per replication

When the block size is 2^{n-1} we have two blocks per replication. This is the simplest case. We choose an unimportant interaction and write down its contrast. We then split each replication into blocks using the signs in this contrast.

The analysis of the results is straightforward. We simply discard the confounded interaction (assuming it to be negligible). The between treatments sum of squares is reduced by the contrast sum of squares of the confounded interaction and its degrees of freedom are reduced by 1.

Example 1 : As before, consider a 2^3 experiment, with 8 blocks of 4 units each and ABC confounded in all replications.

Each of the effects A, B, C, AB, AC, BC has a sum of squares on 1 df so the between treatments sum of squares has 6 df. The between blocks sum of squares, calculated from the block totals, has 7 df. This sum of squares *includes* ABC.

The total sum of squares has 31 df and the residual sum of squares, obtained by subtraction, has 31-7-6=18 degrees of freedom.

Example 2 : Consider a 2^4 experiment with 2 replications in 4 blocks of 8 units. ABCD is confounded and we deem all three-way and the 4-way interaction to be negligible.

Design (Replication 1):

Block 1	(1)	ab	ac	ad	bc	bd	cd	abcd
Block 2	a	b	c	d	abc	abd	acd	bcd

Replication 2 is similarly split between blocks 3 and 4. The order of the blocks and the order of the treatments within a block are randomized.

The important effects are A, B, C, D, AB, AC, AD, BC, BD, CD, giving 10df for the between treatments sum of squares. (This ignores 3- and 4-way interactions).

There are 4 blocks so the between-blocks sum of squares has 3 df and this includes the confounded effect ABCD.

The total sum of squares has 31 df so the residual sum of squares (by subtraction) has 31-10-3=18 df. Note that the residual sum of squares includes the 3-way interactions (4df) which could have been separated out if required.

18.2.2 Case 2: Smaller block fractions

Confounding can be used whenever the block size is of the form 2^{n-p}. For $p = 2$ we have $2^2 = 4$ blocks per replication.

Consider a 2^4 experiment in blocks of size 4 and the following two possible designs.

Design 1 confounds ABC and BCD.

					ABC	BCD
Block 1	(1)	bc	abd	acd	−	−
Block 2	a	abc	bd	cd	+	−
Block 3	ab	ac	d	bcd	−	+
Block 4	b	c	ad	abcd	+	+

Design 2 confounds ABCD and BCD.

					ABCD	BCD
Block 1	a	abc	abd	acd	−	−
Block 2	(1)	bc	bd	cd	+	−
Block 3	b	c	d	bcd	−	+
Block 4	ab	ac	ad	abcd	+	+

We can see that, in Design 2, the main effect A has also been confounded with blocks. When we use four blocks per replication, 3df are involved and hence 3 different effects are confounded. We need to identify the 3 effects before planning the experiment. This is done as follows.

Choose two effects to be confounded. Then the third effect is obtained by "multiplying" these effects together and cancelling those which appear twice. For example:

Design 1 : ABC × BCD = ABBCCD = AD.

Design 2 : ABCD × BCD = ABBCCDD = A.

This works because "multiplying" effects multiplies their signs.

If we had 8 blocks per replication then there would be 7 degrees of freedom between blocks. The blocks could be defined by 3 contrasts but a further $7 - 3 = 4$ effects would also be confounded.

18.3 Confounding − General principles

18.3.1 Rules

1. Choose interactions to confound carefully. We must ensure that *all* confounded interactions are negligible and that the aims of the experiment are met. In particular, check for hidden confounding.

2. For blocks of size 2^{n-p}, split each replication using the signs of p interactions. Randomise the order of the blocks and the order of treatments within blocks.

3. In the analysis, take care to avoid double-counting. The confounded effects should be removed from the between-treatments sum of squares since they will be counted in the between-blocks sum of squares.

18.3.2 Steps in the analysis

Suppose that there are 2^{n-p} units per block and so 2^p blocks per replication and $2^p r$ blocks altogether.

1. Calculate a total sum of squares on $N - 1$ degrees of freedom, where $N = 2^n r$.

2. Calculate a between-blocks sum of squares from the block totals on $2^p r - 1$ df.

3. Calculate an individual sum of squares for each required effect (e.g. for main effects and two-way interactions). You can use the relevant parts of the sign table or use Minitab, including only the required terms in the model.

4. Calculate the residual sum of squares and its df by subtraction. The residual will include interactions which are not confounded but regarded as negligible.

18.3.3 Checks and shortcuts in design

Principal block Define the *principal block* to be the fraction of the design which contains the treatment combination (1).

1. A treatment combination in the principal block and a confounded interaction always have an even number of letters in common.

   For example, consider the designs in 18.2.2.

2. The principal block is closed under "multiplication" (where $a^2 = b^2 = c^2 = \cdots = 1$). Hence, if two members are known a third can be generated.

   For example, consider block 1 in design 1 in 18.2.2. We see that $bc \times abd = acd$.

3. All other blocks can be generated from the principal block by "multiplying" through by any treatment combination not in the principal block.

   For example, consider design 1 in 18.2.2. Block 2 is Block 1 multiplied by a. Block 3 is Block 1 multiplied by ab. Block 4 is Block 1 multiplied by b.

18.4 Example

Two replications of a 2^5 experiment, with factors A, B, C, D, E, are arranged in blocks of eight. That is, each block is a quarter-replication and there are eight blocks. We would like to measure and test all main effects and 2-way interactions.

Confounding: We need two contrasts to confound and these will imply a third by multiplication.

Suppose that we choose to confound ABCDE, which seems, at first glance, an obvious choice. If we also choose a 4-way interaction, e.g. ABCD, then we will also confound a main effect, e.g. E, since ABCDE × ABCD = E. If we choose a 3-way interaction, e.g. ABC, then we will also confound a 2-way interaction, e.g. DE, since ABCDE × ABC = DE.

If we want to measure all main effects and 2-way interactions then we can not confound ABCDE. It is better to pick two 3-way interactions which only share 1 symbol and then the third confounded effect is a 4-way interaction.

Let us choose to confound ABC, CDE and therefore ABDE (= ABC × CDE).

The design (before randomisation) is therefore as follows.

Blocks	ABC	CDE								
1, 5	+	+	c	abc	ad	bd	ae	be	cde	$abcde$
2, 6	+	−	a	b	cd	$abcd$	ce	$abce$	ade	bde
3, 7	−	+	ac	bc	d	abd	e	$abde$	$acde$	$bcde$
4, 8	−	−	(1)	ab	acd	bcd	abe	ace	bce	de

The data are as follows.

Treat.	Block	Obs.	Treat.	Block	Obs.	Treat.	Block	Obs.	Treat.	Block	Obs.
(1)	4	48.9	e	3	51.6	(1)	8	43.1	e	7	48.8
a	2	50.3	ae	1	54.6	a	6	39.6	ae	5	44.1
b	2	55.0	be	1	65.1	b	6	62.3	be	5	69.2
ab	4	61.4	abe	3	68.2	ab	8	60.6	abe	7	73.2
c	1	64.5	ce	2	57.4	c	5	46.7	ce	6	60.5
ac	3	50.4	ace	4	61.0	ac	7	56.1	ace	8	67.6
bc	3	51.5	bce	4	63.7	bc	7	57.8	bce	8	60.9
abc	1	67.5	$abce$	2	74.0	abc	5	66.3	$abce$	6	79.0
d	3	62.8	de	4	49.6	d	7	61.2	de	8	57.7
ad	1	66.6	ade	2	46.4	ad	5	68.6	ade	6	44.1
bd	1	60.8	bde	2	59.2	bd	5	67.8	bde	6	57.7
abd	3	73.6	$abde$	4	57.2	abd	7	58.6	$abde$	8	68.0
cd	2	65.3	cde	1	66.9	cd	6	55.1	cde	5	56.9
acd	4	55.5	$acde$	3	61.7	acd	8	67.2	$acde$	7	61.9
bcd	4	53.7	$bcde$	3	70.0	bcd	8	52.1	$bcde$	7	55.8
$abcd$	2	65.7	$abcde$	1	77.2	$abcd$	6	70.0	$abcde$	5	75.3

18.5 Partial confounding (a brief introduction)

When an experiment which involves confounding is replicated more than once, we need not confound the same interactions in each replication.

Example A 2^3 experiment with 4 replications, in 8 blocks.

Replication	Block	Treatment combinations				Confounds
1	1	(1)	ab	ac	bc	ABC
	2	a	b	c	abc	
2	3	(1)	ab	c	abc	AB
	4	a	b	ac	bc	
3	5	(1)	ac	b	abc	AC
	6	a	c	ab	ac	
4	7	(1)	bc	a	abc	BC
	8	b	c	ab	ac	

Note that we would randomize the order of treatment combinations within blocks and the order of the blocks.

The analysis is similar to that for completely confounded designs, but main effects and interactions can only be estimated from replications in which they are not confounded.

An advantage of partial confounding is that more effects can be measured and tested. In experiments with very small block sizes (e.g. 2) this technique is essential.

The major disadvantage is that, although more effects are measured, each is measured from fewer observations and therefore with less precision.

19 Fractional Replication

19.1 Introduction

In large factorial experiments ($n \geq 6$) even a single replication may cause difficulties by its size. Fractional replication sometimes allows us to use only a fraction (2^{-p}) of a full replication and still retain all important information.

19.2 Illustration 1

This example is not suitable for use in practice but it shows the principles and consequences of the technique.

Consider a 2^3 experiment on 4 experimental units. Suppose that we choose to use only the treatment combinations for which ABC gets +, that is a, b, c, abc.

The sign table is as follows. The constant I is included to help to clarify what is happening.

	I	A	B	AB	C	AC	BC	ABC
a	+	+	−	−	−	−	+	+
b	+	−	+	−	−	+	−	+
c	+	−	−	+	+	−	−	+
abc	+	+	+	+	+	+	+	+

The header row above is under the spanning label "Effect".

Aliases : A glance at this table shows that some effects have identical contrasts. Two effects which have exactly the same contrast are said to be *aliases* (or to be *aliased* with each other). Clearly, if two effects are aliases, we can not distinguish one from the other.

Looking at the table we see that

- C and AB are aliases,

- B and AC are aliases,
- A and BC are aliases

We also see that ABC gets a + everywhere, the same as I. So ABC can not be measured.

In a full replicate, there are 2^n treatment combinations and, if we include the 1 df associated with the constant I, there are 2^n df. Because we are using a half-replicate, the number of degrees of freedom, including the 1 df for the constant (I), is halved. Therefore one effect (indicated by the defining contrast) becomes unmeasurable and every other effect becomes aliased with another.

Defining contrast : Because the signs for ABC are the same as those for I, we say that the *defining contrast* of this half-replicate is ABC $\equiv$ I.

Note that the complementary half-replicate

$$(1) \quad ab \quad ac \quad bc$$

has defining contrast ABC $\equiv$ $-$I, (since the signs for ABC are minus those for I). The aliases in this half-replicate are effectively the same except for a sign change, i.e. A $\equiv$ $-$BC, B $\equiv$ $-$ AC, C $\equiv$ $-$ AB.

The choice of which half-replicate to use would usually be made at random.

19.3   Illustration 2 (more practical)

Consider a 2^6 experiment on 32 units. This is a half-replicate. The 3-way and higher-order interactions are assumed to be negligible.

Consider using the defining contrast ABCDEF = I. (That is, we use those treatment effects where the contrast for the 6-way interaction ABCDEF gets a +. A moment's thought shows that, since 6 is even, we must use those treatment combinations which have symbols containing an even number of the letters a, b, c, d, e, f). The treatment combinations used would be as follows.

(1)	ab	ac	ad	ae	af	bc	bd
be	bf	cd	ce	cf	de	df	ef
abcd	abce	abcf	abde	abdf	abef	acde	acdf
acef	adef	bcde	bcdf	bcef	bdef	cdef	abcdef

We can find an alias by "multiplying" through the defining contrast. For example:

$$
\begin{aligned}
I &\equiv ABCDEF \\
A \times I &\equiv A \times ABCDEF \\
A &\equiv BCDEF
\end{aligned}
$$

Similarly, every main effect will be aliased with a 5-way interaction.

Each 2-way interaction is aliased with a 4-way interaction. For example:

$$
\begin{aligned}
AB \times I &\equiv AB \times ABCDEF \\
AB &\equiv CDEF
\end{aligned}
$$

Hence, if all 4-way, 5-way and 6-way interactions are negligible, we can measure all main effects and 2-way interactions. If the 3-way interactions are also negligible, these can be pooled to form a residual sum of squares on 10 df. (There are twenty 3-way interactions but they are aliased in pairs so there are only ten contrasts).

19.4 Outline of analysis

1. Calculate the total sum of squares from all observations (in Illustration 2 there are 31 df).

2. Calculate the non-negligible effects directly, e.g. using a sign table. Check that none is aliased with another non-negligible effect. (In Illustration 2 there are 6 main effects and 15 two-way interactions and these require 21 df).

3. Calculate the residual sum of squares by subtraction. (In Illustration 2 this has $31 - 21 = 10$ df).

19.5 Example

Consider a half-replication of a 2^4 experiment with factors A, B, C, D. We are interested in the main effects and the AC interaction. All other interactions are considered to be negligible. We can use the defining contrast

$$\text{ABCD} \equiv \text{I}$$

Then each main effect is aliased with a 3-way interaction. The 2-way interaction of interest is aliased with BD. In fact

$$
\begin{aligned}
\text{AB} &\equiv \text{CD} \\
\text{AC} &\equiv \text{BD} \\
\text{AD} &\equiv \text{BC}
\end{aligned}
$$

So, we could measure all three of the 2-way interactions involving any one factor (provided we could consider the others to be negligible). However this would leave no degrees of freedom for the residual.

The treatments and observations are as follows.

(1)	ab	ac	bc	ad	bd	cd	abcd
2.65	4.00	2.57	2.81	3.45	3.26	1.86	3.62

19.6   Smaller fractions

It is possible to have 1/4-replicates, 1/8-replicates etc. For a 2^{-p}-replicate of a 2^n experiment (a "2^{n-p} design"), we need p defining contrasts to define the fraction used (e.g. for a 1/2-replicate we need one defining contrast, for a 1/4-replicate we need two and so on). However care must be taken in choosing the defining contrasts since any two will imply a third which can be found by multiplication. This lead to $2^p - 1$ defining contrasts in total. For example, using ABCD $\equiv$ I and BCDE $\equiv$ I would imply AE $\equiv$ I. This third defining contrast would mean that main effects A and E would be aliased.

The general principles for the analysis remain the same.

19.7   Final comments

We can construct designs involving both fractional replication and confounding. For example, we could have a half-replicate of a 2^6 experiment split between two blocks of 16 experimental units each.

We can also have 3^n experiments, where each factor has three levels and we might use 1/3-replicates etc.

There is much more that can be said about experimental design but that is all we have room for in this module.